Supplementary Information for

Unraveling Mechanisms of Protein Encapsulation and Release in Coacervates via Molecular Dynamics and Machine Learning

Yiwei Wang,‡[a] Rongrong Zou,‡[a] Yeqiang Zhou,[a] Yi Zheng,[a] Chuan Peng,[a] Yang Liu,[a] Hong Tan,[a] Qiang Fu,[a] and Mingming Ding*[a]*

[a] College of Polymer Science and Engineering, State Key Laboratory of Polymer Materials Engineering, Sichuan University, Chengdu 610065, China.

*Corresponding author
Mingming Ding; E-mail: dmmshx@scu.edu.cn or dmmshx@163.com
Yang Liu; E-mail: liuyang_leon@scu.edu.cn

Containing Supplementary Methods, Supplementary Text 1-3, Supplementary Figures List, Supplementary Figs 1-49 and Supplementary Tables 1-4.

Supplementary Methods

In all of our simulations, the all-atom (AA) models of PLys and PGlu were established using Pymol package[1]. All of the models were mapped to coarse-grained (CG) models using Martinize2[2]. The BSA, Saporin, Lysozyme, EGFP and Actin used in the article were obtained from the Protein Database (PDB) and the ID codes of proteins are 4f5s, 1qi7, 1gwd, 6fll and 1nwk respectively. 0.05 protein represents 1 protein: 2000 amino acids residues to quantify proteins in one system[3]. PLys, PGlu, BSA and 0.15M sodium chloride ions were randomly added into a 40 nm × 40 nm × 40 nm water box in all simulation systems, but there were differences in the ingredient adding sequence, BSA content, or secondary structure of poly(amino acid)s in each system. We determined the equilibrium time of the system by identifying the initial time point at which the y-axes of analyses such as SASA, contact number, and number of clusters reach relatively stable values (Fig. S4 S9 S10 S24 S25 S28 S33 S34 S37 S47 and S49†). All systems reached equilibrium within 5 μs. We extended the simulation time to 8 μs to more easily observe the equilibrium of systems.

Snapshots were all captured with VMD[4]. The radial distribution function (RDF), solvent accessible surface area (SASA), end to end distance, contact number, number of cluster (NCluster), energy decomposition analysis and Gibbs energy 2D landscapes were all computed using Gromacs tools 'gmx rdf', 'gmx sasa', 'gmx polystat', 'gmx mindist', 'gmx clustersize', 'gmx energy' and 'gmx sham'. We used the protein or polymer as a whole, used the data from the final 3 μs of the trajectory, and selected whole_mol_com in - selrpos to calculate RDF based on the centroid of the polymer and protein. The positions of the first valleys in the RDFs determine the cut-off points. For calculating contact numbers and cluster numbers, distances shorter than 0.6 nm (the RDF cut-off distance) are regarded as contacts or grouped within the same cluster. Contact maps were generated using homemade Python scripts to visualize the average number of contacts between residues in protein and in coacervate within the system. The identification of residues in BSA with the highest contact frequency with PLys were directly generated by Python code during the computation of the contact maps. The amino acid model of BSA, obtained from the PDB database, was directly annotated in VMD to emphasize residues exhibiting the highest contact frequency. Encapsulation percentage of BSA over time describes the coverage of coacervate molecules on BSA, which was defined as:

$$f_{protein} = c_{protein_{coacervate}} / c_{protein_{nonprotein}} \quad (1)$$

where the c variables represent the numbers of contacts between the CG beads in the two molecule types in the sublabel. The term 'coacervate' here refers to the coacervate phase (including PLys and PGlu molecules and water molecules within the cut-off).

Configurational entropy was calculated using the Schlitter formulation[5], which provides an upper bound S to the true configurational entropy S_{true}:

$$S_{true} < S = \frac{k_B}{2} ln\ det\ [1 + \frac{k_B T e^2}{h^2} D] \quad (2)$$

Here k_B is Boltzmann's constant, e is Euler's number, T is the absolute temperature, and h is Planck's constant divided by 2π. D denotes the $3N \times 3N$ covariance matrix of

mass-weighted atomic Cartesian coordinates:

$$D = < [M^{1/2}(r - < r >)] \otimes [M^{1/2}(r - < r >)] > \quad (3)$$

where r is the coordinate vector of the N atoms or beads considered for the entropy calculation and M is the dimensional diagonal matrix containing the masses of these particles. The angle bracket represents the ensemble average. To study the internal degree of freedom to the entropy, the molecular configurations were superimposed via a translational and rotational least-squares fit, thus excluding the rotational and translational motions. The molecular trajectory was superimposed via both translational and rotational least-square fit to the initial structure (the initial configuration is well equilibrated).

Principal Component Analysis (PCA)[6] as an unsupervised machine learning method was applied to reduce dimensionality and extract motion characteristics of poly(amino acid) chains. we initiated the process by standardizing the dataset, ensuring that each feature has a mean of 0 and a standard deviation of 1. The covariance matrix of the standardized data was computed to assess the correlations between features. Following this, we determined the optimal directions for data transformation by solving for the eigenvalues and eigenvectors of the covariance matrix that we selected the eigenvectors corresponding to the top 35 largest eigenvalues to form the transformation matrix. We finally project the original motion data into a new coordinate system, completing the dimensionality reduction of the data.

A machine learning clustering approach known as Density-Based Spatial Clustering of Applications with Noise (DBSCAN)[7] was applied in our work. DBSCAN is a density-based spatial clustering algorithm that can identify and handle noise and outliers in datasets. Initially, we calculated the truncation radius between PLys and PGlu based on their position of the first valley in RDF and selected it as the neighborhood radius Epsilon ε = 0.6 nm**.** In order to ensure that each cluster contains at least one PLys and one PGlu, and since the content we feed is three-dimensional coordinates, we assigned a minimum point value (MinPts) of 4. Both ε and MinPts significantly influence the clustering results of DBSCAN. Subsequently, we fed the coordinates of all PLys and PGlu beads as data points and calculated the number of neighboring points within the ε neighborhood for each data point. If a data point had at least 3 neighboring points within its ε-neighborhood, we marked it as a core point. We then identified all points that were directly density-reachable, forming a cluster. This process continued until no new points could be added to any cluster. Finally, the number of clusters was calculated. We implemented this process through Python.

Set up systems: BSA encapsulation in coacervate and the impact of ingredient adding sequence on coacervate formation and encapsulation rate. When converting the AA models of PLys and PGlu into CG models using the Martinize2, we opted the coil conformation as the secondary structure of the poly(amino acid)s[8]. Taking the 0.2BSA-LG system as an example, briefly introduce the steps of systems construction. PLys, BSA and 0.15 M sodium chloride ions were randomly added into a 40 nm × 40 nm × 40 nm box and solvate them with Martini's ordinary water beads. This was followed by approximately 5 μs of simulation to achieve simple mixing (STEP1). After

removing the solvent and salt ions from the system, we added PGlu. We reintroduced the solvent and maintained the salt concentration at 0.15 M. Subsequently, an approximately 8 μs simulation was conducted to form and stabilize the coacervates (STEP2). The modeling of systems with other content of BSA and other ingredient addition sequences can be obtained from Table S1.

Set up systems: the impact of ingredient addition sequence on other protein systems. In this section, we select different proteins (Saporin, Lysozyme, EGFP, Actin) to add to the simulation systems. The modeling of systems involving different proteins and ingredient addition sequence on coacervate formation can be referenced from Table S2. Given that the encapsulation efficiency of proteins decreases with an increase in protein content, we opted for systems with lower protein concentrations for the analysis of different protein systems (0.1Protein), aiming to observe more significant variations in encapsulation efficiency. We also chose to simulate systems with different proteins using the same three component addition sequences as in the first section.

Set up systems: the influence of secondary structure in coacervates on protein encapsulation. The establishment of the simulation systems were similar to the LG systems in the first section, only changing the secondary structure of PLys and PGlu from coil to β-sheet. The modeling of systems with different content of BSA can be obtained from Table S3.

Set up systems: pH-induced dissolution of coacervates and protein release. We maintained 0.05 BSA concentration in our simulations at different pH value, which is extracted from the experimental setting[3], and we opted the coil conformation as the secondary structure of the poly(amino acid)s[8]. The establishment of systems was similar to LG systems in the first section. Numbers of PGlu and PLys added to systems were kept constant. We assume that PGlu exists only in its fully protonated (PGlu+H) and fully deprotonated (PGlu-H) states, with a pKa of 4.25. Using the equation pH = pKa + log (A-/HA), we can determine the ratio of these states at different pH levels. When pH=2, PGlu is predominantly in the PGlu+H state (99.5%), and no further modification of the beads was conducted at this pH. And the number of PGlu+H and PGlu-H species varied at different pH levels as indicated in Table S4.

Supplementary Text

Text S1: Analysis of high-frequency contact sites between PGlu and BSA

Taking systems with 0.4 BSA content as examples, we annotated the highest frequency contact sites for BSA/PLys and BSA/PGlu. We observed that the contact sites between BSA and PLys predominantly cluster around Glu and Asp (carry opposite charges to PLys) residues and their adjacent residues as shown in Fig. S7a. Similarly, the contact sites between BSA and PGlu mainly concentrate around Lys and Arg (carry opposite charges to PGlu) residues and their neighboring residues as shown in Fig. S7b This suggests that it is the electrostatic interactions that leads to the occurrence of the encapsulation process and the small charged patches on the BSA surface significantly influence the encapsulation in coacervates.

Text S2: Restricted PC2 motion of PGlu in GL systems

We also detected differences in the distributions of PC2 projections for PGlu chains caused by the ingredient adding sequence on coacervate formation in the GL systems. The PC2 distribution of PGlu is narrower than that in the other systems, indicating that its twisting was restricted (PC2 motion pattern as shown in Fig. S19a-S19e) within the GL systems. Due to the fact that the encapsulation efficiency of BSA protein is less affected by PGlu than PLys, even if PGlu movement is restricted in the GL systems, it is reasonable that no significant improvement in protein encapsulation efficiency was observed in the systems.

Text S3: Schilitter entropy analysis of PGlu

During the Martini 3.0 coarse-grained modeling of PGlu, its side chain consists of only two beads, fewer than the three beads found in the PLys side chain. This structural simplification renders PGlu less susceptible to perturbations in the simulation environment, resulting in a smaller entropy fluctuation as shown in Fig. S22.

Supplementary Figures List

Supplementary Figures

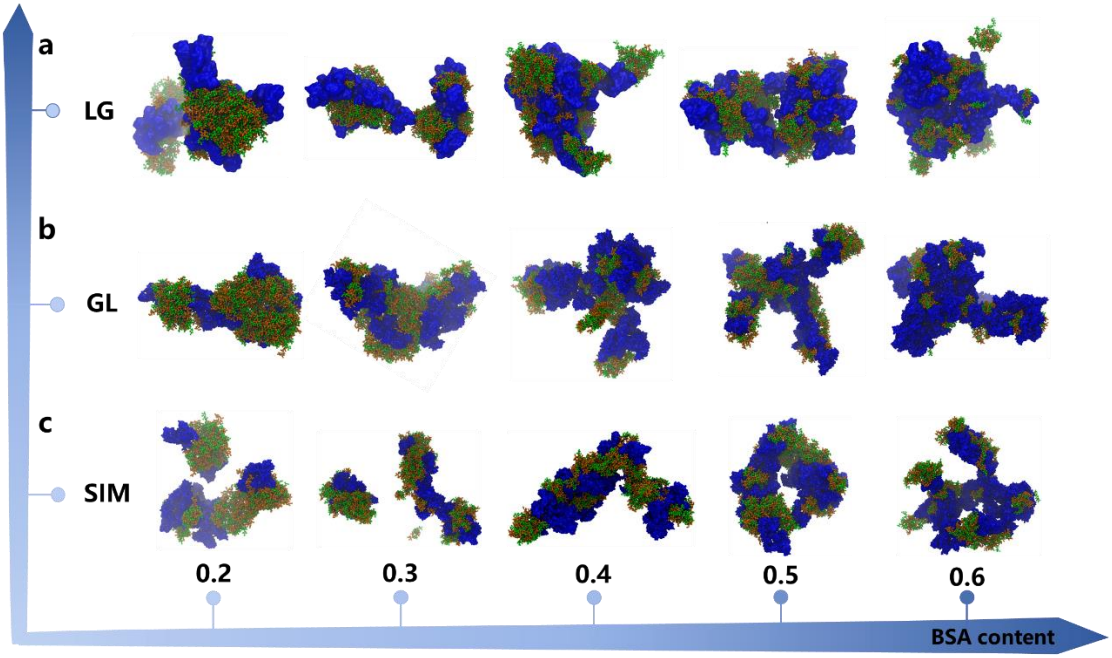

Fig. S1 Snapshots of simulations of BSA within coil systems. The snapshots of the simulations in (a) LG systems, (b) GL systems, (c) SIM systems, and snapshots from left to right represent systems with additive BSA content. Orange represents PLys, green represents PGlu, and dark blue represents BSA. Water and ions are not shown for clarity.

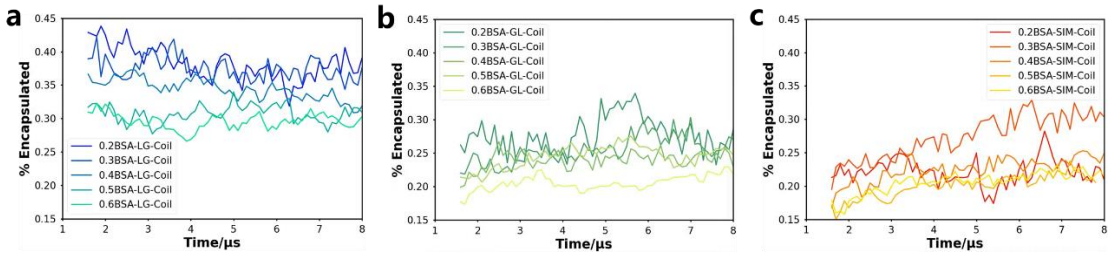

Fig. S2 The encapsulation percentage of coacervates within coil systems. The encapsulation percentage of BSA protein within coacervate as functions of time in (a) LG systems, (b) GL systems, (c) SIM systems.

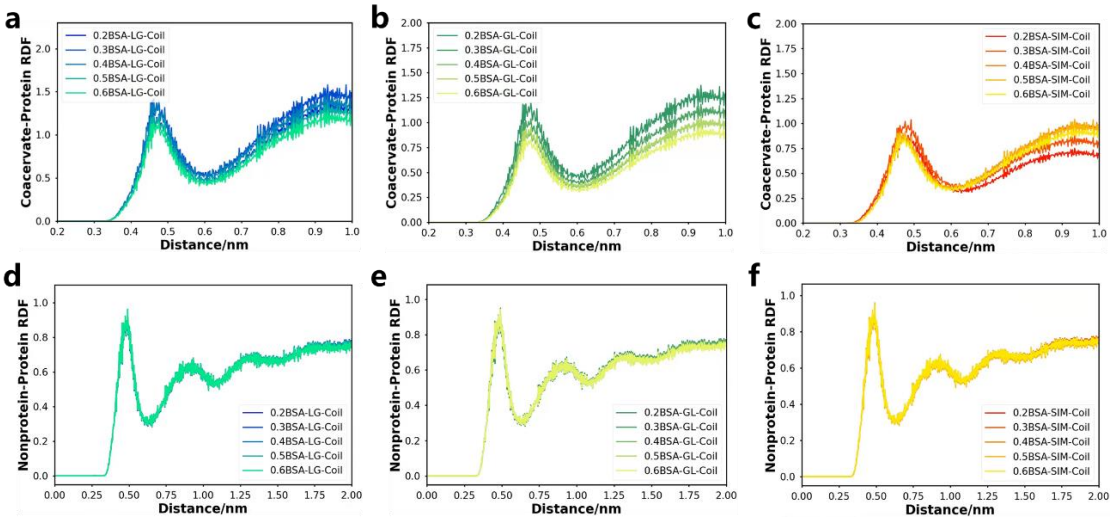

Fig. S3 RDF analysis between various components within coil systems. The RDF between coacervates and protein in (a) LG systems, (b) GL systems, (c) SIM systems. The RDF between nonprotein and protein in (d) LG systems, (e) GL systems, (f) SIM systems.

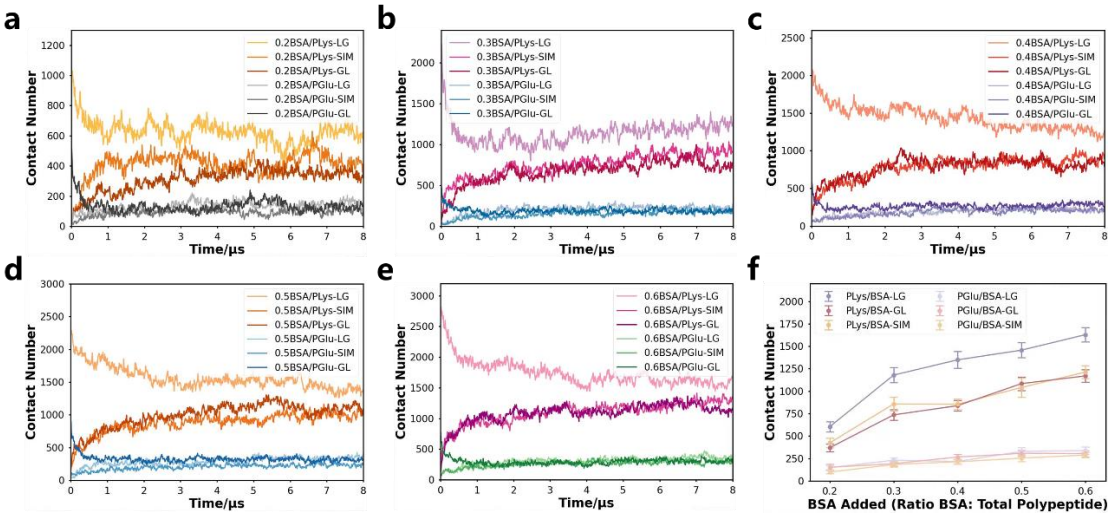

Fig. S4 Contact number of PLys/BSA and PGlu/BSA within coil systems. The contact number between different components over time under different BSA ratios (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA. (f) Contact number of PLys/BSA and PGlu/BSA in various systems.

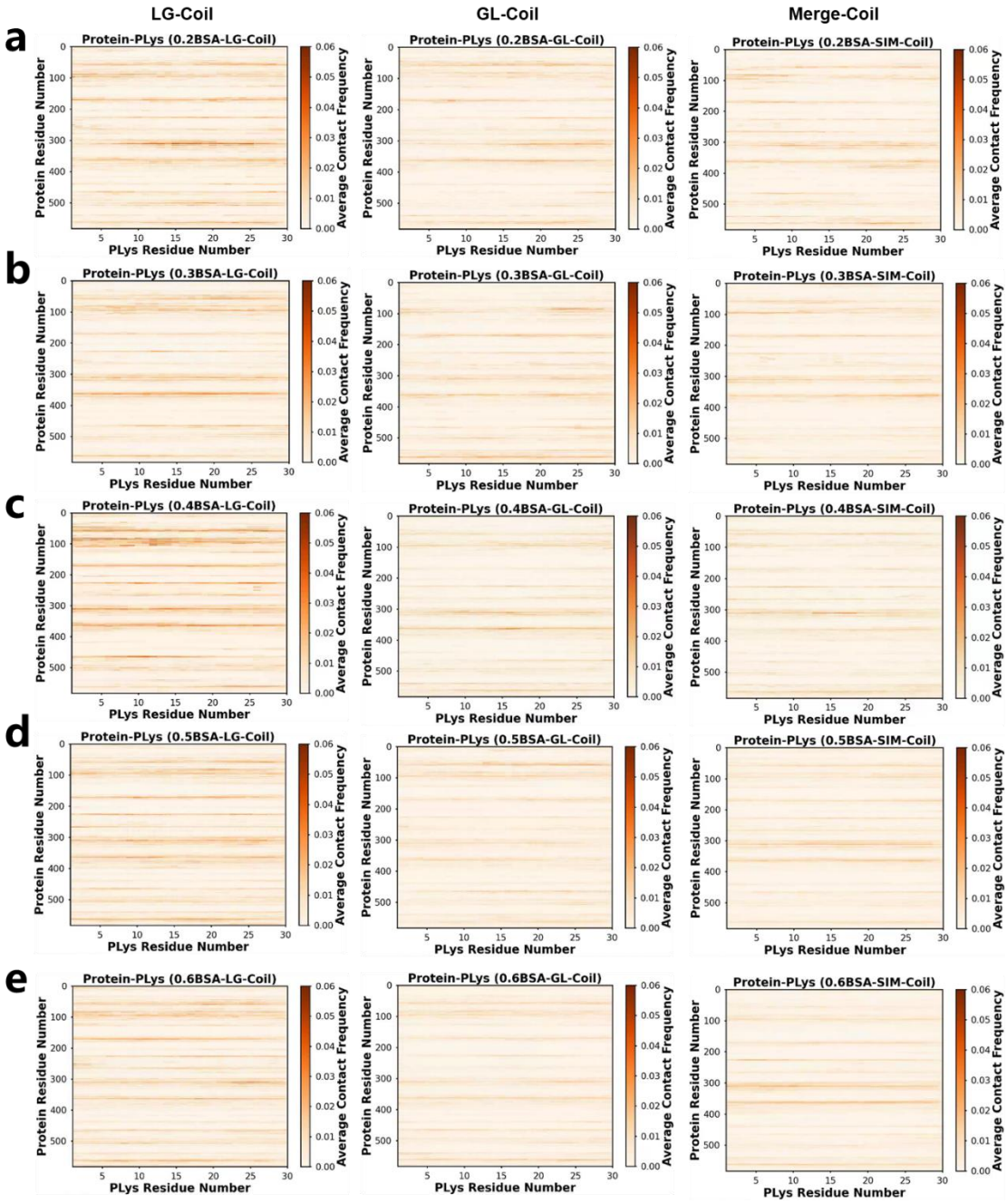

Fig. S5 Contact maps between BSA and PLys within coil systems. (a-e) illustrate contact maps between BSA protein and PLys in systems with different BSA ratio. And illustrations from left to right represent LG, GL and SIM systems, respectively.

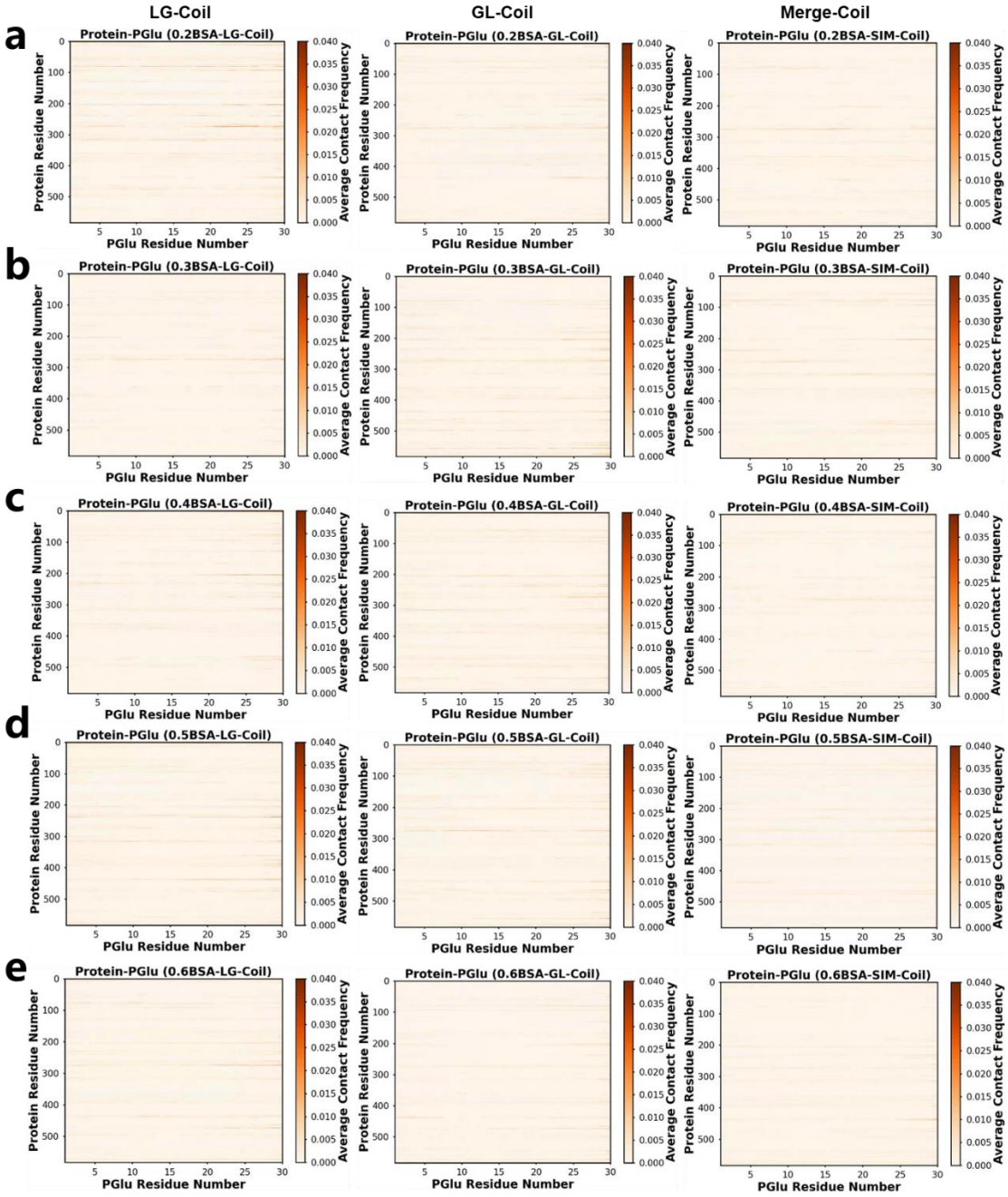

Fig. S6 Contact maps between BSA and PGlu within coil systems. (a-e) illustrate contact maps between BSA protein and PGlu in systems with different BSA ratio. And illustrations from left to right represent LG, GL and SIM systems, respectively.

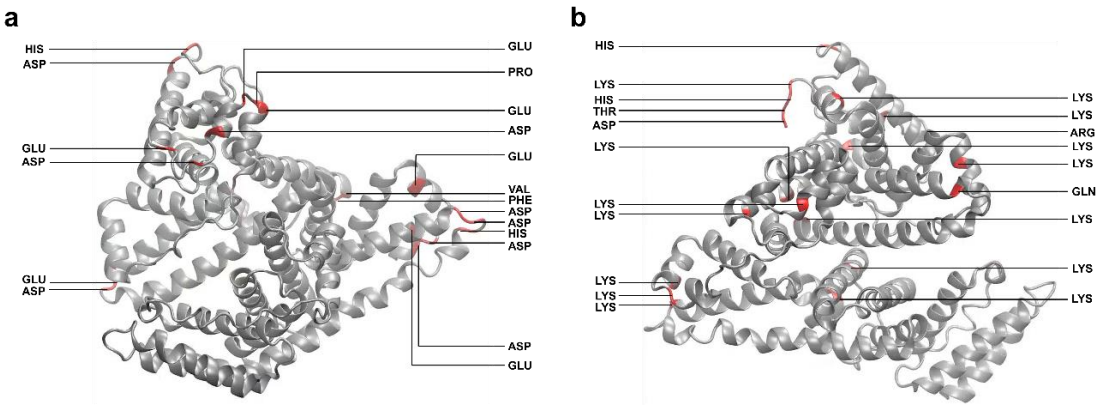

Fig. S7 The highlight of residues in BSA that have the highest contact frequency with poly(amino acid)s within coil systems. (a) The highlight of residues in BSA that have the highest contact frequency with PLys in the 0.4BSA-LG-Coil system. (b) The highlight of residues in BSA that have the highest contact frequency with PGlu in the 0.4BSA-LG-Coil system.

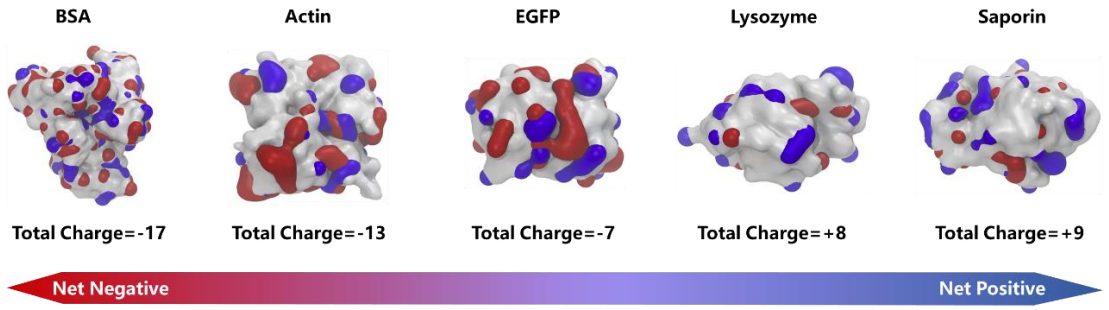

Fig. S8 The charge distribution of proteins along with patches of residues on the proteins surface. Structural rendering of the model proteins BSA, Actin, EGFP, Lysozyme and Saporin highlighting the location and distribution of charged residues. Negatively-charged glutamate and aspartate residues are shown in red. Positively-charged lysine, and arginine are shown in blue.

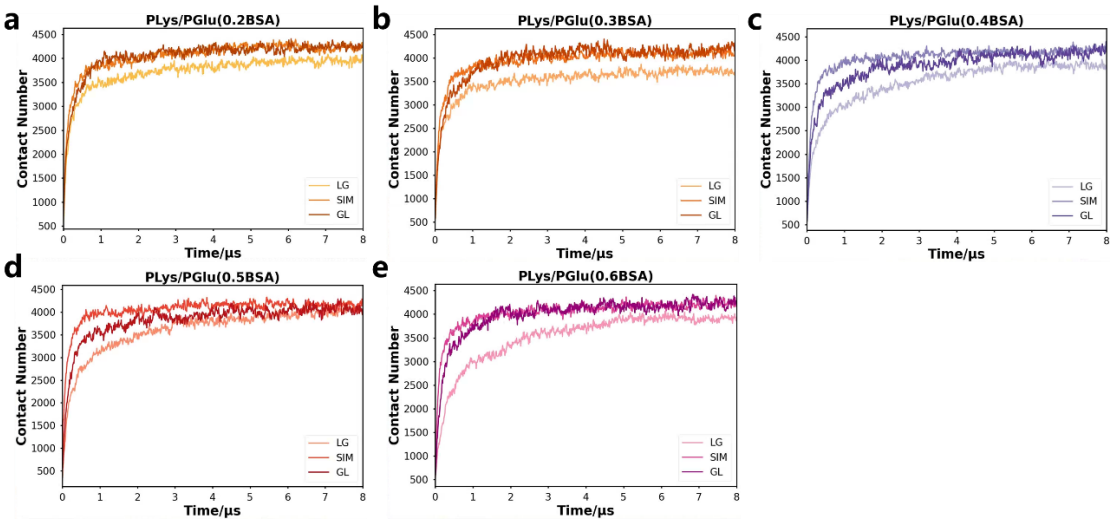

Fig. S9 Contact number between PLys and PGlu within coil systems. The contact number between PLys and PGlu over time under different BSA ratios (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA.

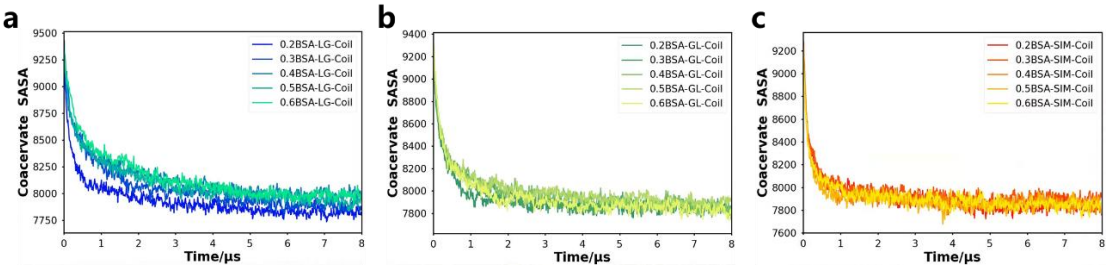

Fig. S10 SASA of coacervates within coil systems. The simulation time evolution of the SASA of coacervates in (a) LG systems, (b) GL systems, (c) SIM systems.

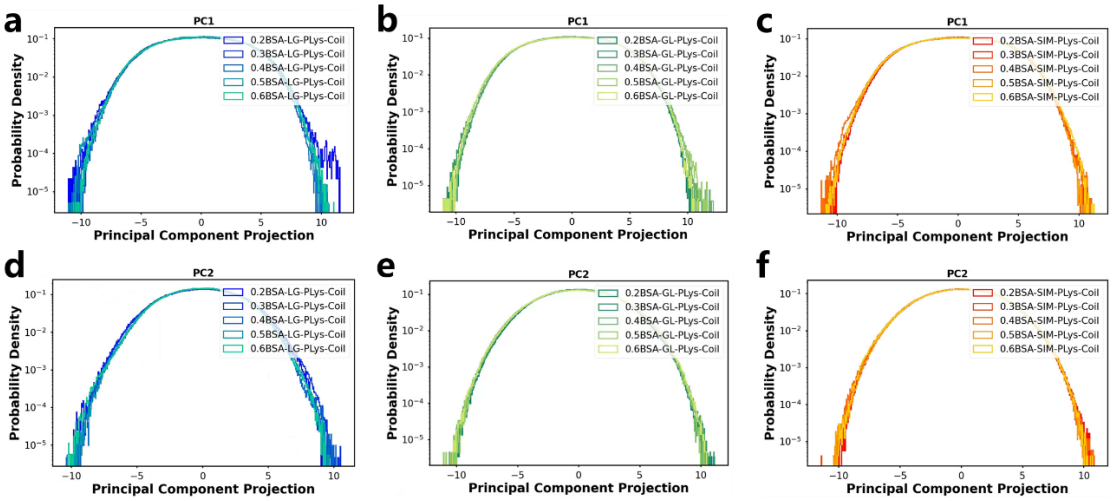

Fig. S11 The effect of BSA content on distributions of PC1 and PC2 projections for PLys within coil systems. (a) PC1 of PLys in LG systems, (b) PC1 of PLys in GL systems, (c) PC1 of PLys in SIM systems, (d) PC2 of PLys in LG systems, (e) PC2 of PLys in GL systems, f PC2 of PLys in SIM systems.

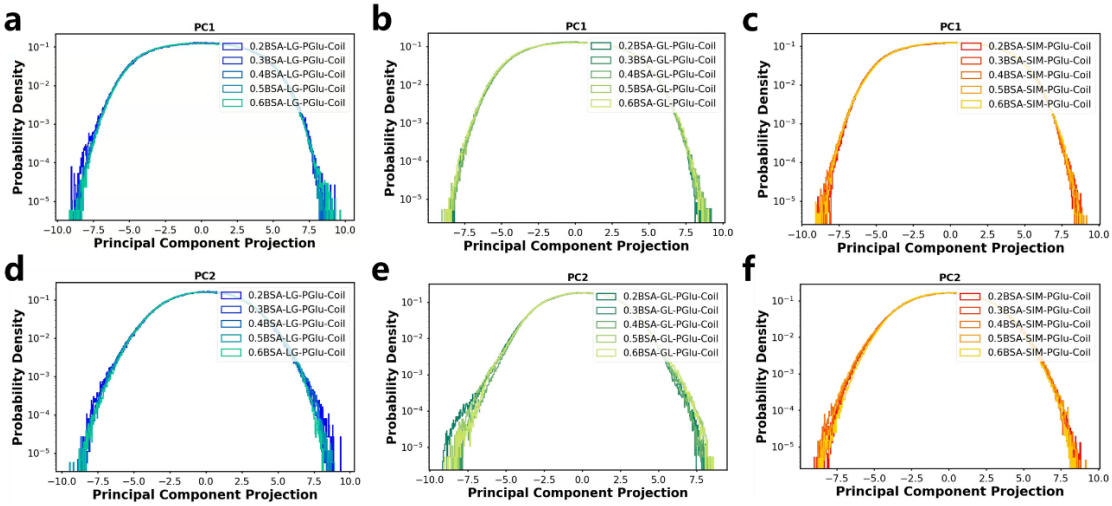

Fig. S12 The effect of BSA content on distributions of PC1 and PC2 projections for PGlu within coil systems. (a) PC1 of PGlu in LG systems, (b) PC1 of PGlu in GL systems, (c) PC1 of PGlu in SIM systems, (d) PC2 of PGlu in LG systems, (e) PC2 of PGlu in GL systems, (f) PC2 of PGlu in SIM systems.

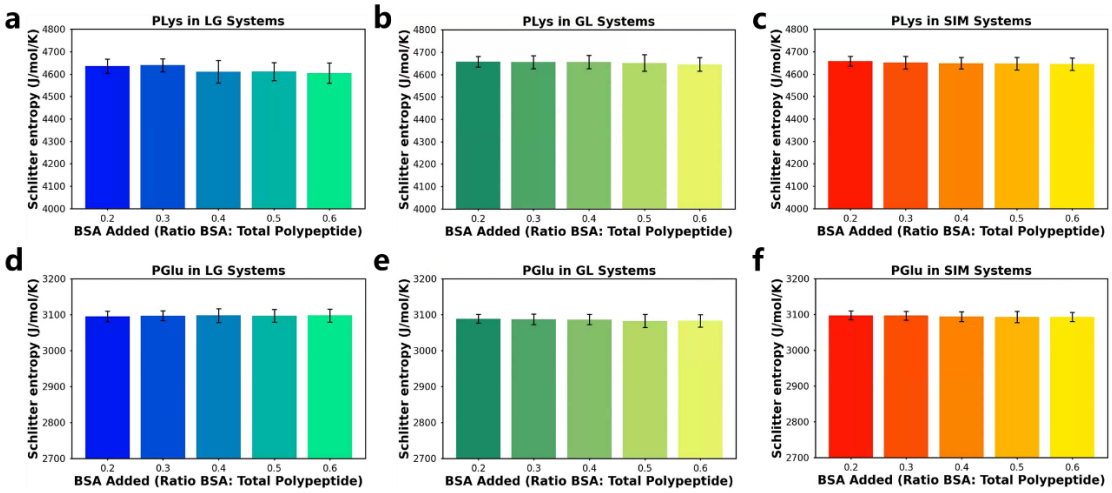

Fig. S13 The Schilitter entropy of PLys and PGlu changes with different BSA content within coil systems. The variation in the Schlitter entropy of PLys in (a) LG systems (b) GL systems (c) SIM systems. The variation in the Schlitter entropy of PGlu in d LG systems e GL systems f SIM systems.

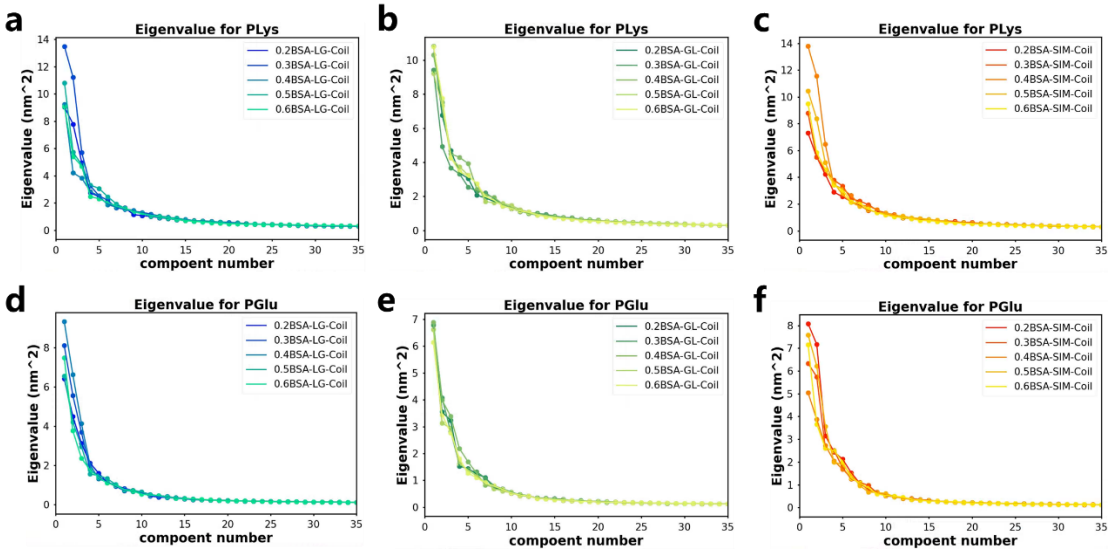

Fig. S14 Eigenvalues of the covariance matrix within coil systems. The eigenvalues of the covariance matrix of (a) PLys in LG systems, (b) PLys in GL systems, (c) PLys in SIM systems, (d) PGlu in LG systems, (e) PGlu in GL systems, (f) PGlu in SIM systems.

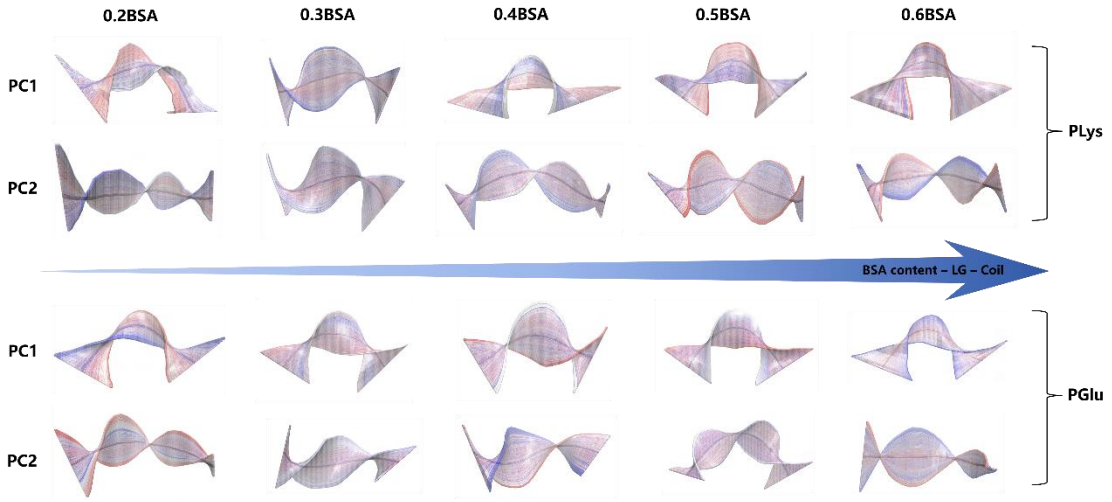

Fig. S15 The PC1 and PC2 motions of PLys and PGlu within LG coil systems. The PC1 and PC2 motions of PLys and PGlu change from blue to white to red over time, with the mean position represented in gray.

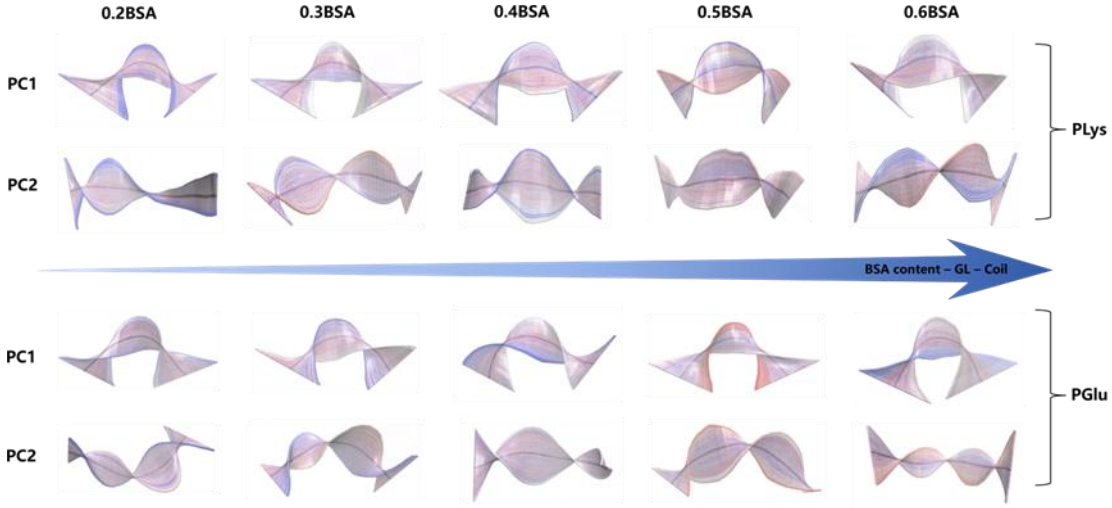

Fig. S16 The PC1 and PC2 motions of PLys and PGlu within GL coil systems. The PC1 and PC2 motions of PLys and PGlu change from blue to white to red over time, with the mean position represented in gray.

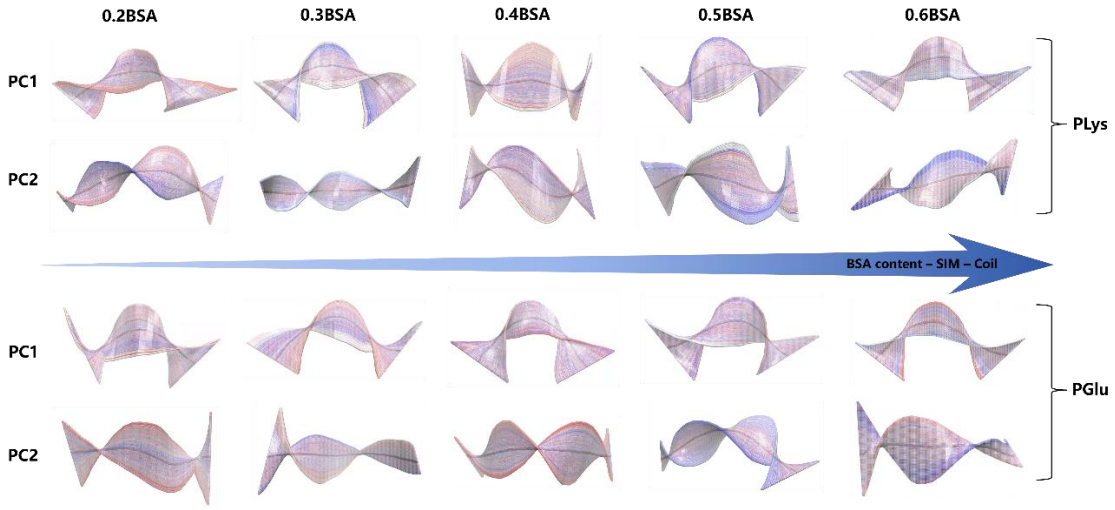

Fig. S17 The PC1 and PC2 motions of PLys and PGlu within SIM coil systems. The PC1 and PC2 motions of Plys and PGlu changes from blue to white to red over time, with the mean position represented in gray.

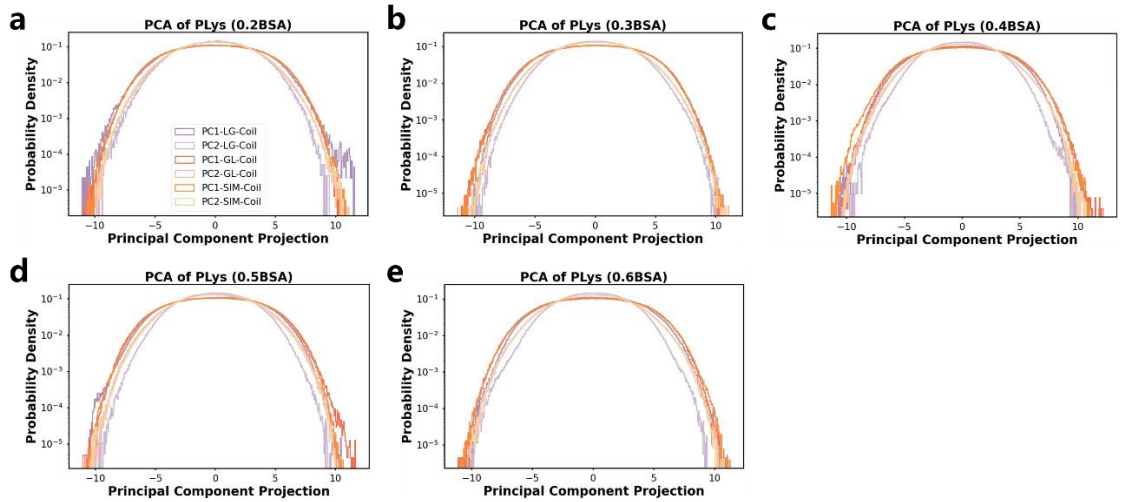

Fig. S18 The principal component analysis of PLys within coil systems. The PC1 and PC2 corresponding principal component projection values of PLys in various systems with different ingredient adding sequences on coacervate formation and different BSA content: (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA.

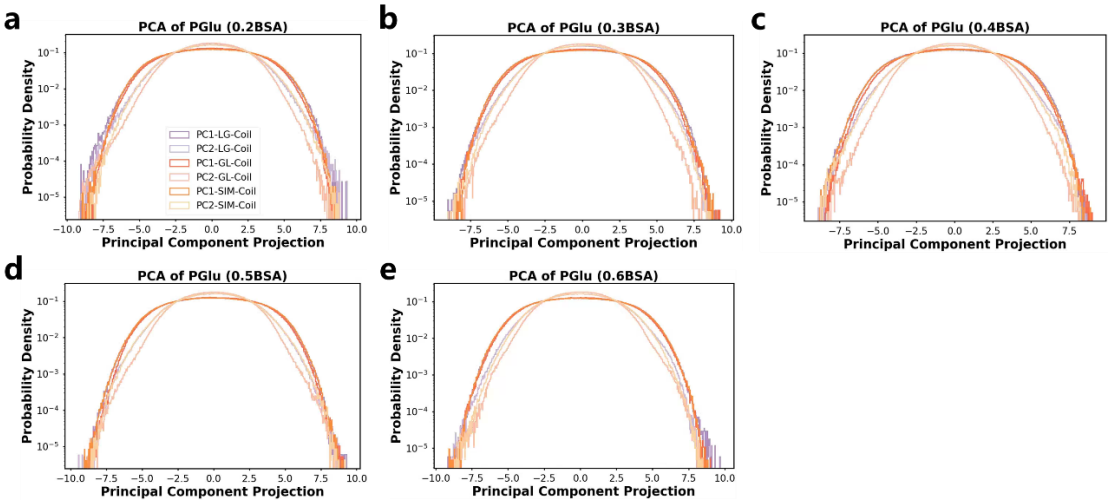

Fig. S19 The principal component analysis of PGlu within coil systems. The PC1 and PC2 corresponding principal component projection values of PGlu in various systems with different ingredient adding sequences on coacervate formation and different BSA content: (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA.

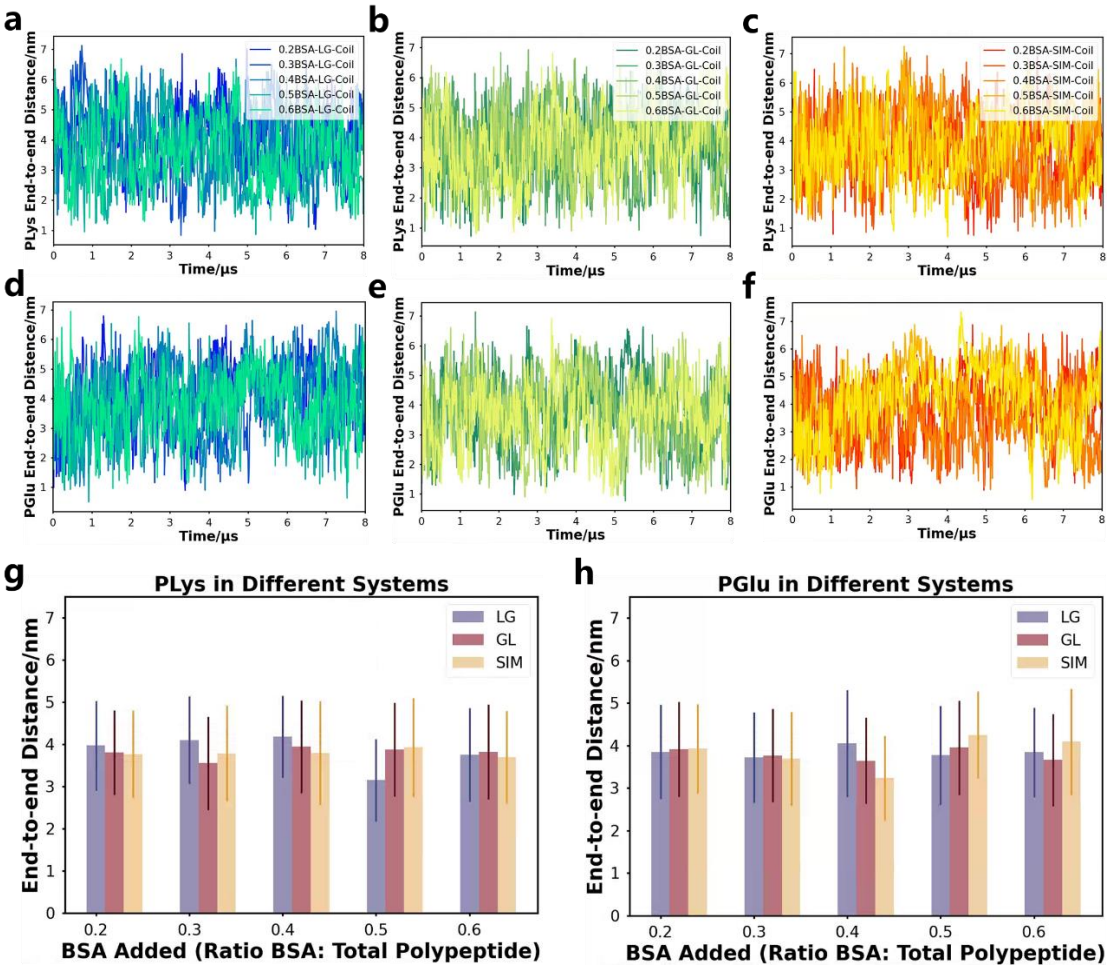

Fig. S20 End-to-end distances of PLys and PGlu within coil systems. (a) and (d) show the end-to-end distance of PLys and PGlu chains as functions of time in LG systems respectively. (b) and (e) show the end-to-end distance of PLys and PGlu chains as functions of time in GL systems respectively. (c) and (f) show the end-to-end distance of PLys and PGlu chains as functions of time in SIM systems respectively.

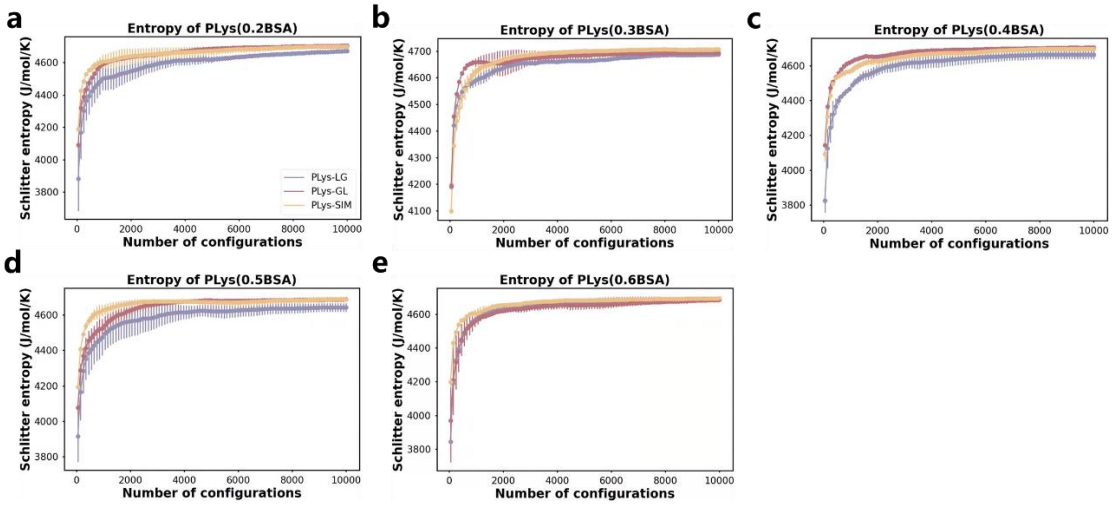

Fig. S21 The Schilitter entropy of PLys changes within coil systems. The variation in the Schlitter entropy of PLys, across different ingredient adding sequence on coacervate formation under varying BSA content: (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA.

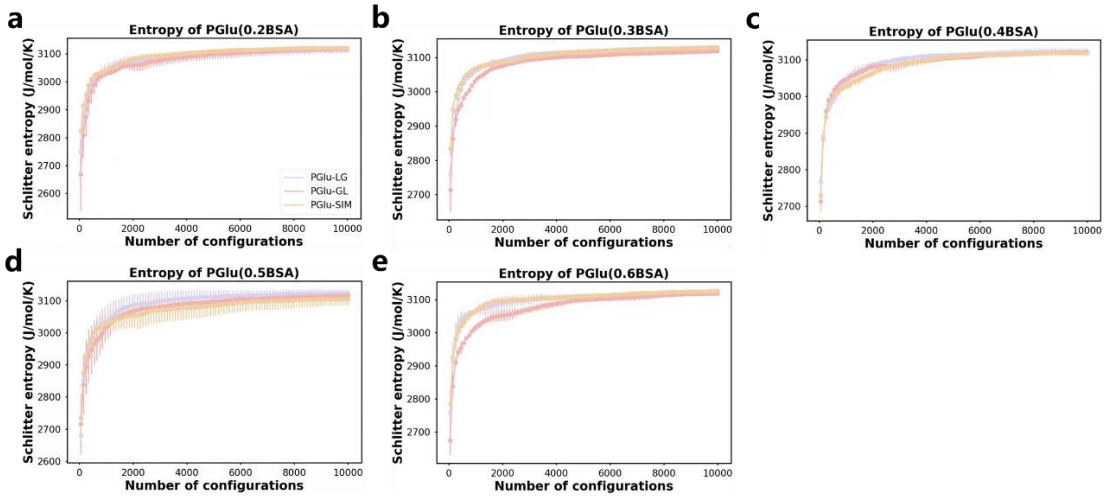

Fig. S22 The Schilitter entropy of PGlu changes within coil systems. The variation in the Schlitter entropy of PGlu, across different ingredient adding sequence on coacervate formation under varying BSA content: (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA.

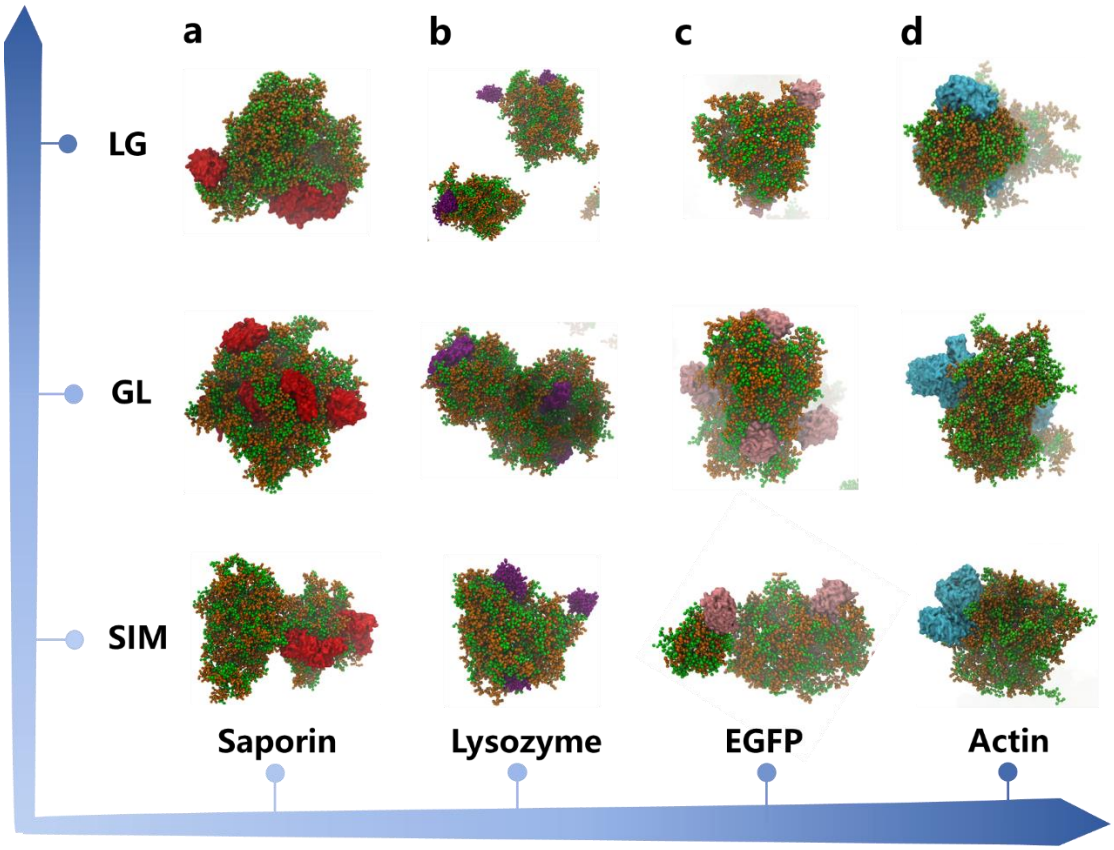

Fig. S23 Snapshots of simulations of different protein systems. The snapshots of the simulations in (a) Saporin systems, (b) Lysozyme systems, (c) EGFP systems, (d) Actin systems. Orange represents PLys, green represents PGlu, red represents Saporin, purple represents Lysozyme, pink represents EGFP, and light blue represents Actin. Water and ions are not shown for clarity.

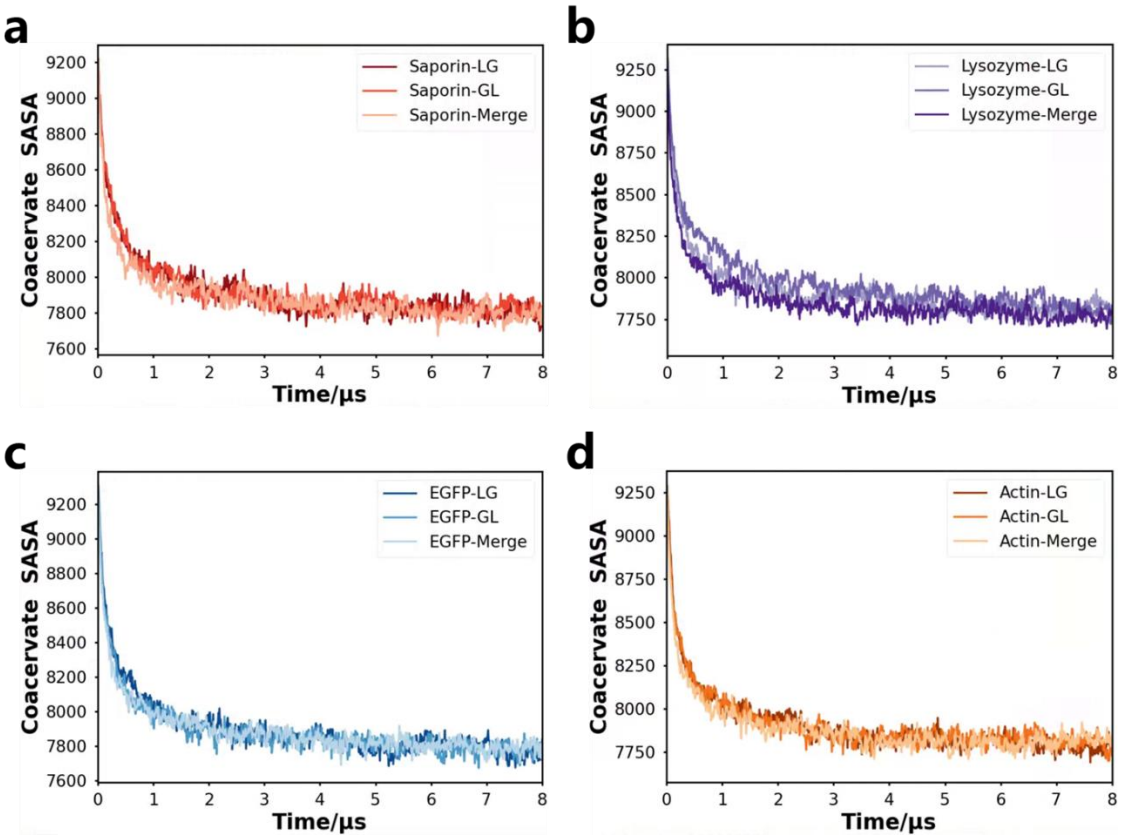

Fig. S24 SASA of coacervates in systems with different type of protein. The simulation time evolution of the SASA of coacervates in (a) Saporin systems, (b) Lysozyme systems, (c) EGFP systems, (d) Actin systems.

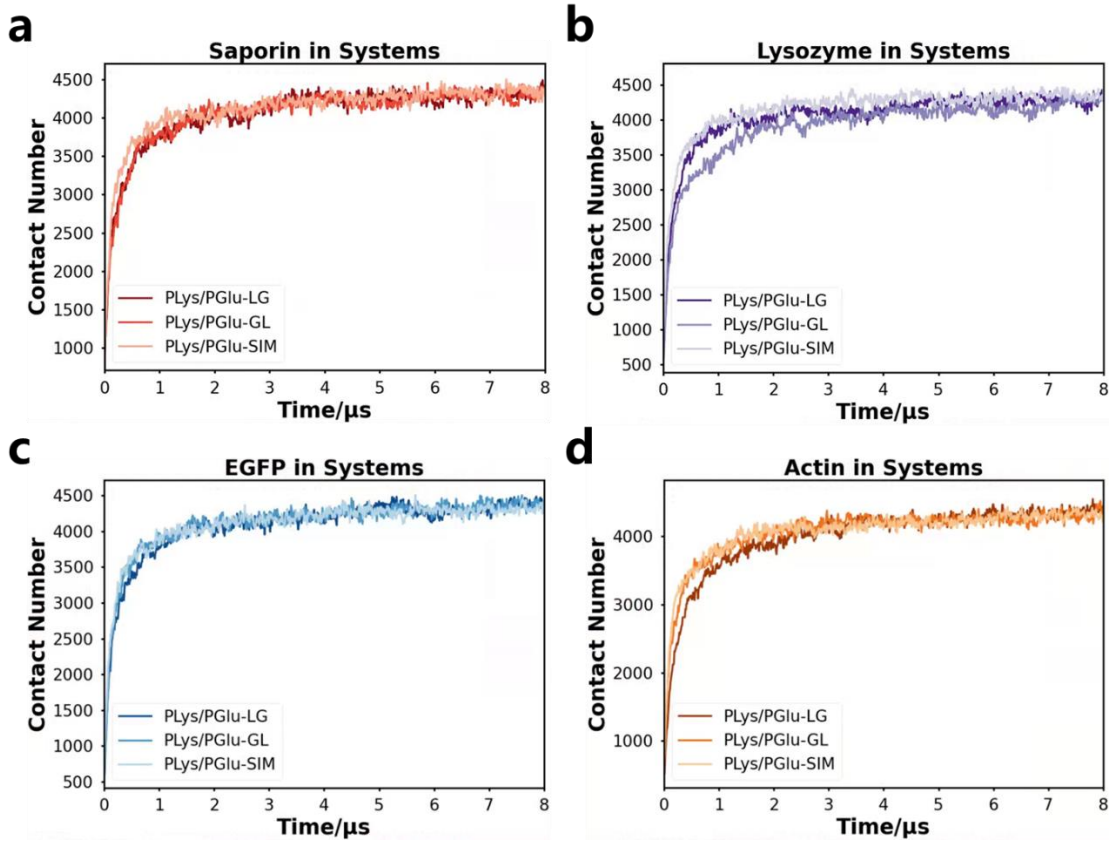

Fig. S25 Contact numbers between PLys and PGlu in systems with different type of protein. The contact number between PLys and PGlu over time under (a) Saporin systems, (b) Lysozyme systems, (c) EGFP systems, (d) Actin systems.

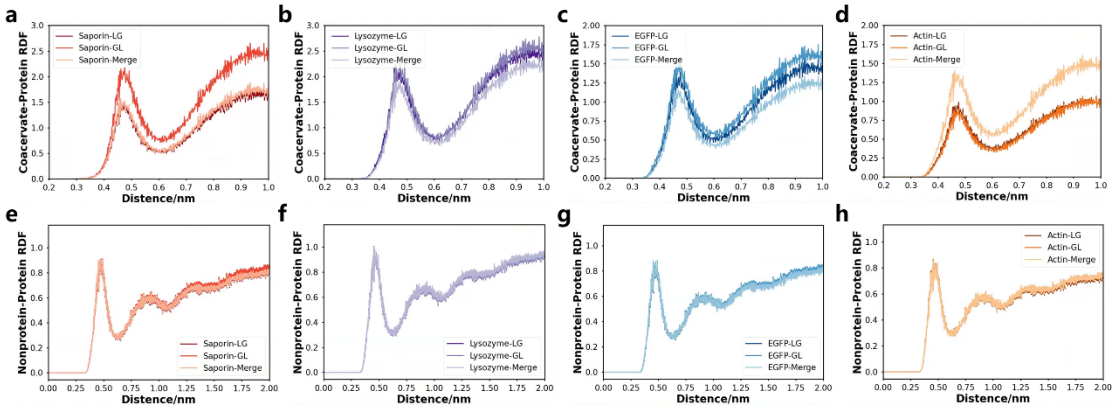

Fig. S26 RDF analysis between various components in systems with different type of protein. The RDF between coacervates and protein in (a) Saporin systems, (b) Lysozyme systems, (c) EGFP systems, (d) Actin systems. The RDF between nonprotein and protein in (e) Saporin systems, (f) Lysozyme systems, (g) EGFP systems. (h) Actin systems.

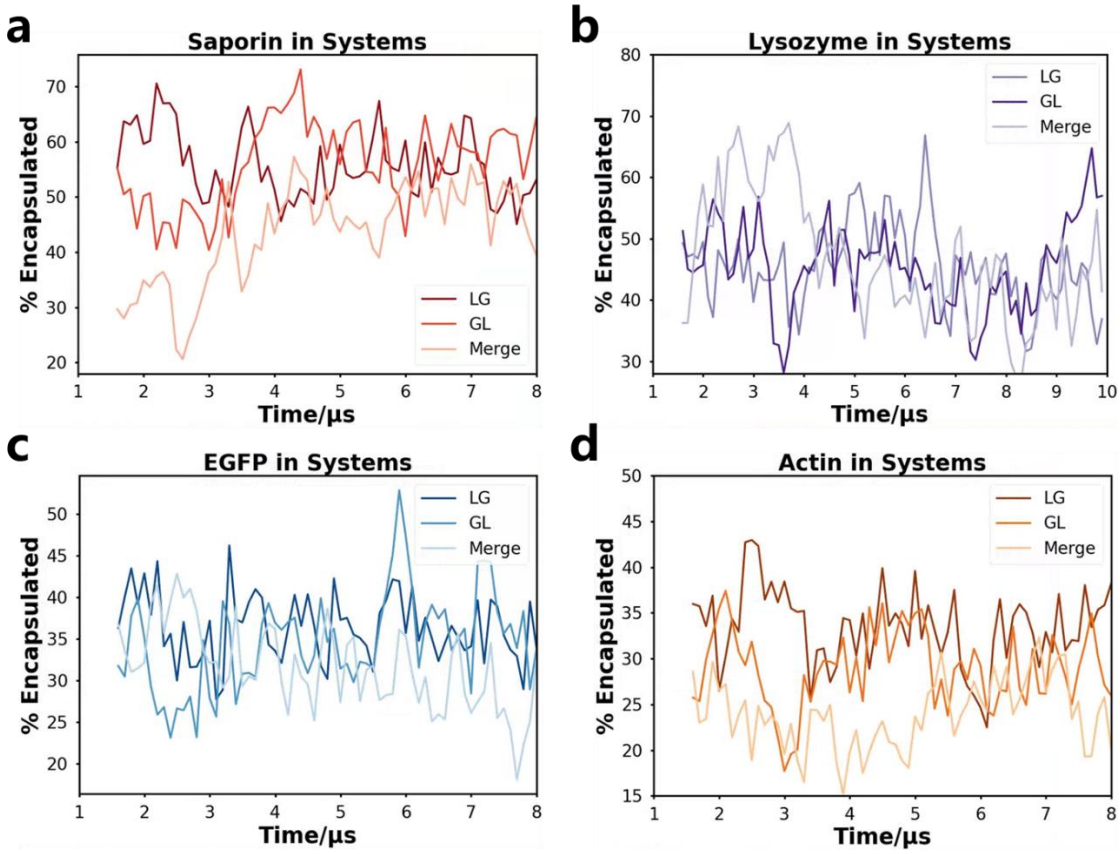

Fig. S27 The encapsulation percentage of coacervates in systems with different type of protein. (a) Saporin, (b) Lysozyme, (c) EGFP, (d) Actin.

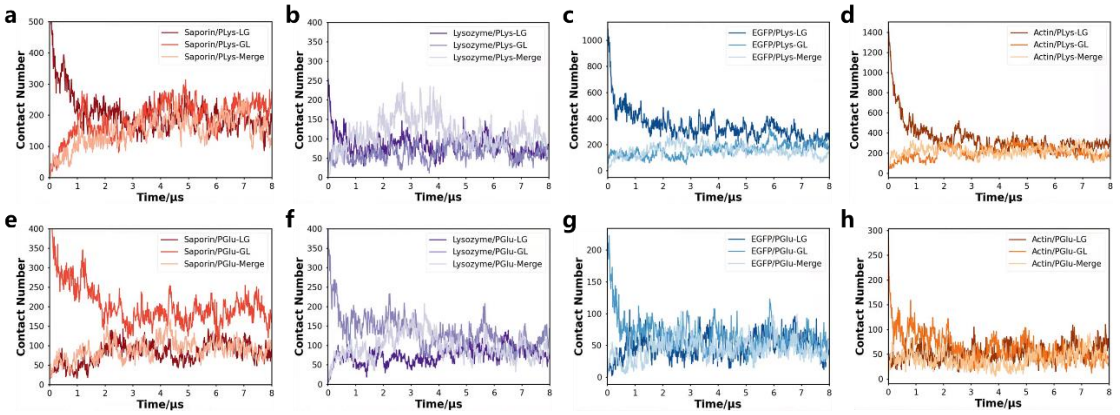

Fig. S28 Contact numbers of Protein/PLys and Protein/PGlu in systems with different type of protein. The contact number between protein and PLys over time under (a) Saporin systems, (b) Lysozyme systems, (c) EGFP systems, (d) Actin systems. The contact number between protein and PGlu over time under (e) Saporin systems, (f) Lysozyme systems, (g) EGFP systems, (h) Actin systems.

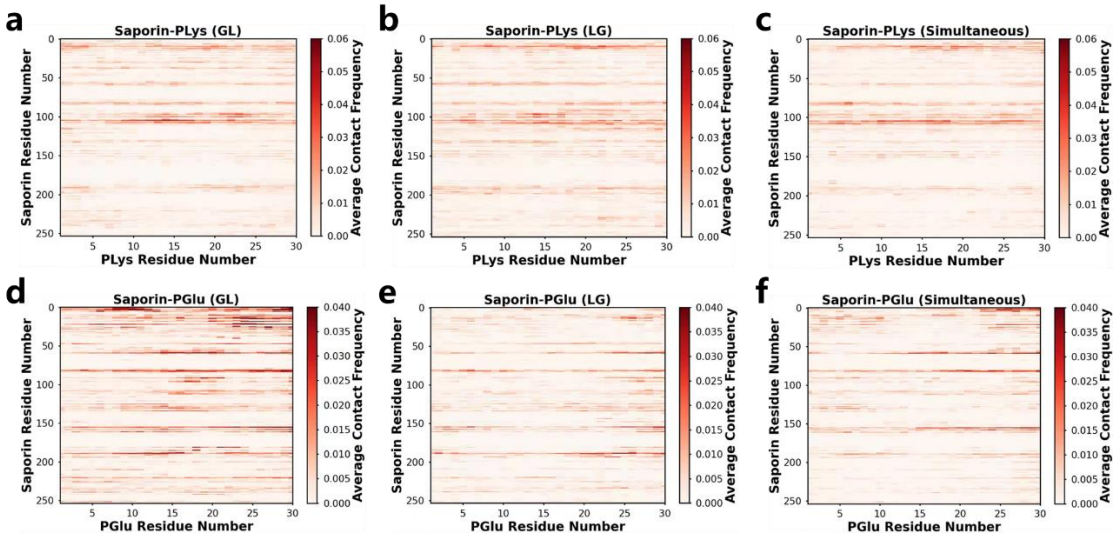

Fig. S29 Contact maps between Saporin and Poly(amino acid) in systems. Contact map between Saporin and PLys in (a) GL, (b) LG, (c) SIM systems. Contact map between Saporin and PGlu in (d) GL, (e) LG, (f) SIM systems.

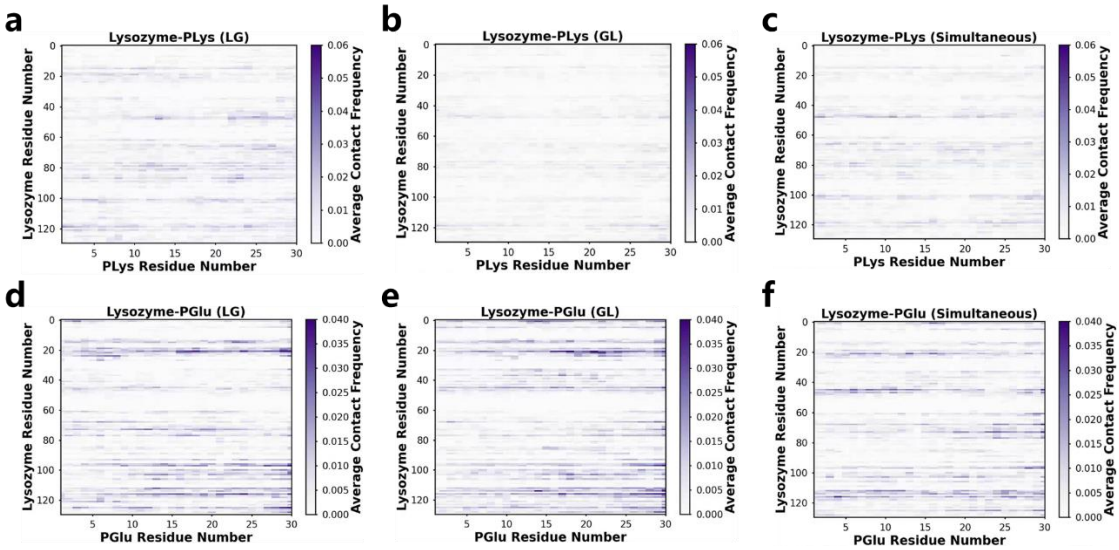

Fig. S30 Contact maps between Lysozyme and Poly(amino acid) in systems. Contact map between Lysozyme and PLys in (a) GL, (b) LG, (c) SIM systems. Contact map between Lysozyme and PGlu in (d) GL, (e) LG, (f) SIM systems.

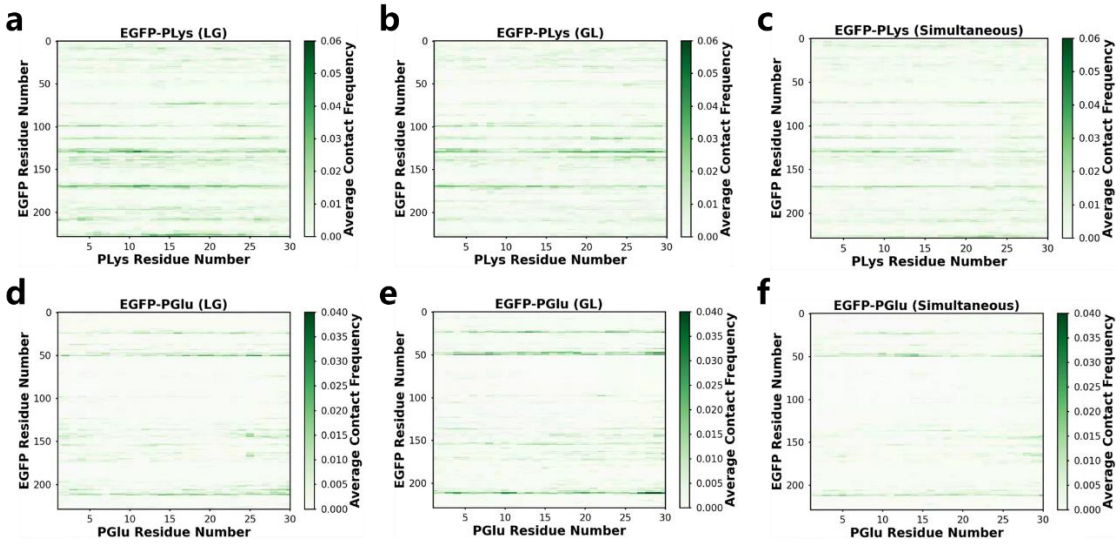

Fig. S31 Contact maps between EGFP and Poly(amino acid) in systems. Contact map between EGFP and PLys in (a) GL, (b) LG, (c) SIM systems. Contact map between EGFP and PGlu in (d) GL, (e) LG, (f) SIM systems.

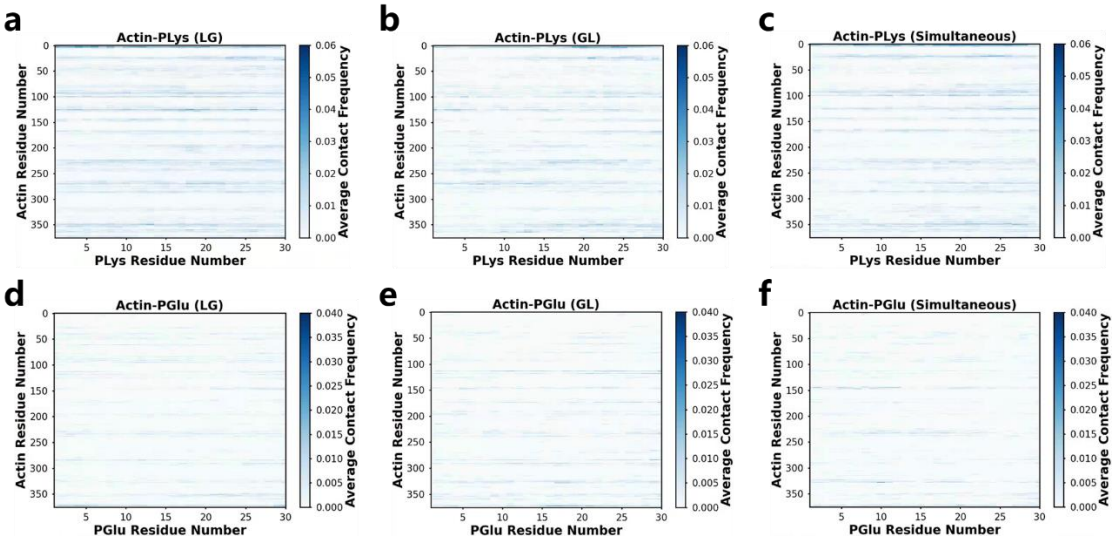

Fig. S32 Contact maps between Actin and Poly(amino acid) in systems. Contact map between Actin and Plys in (a) GL, (b) LG, (c) SIM systems. Contact map between Actin and Pglu in (d) GL, (e) LG, (f) SIM systems.

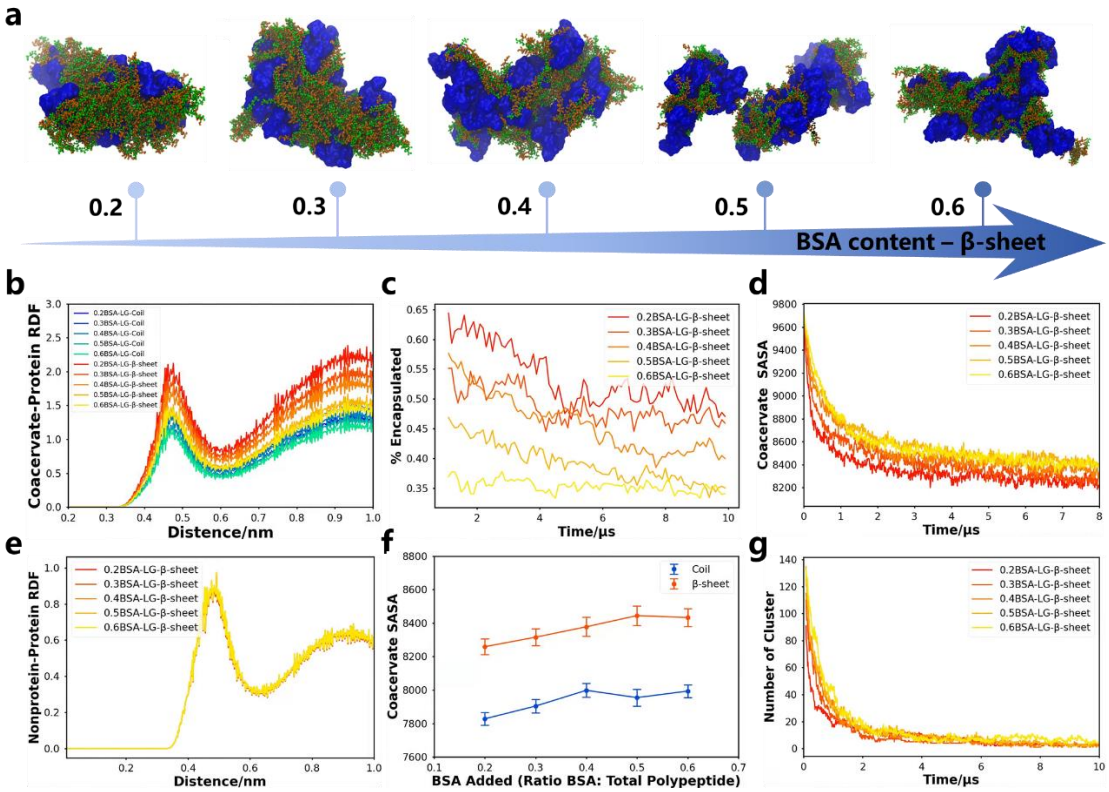

Fig. S33 Secondary structure's effects on coacervate configuration. (a) Snapshots of the final states of β-sheet systems at different BSA ratios. (b) The RDF between coacervates and BSA protein. (c) The encapsulation percentage of BSA protein within coacervate as functions of time. (d) The simulation time evolution of the solvent-accessible surface area (SASA) of coacervates. (e) The RDF between BSA protein and other molecules. (f) The SASA of coacervates examined for random coil and β-sheet systems, with different ratios of BSA content. (g) Cluster number of coacervates as functions of time.

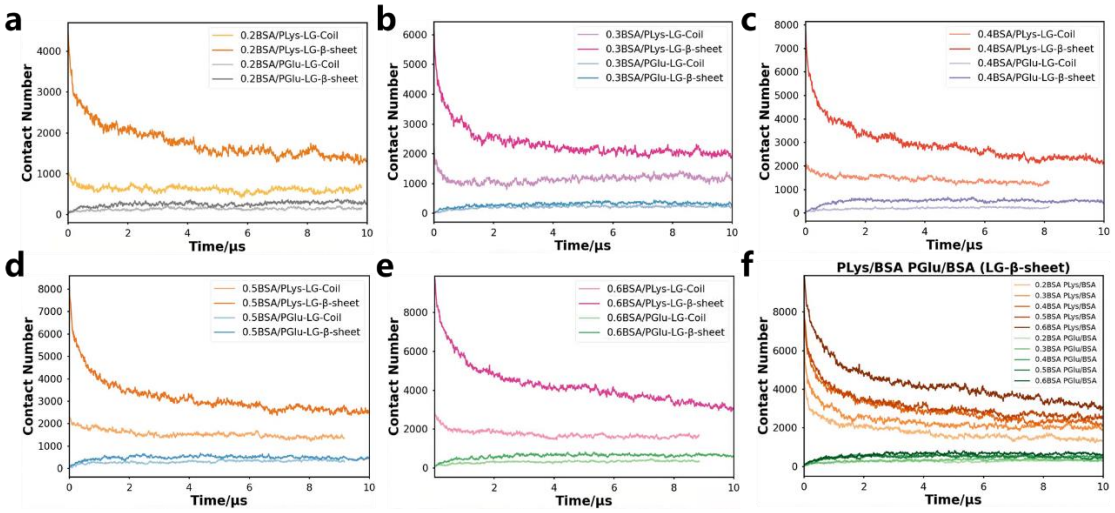

Fig. S34 Contact numbers of PLys/BSA and PGlu/BSA within coil and β-sheet systems. The contact number between different components over time under different BSA ratios (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA. (f) The evolution of the contact number of PLys/BSA and PGlu/BSA over time is analyzed and compared in β-sheet secondary structure systems.

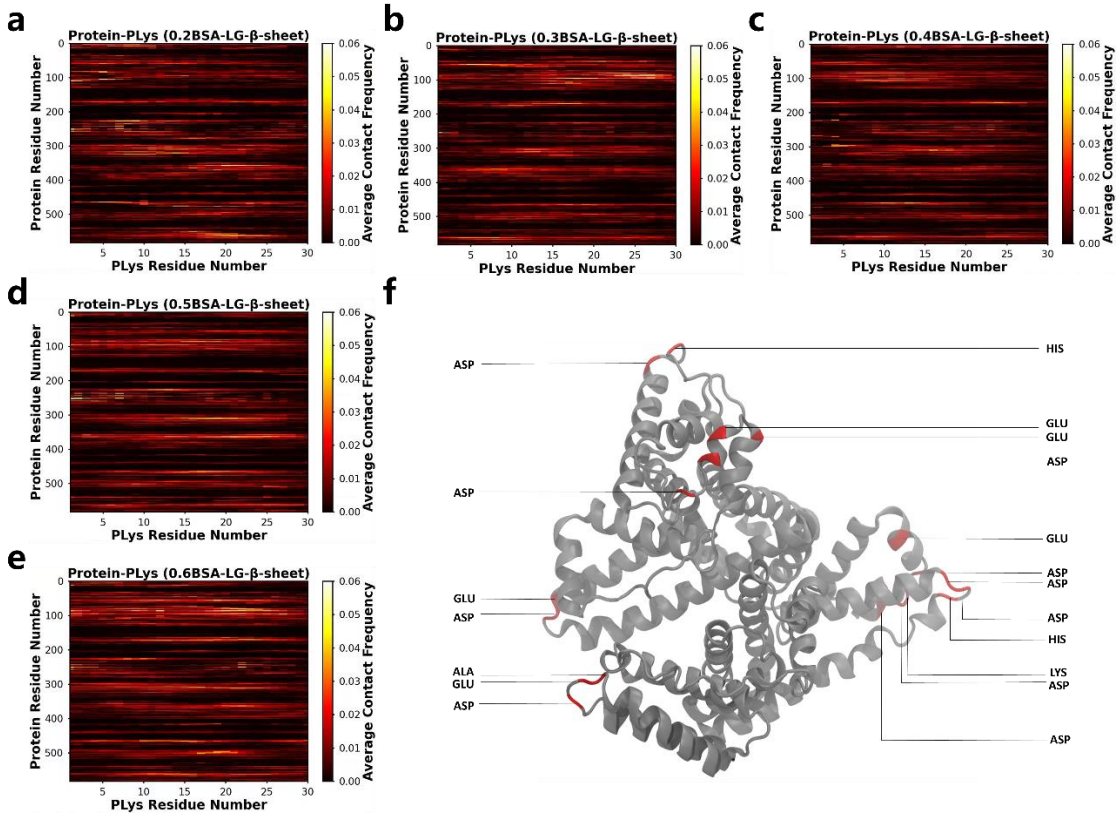

Fig. S35 Contact maps between BSA and PLys within β-sheet systems. (a-e) illustrate contact maps between BSA protein and PLys in systems with different BSA ratio. (f) Residues in BSA with top contact frequency, exemplified by the 0.2BSA-LG-β-sheet system.

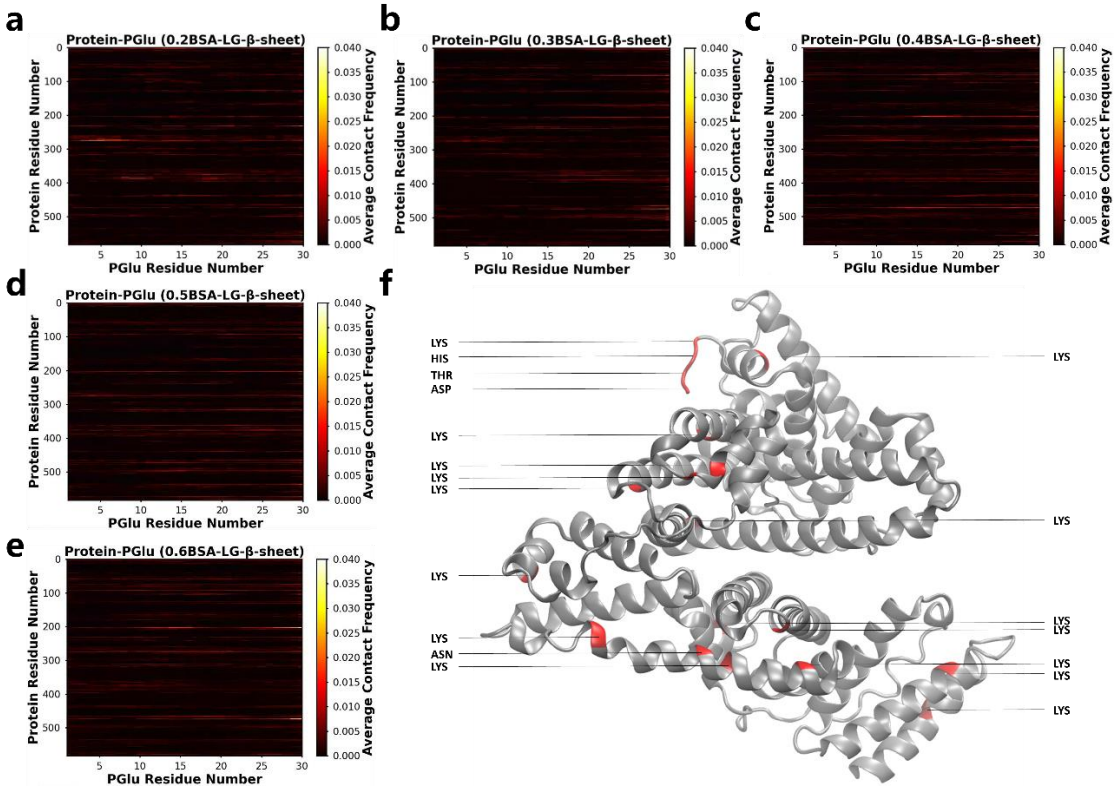

Fig. S36 Contact maps between BSA and PGlu within β-sheet systems. (a-e) illustrate contact maps between BSA protein and PGlu in systems with different BSA ratio. (f) Residues in BSA with top contact frequency, exemplified by the 0.2BSA-LG-β-sheet system.

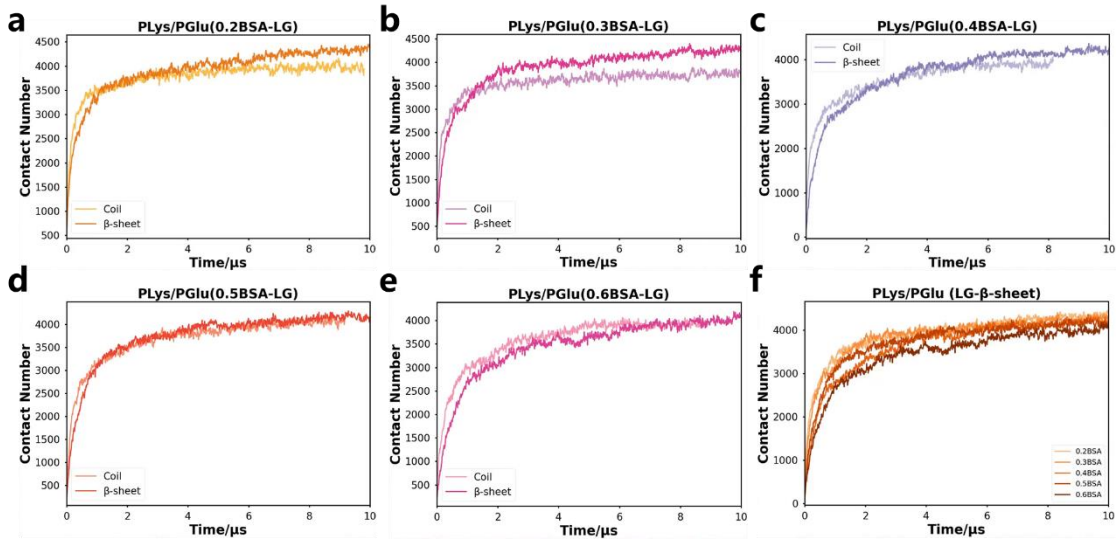

Fig. S37 Contact numbers between PLys and PGlu within coil and β-sheet systems. The contact number between different components over time under different BSA ratios (a) 0.2BSA, (b) 0.3BSA, (c) 0.4BSA, (d) 0.5BSA, (e) 0.6BSA. (f) The evolution of the contact number between PLys and PGlu over time is analyzed and compared in β-sheet secondary structure systems.

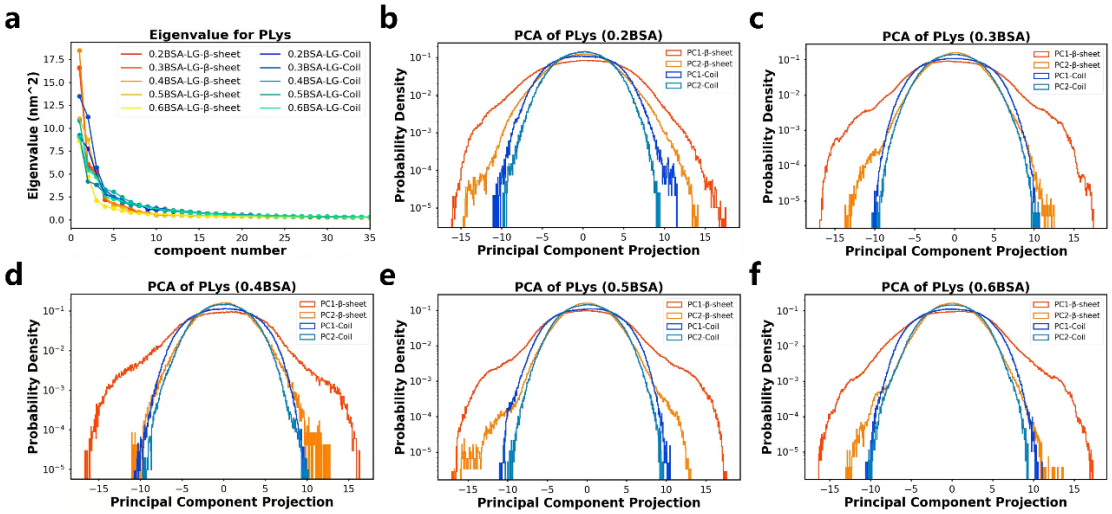

Fig. S38 Principal component analysis of the PLys within coil and β-sheet systems. (a) shows the eigenvalues of the covariance matrix in different systems. The PC1 and PC2 corresponding principal component projection values of PLys in various secondary structural systems under different BSA content (b) 0.2BSA (c) 0.3BSA (d) 0.4BSA (e) 0.5BSA (f) 0.6BSA.

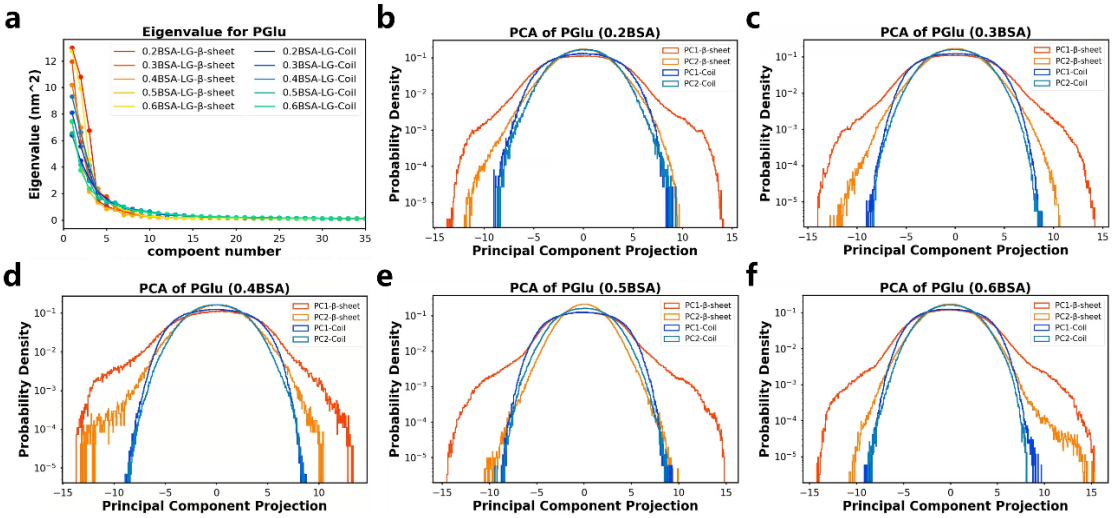

Fig. S39 Principal component analysis of the PGlu within coil and β-sheet systems. (a) shows the eigenvalues of the covariance matrix in different systems. The PC1 and PC2 corresponding principal component projection values of PGlu in various secondary structural systems under different BSA content (b) 0.2BSA (c) 0.3BSA (d) 0.4BSA (e) 0.5BSA (f) 0.6BSA.

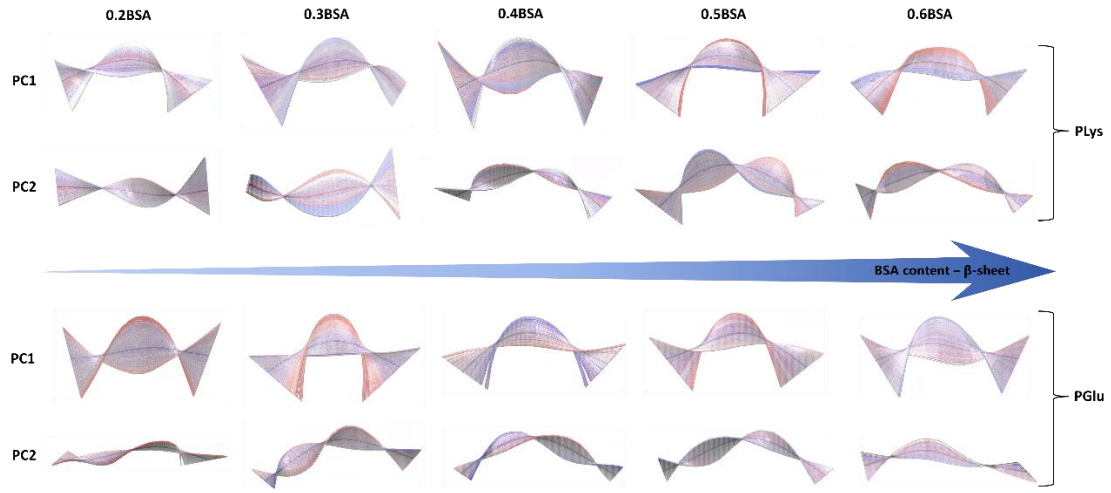

Fig. S40 The PC1 and PC2 motions of PLys and PGlu within β-sheet systems. The PC1 and PC2 motions of PLys and PGlu changes from blue to white to red over time, with the mean position represented in gray.

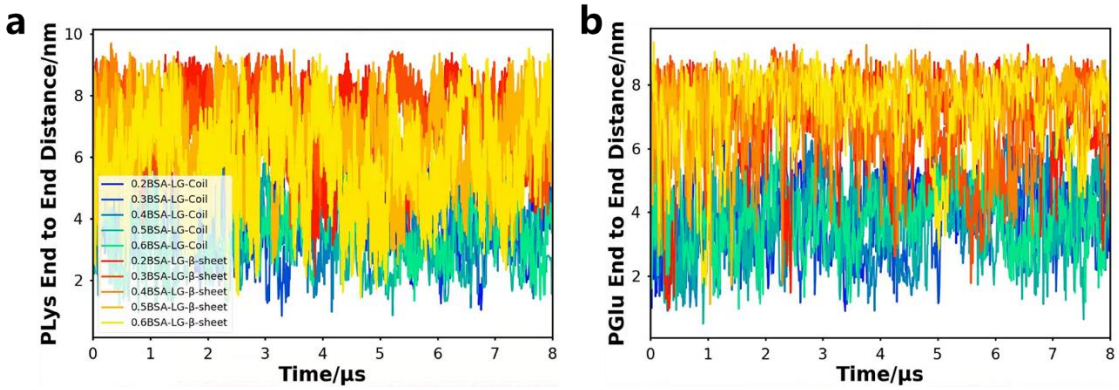

Fig. S41 End-to-end distances of PLys and PGlu within coil and β-sheet systems. (a) and (b) show the end-to-end distance of PLys and PGlu chains as functions of time.

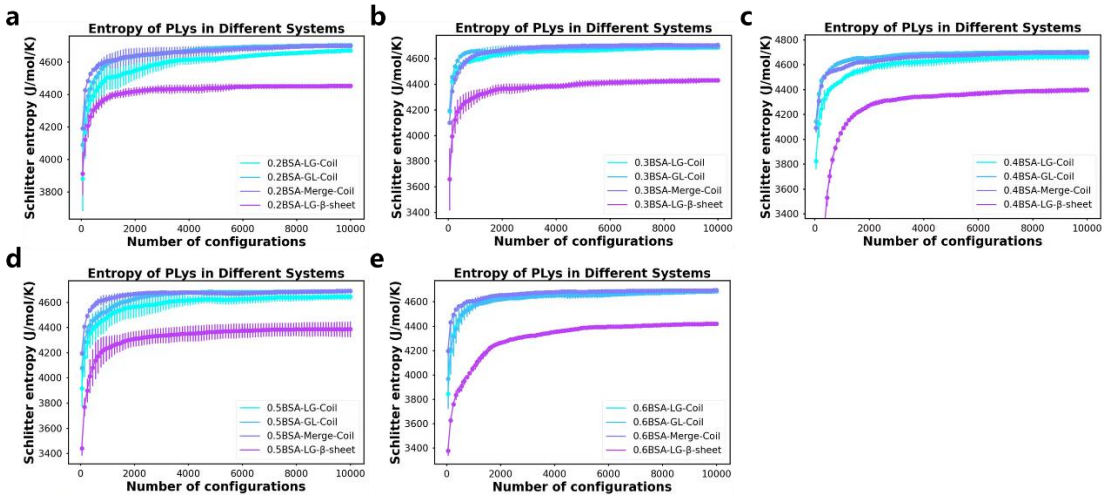

Fig. S42 The Schilitter entropy of PLys changes within coil and β-sheet systems. The variation in the Schlitter entropy of PLys, across different secondary structures and addition sequences under varying BSA content: (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA.

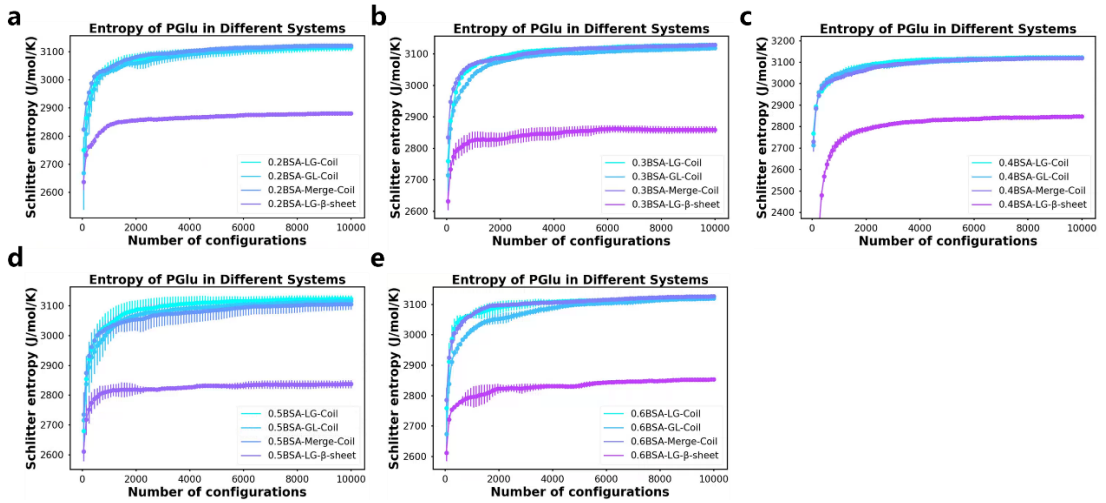

Fig. S43 The Schilitter entropy of PGlu changes within coil and β-sheet systems. The variation in the Schlitter entropy of PGlu, across different secondary structures and addition sequences under varying BSA content: (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA.

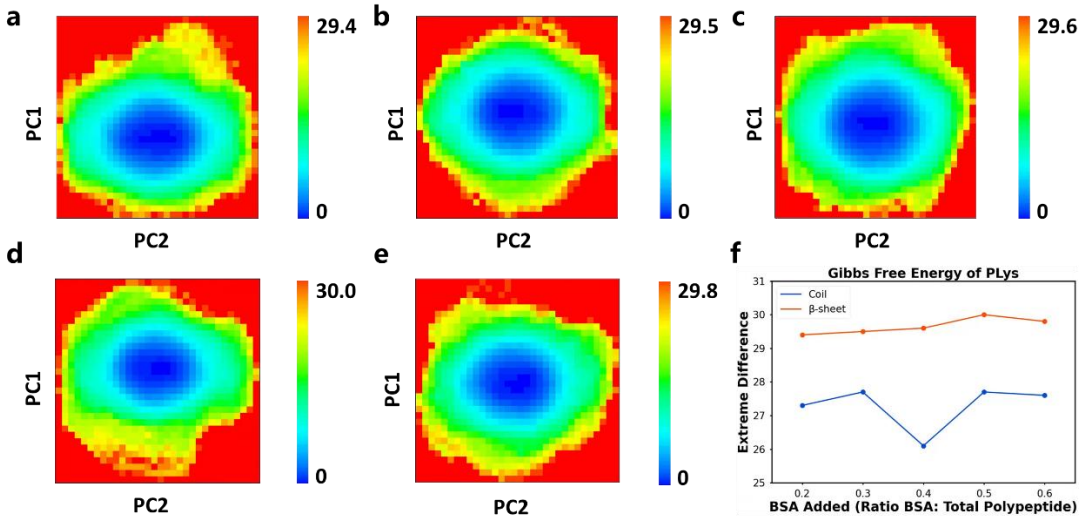

Fig. S44 Gibbs free energy of PLys within β-sheet systems. Gibbs free energy 2D landscapes of PLys under different BSA content (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA. (f) The difference in the Gibbs free energy extreme of PLys for random coil and β-sheet systems.

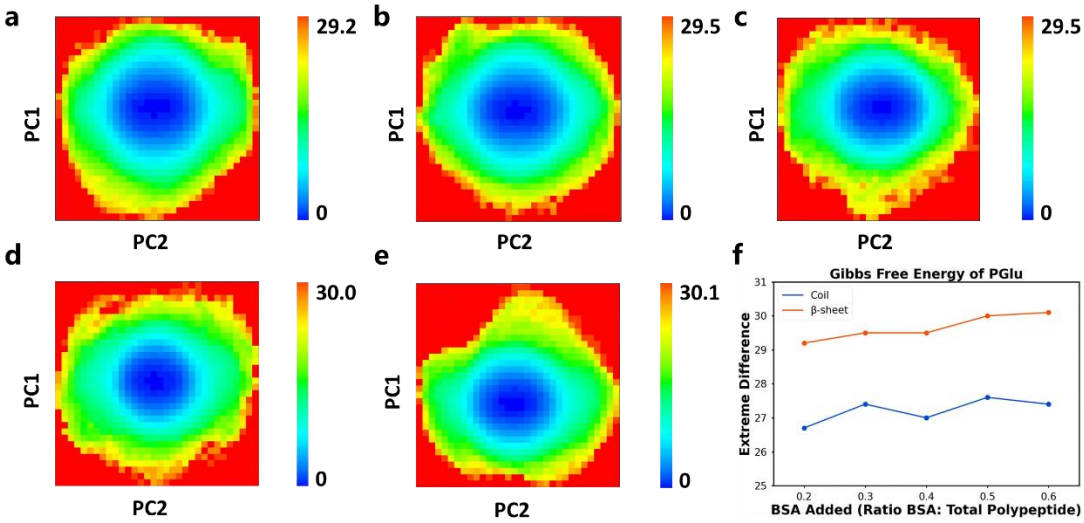

Fig. S45 Gibbs free energy of PGlu within β-sheet systems. Gibbs free energy 2D landscapes of PGlu under different BSA content (a) 0.2BSA (b) 0.3BSA (c) 0.4BSA (d) 0.5BSA (e) 0.6BSA. (f) The difference in the Gibbs free energy extreme of PGlu in random coil and β-sheet systems.

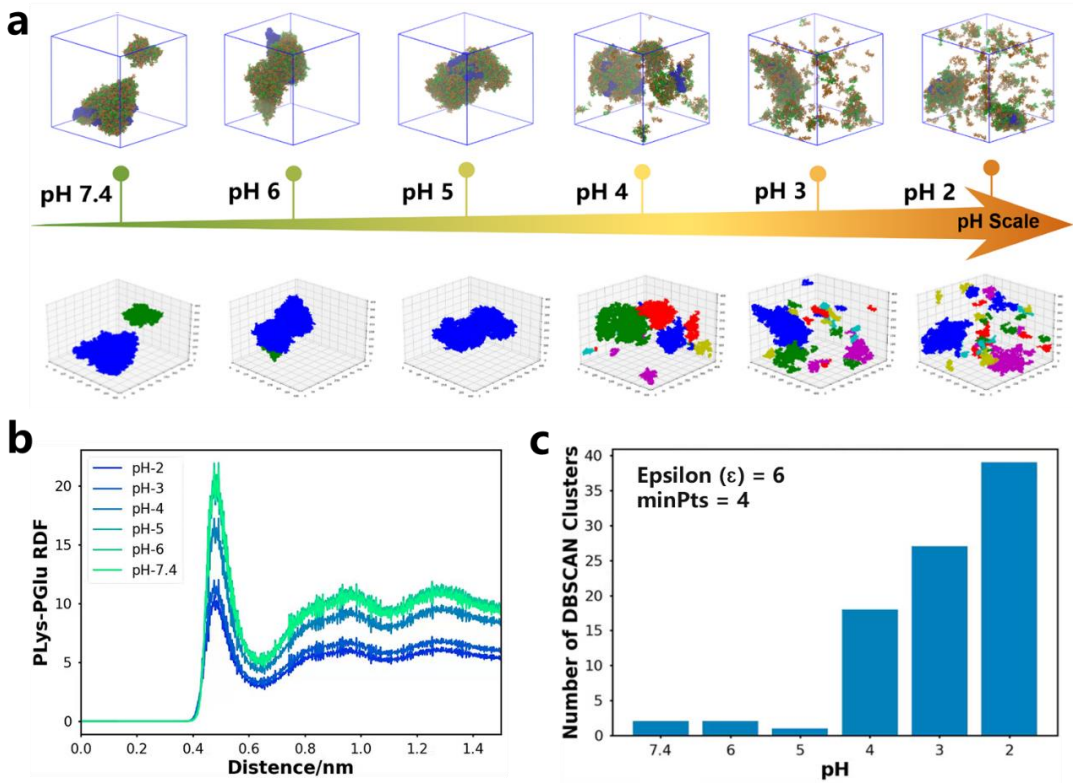

Fig. S46 DBSCAN analysis at different pH values. (a) Up: Snapshots of the final state of 0.05BSA systems after about 10 μs of CG MD simulations at different pH values. Down: The DBSCAN algorithm clustering models. (b) The radial distribution function between PLys and PGlu. c The number of clusters obtained from DBSCAN algorithm clustering analysis at different pH values.

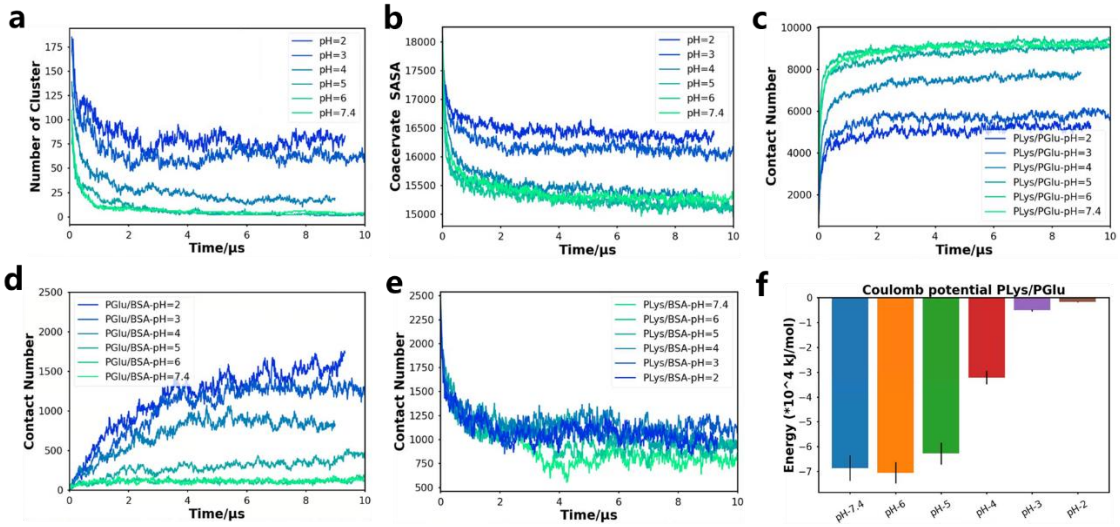

Fig. S47 Quantification of coacervate phase separation over time and the coulomb potential at different pH values. (a) Number of the coacervate clusters. (b) The SASA of the coacervates. (c) The contact number between PLys and PGlu. (d) The contact number between PGlu and BSA. (e) The contact number between PLys and BSA. (f) The Coulomb potential between PLys and PGlu at different pH values.

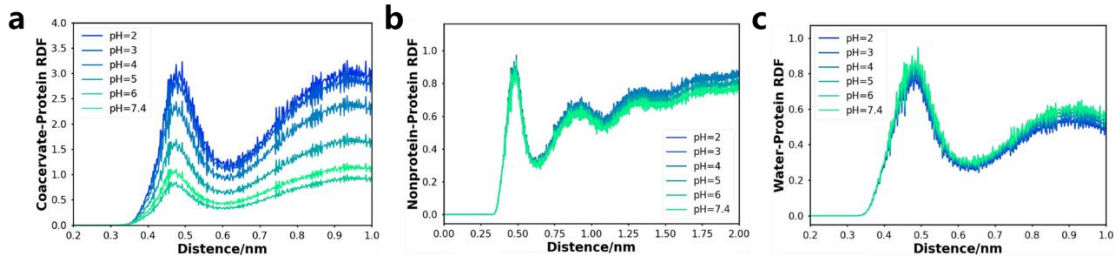

Fig. S48 RDF analysis between various components at different pH values. The radial distribution function between (a) Coacervate (b) Nonprotein (c) Water and BSA at different pH values.

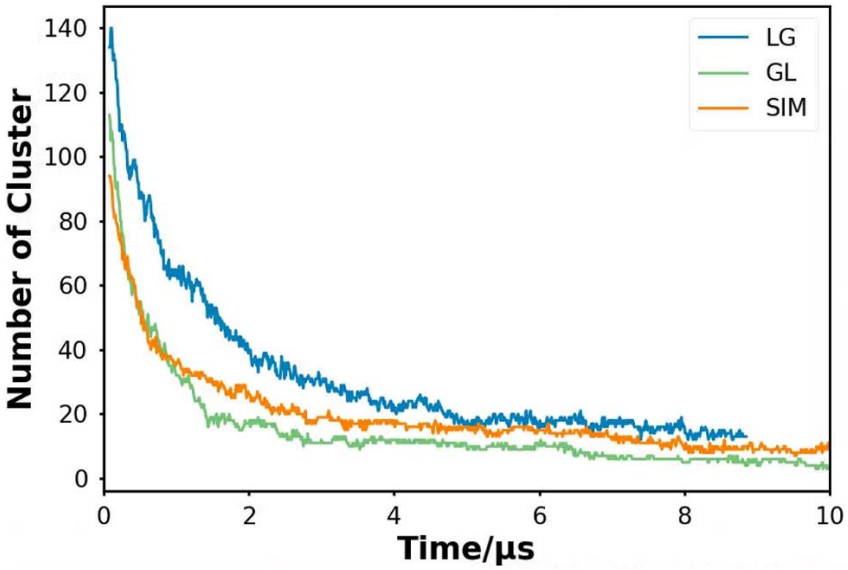

Fig. S49 The time-dependent curves of the number of coacervate clusters in systems with different addition sequences under 0.6 BSA.

Supplementary Tables

Table S1. Modeling of different ingredient addition sequences and BSA content systems

	STEP1[a]: adding ingredients (number of proteins or amino acids)			STEP2[b]: adding ingredients (number of proteins or amino acids)		
	BSA	PLys	PGlu	BSA	PLys	PGlu
0.2BSA-LG	12	100	-	-	-	100
0.3BSA-LG	18	100	-	-	-	100
0.4BSA-LG	24	100	-	-	-	100
0.5BSA-LG	30	100	-	-	-	100
0.6BSA-LG	36	100	-	-	-	100
0.2BSA-GL	12	-	100	-	100	-
0.3BSA-GL	18	-	100	-	100	-
0.4BSA-GL	24	-	100	-	100	-
0.5BSA-GL	30	-	100	-	100	-
0.6BSA-GL	36	-	100	-	100	-
0.2BSA-SIM	-	-	-	12	100	100
0.3BSA-SIM	-	-	-	18	100	100
0.4BSA-SIM	-	-	-	24	100	100
0.5BSA-SIM	-	-	-	30	100	100
0.6BSA-SIM	-	-	-	36	100	100

[a] Simulated for approximately 5 μs.

[b] Simulated for approximately 8 μs.

Table S2. Modeling of different ingredient addition sequences and proteins systems

	STEP1[a]: adding ingredients (number of proteins or amino acids)			STEP2[b]: adding ingredients (number of proteins or amino acids)		
	Protein	PLys	PGlu	Protein	PLys	PGlu
Saporin-LG	6	100	-	-	-	100
Lysozyme-LG	6	100	-	-	-	100
EGFP-LG	6	100	-	-	-	100
Actin-LG	6	100	-	-	-	100
Saporin-GL	6	-	100	-	100	-
Lysozyme-GL	6	-	100	-	100	-
EGFP-GL	6	-	100	-	100	-
Actin-GL	6	-	100	-	100	-
Saporin-SIM	-	-	-	6	100	100
Lysozyme-SIM	-	-	-	6	100	100
EGFP-SIM	-	-	-	6	100	100
Actin-SIM	-	-	-	6	100	100

[a] Simulated for approximately 5 μs.

[b] Simulated for approximately 8 μs.

Table S3. Modeling of β-sheet systems

	STEP1[a]: adding ingredients (number of proteins or amino acids)			STEP2[b]: adding ingredients (number of proteins or amino acids)		
	BSA	PLys	PGlu	BSA	PLys	PGlu
0.2BSA-LG-β-sheet	12	100	-	-	-	100
0.3BSA-LG-β-sheet	18	100	-	-	-	100
0.4BSA-LG-β-sheet	24	100	-	-	-	100
0.5BSA-LG-β-sheet	30	100	-	-	-	100
0.6BSA-LG-β-sheet	36	100	-	-	-	100

[a] Simulated for approximately 5 μs.

[b] Simulated for approximately 8 μs.

Table S4. Modeling of different pH value systems

	STEP1[a]: adding ingredients (number of proteins or amino acids)			STEP2[b]: adding ingredients (number of proteins or amino acids)			
	BSA	PLys	PGlu	BSA	PLys	PGlu+H	PGlu-H
pH-7.4	6	200	-	-	-	200	0
pH-6	6	200	-	-	-	198	2
pH-5	6	200	-	-	-	170	30
pH-4	6	200	-	-	-	72	128
pH-3	6	200	-	-	-	11	189
pH-2	6	200	-	-	-	1	199

[a] Simulated for approximately 5 μs.

[b] Simulated for approximately 8 μs.

Reference

1 W. L. DeLano, *Protein Crystallogr*, 2002, **40**, 82-92.

2 P. C. Kroon, F. Grünewald, J. Barnoud, M. van Tilburg, P. C. Souza, T. A. Wassenaar and S.-J. Marrink, *arXiv preprint arXiv:2212.01191*, 2022.

3 K. A. Black, D. Priftis, S. L. Perry, J. Yip, W. Y. Byun and M. Tirrell, *ACS Macro Lett.*, 2014, **3**, 1088-1091.

4 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38.

5 J. Schlitter, *Chem. Phys. Lett.*, 1993, **215**, 617-621.

6 A. Maćkiewicz and W. Ratajczak, *Comput. Geosci.*, 1993, **19**, 303-342.

7 M. Ester, H.-P. Kriegel, J. Sander and X. Xu, *In kdd,* 1996, **96**, 226-231.

8 S. L. Perry, L. Leon, K. Q. Hoffmann, M. J. Kade, D. Priftis, K. A. Black, D. Wong, R. A. Klein, C. F. Pierce III and K. O. Margossian, *Nat. Commun.*, 2015, **6**, 6052.