Spatiotemporal landscape for the sophisticated transformation of protein assemblies defined by multiple supramolecular interactions

Long Li, Zhen Li, Ziying Wang, Shuyu Chen, Rongying Liu, Xuyang Xu, Zhi Zhang, Linfei Ye, Yu Ding, Quan Luo, et al.

Spatiotemporal landscape for the sophisticated transformation of protein assemblies defined by multiple supramolecular interactions

Long Li[1], Zhen Li[1], Ziying Wang[2], Shuyu Chen[3], Rongying Liu[1], Xuyang Xu[1], Zhi Zhang[1], Linfei Ye[1], Yu Ding[2], , Quan Luo[4,5], Sheng Cao[6], Lei Zhang[3], Anne Imberty[7], Guosong Chen[1,8]*

1. The State Key Laboratory of Molecular Engineering of Polymers and Department of Macromolecular Science, Fudan, University, Shanghai 200433, China

2. State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China

3. MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, School of Science, Xi'an Jiaotong University, Xi'an 710049, China

4. State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry, Jilin University, Changchun, 130012, China

5. Key Laboratory of Emergency and Trauma, Ministry of Education, College of Emergency and Trauma, Hainan Medical University, Haikou 571199, China

6. CAS Key Laboratory of Special Pathogens, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, 430071, China

7. Université Grenoble Alpes, CNRS, CERMAV, Grenoble 38000, France

8. Multiscale Research Institute for Complex Systems, Fudan University, Shanghai 200433, China

* E-mail: guosong@fudan.edu.cn (G. C.)

Abstract

Precise protein assemblies not only constitute a series of living machineries but also provide an advanced class of biomaterials. Previously, we developed the inducing ligand strategy to generate various fixed protein assemblies, without the formation of noncovalent interactions between proteins. Here, we demonstrated that controlling the symmetry and number of supramolecular interactions introduced on protein surfaces could direct the formation of unspecific interactions between proteins and induce various nanoscale assemblies, including coiling nanowires, nanotubes, and nanosheets, without manipulation of the protein's native surfaces. More importantly, these nanoscale assemblies could spontaneously evolve into more ordered architectures, crystals. We further showed that the transformation from the introduced supramolecular interactions to the interactions formed between proteins was crucial for pathway selection and outcomes of evolution. These findings reveal a transformation mechanism of protein self-assembly that has not been exploited before and may provide an approach to generate complex and transformable biomacromolecular self-assemblies

INTRODUCTION

Precise protein assemblies constitute a multitude of cellularmachineries[1] such as, zero-dimensional viral capsid engaging in delivering genomes towards host cells,[2] one-dimensional microtubule capable of transporting cooperation and transformation of these multidimensional assemblies constitute the complex and harmonious living system.[1-4] Furthermore, these fascinating structures have inspired researchers to fabricate protein assemblies in various morphologies, leading to a variety of quite practical outcomes.[5,6] For example, artificial protein architectures including nanocages, filaments, nanotubes, nanosheets and 3D crystals have been successfully fabricated to diverse applications, including drug delivery, cellular processes manipulation, or electron transfer.[7-12] Notably, non-covalent interactions between proteins, namely, protein-protein interactions (PPIs), containing multiple weak interactions such as electrostatic interactions, hydrophobic interactions and hydrogen bonds, have been identified as a determinant of great importance for the formation and transformation of biological machinery and the generation of artificial protein assemblies.[13-15] Thus, modulating the protein-protein interactions is a practical approach to fabricate protein assemblies for integrating diverse functions and properties into a system that replicates or even surpasses the functions of their naturally occurring counterparts. Yet, countless non-covalent PPIs are distributed over large and often discontinuous protein surfaces, which makes it very difficult and complex to directly regulate the formation of PPIs. Therefore, two major approaches have been developed to design artificial protein assemblies by reengineering the specific PPIs. The first is focusing on electrostatic interaction,[16,17] metal coordination,[8,18-20] computational interface design,[11,14,21] and symmetric protein fusion,[22,23] featuring reconstituted specific protein-protein interactions on proteins. The other strategy is to introduce additional interactions in place of specific PPIs, such as supramolecular host−guest interactions,[24-26] and DNA pairing.[12,27] These strategies have been demonstrated successful for the fabrication of artificial protein assemblies, the generation of novel biomaterials, and even the manipulation of cellular behaviors.,[18 7] However, in most of these strategies, strong interactions are involved and designed to fabricate a very fixed and stable protein architecture. And, in most cases, it means that a single engineering-intensive design on proteins results in only one unchangeable assembly,[5,6] which is different from natural protein assemblies with transformable or dynamic properties through the formation of diverse non-covalent interactions between proteins.[1-4] As such, an approach that can readily assemble protein into various architectures by directing the formation of non- covalent interactions between proteins rather than redesigning or replacing specific PPIs is highly valuable and unexplored.

Recently, we proposed the inducing ligand strategy that dual supramolecular interactions consisting of protein- carbohydrate interaction and π−π stacking of rhodamine B (RhB) were employed to array proteins into various assemblies.[28-30] In this work, for the first time, we demonstrate that the introduction of appropriate inducing ligands on protein surface could guide the generation of the unspecific interactions between proteins without changing any condition, which is named as induced protein-protein interactions (iPPIs). The symmetry and number of the binding sites were spatially controlled on the same protein; variations of the spacer of inducing ligand were designed to control the follow-up protein contact during the self- assembly process. It resulted in protein assemblies at various dimensions including coiling nanowires, nanotubes, and nanosheets. More interestingly, these ordered nanostructures could spontaneously evolve into macroscopic crystal architectures without the requirement for crystallization

screening or protein precipitants, resulting in six diffraction-quality crystals. Via tracking the evolutionary process, it was demonstrated that the transformation from the introduced supramolecular interactions to the generated iPPIs were critical for pathway selection and outcome. Furthermore, with diversity of the introduced supramolecular interactions and the formed iPPIs, these assemblies and crystals showed very different stability to temperature, urea and cell culture environment. Endocytosis study further confirmed that assemblies had a significant increase in the rate of uptake by cells than crystals, but crystals could keep releasing fluorescent molecules for a long time, which may be applied to the design of drug delivery vehicles and bioimaging. These findings suggest a new approach to construct various dimensional protein assemblies by combining controlled supramolecular interactions on protein surface with induced protein-protein interactions. And, the revealed two-stage transformation mechanism of assemblies might offer numerous opportunities for the fabrication of transformable complex or even dynamic biomacromolecular architectures.

Results and Discussion

Design of introduced supramolecular interactions on protein

In our previous work, the introduced Supramolecular interactions, namely, ligand-ligand interactions, was identified as a determinant to assemble lectins such as concanavalin A (ConA) and Soybean agglutinin (SBA) into crystal framework and nanotube structures, respectively. Yet, in those works, no interactions between those homotetrameic proteins were observed, and those proteins with four immutable carbohydrate-binding sites can only provide the fixed density of LLIs on proteins as 1/25KDa and 1/27KDa (the ratio of number of binding sites to molecular weight).[28,31] Here, we chose a variant of the lectin from Ralstonia solanacearum (RSL) that was engineered as a monomeric β-propeller by adding linker peptides into a trimeric protein as a model protein with Mw of 29 kDa.[32] Native RSL is a C3-symmetric trimer with six carbohydrate binding sites, in which every binding sites cannot be manipulated individually. While this monomeric RSL (mRSL) has several advantages to precisely control the supramolecular interactions introduced on protein. Firstly, the number of bound ligands could be controlled through engineering of the active binding sites on mRSL since it was demonstrated that Arg to Ala mutation in fucose binding site abolished ligand binding.32 Secondly, the crystal structure of mRSL with fucose ligand (PDB code 4CSD)32 demonstrated planarly distributed six fucose-binding sites adopting a regular hexagonal packing (hexavalent, C6 symmetry) that contains regular triangle (trivalent, C3 symmetry) and rectangle (divalent, C2 symmetry) in geometry (Figure 1a).

At last, the controlled binding sites on the mRSL allow the density of LLIs on protein surface to be regulated from 2/29KDa to 6/29KDa, resulting in the several fold higher. worth to mention that native RSL is a high symmetric protein appropriate for generating regular assemblies.[33-35] This advantage has been used in generating porous crystalline materials by crystallization screening,[25,26,33-35] and even in one case, RSL is capable of forming crystals in acidic buffer with a narrow pH range.[26] The variable crystalline architectures affected by different solutions show that there is no specific PPIs between two RSLs to dominate the self-assembly outcomes, which supports the selection of RSL for our current purpose, i.e. employing supramolecular interactions based on inducing ligands to guild the self-assembly process.

The different variants were named as RSL-m (m represents the number of carbohydrate binding sites, m= 2, 3, 6). Furthermore, to control the

spatial distance between proteins, three ligands containing α-fucopyranoside and RhB with different ethylene oxide repeating units as spacers (FnR, n = 2, 3, 4), were synthesized (Figure 1b, details in the Supporting Information Scheme 1 and Figure S38-49). Such ligands possess three functions: (a) they bind to the protein through specific recognition between fucoside and RSL; (b) the spacer can be varied to adjust the spatial distance between two neighboring proteins with no effect on the dimerization of RhB (Figure S2); (c) the noncovalent and dynamic dimerization of RhB is a suitable interaction to generate iPPIs in order to afford diverse protein stacking, providing more conformational flexibility than covalent bond in a dynamic self-assembly system (Figure 1c).

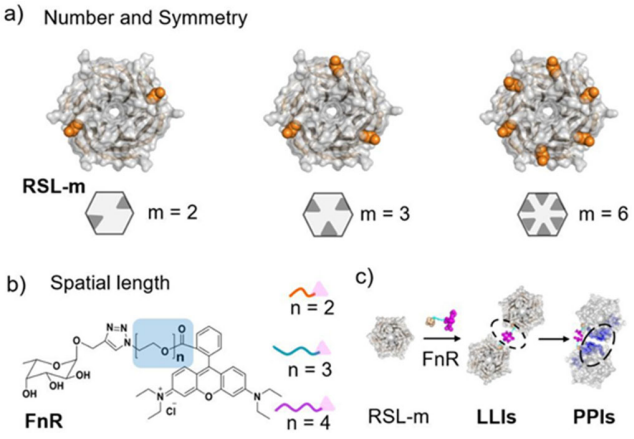

Figure 1. *Design of multiple supramolecular interactions. a) Illustration for controlling the number and symmetry of supramolecular interactions on protein, and representative views of model protein. (Structure adapted from the crystal structure, PDB ID: 4CSD). b) Chemical structure of chemical ligand FnR, and the ethylene oxide repeating units to regulate the spatial length. c) Schematic illustration for the generation of protein-protein interactions by RhB dimer*

To further confirm this, the all-atom molecular dynamics (MD) simulation was employed to examine whether the dimeric ligand can induce the production of PPIs. Taking the F4R as an example, two RSLs connected by the stretched dimer of F4R was considered as the initial state (Figure S3a). After 200 ns running, the proteins moved close to each other, creating PPIs consisting of five hydrogen bonds occurred close to the *inducing ligand*-binding sites (Figure S3b). The simulation results demonstrated that binding to the dimerized ligands might induce the generation of non-covalent interactions between two RSLs. Simulations were run with three ligands of various spacers, F2R, F3R, and F4R, than that of those homotetrameic proteins (Figure S1). It is resulting in dimeric distances of ~17 Å, ~18 Å, and ~23 Å (measured between the C1 atoms of α-fucopyranoside, Figure S3c), respectively. It exhibited the possibility of the various spacers of ligands to adjust the spatial distance.Together, the number and symmetry of binding sites were

Generating diverse protein assemblies in different dimensions.

The three purified protein were demonstrated to exist as a monomer in solution via the Size Exclusion Chromatography (SEC) at concentration up to 60 mg/mL (Figure S4). The expected binding stoichiometry ofthe different RSL-m and affinity towards fucosides were Experimentally confirmed by Isothermal Titration Calorimetry (ITC) (Figure S5). Self-assembly experiments were performed with RSL-m (m = 2, 3, 6) (1 mg/mL) and FnR ligands (n = 2, 3, 4), which were added with the molar ratio of 2 to 1 (calculated as number of binding sites) in Tris buffer before incubation at 4 °C.

The UV−vis (UV) and Circular dichroism (CD) spectroscopies were employed to check the LLIs that were directly provided by RhB dimerization. For example, RSL-2, RSL-3 and RSL-6 (1 mg/mL) were mixed with F3R (0.4 mM) and samples were characterized after 24 h incubation. As shown in UV spectra (Figure 2a), an increase

of the signal corresponding to dimeric RhB (580 nm and 525 nm) was observed.28,31,36 Together with decrease of the signal of monomeric RhB at 560 nm, it confirmed the establishment of LLIs. To quantitatively compare the strength of LLIs in various RSLs, the dimeric signal at 580 nm was checked in detail (inset of Figure 2a). It suggested that RhB dimerization varies as RSL-6 > RSL-3 > RSL-2, indicating stronger LLIs with the increase of the number of binding sites. It also agreed with the reduced trend of monomeric RhB signal at 560 nm. Akin to the UV observation, the corresponding CD spectra confirmed the increase of RhB dimerization as RSL-6 > RSL-3 > RSL-2 (Figure 2a). The corresponding dimerized RhB signal found both in UV-vis and CD spectra confirmed the LLIs connecting proteins, with the latter providing the achiral RhB a chiral environment. Meanwhile, Transmission Electron Microscope (TEM) and Atomic Force Microscope (AFM) were used to observe the assembly of proteins at 24 h. The results of TEM and AFM showed that RSL-2 and RSL-3 gave nanoscale tubes and coiling wires respectively, while RSL-6 afforded massive disordered precipitation (Figure 2b, S6). Dynamic Light Scattering (DLS) results further confirmed the increased hydrodynamic size of aggregated RSL-m induced by LLIs in solution (Figure S7). The conversion of protein (Cv^p) was further employed to evaluate the degree of aggregation. After centrifuging at 18000 rpm for 30 min at 4 °C to remove the assembled or aggregated proteins, the Cv^p was calculated by the residual protein in supernatant and the concentration of protein was measured by Bradford method. The gradually increased Cv^p in RSL-2, RSL-3 and RSL-6 (Figure S8), confirmed that more proteins were transformed into assemblies or aggregates with the increased binding sites. Together, the results of TEM and DLS were consistent with the results from CD and UV, showing that when the spacer of ligand was fixed, the number of LLIs and their symmetry, determined the aggregation of RSLs to ordered microstructures in solution (Figure 2c).

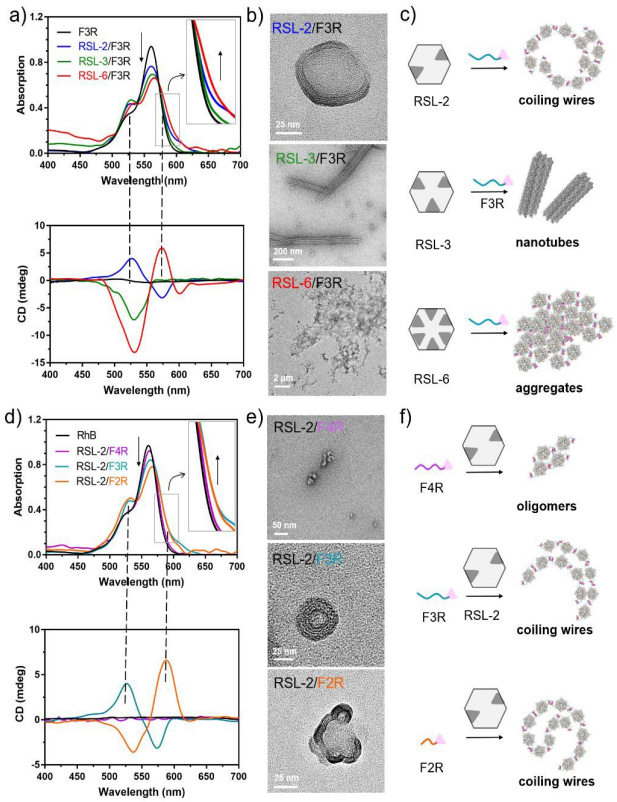

Figure 2. *The number of introduced supramolecular interactions and spacer of ligands regulate the self-assembly of proteins. a) UV−vis and CD spectra of RSL-2, RSL-3, and RSL-6 mixed with F3R. b) TEM images of RSL-2, RSL-3 and RSL-6 assemblies induced by F3R. c) Illustration of the self-assembly of proteins regulated by the quantity of introduced supramolecular interactions. d) UV−vis and CD spectra of F2R, F3R, and F4R mixed with RSL-2. e) TEM images of RSL-2 assemblies induced by F2R, F3R, and F4R. f) Illustration of the effect that spacer of ligands regulated on the self-assembly of proteins*

To inspect the effect of spacer of inducing ligands, FnRs with distinct spacers were tested on the same protein (RSL- 2). From UV, the increase of signal of dimeric RhB at 580 nm and 525 nm, and decrease of signal of monomeric RhB at 560 nm, confirmed the presence of LLIs in F2R and F3R but not in F4R. The strength of LLIs

provided by various spacers, could be categorized from the increase of dimerization as F2R> F3R> F4R. Furthermore, the CD spectra also confirmed the increased signal of dimeric ligand as F2R > F3R > F4R at 580 nm and 525 nm (Figure 2d). It revealed that the LLIs increased as the spacer become shorter. On the other hand, the combination of TEM and DLS results showed that the mixture of RSL-2/F2R gave rise to coiling wires reminiscent of that of RSL-2/F3R, while the mixture of RSL- 2/F4R afforded no obvious assemblies (Figure 2e, S9). Those results indicated that the spacer of ligands impacted on the aggregation of RSL. Similar, when RSL-3 and RSL-6 was used to test ligands with different spacers, dimerization of ligands F2R and F3R was confirmed by the signal of RhB dimer at 580 nm in UV-vis and CD spectra (Figure S10, S11), but was not found of RSL-3 with F4R. The RSL-3 showed nanotubes (RSL-3/F2R) and no apparent assemblies (RSL- 3/F4R) (Figure S11). While RSL-6 assembled as random aggregates (RSL-6/F2R) and nanosheets (RSL-6/F4R) (Figure S6, S10), indicating more contact between proteins in the former. Furthermore, the degree of aggregation was revealed by the conversion of RSL-2. The Cv^p was significantly increased with the shorter spacer as F2R > F3R > F4R (Figure S12), suggesting that stronger aggregation was generated by the shorter spacer. These results further confirmed that, when the number of binding sites, i.e. number of LLIs, was fixed, the spacer of ligands could further regulate the aggregation of RSLs to ordered microstructures in solution (Figure 2f). Protein packing in assemblies. To further confirm whether these assemblies were mainly maintained by the dimeric RhB, these assemblies were incubated with 1mM β disappeared CD signal in all samples indicated the dissociation of the assemblies (Figure S13). This was also confirmed by the TEM observation (Figure S14). Therefore, based on the planar distribution of binding sites, the possible models were built according to the number and symmetry of the introduced supramolecular interactions. The planar models proposed a molecular arrangement where seven proteins constituted a close-packed regular hexagon in RSL-6, and two neighboring proteins formed linear structure in RSL-2. For RSL-3, two packing patterns could be generated, including a hollow packed regular hexagon constituted by six proteins and a close-packed structure similar to the pattern of RSL-6 (Figure S15).

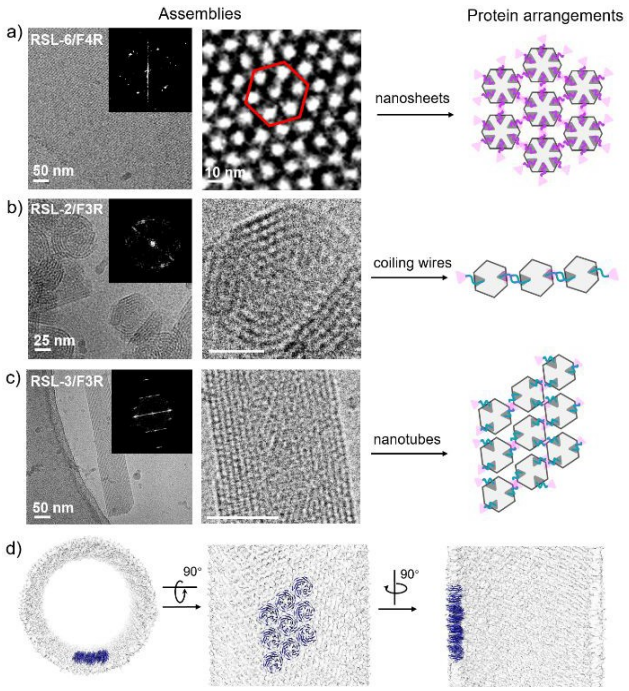

Figure 3. *Protein arrangements in assemblies. a) Cryo-EM image, 2D class averages, and protein arrangements of RSL- 6/F4R nanosheets. b) Cryo-EM images and protein arrangements of RSL-2/F3R coiling wires. c) Cryo-EM images and protein arrangements of RSL-3/F3R nanotubes. d) Three- dimensional (3D) reconstruction of RSL-3/F3R nanotubes. Corresponding density maps of nanotube have been deposited in EMDB with accession codes EMD-34879. Fourier transform was shown in the inset*

We deemed that, if the packing of RSL was controlled by the LLIs, it would adopt the molecular arrangement in the model. To confirm

this, Cryo-electron microscopy (Cryo-EM) was employed to investigate the internal arrangement of the protein. The results showed that RSL-6/F4R nanosheets exhibited typical lattice pattern (Figure 3a). More importantly, the result of 2D class averages further indicated that the arrangement of RSL-6 (close-packed regular hexagon) was indeed the same as its model (Figure 3a). For RSL-2, the coiling wire structure was clearly observed by Cryo-EM (Figure 3b, S16), which was consistent with the results of negative staining TEM (Figure 2e). And the coiling wire structure of multiple proteins linked together showed that the protein arrangement in RSL-2 assemblies also agreed with its models (two proteins connected by two ligands) (Figure 3b, S16). For RSL-3, although the observation of Cryo-EM showed some nanosheet features in the multi-layer assemblies with good Fast Fourier Transform (Figure 3c, S16), the result of cryogenic computed tomography well supported that these assemblies were multi-layered nanotubes (Video 1). Furthermore, taking the RSL-3/F2R assemblies as an example, 3D reconstruction was employed to investigate the internal protein packing. Via fitting the RSL proteins into the reconstructed map of single layer, the result confirmed that the packing of RSL-3 was a close-packed pattern instead of a hollow packed regular hexagon (Figure 3c, 3d). Together, combining with the results of UV and CD spectra, the observation of Cryo-EM suggested that the RSL arrangements in assemblies were determined by the LLIs on the proteins.

Transformation of protein assemblies into crystals.

Although the assemblies and arrangements of RSL were dominated by the LLIs on the protein, the compact protein arrangements in these assemblies indicated the possibility of generating iPPIs. Therefore, these samples were incubated for extended periods to find out whether non-covalent interactions between proteins could be induced and affect the morphologies of assemblies. To our surprise, various protein single crystals with magenta color and well- defined morphologies were observed in these samples (Figure 4a). Confocal Laser Scanning Microscope (CLSM) was employed to confirm that the magenta appearance of these crystals emerged from the presence of RhB in crystals (Figure S17). In morphology, the RSL-6/F4R crystal could be defined as rectangle, the crystals of RSL-3/F2R were obtained with two forms, i.e. long strip (form a) and polygon (form b), the RSL-3/F3R, RSL-2/F2R and RSL-2/F3R could be defined as granular, rhombus, and cuboid, respectively (Figure 4a, S17, S18). The other combinations failed to generate crystals (Figure 4b). The diverse outcomes revealed that the ordered assemblies were not the ultimate steady state, but rather served as precursors for crystal formation. Furthermore, combining the observation of Cv^p and morphologies, a protein phase diagram was tentatively mapped to reveal the relationship between the LLIs and the transformation from assemblies to crystals (Figure 4b). Since the Cv^p positively correlated with the quantity of binding sites, but negatively correlated with the size of the spacer (Figure S19), crystallization could not be afforded in the regime where the proteins induced by LLIs preferentially exists as monomeric state or precipitation. And only the appropriate LLIs were able to introduce proteins into assemblies, and further transformed them into crystals.

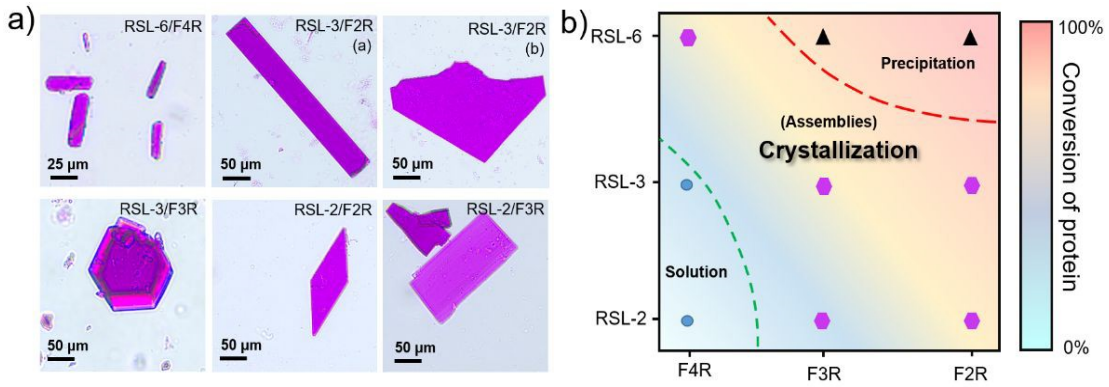

Figure 4. *Ordered assemblies evolved into crystals by prolonged incubation time. a) Optical micrographs of various protein single crystals forms. The crystals were generated by mixing various proteins and ligands, and then incubated at 4 °C for six months. b) The assembled protein phase diagrams. The conversion of protein (Cv^p) was calculated by the residual protein in solution*.

Table 1. Summary of Crystal Properties for RSL-m/FnR Single Crystals.

Sample	PDB ID	space group	*a, b, c* (Å)	*α, β, γ* (°)	resolution (Å)	Z^b
RSL-6+F4R	7E7W	P1[a]	46.56 46.55 103.25	90.04 89.92 119.98	4.10	3
RSL-3+F2R (a)	7E7R	$P2_12_12_1$	67.42 81.95 97.11	90.00 90.00 90.00	1.82	2
RSL-3+F2R (b)	7E7T	P12 1	44.34 88.55 67.89	90.00 93.83 90.00	1.98	2
RSL-3+F3R	7E7N	P1	44.30 45.63 46.84	94.12 117.04 117.20	1.50	1
RSL-2+F2R	7E7U	$P2_12_12_1$	47.80 79.47 152.46	90.00 90.00 90.00	2.01	2
RSL-2+F3R	7E7V	$P12_11$	43.49 161.51 54.13	90.00 91.53 90.00	1.61	3

[a] *A "pseudo hexagonal" due to internal pseudo symmetry of RSL β-propeller.* [b] *The number of monomer per asymmetric unit.*

According to this, we attempted to generate crystals in those samples that did not form protein crystals, such as RSL-6/F2R, RSL-6/F3R and RSL-3/F4R, by adjusting the aggregation state of proteins induced by LLIs. The concentration of ligands as another factor that could modulate the RhB dimerization was further investigated.36 Via mixing RSL-6 with F4R at different molar ratios, only aggregates formed when the ratio of RSL-6 to F4R was below 3:2, and the size and completion of the crystal were increased with the increased ligand concentration (Figure S20), suggesting that the concentration of ligands could regulate the aggregation state of proteins. Furthermore, via investigating the Cv^p in RSL-6 and RSL-3 induced by different ligand concentrations at 24 h, it was observed that the Cv^p showed a linear function of the ligand concentration when lower than 0.2 mM in RSL-6 and 0.25 mM in RSL-3 (Figure S21). Thereby, the crystals of RSL-6/F2R and RSL- 6/F3R were successfully obtained by adjusting the molar ratio of protein to ligand with 1:1, and the crystal of RSL- 3/F4R was also obtained by increasing the molar ratio of protein to ligand with 1:4 (Figure

S22, S23). Although, the crystallographic data could not be obtained by X-ray diffraction because of the small crystal size, distinct crystal stripes were observed by high resolution TEM (Figure S24). It further proved that fine-tuning the intensity of LLIs to induce protein into the appropriate assembly or aggregates was able to introduce the proteins into crystal architectures, as illustrated in Figure 4b.

To further investigate how the assemblies evolved into crystals, data sets of six crystals of lectin/LLIs complexes were obtained by X-ray diffraction and the structures were solved by molecular replacement (Table 1, S1). When taking monomer A of RSL-3/F3R as a reference, a root-mean- square deviation (rmsd) of 0.45 Å, 0.18 Å, 0.21 Å, 0.20 Å and 0.25 Å for all atoms from each structures demonstrated the relative rigidity of the RSL protein (Figure S25), suggesting that the mutated binding sites and the addition of ligand totally did not affect its conformation. By classifying the packing as the protein array related with ligands, the representative repeating units were extracted from the solved structures (Figure 5 middle). The position of ligands in RSL-6/F4R crystal was inferred from the complex between RSL-6 and fucose (PDB code 4CSD) due to the low resolution. As shown in Figure 5a, RSL-6/F4R displayed a hexagonal dense packing similar to the molecular arrangement of its assembly. Furthermore, along the ac face of crystal packing, proteins were stacked layer by layer. These characteristics suggested that the RSL-6/F4R crystals could be constructed from the nanosheets. Similarly, for the RSL-2 and RSL-3 crystals, taking RSL-2/F3R and RSL-3/F3R as examples, the crystal packing of RSL-2 where two proteins connected with a pair of F3R form linear arrangements (Figure 5b), can also be generated from the assembly packing. Furthermore, a hexagonal dense packing was observed in RSL-3/F3R crystal, and the layer-by-layer stacking along the ab face similar to the multilayer structures in Cryo-EM (Figure 5c). These similar properties suggested that these crystalline protein arrangements were originated from their assembled patterns

.

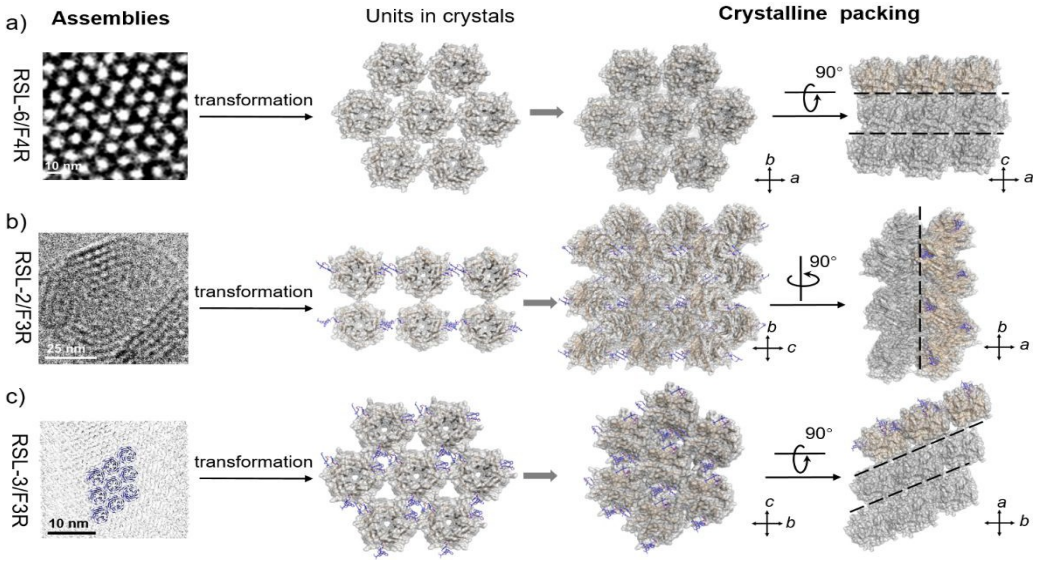

Figure 5*. The characteristics of protein packing in crystals. Packing of protein in assemblies (left) and in crystals with representative unit (middle) and global crystal packing (right) for a) RSL-6/F4R, b) RSL-2/F3R, and c) RSL-3/F3R.*

iPPIs underpinned the crystal architectures.

Interestingly, closer inspection showed that all the LLIs could only exist in the two-dimensional single molecular layer, instead of the three-dimensional crystals (Figure 5 right, S26), suggesting the possibility of iPPIs instead of LLIs as an interaction to underpin the generation of crystal architectures. Considering the resolution of crystal structures, the iPPIs and LLIs of RSL-2 and RSL-3 were investigated in detail. The dimeric ligand was able to be well-defined by the clear electron density in the crystal of RSL-2/F3R (Figure 6a). For RSL-2/F2R not only the dimeric ligand existed (Figure 6a), but also individual ligand sites were found (Figure S27), implying that the LLIs might be disrupted. Comparing the crystal structures of RSL-2 complexed with two ligands, F3R (full length of F3R ~18 Å) resulted in a distance of ~16 Å between two neighboring RSL-2, and F2R resulted in a distance of ~14 Å with full length of F2R ~17 Å (Figure 6b). This result confirmed that the size of the ligand spacer could affect the spatial distances of proteins. The PPIs in RSL-2 crystals were further studied by the analysis of PISA. The result showed that two separated major protein interfaces occurred close to the inducing ligand-binding sites in both RSL-2/F2R and RSL- 2/F3R complexes (calculated as the solvent accessible, Figure 6c), confirming that the PPIs were induced by the LLIs. Furthermore, comparing the two interfaces in RSL- 2/F2R (PPI' and PPI'') and in RSL-2/F3R (PPI' and PPI''), the two larger interfaces (> 1000 Å2) were identical (PPI', Figure 6c, colored by limon), and the two other ones (approx. 800 Å2) were different (PPI'', Figure 6c, colored by orange or blue). The PPI' in RSL-2/F2R displayed strong interaction that involved five hydrogen bonds (G24-T129, N42-K116, E43-N133, D46-T129 and G68-T129), and two salt bridges (E43-K116, D46-R120) (Figure 6c), which also presented in the PPI' of RSL-2/F3R. The generation of the identically strong PPI' in different crystals of RSL-2 implied that the iPPIs in the process of crystals generation might be stabilized by the strengthening of intermolecular interactions on the contact areas.

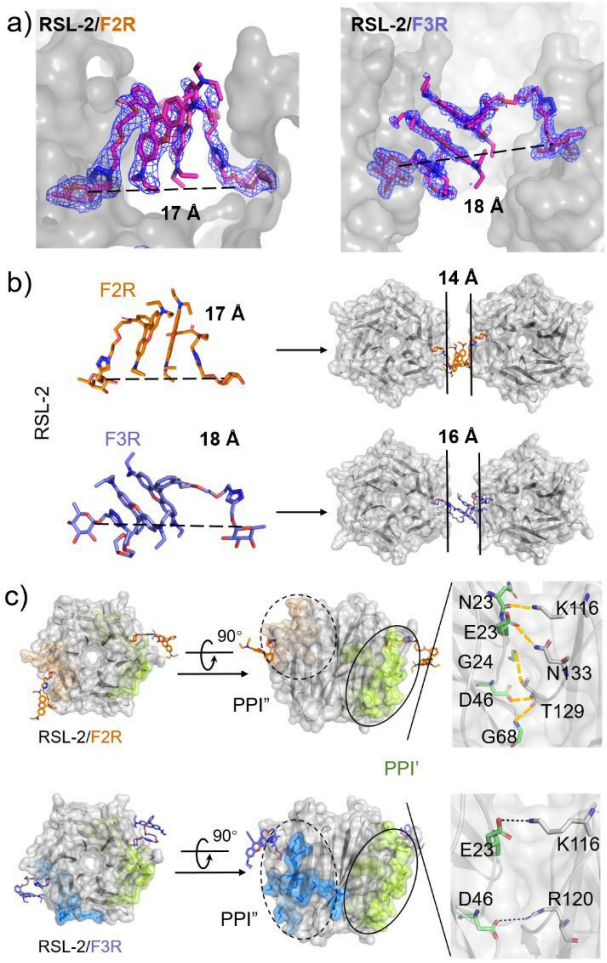

Figure 6*. iPPIs were stronger than LLIs in crystals of RSL-3 and RSL-2. a) The dimeric ligands in RSL-2/F2R and RSL- 2/F3R. The 2Fo−Fc maps contoured at 1σ. b) The spatial distance of dimeric F2R and F3R in crystal structures measured between the C1 atoms of α-fucopyranoside, and the two connected proteins with ligands measured between the Cα atoms of T220 and A85 residues. c) Protein interface analysis of RSL-2/F2R and RSL-2/F3R, both with two major protein interfaces, a conserved one (PPI', colored by limon) and a variable one (PPI'', colored by orange or slate). The identical two PPI's in RSL-2/F2R and RSL-2/F3R contained the same five hydrogen bonds and two salt bridges. Protein interfaces were colored by cyan and limon*

For the crystals of three RSL-3 complexes, electron density was not of sufficient quality to determine the conformation of RhB and ethylene oxide repeating units (Figure S28). While the similar conformational flexibility could not be easily observed for RSL-2 (Figure 6a) with the same ligands (F2R and F3R). Further analysis of RSL-3/F2R crystals type a and type b, showed a much closer distance (~10 Å, ~11 Å) of two contiguous ligands than that in a typical distance for dimeric ligands (~16 Å) (Figure S29, S30). Meanwhile, only three adjacent ligand sites could be found in RSL-3/F3R (Figure S27). Consisting with the individual ligand site existed in RSL-2/F2R, these atypical distances and unlikely dimeric ligand sites suggested the disordered LLIs in crystals. Furthermore, the analysis results of PISA (Protein Interfaces, Surfaces, and Assemblies) showed that the average of protein contact area was two to four-fold more than the theoretical maximum areas of RhB dimerization in all crystals (Table S2). It further suggested that the iPPIs, not the LLIs, were the major interactions underpinning crystals.

Mechanisms for the transformation of assemblies into crystals

Furthermore, the processes of assemblies transformed into crystals were traced by TEM. For RSL- 6/F4R, some aggregates up to 1 μm were first observed after incubation for 30 min. And after 4 h incubation, the aggregates transformed into nanosheets (Figure 7a), which was consistent with the result of DLS (Figure S31). Larger lamella and typical protein crystals formed after 96 h incubation (Figure 7a). Meanwhile, the CD signal was traced to check the transformation of LLIs. The CD signals at 580 nm exhibited a sustained increase within 2 days, and then maintained at 10 (Figure S32), showing that the stable state has been generated in RSL-6/F4R, which agreed with the crystal generation at 48 h. And, long-term observations showed that these crystals were the final steady state. For the transformation of RSL-3 and RSL-2, taking the RSL- 3/F3R and RSL-2/F3R as examples, random aggregates of RSL-3 were induced by F3R within a short time, and their size increased with incubation time (Figure 7b), which was further confirmed by DLS results (Figure S33). Then, the multi-layered nanotubes emerged, and remained stable for 2 months. Importantly, those well-defined tubes disintegrated into large aggregates with longer incubation, and crystal nucleation with corresponding pattern was observed by TEM (Figure 7b). The crystalline formation of RSL-2/F3R also began with formation of aggregates after 30 min. Its size continuously increased until some coiling wires emerged after 12 h incubation (Figure S34), agreeing with the DLS results (Figure S33). The coiling wires further aggregated after 1 month, and the nucleation restarted with further incubation (Figure S34). Furthermore, the CD signals of RSL-3/F3R and RSL-2/F3R showed rapid increase within 4 h before reaching equilibrium, and remained stable within 2 days (Figure S32). Combining with the TEM observation, it suggested that the stable assemblies had been generated by the induction of LLIs. And then, the CD signal was decreased after 2 months incubation, and further declined after 4 and 6 months, respectively (Figure S32), suggesting the weakening of LLIs with long incubation time. Interestingly, although the assembly arrangement and crystalline packing of RSL-6/F4R and RSL-3/F3R were similar (Figure 5), the transformation period from assemblies to crystals was very different. Considering that the crystal nucleation of RSL-3 passed through the nanotubes disintegrated into large aggregates, while the nucleation of RSL-6 was formed directly from its assemblies, it further indicated that the assembly or aggregation state induced by different LLIs was crucial to the

subsequent crystal formation, as illustrated in Figure 4b. Together, this process consisted of two stages that occurred sequentially from assembly generation to crystal formation (Figure 7c). In the first stage, the LLIs dominated the generation of assemblies, as revealed by the TEM, UV-vis, CD (Figure 2), and Cryo-EM (Figure 3). Whereas the iPPIs determined the formation from assemblies to crystals, as evidenced by the apparent protein contact areas and PPIs in the crystal structures (Figure 6).

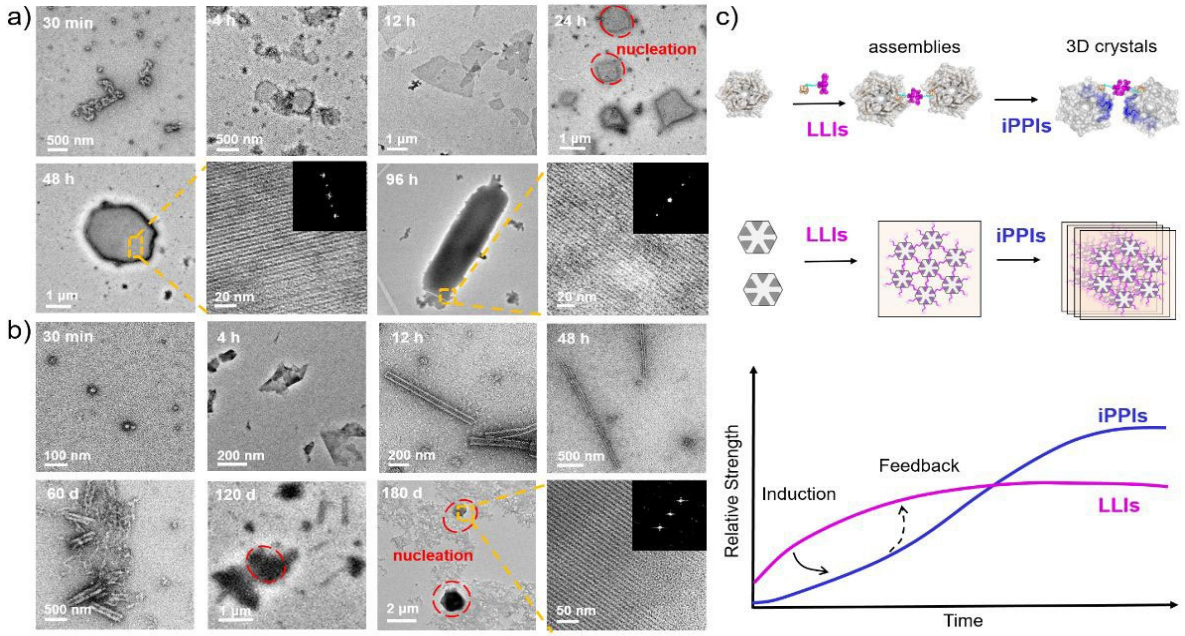

Figure 7. *The process of assemblies evolved into crystals. a) Time tracking of RSL-6/F4R crystal formation by TEM. b) Time tracking of RSL-3/F3R crystal formation by TEM. Fourier transform was shown in the inset. c) Illustration of the evolution from assemblies to crystals by the transformation between LLIs and iPPIs.*

LLIs and iPPIs modulated the properties of assemblies and crystals

The diverse LLIs and iPPIs implied that these assemblies and crystals could have different characteristics. Therefore, these assemblies and crystals collected by centrifugation were incubated with macrophages RAW 264.7, and CLSM was employed to confirm the presence of red fluorescence after 24 hours. The results suggested that all assemblies and crystals maintained excellent bioimaging capabilities and cellular uptake (Figure 8a, S35). For all assemblies and crystals, time-dependent experiments showed that the degree of endocytosis gradually increased with increasing incubation time (Figure S36, S37a). Furthermore, the nanoscale assemblies showed rapid endocytosis by cells, as evidenced by a significant enhancement of fluorescence intensity within 4 h (Figure 8a, S37a). For macroscopic crystals, they could only be endocytosed by cells after solubilization in a physiological environment, resulting in enhanced fluorescence intensity after prolonged incubation (Figure 8a, 8b). When the same amounts of crystals were incubated with cells, the RSL-6/F4R crystal showed minimal fluorescence intensity (Figure 8b), suggesting that it was more stable than other crystals in physiological environment. To further test their stability, these assemblies and crystals were treated with different concentration of urea, or

incubated at 50 °C. The results showed that all assemblies were dissolved into 1 mol/L urea (Figure S37b), while only 20%~50% of all crystals were dissolved at the same concentration (Figure 8c). It suggested that although both had the same composition, the crystals had a significant enhancement in urea resistance. Similarly, all assemblies were disintegrated by heating for 20 min (Figure S37c), but the crystals did not dissolve completely until 35 min (Figure 8d). And, the RSL-6/F4R crystal exhibited the best urea and heat resistance (Figure 8c, 8d). Combining the contribution of LLIs and iPPIs in the generation of these ordered structures, these results demonstrated that the crystals underpinned by iPPIs was more stable than the assemblies supported by LLIs.

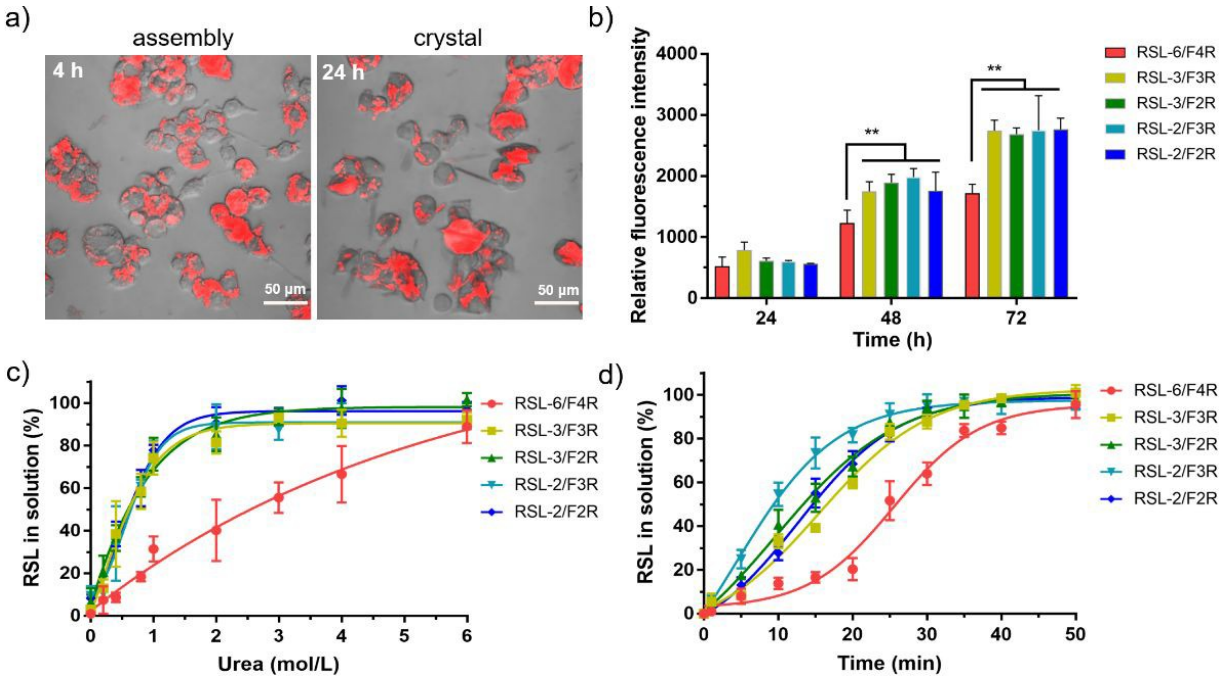

Figure 8. *Stability of assemblies and crystals as defined by LLIs and iPPIs. a) Confocal fluorescence microscopy images of macrophages incubated with the RSL-6/F4R assembly and crystal. b) Graph showing the time dependence of crystals incubated with macrophages. c) Dissolution curves of crystals at different urea concentrations. d) Time-dependent functions of crystals treated at 50°C. Data are shown as the mean ± SEM of three independent experiments. Key: **, $p < 0.01$.*

CONCLUSION

In summary, this work presents a successful new approach to induce the interactions between proteins and generate protein assemblies via using supramolecular interactions generated by the small molecular ligand. More importantly, the number and symmetry of the introduced supramolecular interactions and the spacer of ligands are feasibly designed handles to trigger diverse non-covalent interactions between proteins and modulate morphologies of protein assemblies based on the same protein. Crucially, these assemblies could spontaneously evolve into more ordered crystal structures, via the transformation of introduced supramolecular interactions and induced protein-protein interactions. Not only are morphologically

diverse, these assemblies and crystals also exhibit differences in rates of cellular uptake, urea resistance, and thermal stability. Due to the excellent luminous efficiency of rhodamine B, these assemblies and crystals may be highly available for bioimaging. Furthermore, differences in cellular uptake rates can also potentially be applied to the design of drug delivery vehicles with controllable release times. At last, without the requirement of crystallization screening and specific agents, the raveled mechanism of well-defined nanostructures transformed into crystals may open an alternative avenue to the development of protein- based drug-delivery systems, transformable and dynamic protein architectures, protein crystalline materials, and macromolecular crystallography.

ACKNOWLEDGEMENTS

G.C. thanks National Key Research and Development Program of China (2022YFB3804401) and NSFC/China (Nos. 52125303, 51721002 and 21975047) for financial support. Q.L. thanks the Key Laboratory of Emergency and Trauma (Hainan Medical University), the Ministry of Education (Grant No. KLET-202012) for financial support. The authors would like to thank the Shanghai Synchrotron Radiation Facility (SSRF, China) for x-ray diffraction. This work is also supported by the Shanghai Municipal Science and Technology Major Project (grant no. 2018SHZDZX01) and ZJ

REFERENCE

(1) Pieters, B. J. G. E.; van Eldijk, M. B.; Nolte, R. J. M.; Mecinovic, J. Natural supramolecular protein assemblies. *Chem. Soc. Rev.* **2016**, *45*, 24-39.

(2) Alonso, J. M.; Glrzny, M. Ł.; Bittner, A. M. The physics of tobacco mosaic virus and virus-based devices in biotechnology. *Trends Biotechnol.* **2013**, *31*, 530–538.

(3) Rachand, G. D.; Snoerke, E. D.; Stevens, M. J. Microtubule-Based Nanomaterials: Exploiting Nature's Dynamic Biopolymers. *Biotechnol. Bioeng.* **2015**, *112*, 1065–1073.

(4) Sleytr, U. B.; Schuster, B.; Egelseer, E. M.; Pum, D. S-Layers: Principles and Applications. *FEMS Microbiol. Rev.* **2014**, *38*, 823–864.

(5) Sun, H.; Luo, Q.; Hou, C.; Liu, J. Nanostructures based on protein self-assembly: From hierarchical construction to bioinspired materials. *Nano Today* **2017**, *14*, 16-41.

(6) Zhu, J.; Avakyan, N.; Kakkis, A.; Hoffnagle, A. M.; Han, K.; Li, Y.; Zhang, Z.; Choi, T. S.; Na, Y.; Yu, C. J.; Tezcan, F. A. Protein Assembly by Design. *Chem. Rev.* **2021**, *121*, 13701-13796.

(7) Ben-Sasson, A. J.; Watson, J. L.; Sheffler, W.; Johnson, M. C.; Bittleston, A.; Somasundaram, L.; Decarreau, J.; Jiao, F.; Chen, J.; Mela, I.; Drabek, A. A.; Jarrett, S. M.; Blacklow, S. C.; Kaminski, C. F.; Hura, G. L.; De Yoreo, J. J.; Kollman, J. M.; Ruohola-Baker, H.; Derivery, E.; Baker, D. Design of biologically active binary protein 2D materials. *Nature* **2021**, *589*, 468-473.

(8) Brodin, J. D.; Smith, S. J.; Carr, J. R.; Tezcan, F. A. Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. *J. Am. Chem. Soc.* **2015**, *137*, 10468–10471.

(9) Dai, J.; Knott, G. J.; Fu, W.; Lin, T. W.; Furst, A. L.; Britt, R. D.; Francis, M. B. Protein-Embedded Metalloporphyrin Arrays Templated by Circularly Permuted Tobacco Mosaic Virus Coat Proteins. *ACS Nano* **2021**, *15*, 8110-8119.

(10) Hirayama, S.; Oohora, K.; Uchihashi, T.; Hayashi, T. Thermoresponsive Micellar Assembly Constructed from a Hexameric Hemoprotein Modified with Poly(N-isopropylacrylamide) toward an Artificial Light-Harvesting System. *J. Am. Chem. Soc.* **2020**, *142*, 1822-1831.

(11) Shen, H.; Fallas, J. A.; Lynch, E.; Sheffler, W.; Parry, B.; Jannetty, N.; Decarreau, J.; Wagenbach, M.; Vicente, J. J.; Chen, J.; Wang, L.; Dowling, Q.; Oberdorfer, G.; Stewart, L.; Wordeman, L.; Yoreo, J. D.; Jacobs-Wagner, C.; Kollman, J.; Baker, D. De novo design of self-assembling helical protein filaments. *Science* **2018**, *362*, 705–709.

(12) Brodin, J. D.; Auyeung, E.; Mirkin, C. A. DNA-mediated engineering of multicomponent enzyme crystals. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 4564–4569.

(13) Garcia-Seisdedos, H.; Empereur-Mot, C.; Elad, N.; Levy, E. D. Proteins evolve on the edge of supramolecular self-assembly. *Nature* **2017**, *548*, 244-247.

(14) Gonen, S.; DiMaio, F.; Gonen, T.; Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **2015**, *348*, 1365-1368.

(15) Luo, Q.; Hou, C.; Bai, Y.; Wang, R.; Liu, J. Protein assembly: versatile approaches to construct highly ordered nanostructures. *Chem. Rev.* **2016**, *116*, 13571–13632.

(16) Kostiainen, M. A.; Hiekkataipale, P.; Laiho, A.; Lemieux, V.; Seitsonen, J.; Ruokolainen, J.; Ceci, P. Electrostatic Assembly of Binary Nanoparticle Superlattices Using Protein Cages. *Nat. Nanotechnol.* **2013**, *8*, 52–56.

(17) Simon, A. J.; Zhou, Y.; Ramasubramani, V.; Glaser, J.; Pothukuchy, A.; Gollihar, J.; Gerberich, J. C.; Leggere, J. C.; Morrow, B. R.; Jung, C.; Glotzer, S. C.; Taylor, D. W.; Ellington, A. D. Supercharging Enables Organized Assembly of Synthetic Biomolecules. *Nat. Chem.* **2019**, *11*, 204–212.

(18) Song, W. J.; Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014**, *346*, 1525-1528.

(19) Bailey, J. B.; Zhang, L.; Chiong, J. A.; Ahn, S.; Tezcan, F. A. Synthetic Modularity of Protein-Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2017**, *139*, 8160–8166.

(20) Bailey, J. B.; Tezcan, F. A. Tunable and Cooperative Thermomechanical Properties of Protein–Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2020**, *142*, 17265–17270.

(21) Lanci, C. J.; MacDermaid, C. M.; Kang, S. G.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. Computational Design of a Protein Crystal. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7304–7309.

(22) Forse, G. J.; Ram, N.; Banatao, D. R.; Cascio, D.; Sawaya, M. R.; Klock, H. E.; Lesley, S. A.; Yeates, T. O. Synthetic Symmetrization in the Crystallization and Structure Determination of CelA from Thermotoga Maritima. *Protein Sci.* **2011**, *20*, 168–178.

(23) Padilla, J. E.; Colovos, C.; Yeates, T. O. Nanohedra: Using Symmetry to Design Self Assembling Protein Cages, Layers, Crystals, and Filaments. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2217–2221.

(24) Engilberge, S.; Rennie, M. L.; Dumont, E.; Crowley, P. B. Tuning Protein Frameworks via Auxiliary Supramolecular Interactions. *ACS Nano* **2019**, *13*, 10343–10350.

(25) Guagnini, F.; Antonik, P. M.; Rennie, M. L.; O'Byrne, P.; Khan, A. R.; Pinalli, R.; Dalcanale, E.; Crowley, P. B. Cucurbit[7]uril-Dimethyllysine Recognition in a Model Protein. *Angew. Chem., Int. Ed.* **2018**, *57*, 7126-7130.

(26) Ramberg, K. O.; Engilberge, S.; Skorek, T.; Crowley, P. B. Facile Fabrication of Protein–Macrocycle Frameworks. *J. Am. Chem. Soc.* **2021**, *143*, 1896–1907.

(27) Partridge, B. E.; Winegar, P. H.; Han, Z.; Mirkin, C. A. Redefining Protein Interfaces within Protein Single Crystals with DNA. *J. Am. Chem. Soc.* **2021**, *143*, 8925–8934.

(28) Sakai, F.; Yang, G.; Weiss, M. S.; Liu, Y.; Chen, G.; Jiang, M. Protein crystalline frameworks with controllable interpenetration directed by dual supramolecular interactions. *Nat. Commun.* **2014**, *5*, 4634.

(29) Yang, G.; Zhang, X.; Kochovski, Z.; Zhang, Y.; Dai, B.; Sakai, F.; Jiang, L.; Lu, Y.; Ballauff, M.; Li, X.; Liu, C.; Chen, G.; Jiang, M. Precise and Reversible Protein-Microtubule-Like Structure with Helicity Driven by Dual Supramolecular Interactions. *J. Am. Chem. Soc.* **2016**, *138*, 1932–1937.

(30) Hu, R. T.; Yang, G.; Ding, H. M.; Ma, J. B.; Ma, Y. Q.; Gan, J. H.; Chen, G. S. Competition between Supramolecular Interaction and Protein–Protein Interaction in Protein Crystallization: Effects of Crystallization Method and Small Molecular Bridge. *Ind. Eng. Chem. Res.* **2018**, *57*, 6726–6733.

(31) Li, Z.; Chen, S.; Gao, C.; Yang, Z.; Shih, K. C.; Kochovski, Z.; Yang, G.; Gou, L.; Nieh, M. P.; Jiang, M.; Zhang, L.; Chen, G. Chemically Controlled Helical Polymorphism in Protein Tubes by Selective Modulation of Supramolecular Interactions. *J. Am. Chem. Soc.* **2019**, *141*, 19448–19457.

(32) Arnaud, J.; Trondle, K.; Claudinon, J.; Audfray, A.; Varrot, A.; Romer, W.; Imberty, A. Membrane deformation by neolectins with engineered glycolipid binding sites. *Angew. Chem., Int. Ed.* **2014**, *53*, 9267-9270.

(33) Guagnini, F.; Engilberge, S.; Ramberg, K. O.; Pérez, J.; Crowley, P. B. Engineered assembly of a protein–cucurbituril biohybrid. *Chem. Commun.* **2020**, *56*, 360-363.

(34) Guagnini, F.; Engilberge, S.; Flood, R. J.; Ramberg, K. O.; Crowley, P. B. Metal-Mediated Protein–Cucurbituril Crystalline Architectures. *Cryst. Growth Des.* **2020**, *20*, 6983–6989.

(35) Ramberg, K. O.; Guagnini, F.; Engilberge, S.; Wrońska, M. A.; Rennie, M. L.; Pérez, J.; Crowley, P. B. Segregated Protein–Cucurbit[7]uril Crystalline Architectures via Modulatory Peptide Tectons. *Chem. Eur. J.* **2021**, *27*, 14619–14627.

(36) Alig, A. R. G.; Gourdon, D.; Israelachvili, J. Properties of Confined and Sheared Rhodamine B Films Studied by SFA-FECO Spectroscopy. *J. Phys. Chem. B* **2007**, *111*, 95-106.

(37) Yan, X.; Li, J.; Yang, R.; Li, Y.; Zhang, X.; Chen, J. A new photoelectrochemical aptasensor for prion assay based on cyclodextrin and Rhodamine B. *Sens. Actuators B Chem.* **2018**, *255*, 2187–2193.

(38) Mondal, A.; Jana, N. R. Fluorescent detection of cholesterol using β-cyclodextrin functionalized graphene. *Chem. Commun.* **2012**, *48*, 7316–7318.

(39) Sueishi, Y.; Matsumoto, Y.; Kimata, Y.; Osawa, Y.; Inazumi, N.; Hanaya, T. Characterization of group-inclusion complexations of rhodamine derivatives with native and 2,6-*di*-*O*-methylated β-cyclodextrins. *J. Incl. Phenom. Macro.* **2020**, *96*, 365–372.