Precise protein assembly of array structures

Guang Yang, Libin Wu, Guosong Chen* and Ming Jiang

The assembly of proteins into various nano-objects with regular and periodic microstructures, i.e. protein arrays, is a fast-growing field in materials science. Due to the structural complexity of proteins, reports in this field are still quite limited. In this review, we summarize the recent developments in protein array construction by different driving forces, including electrostatic interactions, metal–ligand interactions, molecular recognition and protein–protein interactions. In line with our particular interest, assemblies driven by molecular recognition are particularly explored. Finally, functionalities of the obtained protein arrays are briefly discussed.

1. Introduction

Proteins, one of the most versatile building blocks in nature, are the basic structural and functional components of all kinds of living organisms. Many natural proteins assemble into periodic structures with different dimensions and morphologies via various non-covalent interactions. For instance, some viruses have precise zero dimensional (0D) assembly structures. One dimensional (1D) microtubules, one of the key components of cytoskeletons, have been demonstrated, consisting of long, filamentous and tube-shaped self-assembled architectures with precise packing of proteins, feathering dynamic formation and deformation. Two-dimensional (2D) S-layer arrays formed from membrane regions of Halobacteria species perform ion and molecule sieving, as well as surface recognition. Therefore the investigation of protein assembly behavior, especially the construction of designed protein supramolecular structures, is not only interesting for bio-nanotechnology, aimed at the production of advanced materials for biological applications, catalysis and template reactions etc., but is also important to understand the formation mechanism of natural protein assemblies, which would form a solid basis for developing new treatment strategies for the undesired protein misfolding in Parkinson’s, Alzheimer’s and Huntington’s disease, etc.

In contrast to other biological building blocks, such as DNA, which has been used to construct all kinds of supramolecular structures with high precision and multiple functionalities, proteins are relatively difficult to fabricate due to their structural complexity and heterogeneity. Compared to DNA and peptide, proteins themselves are small nano-objects with a geometric shape. In nature, protein assembly is mainly mediated by protein–protein interactions (PPI), which in turn are governed by

---

Guang Yang completed his Bachelor’s degree at the College of Chemistry and Materials Science at Anhui Normal University, China, in 2012. He then joined Prof. Ming Jiang and Prof. Guosong Chen’s group for his PhD in the same year. Currently, his primary research interest is protein self-assembly and the self-assembly of amphiphilic compounds based on carbohydrate.

Libin Wu completed his Bachelor’s degree at the College of Chemistry and Materials Science at Hebei Normal University, China, in 2011, and his Master’s degree at the College of Chemistry at Nankai University, China, in 2015. He then joined Prof. Ming Jiang and Prof. Guosong Chen’s group for his PhD. Currently, his main research interest is the construction of biodegradable carbohydrate nanoparticles and their applications.
hydrophobic interactions and geometric shape matching, with the help of electrostatic interactions and hydrogen bonding etc. PPI can be used or manipulated to obtain novel self-assembled architectures after elegant structural design by computer modeling. On the other hand, when “exotic” interactions are introduced, i.e. interactions that normally do not exist in natural proteins, including electrostatic interactions, substrate–ligand interactions and metal coordination, the original PPI needs to be artificially adjusted to promote these “exotic” interactions. Moreover, proteins are a different kind of building block to small molecules in supramolecular chemistry. With their regular geometry and size under extremely low polydispersity, they could be employed as colloidal particles with well-controlled surface electronic properties and hydrophobic regions, which could not be achieved with small molecules. Thus the range of possible self-assembled structures produced by the large family of proteins is large, beyond our expectation.

To date, scientists have approached protein self-assembly roughly via either a biological route 11 or a chemical route. 12 In the biological route, genetically engineered proteins are frequently used, with well-designed geometry and contact surfaces. In the chemical route, the self-assembly of the proteins is mediated by specific non-covalent interactions, created by supramolecular chemistry. The exotic interactions include metal–ligand interactions, electrostatic interactions and host–guest molecular recognition. In recent years, these two strategies have been more regularly combined and merged. Sometimes one cannot easily categorise research on protein assembly into one of the two routes, because tools from both methods are now being used. Thus we may say that this field is just a healthy and fast-growing baby from the marriage of chemistry and biology.

Since protein self-assembly is a broad and varied research field, as proteins may self-assemble by themselves or with other components including small molecules, polymers, nanoparticles, etc., quite a few excellent reviews 13–17 have been published in the literature. In this review, we will focus on the fabrication of artificial protein “arrays”, i.e. proteins self-assembled in a periodic and controllable manner. The fabrication of arrays allows proteins to be packed in a dense, oriented and regular way, which also facilitates structural determination by X-ray crystallography and small angle X-ray scattering (SAXS). Protein aggregates packed without obvious regularity and periodicity will not be included. In this review, the protein arrays recently reported in the literature will be divided into four sections by their main driving forces: electrostatic interactions, metal–ligand interactions, molecular recognition, and PPI.

In line with our particular interest, progress in protein arrays made by designed molecular recognition will be particularly explored. Finally, the functionalities achieved by the above assemblies will be briefly addressed.

2. Representative morphologies formed by protein assemblies

In living organisms, proteins, as one of the most important biomacromolecule families, often exist as sophisticated assembled arrays with different morphologies (Fig. 1a), such as 0D rings 18 (Fig. 1a(I)), catenanes, 19 and cages 20 (Fig. 1a(II)), 1D protein tubes 21 (Fig. 1a(III)) or fibers, 22 and 2D protein arrays 23 (Fig. 1a(IV)) with a variety of biological activities. The fabrication of these unique and elaborate structures requires the combination of multiple non-covalent interactions. Inspired by these fascinating regular protein assemblies with indispensable functions in biological systems, great efforts have been made in laboratories to develop protein assembly arrays with different dimensions via designed non-covalent interactions (Fig. 1b). The fabrication of the assembled arrays depends mainly on the highly selective and directional interactions, and the symmetries of the...
selected building blocks. In principal, if a protein unit has two binding sites or interfaces, linear or cyclic protein assemblies may be produced by the connection of units in a head-to-tail or head-to-head manner (Fig. 1b(I) and (II)). This process is similar to polymerization or cyclization in synthetic chemistry. For instance, a series of low dimensional protein arrays have been constructed via different non-covalent interactions by several groups.15,23,24 Meanwhile, 0D protein cages were fabricated by the fusion of protein oligomers with the required symmetry.25 When more advanced protein assembly structures with diverse architectures (such as 2D arrays, 3D arrays, pseudo-1D helical microtubes, etc.) are desired (Fig. 1b(III)–(V)), two aspects need to be considered: the proper selection of non-covalent interactions, and the suitable position of binding sites on the selected proteins. To date, research efforts have been made to build protein assemblies with highly ordered structures on a large scale via different non-covalent interactions from different predesigned protein building blocks. For instance, helical protein microtubes were constructed by utilizing a metal coordination strategy,26 and the specific interactions between saccharide and lectin.27 2D protein arrays were also developed, in which sophisticated supramolecular chemistry strategies were employed.28 Few 3D protein arrays exist in nature, but there is no doubt that the successful construction of 3D protein arrays would be crucial in gaining a deep understanding of protein structure and function, and would also provide new insight into the development of innovative protein materials with desirable porosity.29 Thus, 3D protein arrays formed by different formation mechanisms via different non-covalent interactions have been fabricated.

3. Protein self-assembly driven by electrostatic interactions

Electrostatic interactions have been widely employed for the self-assembly of proteins with other molecules into ordered nanostructures in solution or solid state. The interaction can be easily employed because different proteins have different isoelectric points (pI), which means that at a given pH in aqueous solution, different proteins may possess different net charges. When combining with other materials bearing the opposite net charge, such as linear polymers, dendrimers, micelles and inorganic nanoparticles, the aggregation of the protein and these materials is expected. In the literature, regular morphologies of assembled proteins have been achieved over the past decade using the electrostatic interactions between proteins and other molecules or nanoparticles.30 However, the periodic packing of proteins could not be easily achieved in these structures until very recently, when array structures formed from self-assembled proteins have been achieved. To build such structures, some basic conditions are required. For instance, the size and shape of two or more building blocks with opposite charges need to match well, especially the charge distributions. It is known that the electrostatic interaction between the building blocks can be tuned by the electrolyte concentration and pH of the solution, so the assembly and disassembly of proteins via electrostatic interactions can be controlled by the electrolyte concentration and pH.

Array structures formed from self-assembled proteins driven by electrostatic interactions are limited to several kinds of model proteins. The most common of these is Cowpea chlorotic mottle virus (CCMV), the first sphere-like virus that can be assembled in vitro. The successful fabrication of CCMV in vitro provides an excellent model to better understand the assembly, disassembly and stability of the capsid virus. CCMV (pI ~ 3.8) is stable at pH 3–6, and can form protein particles with an icosahedral capsid structure 28 nm in diameter containing 180 identical subunits.31 The CCMV particle clearly shows negatively charged regions across its surface in neutral solution. In fact, CCMV has been well known as a supramolecular building block for many years, forming delicate self-assembled structures with other components possessing the opposite charge.32 Another building block which has been reported for the assembly of protein arrays is apo ferritin (aFT) from Pyrococcus furiosus. At neutral pH, aFT carries a net negative charge (pI ~ 4.5), which forms a patchy electrostatic potential map on the surface of the protein cage (diameter 12 nm).33 To fulfil the above-mentioned criteria for protein assembly driven by electrostatic interactions, positively charged building blocks with a much smaller size have been used to co-assemble with CCMV or aFT. These building blocks can be either soft or hard, including the dendrimer polyamidoamine (PAMAM), avidin, gold nanoparticles (GNP) etc.

Kostiainen and coworkers made significant contributions to this field by fabricating three dimensional protein arrays with different packing parameters through electrostatic interactions. Several 3D protein arrays have been constructed using CCMV (Fig. 2a) or aFT (Fig. 2b) as the major component, with various smaller positively charged components. In one typical example, tetramer protein avidin (pI 10.5, $M_w = 68k$), with a cross-section of about 7.2 nm and a hydrodynamic diameter of about 5.4 nm, was employed to co-assemble with CCMV.34 Avidin was selected because the structure features four positively charged regions with tetrahedral geometry under neutral conditions. The unit
cell of the array structure formed from CCMV and avidin was confirmed to be bcc (body-centred cubic) by SAXS and high resolution cryogenic electron microscopy (cryo-EM), which showed that the CCMV formed the unit cell frame, leaving empty spaces in a tetrahedral and octahedral arrangement with diameters of 11 nm and 7 nm, respectively (Fig. 2a(I)). The avidin was then located at the edges of the empty spaces, acting as a bridge between the CCMV particles, stabilising them through the electrostatic interactions between the oppositely charged regions in avidin and CCMV. Since the array was mediated by electrostatic interactions, the assembly and disassembly of the array can be reversibly controlled by altering the electrolyte strength of the solution, such as concentration of NaCl. The protein array can only be formed under certain concentrations of NaCl; otherwise, irregular aggregates or no aggregation will be observed. To demonstrate the effect of avidin on the unit cell of the CCMV assembly, PAMAM was employed. PAMAM is a positively charged single molecular nanoparticle with a diameter of $6.5 \text{ nm}$, which is similar to that of avidin. Unlike avidin, the cationic surface of PAMAM is not patchy, and the object can be viewed as a globular, evenly charged nanoparticle. It was found that the co-assembly of aFT and PAMAM formed a structure with a face-centred-cubic (fcc) unit cell $^{35}$ (Fig. 2b), as opposed to the bcc structure of CCMV/avidin. SAXS clearly showed the different packing parameters of the assemblies. This result demonstrated that the open bcc structure of the CCMV/avidin array is due to the patchy interactions, and not an effect of the size ratio between the two proteins. Similarly, both AuNPs (diameter 8.5 nm)$^{36}$ and amphiphilic dendron micelles,$^{37}$ both of which are spheres with evenly positively charged surfaces, formed the same fcc structure with CCMV (Fig. 2a(II) and (III)).

1D structures can also be formed by electrostatic interactions. Liu and coworkers reported a series of 1D arrays made from proteins and synthetic nanoparticles. A ring shaped model protein (SP1) isolated from aspen plants ($P. tremula$),$^{38}$ is a cricoid protein constructed spontaneously from 12 subunits tightly packed together via hydrophobic interactions. The protein ring, with an outer diameter of about 11–12 nm and an inner diameter of about 2–3 nm, showed negative charges on both the top and bottom surfaces of the ring, while mainly positively charged residues were distributed on the sides. The surface electrostatic potential of SP1 is shown in Fig. 3. Such heterogeneous charge distribution gives rise to the possibility of constructing sandwich-like protein arrays by introducing positively charged nanoparticles with a size compatible with the negatively charged areas of SP1. Synthetic particles, including rigid CdSe quantum dots (QD)$^{39}$, soft PAMAM dendrimers$^{40}$ and core cross-linked micelles (CCM),$^{41}$ with positive charges on the surface and a similar size to the SP1 hole, combined with the SP1 in an alternating manner, forming 1D hybrid protein nanoarrays (Fig. 3). These array structures were well characterized by atomic force microscopy (AFM) and transmission electron microscopy (TEM) with negative staining. In the TEM images (Fig. 3(I)–(III)), the short white lines indicated SP1 after negative staining, and

---

**Fig. 2** Electrostatically-mediated 3D protein arrays. (a) Formation and characterization of 3D protein arrays between CCMV and (I) avidin, (II) gold nanoparticles (GNP), and (III) amphiphilic dendron micelles. The TEM images of the arrays are adapted from ref. 34 and 36 with permission from Nature Publishing Group, and ref. 37 with permission from the American Chemical Society. The insets of the TEM images show cartoons of the unit cells. (b) Arrays formed from aFT and PAMAM. The TEM image of the arrays is adapted from ref. 35 with permission from the American Chemical Society. The inset of the TEM image is a cartoon of the fcc unit cell.

**Fig. 3** Electrostatic interaction-directed formation of 1D protein arrays: the charge distribution of SP1 and the formation of different 1D protein arrays between SP1 and CdSe QD, PAMAM and CCM. TEM images of the 1D protein arrays of SP1 with (I) CdSe QD, (II) PAMAM and (III) CCM. The images are adapted from ref. 39–41 with permission from the American Chemical Society.
showed that the distance between two SP1 proteins was in good agreement with the diameter of the synthesized nanoparticles.

4. Protein self-assembly driven by metal–ligand interactions

In nature, metal coordination plays an important role in both the structure and function of proteins, especially in metalloproteins.42,43 Certain types of amino acid residues are capable of binding different metal ions. Compared to electrostatic interactions, metal–ligand interactions are more directional and have high binding ability and good reversibility. Moreover, control of the specific binding sites of metal coordination gives a greater number of possibilities for building novel structures compared to electrostatic interactions. However, proteins are complex building blocks with large size, static conformations and inevitable interfaces. Tuning the inherent interactions between proteins with the introduced metal–ligand interaction is the first obstacle.15 Thus the proper design of the type, strength and position of the metal coordination, as well as the protein structures, including size, shape, surface hydrophobicity and electron density, are crucial to achieving a successful outcome.

In this field, Tezcan and coworkers performed some pioneering work. The building block they used, cytochrome (Cyt), is a common protein material in assembly studies. The group spent several years studying possible oligomers formed by Cyt variant cb562, a stable four-helix bundle with a rigid and cylindrical shape. They modified cb562 with particular ligands which bind with metal ions, including Cu2+, Ni2+, Zn2+ etc., to organize cb562 into several discrete protein oligomers with different shapes via the coordination between the metal ions and the introduced ligands.44–46

Recently the group made a breakthrough, demonstrating that properly introduced Zn2+ binding ligands on the surface of cb562 direct the protein to assemble into 1D helical tubes, and 2D and 3D crystalline arrays.47 The well-designed cb562 protein has two sets of ligands for the Zn2+ ion with different binding abilities, i.e. two metal-chelating bis-histidine motifs (His59/His63 and His73/His77) which bind more strongly, and two pairs of alanine and aspartic acid motifs which bind less strongly. One set of ligands binds one Zn2+. So when 1 equiv. Zn2+ ions (calculated for the strong ligand set) was added into the protein solution, a dimer of cb562 was first formed via the binding of the strong ligands. When the concentration of Zn2+ ions was increased to a higher level (>1.5 equiv.), the rest of the ligands on the cb562 surface with lower binding abilities began to chelate with Zn2+, forming 2D arrays (Fig. 4a). After the 2D array formed, two further directions of assembly were possible: when the Zn2+:cb562 ratio ([cb562] = 50 μM) reached about 100, or pH ~ 8.5, the 2D arrays rolled into helical protein tubes; when the Zn2+:cb562 ratio was lower (2 ≤ ratio ≤ 20) or with a lower pH (5.5 ≤ pH ≤ 7.5), the 2D arrays tended to stack into a 3D structure. Here the concentration of Zn2+ ions controlled the nucleation rate of the 2D assembly. When the nucleation rate was faster (high Zn2+ concentration), the 2D array tended to grow larger and finally rolled into slim 1D protein tubes with a length on the μm scale. When the nucleation rate was slower (low concentration of Zn2+), the stacking of the 2D arrays was faster than their growth, and 3D arrays formed from the stacking of the 2D arrays. TEM images with negative staining and cryo-EM reconstruction showed that the observed 2D packing in the wall of the nanotube and a single layer of the 3D array was almost the same, confirming that they shared the same original 2D array.
which indicated the possibility of this nucleation rate-controlled mechanism. The nucleation rate can also be controlled by pH of the aqueous solution, resulting in similar morphology control. Interestingly, the 3D protein arrays eventually formed a single crystal, the atomic structure of which was resolved by X-ray crystallography, presenting solid evidence for the mechanism. Moreover, the 1D protein nanotubes were found to completely convert to the thermodynamically more stable 2D protein arrays after heating, showing the valuable stimuli-responsive behaviour of the protein assemblies. Finally, rhodamine was incorporated at the surface of cb562. Once the 2D arrays formed, the formation of rhodamine dimers induced the 2D arrays to stack into 3D arrays. Most recently, by using a similar strategy, 2D wavy protein assembly structures were constructed from the V-shaped protein SMAC, which is obtained from the mitochondria of human beings and plays a key role in the apoptosis process.49

Aida and coworkers chose the protein GroEL, a barrel-shaped tetradecameric protein assembly with an inner diameter of 4.5 nm, to construct 1D protein tubes via metal-directed self-assembly. In this work, the opening of the GroEL barrel was site-specifically modified with a number of photochromic spiropyran motifs by a combination of genetic engineering and chemical modification.50 It is known that after irradiation with ultraviolet light (UV) (wavelength <400 nm), spiropyran, which does not chelate Mg2+, can be transformed into merocyanine, which on the contrary, is able to chelate Mg2+ forming a 2 : 1 complex.51 Thus 1D protein tubes were constructed from the modified GroEL in the presence of Mg2+ under UV irradiation. The array of the protein barrel along the axis of the tube could clearly be seen under negative-staining TEM, with the barrels represented by short white lines (Fig. 4b). Upon further visible light irradiation, merocyanine was transformed back to spiropyran, losing the binding ability to Mg2+, so the protein tubes dissociated. The protein tubes were also shown to be responsive to and controllable by ethylenediaminetetraacetic acid (EDTA).52 More importantly, hydrolysis of intracellular adenosine-5-triphosphate (ATP) into adenosine-5′-diphosphate (ADP) was found to induce conformational changes of GroEL protein, which generated a mechanical force that led to disassembly of the tube. This scission occurred with a sigmoidal dependence on ATP concentration, meaning that the tube can differentiate biological environments in terms of the concentration of ATP.53

The strategy of metal-directed protein self-assembly was also skillfully used by Liu and coworkers to construct fascinating protein assemblies. In their reports, glutathione S-transferase (GST), a homodimer which catalyzes the conjugation of reduced glutathione (GSH) to electrophilic substrates for detoxifying endogenous cytotoxic compounds, was selected as a model protein building block. Each monomer of the GST homodimer was modified via genetic engineering techniques with one N-terminal His-tag, which bound with Ni2+ forming a 2 : 1 complex, i.e. two His-tags from different monomers of GST bind with one Ni2+ ion. Thus in the presence of Ni2+, the modified GST formed assemblies via coordination of the His-tag to Ni2+.54 It can be imagined that, when the two His-tags on the surface of GST were positioned at the two asymmetrical N-termini in opposite directions, GST can self-assemble into 1D linear protein arrays due to the coordination of the two His-tags from neighboring protein motifs and Ni2+ (Fig. 5a). The formation of the 1D linear protein array was confirmed by AFM, which showed a nano-array several μm in length and about 4.9 nm in height. It was found that the sphere-like modified GST particles were apparently divided by spacers, revealing that the modified GST dimers were bridged and assembled by the specific interactions between Zn2+ and His-tags. When the two His-tag side chains were placed on the shoulders of the dimer GST, forming a “V” shape from the side view which was perpendicular to the C2 axis of the dimer, protein nanorings were obtained accordingly (Fig. 5b). Diameters of the protein nanorings can be either structurally tuned by altering the size of the “V” shape opening, or tuned by the concentration of NaCl, which adjusted PPIs additional to the Zn2+/His-tag interaction, resulting in different distances between proteins.55

5. Protein self-assembly driven by molecular recognition

In nature, small molecular ligands can be specifically recognized by some proteins, interactions which are crucial to the functions of the proteins. For example, the molecular recognition between carbohydrate and lectin is adopted by pathogens for their adhesion to host tissues. In laboratories, the specific binding between carbohydrates and lectins was well characterized by X-ray crystallography,56 ITC,57 SPR58 and QCM59 etc. Notably, some lectins exist as oligomers, including dimers, trimers, tetramers etc., which could also be controlled in the proper pH range. So some lectin oligomers can be introduced into protein arrays via their specific recognition to synthetic or natural ditoxy carbohydrate ligands. The research was first started at the end of the last century. A beautiful direct evidence of lectin self-assembly mediated by carbohydrate came from Freeman and coworkers.60 Lectin Concanavalin A (ConA), which is a homotetramer with a tetrahedral shape and D2 symmetry under neutral conditions,
was selected as a model protein. There are four sugar binding sites at the four vertices of the ConA tetrahedron, which allow ConA to bind α-D-mannopyranoside or α-D-glucopyranoside just like a sp³ hybridized carbon atom. Upon the addition of a ditopic mannopyranoside, ConA formed diamond-like 3D crystalline protein arrays as a result of the ditopic sugar linking neighbouring ConA units (Fig. 6a). The crystalline array, with a distance of about 6.9 nm between the centers of neighboring protein molecules, was observed by TEM with negative staining. X-ray crystallography analysis of the single crystal assembly was performed under a resolution of 6 Å, revealing a pseudo-cubic orthorhombic unit cell (space group P2₁2₁2₁) with an average cell dimension of 204 Å. The result indicated the potential of native protein molecules to be employed as building blocks for self-assembly without any chemical or biological modification. Several years later, a similar approach was used by Imberty and coworkers. They synthesized a series of ditopic mannopyranosides, in which a branched trisaccharide 2-Man1-3(2-Man1-6)Man (triMan) in Fig. 6b) showed satisfactory results forming protein arrays with lectin BC2L-A, which exists as a homodimer under neutral conditions with two sugar binding sites along the C₂ axis in opposite directions. AFM showed that the mixture of BC2L-A and triMan formed a linear nanowire structure. Moreover, the structure of the assembly was further confirmed by X-ray crystallography, showing the 1D array of BC2L-A. Within the arrays, the two stretched mannopyranoside ends from triMan bind with neighbouring BC2L-A, which linked the lectin in a 1D array.

In the examples using ditopic molecules mentioned above, the two saccharide units were linked via covalent bonds, thus the formation mechanism of the 1D or 3D array was similar to that of condensation polymerization, in which a strictly controlled 1:1 ratio of protein binding sites to sugar motifs was required. This criterion limited the obtained morphologies and the assembly speed as well. In fact, the regular structures achieved in the above examples were the minority of cases. When further variations on the ligand were performed, precipitates of proteins without regular structures were formed. Thus the research field remained quiet for more than a decade. Very recently, Chen and coworkers found that by using the classical binding between ConA and mannopyranoside, but simply replacing the covalent linkage between the ditopic sugar with a non-covalent linkage, i.e. π-π stacking between rhodamine B (RhB), regular protein arrays were achieved. The chemical structures of the key compounds in the study are shown in Fig. 7a.

The dimerization of RhB has been known for decades; it is concentration-dependent and can also be controlled by salt concentration and other factors. Chen et al. found that under a certain concentration (<0.6 mM), dimerization of α-D-mannopyranoside-modified RhB (Rh3Man, Fig. 7a) did not occur. However, when ConA was added to the solution of free Rh3Man, which has a specific binding ability to ConA, crystals of ConA and Rh3Man formed immediately and precipitated from the solution. The crystals were square-shaped, with a diameter on the μm scale. The size of the crystals could even be controlled using different salt concentrations and temperatures, with the largest reaching a size in the range of 100 μm (Fig. 7b). The crystals showed a platelet shape with a height as low as 200 nm. The results of UV-vis and circular dichroism (CD) spectra of the dissolved crystals indicated that dimerization of RhB happened under these low concentrations, facilitated by the binding of ConA with Rh3Man. Results of X-ray crystallography at a resolution of 2 Å indicated that ConA first packed regularly as a 2D lattice structure, which was mediated by the binding of ConA/Rh3Man and the π-π stacking of RhB, i.e. the dimerization of RhB drove the Rh3Man molecule to form a bridge, linking the neighbouring ConA into 2D lattice (Fig. 7c). The 2D lattice then packed regularly into a 3D crystal structure (Fig. 7d and e). The authors emphasized that the π-π stacking of RhB happened only on the surface of ConA, which served as an inducing ligand for the formation of the
protein crystal (Fig. 7f), and this mechanism explains why the conversion of ConA to ConA/Rh3Man assembly has a positive correlation with the concentration of Rh3Man (from 0.1 to 0.5 mM), fixing the concentration of ConA monomer at $1 \times 10^{-4}$ M (Fig. 7g).

By using X-ray crystallography, it was found that in the crystal, ConA/Rh3Man formed a porous framework (Fig. 7h), similar to the protein framework driven by metal coordination. However, by using ligand Rh4Man with slightly longer tether length between ConA and Man, two identical frameworks were achieved in the same crystal structure, which are interpenetrating to each other (Fig. 7i). This is the first finding of the interpenetrating protein framework, which is achieved by delicately adjusting the chemical structure of the small molecular ligand. By using $\pi$–$\pi$ stacking together with lectin/sugar interaction, in this work, it turned out that there are three advantages can be emphasized compared to the previous covalent systems: (1) Morphology of the protein assembly was carefully tuned. (2) Conversion ratio of free proteins to self-assembled ones was dramatically improved with more equivalent of the inducing ligand employed. (3) More importantly, crystallization speed, or self-assembly speed was also accelerated dramatically by using extra amount of the ligand.

Recently, the strategy of using a combination of sugar/lectin interactions and $\pi$–$\pi$ stacking has been extended to more lectins. Chen’s group selected soybean agglutinin (SBA) as a building block, which has four binding sites to N-acetyl-$\alpha$-D-galactosamine (GalNac) or $\alpha$-D-galactopyranoside (Gal). The tetrameric SBA exhibited $D_2$ symmetry (Fig. 8a) with a slightly twisted planar shape, and did not form any tubular structures itself or with covalently connected ditopic GalNac/Gal molecules. (f) One tetrameric SBA model fitted in the electron density map. (g) Four adjacent SBA tetramers with four pairs of R3GN dimers shown as spheres. (h) Close-up view of two pairs of the R3GN dimer. Adapted from ref. 27 with permission from the American Chemical Society.
present within one helical period (Fig. 8d–f). In one protofila-
mament, the neighbouring SBAs were linked via the dimerized
R3GN as a bridge (Fig. 8g and h). The conclusion was not only
supported by computer simulation but also by the results from
UV-vis and CD spectroscopy. Similar to the previous crystalization,
by using a higher ratio of the ligand, the tube length was found to
increase by cryo-EM. Finally the tube was disassembled by the
addition of cyclodextrin, which formed a host-guest complex with
RhB, and re-assembly of the tube was achieved after further
addition of adamantane, which competed with RhB forming a
more stable complex with cyclodextrin.

Non-covalent interactions employed in protein assembly
are of course not limited to π–π stacking. Supramolecular
chemistry provides us with a plethora of interactions, including
hydrogen bonding, host–guest interactions etc., to be considered.
Recently, host–guest interactions based on cucurbit[8]uril (CB[8])
were employed by Liu and coworkers.66 The short oligopeptide
chain Phe–Gly–Gly (FGG) was engineered to the surface of the
GST dimer in opposite directions. It is known that CB[8] forms
1 : 2 complexes with Trp–Gly–Gly (WGG) or Phe–Gly–Gly (FGG),
with binding constants up to 10^9–10^11 M^−2. The high binding
constant makes protein assembly in dilute solution possible. As
designed, a 1D protein array was formed when CB[8] connected
with the FGG group from neighbouring GST. They also found
that the length of the array was controlled by the molar ratio of
CB[8] and FGG–GST dimers.

Ligand-mediated protein assembly has been proved success-
ful not only in solution, but also at a solid–liquid interface.
Protein crystallization at the solid–liquid interface has been a
powerful method for quite a long time,67 where small molecu-
lar ligands induce the formation of a 2D protein crystalline
array. Here we present just one example. Engel and coworkers
reported streptavidin 2D arrays via the specific recognition
between streptavidin and biotin on a highly oriented pyrolytic
graphite (HOPG) surface (Fig. 9).68 Streptavidin is a tetramer
protein with four binding sites, which can specifically interact
with biotin with a high binding affinity of about 10^8 M^−1. In this
work, biotinylated lipid molecules first formed a monolayer at
the air/water surface (Fig. 9a), which induced streptavidin to
pack regularly via the recognition between biotin and streptavidin
(Fig. 9b). Two of the binding sites of streptavidin bound to the
biotin on the lipid monolayer, while the other two binding sites
were immersed in water. Finally, the 2D array of protein and the
monolayer of biotinylated lipid were carefully transferred onto the
HOPG surface for characterization (Fig. 9c and d). The combi-
nation of high resolution AFM and TEM clearly revealed the
formation of a 2D protein array structure (Fig. 9e and f). This
method is an important strategy for the construction of 2D protein
arrays as new nanomaterials.

6. Protein self-assembly driven by PPI

It is well known that PPIs present at the protein–protein inter-
face play a key role in cellular complexity.69 Many cellular events,
such as the mobility of the cell, information communication
and the transport of energy etc., are controlled by transient
PPIs. Additionally, permanent PPIs are fundamental for the
framework of cellular machinery. Now scientists agree that the
contact surfaces of PPI should be complementary and
predominantly nonpolar, with a contact area larger than about
600 Å^2. Meanwhile, the contribution of hydrogen bonds and
salt bridges at the contact rim is negligible.70 In order to mimic
nature, scientists have tried to design PPIs to be the dominant
interaction between proteins, by which method protein assem-
bles with delicate structures were achieved.71 Efforts in this
field have been made by using mutant proteins. Many proteins
in nature exist as oligomers with certain symmetry, i.e. the
proteins are a natural assembly of subunits. The PPIs between
these subunits are normally strong enough to induce them to
assemble into the oligomer. These PPIs have been utilized in
protein assembly aiming at regular structures in the nm or
µm scale. For example, two subunits from different protein
oligomers can be fused, forming a new building block which has
two different PPI interfaces. The new building blocks
will interact with each other via PPI. When the subunits are
carefully selected with certain rotational symmetry, different
architectures could be achieved. Yeates and coworkers have
constructed several kinds of protein assemblies, such as filaments,
cages and layers, by taking advantage of protein engineering and
their inherent PPIs.25,71 Most recently, several protein cubes have
been constructed by fusing a monomer of trimeric aldolase KDPGal
(PDB code: 2V82) and a monomer of dimeric protein FkpA (PDB ID:
1Q6U) together via a short alpha-helical tether (Fig. 10a(I)).71
This new heterodimer formed a cubic cage with octahedral sym-
metry via PPI of the original trimer and dimer (Fig. 10a(II)).
The cage has precise molecular weight, as measured by mass
spectrometry (MS) and tandem mass spectrometry (MS/MS), and a

Fig. 9 Biotin-directed streptavidin array formation on a HOPG surface.
(a–d) The formation process of the streptavidin arrays. (e) AFM image of
the array, scale bar: 50 nm. (f) TEM image of the array, scale bar: 100 nm.
Adapted from ref. 68 with permission from Wiley-VCH.
7. Functionalities achieved by protein self-assemblies

Protein arrays do give immediate access to numerous functionalities because of their almost exact structure and the character of the protein units. With the fast development of electron microscope technology, protein arrays present emerging possibilities for protein structure determination, which may be treated as powerful accompaniment to X-ray crystallography. Benefitting from the abundance of functional groups on the protein surface, which bind with some metal ions, and the regular space within the arrays, the protein arrays can be used as a template for the synthesis of specific inorganic nanoparticles. Protein itself has a lot of functionalities in catalysis, imaging, etc. More interestingly, some groups found that protein arrays have better or even new functionality than the building blocks themselves before assembly. For instance, Liu et al. found that when the GST dimers were linked into 1D nanoarrays, the arrays showed better performance for catalysis than the dimer itself. The linear array of GST–quantum dot hybrids showed better light harvesting performance than the two components individually. Chen et al. found that a SBA protein tube showed a more pronounced immunological effect than the SBA tetramer. Aida et al. utilized the protein tube as a drug or protein carrier for drug delivery, triggered by ATP. Probably the most exciting result came from the 3D crystalline array formed by metal–ligand interactions. Proteins inside the arrays have much higher stability in organic solvent than in the free state, and the protein array also remains stable at high temperatures up to 80–90 °C. Finally, this protein array showed excellent redox activity for the synthesis of inorganic materials.

8. Perspectives

Recent progress has been made in the young field of constructing artificial protein arrays with different dimensions via different kinds of non-covalent interactions. Compared with protein aggregates without obvious regularity and periodicity, the structure of protein arrays with a high degree of order can be identified even at the atomic scale, which on one hand facilitates the determination of the packing mode in arrays, and on the other hand inspires further design and control of the packing. In this article, four kinds of non-covalent methods were introduced in detail: electrostatic interactions, metal–ligand interactions, molecular recognition, and PPI. As we know, in nature protein array formation is driven by multiple non-covalent interactions, however to date the driving forces used to construct protein arrays in laboratories are still limited. One important reason for the limited functionalities achieved by protein arrays thus far is the limited kinds of protein building blocks that have been used. Thus, the next step in this field should be exploration of more non-covalent interactions as driving forces, which could be used simultaneously or synergistically. Meanwhile, the development of more building blocks, especially functional proteins, is also important.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 91527305, 91227203, 51322306 and GZ962),