ChemPubSoc Europe
DOI: 10.1002/cphc.201402722
CHEMPHYSCHEM Minireviews

Self-Assembled Cage-Like Protein Structures

Rindia M. Putri, Jeroen J. L. M. Cornelissen, and Melissa S. T. Koay*[a]

Proteins and protein-based assemblies represent the most structurally and functionally diverse molecules found in nature. Protein cages, viruses and bacterial microcompartments are highly organized structures that are composed primarily of protein building blocks and play important roles in molecular ion storage, nucleic acid packaging and catalysis. The outer and inner surface of protein cages can be modified, either chemically or genetically, and the internal cavity can be used to template, store and arrange molecular cargo within a defined space. Owing to their structural, morphological, chemical and thermal diversity, protein cages have been investigated extensively for applications in nanotechnology, nanomedicine and materials science. Here we provide a concise overview of the most common icosahedral viral and nonviral assemblies, their role in nature, and why they are highly attractive scaffolds for the encapsulation of functional materials.

1. Introduction

The controlled assembly of highly ordered structures, architectures and systems, using simple molecular building blocks has been of growing interest for applications in nanotechnology, synthetic biology and material science.[1–13] At the interfaces between chemistry, biology and engineering, proteins and other biomolecules that can reversibly assemble have been shown to be extremely useful for controlling the self-assembly and self-organization of molecules on the nanometer scale. A classic example is that of protein cages, such as viruses, bacteriophages, bacterial microcompartments, ferritins and heat shock proteins, which are highly abundant in nature and play crucial roles in cargo delivery, molecular storage and catalysis.[14–22] Virus capsids and bacteriophages in particular have been of great interest owing to their unique mechanical and physical properties in nature: they must be sufficiently stable to protect their genome in the extracellular environment, yet sufficiently unstable in order to release their genome into host cells. For example, bacteriophages such as ϕ29 and λ, which encapsulate highly packed DNA genomes, are capable of withstanding internal pressures of up to tens of atmospheres without rupturing. This remarkable feature has sparked a significant amount of research into the mechanism of genome release, driving developments particularly in the fields of atomic force microscopy nanoindentation, which has since proven to be a highly powerful technique to study the mechanical properties of protein cages and viruses. For their applications in nanotechnology, the outer and inner surface of protein cages can be modified, either chemically or genetically, and the internal cavity can be used to template, store and arrange molecular cargo within a defined space (Figure 1). Importantly, because many viruses and bacteriophages are composed of multiple copies of identical proteins that assemble into highly symmetrical structures, functionalization at a single amino acid position of the monomer is translated over the entire assembly, (i.e. for a triangulation number $T=3$ virus capsid, one mutation of the monomer leads to 180 identical mutations that are positioned symmetrically over the entire icosahedron). This can be highly advantageous for applications, particularly in cell recognition and molecular targeting, because a single modification can introduce multivalency effects. Although the library of assemblies is highly diverse, spanning a range of size and length dimensions of 14–880 nm, we provide here an overview of the most

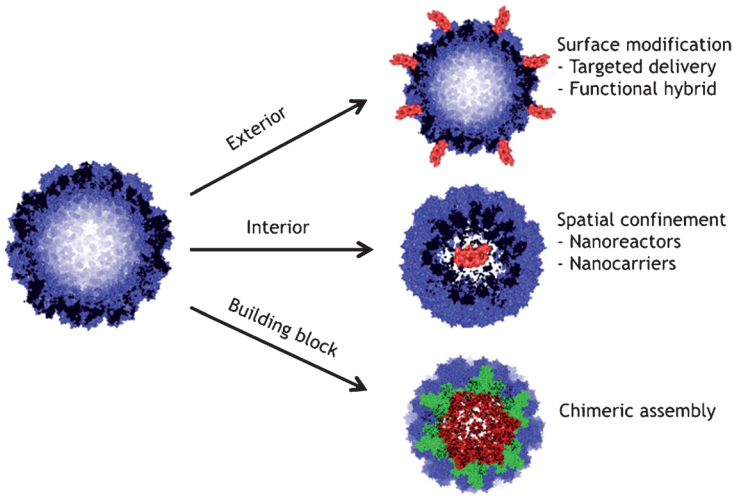


Figure 1. Protein cages as versatile platforms. The exterior can be functionalized with small molecules, ligands, or peptides, whereas the interior can be used for the encapsulation of molecular cargo. In some cases, the protein cage can be disassembled and used as controlled self-assembling building blocks.

[a] R. M. Putri, Prof. J. J. L. M. Cornelissen, Dr. M. S. T. Koay
Department of Biomolecular Nanotechnology
MESA+ Institute for Nanotechnology
University of Twente, P.O. Box 17
7400AE Enschede (The Netherlands)
E-mail: m.s.t.koay@utwente.nl

common icosahedral protein cages, with a particular focus on viruses and encapsulins. We discuss their self-assembly properties and their potential applications in material science. Note: Here, we define self-assembling protein cages as naturally occurring cages that spontaneously assemble into highly symmetrical, monodisperse particle-like structures, without the assistance of additional helper proteins or peptides. In many cases, assembly is driven by changes in pH and/or ionic strength.

2. Bacteriophages and Viruses

Bacteriophages are viruses that specifically target and infect bacterial cells, and one of the most abundant types of organisms found in nature. Inherently programmed to infect bacteria, bacteriophages exist in various shapes and morphologies and typically have one of two different life cycles: the active lytic cycle in which bacteriophages infect and kill their host cells, or the dormant lysogenic cycle during which bacteriophages are integrated and co-exist in their host genome until activated.[16] The so-called tailed bacteriophages are the most common class found and are composed of a head that contains genetic material and a tail that plays an important role during host invasion. In comparison, isometric bacteriophages lack the tail component and are highly reminiscent of icosahedral viruses (Figure 2). Due to their ability to target specific receptors on the surface of bacteria, bacteriophages have been investigated extensively as molecular vectors for gene transfer, as diagnostic tools and as novel therapeutic agents.[24–28] In nanotechnology, the outer protein shell of bacteriophages have served as highly versatile building blocks, and have found application in peptide (phage) display and as scaffolds for the fabrication of nanowires, electronic and photochemical devices.[25,29,30]

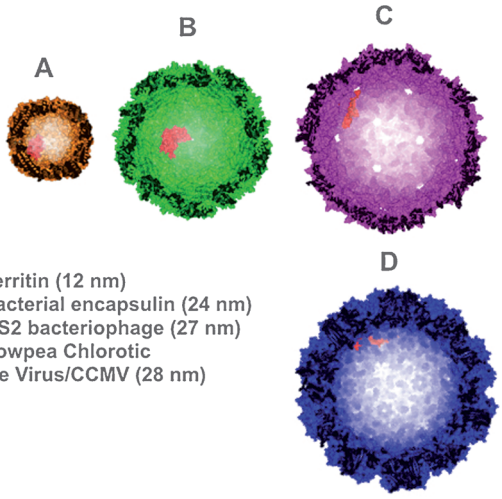
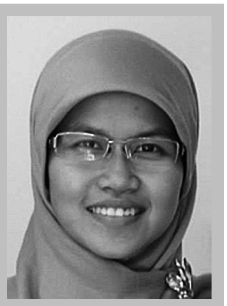


Figure 2. Size and morphology of the most commonly used protein cages in nanotechnology (protein cages are not drawn to scale).

Icosahedral viruses and bacteriophages are assembled according to the Caspar–Klug quasi-equivalence theory in which $60N$ subunits (where N is defined as the triangulation T number) are symmetrically arranged as pentamers and hexamers to form the closed icosahedron shell. The smallest assembly is composed of 60 protein subunits arranged as 12 pentamers to form a $T=1$ capsid, however many other viruses also exist in nature with $T=3$, 4, 7, 12, and 13.[23] Non-enveloped plant and bacterial viruses are two-component systems composed of an outer protein shell (capsid) that surrounds its genomic material (RNA/DNA). In some cases, the capsid can be disassembled in vitro and the genomic RNA/DNA cargo can be precipitated and removed. The capsid proteins can often be reassembled by changing the pH and/or ionic strength to form empty icosahedral structures that resemble the morphology of native viruses.[31] Devoid of their native RNA material, foreign cargo can be effectively encapsulated inside virus capsids during self-as-

Rindia M. Putri received her B.Sc. in chemistry from Institut Teknologi Bandung (ITB), Indonesia in 2011. In 2013, she obtained her masters degrees in chemistry and chemical engineering from ITB and the University of Twente, respectively. She is currently a PhD student in the Biomolecular Nanotechnology group under the direction of Dr. Koay and Prof. Cornelissen. Her research interest is the engineering of protein-based bacterial compartments as biocompatible nanoplatforms.



Prof. Jeroen J. L. M. Cornelissen is Professor of Biomolecular Nanotechnology at the University of Twente. He obtained his Ph.D. in 2001 from the University of Nijmegen, the Netherlands, under the supervision of Prof. Roeland Nolte, and did postdoctoral work at the IBM Almaden Research Center, San Jose, CA. In 2002, he returned to Nijmegen as an Assistant Professor before taking up his present position in 2009. His research interests are in the fields of protein–polymer

Dr. Melissa S. T. Koay completed her Ph.D. at the Bio21 Institute and University of Melbourne, Australia, in the group of Prof. A. G. Wedd. In 2006, she worked in the groups of Prof. W. Lubitz and Prof. W. Gärtner at the Max Planck Institute for Bioinorganic Chemistry, Mülheim an der Ruhr, Germany. She moved to the Eindhoven University of Technology in 2008 to work in the groups of Dr. M. Merkx and Prof. E. W. Meijer. Since 2011, she has been an Assistant Professor at the University of Twente. Her research focuses on the design of protein-based compartments for the synthesis of biologically active molecules.

sembly, resulting in filled virus-like hybrid assemblies, the morphology of which resembles those of native viruses.[32] This ability to reversibly assemble and disassemble protein capsules serves as a powerful approach for the encapsulation of polymers, inorganic templates, small molecules and dyes, proteins and enzymes for all applications in nanotechnology, nanomedicine and materials science (Figure 3).[22,33–37] This versatile approach of cargo loading is of particular interest in the field of chemical physics because the physical and mechanical properties of the protein assemblies are manipulated and tuned by the hardness of the encapsulated cargo (e.g. quantum dots/nanoparticles vs. DNA).

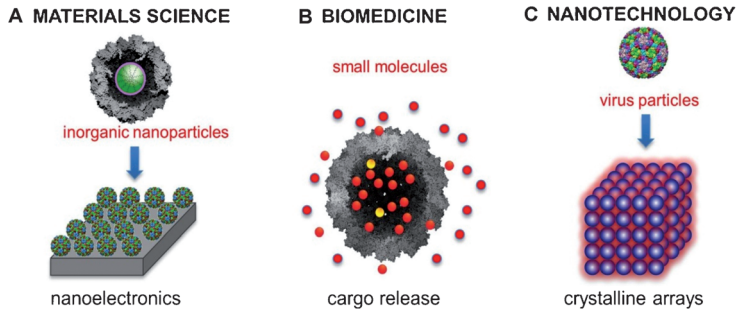


Figure 3. Viruses and bacteriophages for applications in A) materials science for the fabrication of electronic devices, B) biomedicine as cargo delivery vehicles and C) nanotechnology for higher-order assemblies.

2.1. Bacteriophages

MS2 is a small enteric RNA bacteriophage that belongs to the *Leviviridae* family.[38] The native icosahedral MS2 virion is composed of three components: the virus coat protein, a single copy of the maturation protein (protein A) and single-stranded RNA. Interestingly, the MS2 capsid can be reassembled in vitro using a sequence-specific RNA stem loop and acid-disassembled capsid dimers to form 27 nm capsids that are extremely resilient to various solvents, pH values 3–10 and temperatures up to 55 °C, making them highly versatile building blocks for self-assembly.[26,33,35,39–41] Another enteric RNA bacteriophage that also infects *Escherichia coli* is the Qβ bacteriophage, which assembles 90 dimers around a 4.2 kb single strand of RNA to form icosahedral virus capsids of 25 nm in diameter.[42,43] Unlike MS2, which requires the stem-loop structure for assembly, Qβ incorporates helper proteins—the A1 protein and a maturation protein A2—for the formation of infectious virus particles.[42,43] Unlike other viruses and bacteriophages, the assembly of the Qβ bacteriophage depends on the formation of disulfide bonds, which are crucial for covalently linking the monomeric Qβ protein subunits.

Unlike MS2 and Qβ bacteriophages, which encapsulate single-stranded RNA, the P22 bacteriophage encapsulates double-stranded DNA. Procapsid assembly is dependent on the presence of scaffolding proteins, which are co-assembled with monomeric virus coat proteins to form the final $T=7$ icosahedron, consisting of 420 coat proteins and approximately 300 scaffolding proteins. Unlike other bacteriophages, the P22 procapsid exhibits irreversible morphological changes upon heating. Upon heating at 65 °C, the procapsid shell expands from 58 to 64 nm to form the so-called expanded mature capsid. Expansion of the shell causes the simultaneous release of the scaffold proteins from the capsid interior. By further heating to 75 °C, 12 pentamers of the protein shell are irreversibly released, creating 10 nm holes in the so-called "wiffle-ball" capsid assembly.[18,44,45]

2.2. Icosahedral Viruses

Cowpea chlorotic mottle virus (CCMV) is a single-stranded RNA plant virus from the *Bromoviridae* family. The capsid is assembled from 90 dimeric coat protein subunits to form $T=3$ icosahedral capsids with an outer diameter of 28 nm.[19,20,23] The C terminus (residues 186–190) is essential for capsid assembly and forms noncovalent interactions between individual virus protein monomers. The N terminus (residues 1–26) contains nine arginine residues and one lysine residue, and serves as a positively charged template that interacts with the negatively charged viral RNA.[19,23,46] CCMV exhibits well-studied in vitro pH- and salt-driven assembly and disassembly pathways that are similar to those of other icosahedral viruses, such as the brome mosaic virus (BMV) and cowpea mosaic virus (CPMV).[23,47,48] The structures of the native and swollen forms of CCMV have been determined to 3.2 Å resolution and provide a detailed insight into the disassembly and reassembly process.[20] At pH 5.0, the native CCMV is highly stable. Upon raising the pH to greater than 6.5, the native CCMV capsid swells, leading to the formation of 60 pores that are approximately 2 nm in size. At pH 7.5, the stabilizing interactions are disrupted and the CCMV capsid disassembles into 90 dimers and the single-stranded RNA cargo is released. Devoid of its RNA cargo, CCMV is able to form empty virus-like particles at pH 5.0. CCMV has been studied extensively for the encapsulation of various functional materials, such as enzymes, nanoparticles, proteins and polymers, as the size of its assembly could be tuned to form either $T=1$, 2 or 3 assemblies depending on the length and flexibility of the cargo.[3,10,12,31,33,49–53]

3. Nonviral Assemblies

Compared to virus-based assemblies, nonviral protein cages exhibit a relatively narrow distribution of sizes. The most common class of nonviral protein cages are ferritins and ferritin-like proteins; other examples include chaperones, heat shock proteins, lumazine synthases, and more recently, bacterial encapsulins.[10,13,33,34] Despite their similarity in size and morphology, nonviral assemblies are highly abundant in nature and play crucial roles in molecular storage, catalysis and in confining metabolic pathways. In the following section, we focus on ferritins and bacterial encapsulins.

3.1. Ferritins

Iron is an essential trace element that has a crucial role in cellular oxygen transport and electron transfer.[14] However excess iron is also intrinsically toxic due to its ability to catalyze hydroxyl radical formation by Fenton redox chemistry, which leads to cellular oxidative stress. For these reasons, ferritins and the related Dps (DNA-binding proteins from starved cells) play an important role in maintaining intracellular iron levels, by regulating iron storage and release in almost all living organisms.[15,54] Ferritins are a superfamily of spherical protein cages composed of either 12 (mini-ferritins) or 24 protein subunits (maxi-ferritins); the biochemical and structural properties of all ferritins are highly conserved (Figure 4A). Although ferritins contain an insufficient number of subunits to follow the 60 *N* Caspar–Klug quasi-equivalence theory, their structural features are highly reminiscent of icosahedral viruses. Mini-ferritins have an outer diameter of approximately 9 nm and have a hollow cavity approximately 5 nm in diameter that can store up to 500 Fe ions.[21,55,56] Maxi-ferritins have an outer diameter of 12 nm, an inner diameter of 8 nm, and can store up to 4500 Fe ions.[57] Upon entry into the ferritin cage, Fe^{2+} is oxidized to Fe^{3+} by reaction with either O_2 or H_2O_2. Once the inner cavity is filled, a core of hydrous ferric oxide mineral in the central cavity is formed. The pH-driven in vitro disassembly and reassembly of maxi-ferritin has been studied in detail: at pH 6.7, the protein shell disassembles into dimers at ionic strengths below 200 mM NaCl; it can be reassembled when the ionic strength is raised above 600 mM NaCl.[58] Alternatively, ferritins can be disassembled into its dimers by lowering the pH to less than 2, and reassembled by increasing the pH to 7.5.[59]

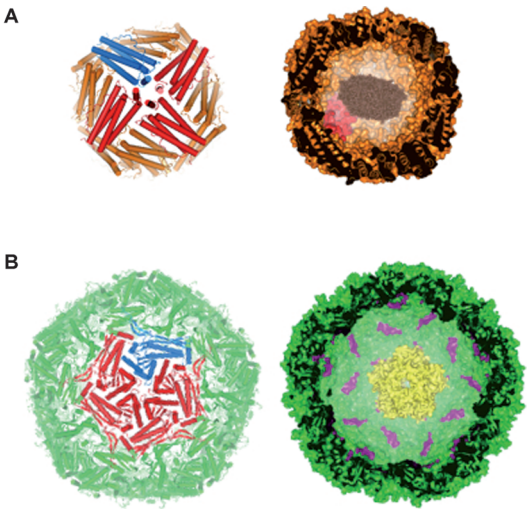


Figure 4. Structural representation of nonviral protein-based assemblies. A) Structure of the ubiquitous ferritin cage. Iron, as Fe^{2+}, enters the central cavity where it is oxidized and then mineralized into a hydrous ferric oxide core (brown). B) Structure of encapsulin from the bacterium *Thermotoga maritima*. The protein cage is composed of 12 pentamers, whereby the monomer (blue) and one pentamer (red) are shown. The protein cargo (yellow) occupies the internal cavity as a result of interactions with a conserved peptide sequence (purple).

3.2. Bacterial Encapsulins

Originally referred to as linocin-like proteins, bacterial encapsulins have been identified in organisms including the hyperthermophilic bacterium *Thermotoga maritima*, which grows at an optimal temperature of 80 °C, and *Brevibacterium linens*, a gram-positive bacterium that is often used for the ripening of cheese.[17,60] The protein shell of encapsulins is composed of 60 monomers that self-assemble in vivo into highly stable 20–24 nm icosahedral structures. In the case of *B. linens*, the internal cavity accommodates dye-decolorizing peroxidase (DyP), whereas the encapsulin from *T. maritima* houses ferritin-like protein (Flp).[60] Interestingly, a conserved peptide was identified on the C terminus of both DyP and Flp, and is thought to be essential for directing enzyme encapsulation inside the protein shell during in vivo assembly (Figure 4B).[17,60,61] Although research involving encapsulins is only in its infancy, these proteins are particularly fascinating as they exhibit several virus- and bacteriophage-like properties: the outer shell assembles into $T=1$ particles according to the Caspar–Klug quasi-equivalence theory, and cargo encapsulation is also similar to that of viruses and bacteriophages (i.e. it is a template-driven process). Interestingly, there is no significant sequence homology to virus-based proteins, and no other genes of viral origin were found in the vicinity of the encapsulin genes, suggesting that they are actually nonviral entities.[60]

4. Template-Driven Self-Assembly

The ability to load viruses and bacteriophages with foreign cargo is of significant interest, particularly for applications in nanomedicine, such as cargo delivery vehicles for small drugs, contrast agents and/or fluorescent dyes, and also for applications in nanoelectronics and optics through the use of monodisperse nanoparticles for enhanced surface plasmon resonance effects.[26,29,49,55,59,62–67] Pioneering work in this research area initially focused on mineralization inside ferritins. In ferritin, hydrophilic and hydrophobic channels provide the means by which iron and other metal ions can diffuse and penetrate into the ferritin core and accumulate. For example, in vitro mineralization can be achieved at room temperature by incubating the empty intact ferritin shell (apoferritin) with aqueous Fe^{2+} solutions.[68,69]

Following the success of in situ nanoparticle mineralization in ferritins and protein cages, research in this area has extended towards directing other synthetic cargo to the interior. In particular, the MS2 bacteriophage has proven to be a highly versatile platform. During the assembly process, 32 pores 2 nm in size are formed around the 27 nm capsid structure. By introducing site-specific cysteine mutations on the interior of the capsid, small molecules such as taxol, Gd-based MRI contrast agents, and fluorescent dyes could be attached.[28,62,66,70]

Further advances have brought about the design of templates and linkers, such as DNA, peptides and scaffolding proteins, that can tether and direct cargo to the interior of protein cages. The advantage of these strategies is that the desired

cargo can be easily exchanged for another, offering a much more modular encapsulation approach.

4.1. DNA/RNA-Driven Self-Assembly

Inspired by the natural loading of DNA and RNA in viruses and bacteriophages, the use of nucleic-acid-based linkers and their mimics has become increasingly popular to direct cargo loading to the interior of self-assembled protein cages. For example, the use of RNA and DNA aptamers has been shown to be a powerful approach that allows a high degree of control and selectivity over in vivo cargo loading into Qβ bacteriophages. Aptamers are molecules of RNA or DNA with short sequences that are identified from in vitro screening. Similar to phage display, RNA and DNA aptamers are selected from random-sequence libraries and their sequences are optimized for high-affinity ligand binding. Aptamers are predominantly unstructured in solution, however, upon association with the target ligand, aptamers fold into well-ordered scaffold structures.[71] Using this "hook–bait" approach, Fiedler et al. cleverly designed an RNA hairpin aptamer that could be used as a baited hook to direct the in vivo packaging of functional cargo into Qβ bacteriophages. In this strategy, the desired cargo is engineered with an arginine-rich peptide (Rev) derived from HIV-1.[43] Isopropyl β-D-thiogalactopyranoside induction promotes the co-expression of the Qβ coat protein, Rev-tagged cargo enzyme, and bifunctional mRNA. The Rev tag binds favorably to the 5′ end of the α-Rev RNA aptamer. At the 3′ end of the same RNA aptamer, the hairpin Qβ genome acts as an anchor, tethering itself to the interior of the Qβ monomers (Figure 5A). This approach was used for the encapsulation of aspartate dipeptidase peptidase E (PepE), firefly luciferase (Luc) and a thermostable mutant of Luc (tsLuc) inside bacteriophage Qβ.[40]

Another approach involves functionalization of short, synthetic DNA strands to desired cargo proteins for the in vitro encapsulation into virus-like assemblies. In this case, the negatively charged DNA serves as a template to promote protein encapsulation inside CCMV during self-assembly (Figure 5B).[72] By designing complementary DNA strands, this strategy proved to be a highly modular approach for the co-encapsulation of a two-enzyme pathway into a single capsid. In this study, a short DNA oligonucleotide was covalently coupled to either the lysine or cysteine residues of one enzyme, and the complementary oligonucleotide was coupled to the second enzyme. DNA hybridization promoted heteromeric enzyme–enzyme association and co-encapsulation. Encapsulation of two-enzyme cascades has been of particular interest in recent years for understanding how (bio)chemical reactions take place in confined spaces (i.e. in cells) and how the "excluded volume effect" influences enzyme kinetics. Note: Minton's excluded volume effect predicted that compared to larger volume in bulk system, the smaller volume occupied by biomolecules inside a cell greatly alters the activity of an enzyme, whereby the smaller available volume due to confinement results in a higher effective concentration of biomolecules within organelles, thereby influencing the thermodynamic activity and kinetic of reactions.

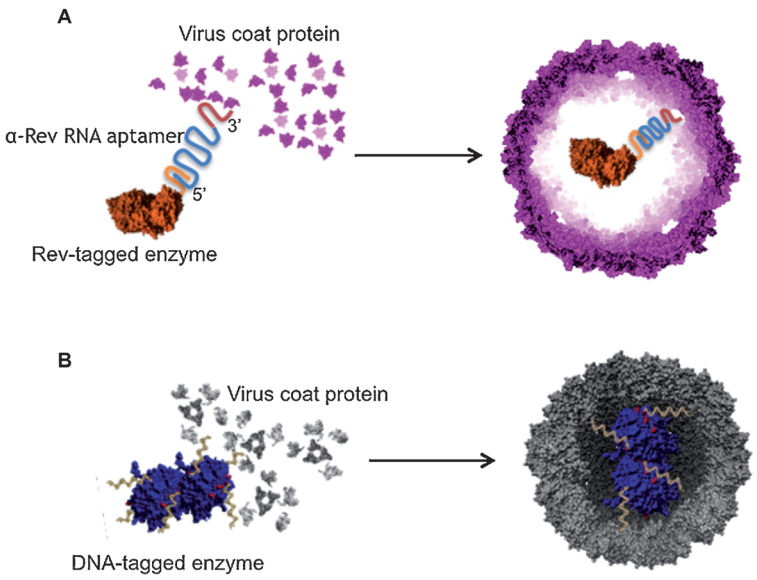


Figure 5. Representation of two strategies used to direct proteins and/or enzymes towards the interior of protein cages. A) RNA aptamer technology promotes a "hook and bait" type interaction between the virus coat protein and the desired cargo. B) Short DNA oligonucleotides can be covalently coupled to target enzymes. The DNA drives the self-assembly of the virus into the capsid form, and simultaneously directs the encapsulation of the cargo.

4.2. Peptide-Driven Self-Assembly

In contrast to bacteriophages MS2 and Qβ, the bacteriophage P22 is slightly unusual in that it relies on a scaffold protein for self-assembly. In the absence of a scaffolding protein, the bacteriophage subunits form aberrant structures.[73] In the case of P22, a truncated version of the scaffold protein (amino acids 141–303) fused to a target protein was shown to be sufficient to encapsulate fluorescent proteins and enzymes during in vivo assembly.[44,74–77] After the capsid is heated to the expanded mature form, the scaffold proteins can be removed by enzymatic digestion of the linker between the scaffold and the cargo. This serves as a particularly attractive strategy for the development of nanoreactors, as the cargo is free to diffuse yet remains within the confines of a localized environment (Figure 6A). For example, the P22 bacteriophage has been used to encapsulate enzymes, proteins and polymers to study the effects of confinement on kinetics and for controlled polymerization, respectively.[44,74–77]

"Coiled-coil" peptide motifs are often found in nature as part of a DNA-binding domain in various transcription factors, and have been implicated as key components in gene regulation.[78] In CCMV, heterodimerizing E–K coiled-coil peptides[79] were engineered by incorporating an extended, positively charged, lysine-rich (K-coil) α-helix arm to the N terminus of the CCMV capsid protein. Introduction of the complementary

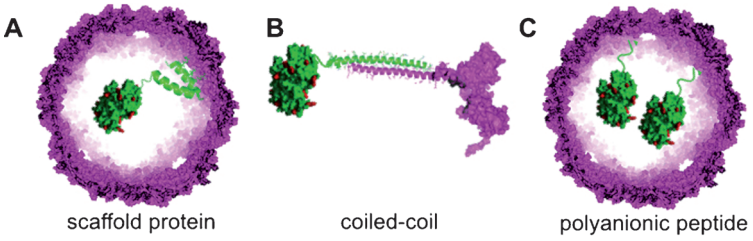


Figure 6. The three peptide-based strategies used to direct the encapsulation of biological cargo into protein cages. A) Scaffold proteins, B) coiled-coil heterodimerizing peptides, and C) polyanionic peptides that have been genetically engineered to the target enzyme. In each case, specific interactions with the virus coat protein promote encapsulation into the interior.

negatively charged glutamic-acid-rich (E-coil) α helix to the C terminus of EGFP and lipase B from *Pseudozyma antarctica* (PalB) promoted the formation of a heterodimeric complex (Figure 6B). Upon lowering the pH to 5.0, encapsulation of the desired fluorescent protein or enzyme cargo during CCMV assembly could be effected. By varying the stoichiometric ratios, it was shown that up to 15 EGFP molecules could be encapsulated in CCMV. However it should be noted that despite the success of coiled-coil peptides in directing EGFP cargo inside CCMV, a detailed study revealed a complex assembly process, which perturbed the loading efficiency. At pH 7.5, the EGFP–CCMV heterodimeric complex is stable. However, during in vitro assembly at pH 5.0, the E–K heterodimeric coiled-coil reportedly dissociates into E3 homotrimers and K monomers.[80,81] Recently, a more efficient approach was reported, whereby the CCMV monomer was fused through a flexible peptide linker to a GFP variant. Using this approach, stoichiometric loading of up to 20 GFP molecules could be accomplished.[52]

The native MS2 bacteriophage relies on a targeting RNA hairpin for encapsulation of its genomic RNA. Tullman-Ercek and co-workers reported the covalent attachment of either 1) a negatively charged polymer [DNA, RNA or poly(acrylic acid)], 2) a polyanionic peptide (16-mer) linker, or 3) a negatively charged protein, to target non-native (foreign) cargo to the interior of MS2 bacteriophage (Figure 6C).[40] Such strategies are somewhat reminiscent of those developed for CCMV, although MS2 assembly was reported to be highly dependent on the presence of a stabilizing osmolyte, trimethylamine *N*-oxide (TMAO). TMAO was shown to reduce aggregation of the capsid proteins during assembly, and was used successfully for the encapsulation of a GFP derivative and *E. coli* alkaline phosphatase.

For the bacterial encapsulins, a conserved peptide was found on the C terminus of the native DyP (*B. linens*) and Flp (*T. maritima*) enzymes. This C-terminal extension was thought to be essential for directing enzymes inside the protein shell during in vivo assembly. Recently, we reported the use of the C-terminal peptide to direct non-native cargo with high fidelity and exceptional loading accuracy.[37] Although little is known about the actual in vivo assembly process, cargo loading appears to be a highly efficient and concerted process.

5. Longer Range Ordering of Protein Assemblies

The fabrication of 2D monolayer and/or 3D arrays has been of growing interest for applications in electron and X-ray crystallography, nanoelectronics, molecular devices and as functional hybrid materials.[26,53,82–88] Several examples using atoms, nanoparticles, colloids and quantum dots are known,[82,83,89–93] although there is still a significant lack of control over the size, positioning and organization of molecular components (particles), all of which play important roles in the formation of 2D and 3D crystalline arrays.[83] For example, strong covalent binding has a tendency to promote dense but rather irregular assemblies, whereas weaker interactions fail to induce any clustering or longer range ordering at all. In recent years, protein cages have emerged as ideal molecular building blocks: firstly, the internal cavity serves as an excellent template for the in situ synthesis of highly monodisperse nanoparticles or quantum dots. Secondly, the high degree of symmetry of the exterior is ideal for controlling longer range ordering on surfaces.[26,85,86] 2D protein arrays are of particular interest for sensor, diagnostic and vaccine applications as they allow for the high-density display of peptides and proteins on their exterior surfaces. Similarly, 2D and 3D arrays also enable the periodic organization and precise arrangement of (protein-based) particles into semicrystalline superlattices for applications in plasmonic, optoelectronic, magnetic, and semiconductor devices.

5.1. 2D Arrays

Similar to traditional atomic or colloid-based crystalline arrays, 2D protein cage arrays can be formed either by exploiting the electrostatic interactions between the protein cage and surface substrate,[83,94,95] or by tethering the protein to the substrate using specific binding peptides.[95–98] Here, we focus on the use of electrostatic interactions[99] and peptide binding strategies for the formation of higher-order assemblies.

In an initial study, Douglas and co-workers reported a strategy to impart directional monolayer formation of CCMV on gold substrates by introducing a single cysteine point mutation on the CCMV surface. However, owing to the high degree of symmetry of CCMV, a single cysteine mutation per monomer translates into 180 surface-exposed cysteine residues, which can then lead to uncontrolled aggregation. Indeed, control experiments with CCMV bearing the cysteine mutation demonstrated the formation of large irregular aggregates when bound to the gold surface. To gain control over aggregation, CCMV was bound through the cysteine thiol to an activated thiol resin and the remaining surface-exposed cysteine residues were alkylated using iodoacetic acid. Removal of the CCMV from the resin led to the formation of so-called "symmetry broken" CCMV particles bearing one free cysteine residue, which could then be used to form a discrete monolayer on gold surfaces.

Yamashita and his colleagues have reported impressive work involving the use of ferritin- and Dps- (a DNA-binding protein from nutrient starved cells) templated assemblies for the fabrication of metal oxide semiconductors. In one example, a 2D hexagonally close-packed (HCP) array of ferritin molecules was assembled by displaying a carbonaceous-specific binding peptide (DYFSSPYYEQLF) on the exterior surface of the ferritin cage.[97] In subsequent work, the same authors reported a novel technique for the controlled growth of crystalline ferritin arrays using a three-phase (solid–liquid–gas) contact line, which promoted crystal growth from a specific nucleation point (an important factor for the formation of highly crystalline arrays).[83]

5.2. 3D Lattice Arrays

The development of 2D surface arrays continues to be of great interest, and recently CCMV and ferritin have been investigated as "atomic-like" particles to direct the self-assembly of 3D binary superlattices.[100] The exterior surface of both CCMV and ferritin carry a net negative charge that is localized in certain so-called "electrostatically patchy" regions. Using the native protein cages (i.e. CCMV filled with RNA, and ferritin cages filled with iron oxide nanoparticles), two different superlattices were assembled through electrostatic interactions with complementary positively charged gold nanoparticles. The superlattices, composed of four CCMV assemblies and 32 gold nanoparticles, form an unprecedented crystalline structure of AB8 face-centered cubic (FCC) symmetry. Interestingly, 3D binary superlattice structures could be assembled by tuning both the Debye screening length and pH.

In subsequent work, CCMV was co-assembled with avidin to form heterogeneous crystals.[96] Avidin is a tetrameric glycoprotein that binds biotin with high selectivity and affinity, and has been used extensively to promote protein–protein interactions for applications in chemical biology. In this example, CCMV and avidin bearing oppositely charged surfaces were assembled into binary crystal arrays by exploiting (patchy) electrostatic interactions. This approach allows crystal formation to occur without the need to covalently modify the protein cages. Because the interactions are driven by electrostatics, the crystals were also responsive to external stimuli, such as changes in pH and electrolyte, thereby demonstrating the potential of these systems in 3D assemblies.

6. Conclusion

We have presented a brief overview of some of the most common icosahedral protein cages, their self-assembly properties and their potential applications in material science. Nature has designed, engineered and evolved a diverse toolbox of viral and nonviral protein cages that have particular functions in molecular storage, biological catalysis, gene delivery, detoxification and mineralization. These assemblies have much potential for applications in nanotechnology as "smart" cargo carriers for targeted delivery, as "soft" nanoparticles for enhanced imaging, mediating plasmon effects and/or as fluorescence sensors. More recently, various techniques have been developed to enable the periodic organization and precise arrangement of (protein-based) particles into 2D or 3D arrays. However, creating protein-based assemblies with specific orientation and chemical properties still remains highly challenging. With some highly promising examples, in the coming years, it will be possible to design protein arrays that reversibly assemble, on command, into complex materials with unique mechanical and physical properties.

Acknowledgements

R.M.P. gratefully acknowledges the Indonesia Endowment Fund for Education (LPDP) for funding her doctoral studies.

Keywords: molecular storage · protein arrays · protein cages · self-assembly · virus-like particles

[1] R. P. Brinkhuis, F. P. J. T. Rutjes, J. C. M. van Hest, *Polym. Chem.* **2011**, *2*, 1449.
[2] A. de La Escosura, R. J. M. Nolte, J. J. L. M. Cornelissen, *J. Mater. Chem.* **2009**, *19*, 2274.
[3] C. E. Flynn, S.-W. Lee, B. R. Peelle, A. M. Belcher, *Acta Mater.* **2003**, *51*, 5867.
[4] J. D. Keasling, *ACS Chem. Biol.* **2008**, *3*, 64.
[5] Z. Liu, J. Qiao, Z. Niu, Q. Wang, *Chem. Soc. Rev.* **2012**, *41*, 6178.
[6] M. Marguet, C. Bonduelle, S. Lecommandoux, *Chem. Soc. Rev.* **2013**, *42*, 512.
[7] T. Marks, R. Sharp, *J. Chem. Technol. Biotechnol.* **2000**, *75*, 6.
[8] R. J. R. W. Peters, I. Louzao, J. C. M. van Hest, *Chem. Sci.* **2012**, *3*, 335.
[9] P. Ball, *Nanotechnology* **2005**, *16*, R1.
[10] C. M. Soto, B. R. Ratna, *Curr. Opin. Biotechnol.* **2010**, *21*, 426.
[11] B. Städler, A. D. Price, R. Chandrawati, L. Hosta-Rigau, A. N. Zelikin, F. Caruso, *Nanoscale* **2009**, *1*, 68.
[12] N. F. Steinmetz, *Nanomedicine: NBM* **2010**, *6*, 634.
[13] M. Brasch, J. J. L. M. Cornelissen, M. S. T. Koay in *Synthetic Biology, Vol. 1* (Eds.: M. Ryadnov, L. Brunsveld, H. Suga), Royal Society of Chemistry, Cambridge, **2014**, p. 230.
[14] N. C. Andrews, P. J. Schmidt, *Annu. Rev. Physiol.* **2007**, *69*, 69.
[15] P. Arosio, S. Levi, *Free Radical Biol. Med.* **2002**, *33*, 457.
[16] M. R. J. Clokie, A. D. Millard, A. V. Letarov, S. Heaphy, *Bacteriophage* **2011**, *1*, 31.
[17] J. Corchero, J. Cedano, *Microb. Cell Fact.* **2011**, *10*, 92.
[18] K. N. Parent, R. Khayat, L. H. Tu, M. M. Suhanovsky, J. R. Cortines, C. M. Teschke, J. E. Johnson, T. S. Baker, *Structure* **2010**, *18*, 390.
[19] J. A. Speir, J. E. Johnson, *Curr. Opin. Struct. Biol.* **2012**, *22*, 65.
[20] J. A. Speir, S. Munshi, G. Wang, T. S. Baker, J. E. Johnson, *Structure* **1995**, *3*, 63.
[21] S. Stefanini, S. Cavallo, B. Montagnini, E. Chiancone, *Biochem. J.* **1999**, *338*, 71.
[22] E. C. Theil, R. K. Behera, T. Tosha, *Coord. Chem. Rev.* **2013**, *257*, 579.
[23] J. E. Johnson, J. A. Speir, *J. Mol. Biol.* **1997**, *269*, 665.
[24] C. E. Ashley, E. C. Carnes, G. K. Phillips, P. N. Durfee, M. D. Buley, C. A. Lino, D. P. Padilla, B. Phillips, M. B. Carter, C. L. Willman, C. J. Brinker, C. C. Jdo, B. Chackerian, W. Wharton, D. S. Peabody, *ACS Nano* **2011**, *5*, 5729.
[25] K. N. Avery, J. E. Schaak, R. E. Schaak, *Chem. Mater.* **2009**, *21*, 2176.
[26] S. L. Capehart, M. P. Coyle, J. E. Glasgow, M. B. Francis, *J. Am. Chem. Soc.* **2013**, *135*, 3011.
[27] B. Chackerian, J. d. C. Caldeira, J. Peabody, D. S. Peabody, *J. Mol. Biol.* **2011**, *409*, 225.
[28] E. W. Kovacs, J. M. Hooker, D. W. Romanini, P. G. Holder, K. E. Berry, M. B. Francis, *Bioconjugate Chem.* **2007**, *18*, 1140.
[29] E. Gazit, *FEBS J.* **2007**, *274*, 317.

[30] M. Murugesan, G. Abbineni, S. L. Nimmo, B. Cao, C. Mao, *Sci. Rep.* **2013**, *3*, 1820.
[31] W. F. Ochoa, A. Chatterji, T. Lin, J. E. Johnson, *Chem. Biol.* **2006**, *13*, 771.
[32] In some cases, non-native rod-like virus assemblies can be formed, particularly in the presence of a rigid cargo template, such as double-stranded DNA, peptide amphiphiles or other rod-like structures.
[33] T. Douglas, M. Allen, M. Young in *Biopolymers Online*, Wiley-VCH, Weinheim, **2005**.
[34] M. L. Flenniken, M. Uchida, L. O. Liepold, S. Kang, M. J. Young, T. Douglas in *Viruses and Nanotechnology, Vol. 327* (Eds.: M. Manchester, N. Steinmetz), Springer, Berlin, Heidelberg, **2009**, p. 71.
[35] M. G. Mateu, *Protein Eng. Des. Sel.* **2011**, *24*, 53.
[36] J. Rong, Z. Niu, L. A. Lee, Q. Wang, *Curr. Opin. Colloid Interface Sci.* **2011**, *16*, 441.
[37] W. F. Rurup, J. Snijder, M. S. T. Koay, A. J. R. Heck, J. J. L. M. Cornelissen, *J. Am. Chem. Soc.* **2014**, *136*, 3828.
[38] O. Rolfsson, K. Toropova, V. Morton, S. Francese, G. Basnak, G. S. Thompson, S. W. Homans, A. E. Ashcroft, N. J. Stonehouse, N. A. Ranson, P. G. Stockley, *Comput. Math. Methods Med.* **2008**, *9*, 339.
[39] C. D. Anobom, S. C. Albuquerque, F. P. Albernaz, A. C. Oliveira, J. L. Silva, D. S. Peabody, A. P. Valente, F. C. L. Almeida, *Biophys. J.* **2003**, *84*, 3894.
[40] J. E. Glasgow, S. L. Capehart, M. B. Francis, D. Tullman-Ercek, *ACS Nano* **2012**, *6*, 8658.
[41] S. M. B. Lima, A. C. Q. Vaz, T. L. F. Souza, D. S. Peabody, J. L. Silva, A. C. Oliveira, *FEBS J.* **2006**, *273*, 1463.
[42] S. D. Brown, J. D. Fiedler, M. G. Finn, *Biochemistry* **2009**, *48*, 11155.
[43] J. D. Fiedler, S. D. Brown, J. L. Lau, M. G. Finn, *Angew. Chem. Int. Ed.* **2010**, *49*, 9648; *Angew. Chem.* **2010**, *122*, 9842.
[44] J. Lucon, S. Qazi, M. Uchida, G. J. Bedwell, B. LaFrance, P. E. Prevelige, T. Douglas, *Nat. Chem.* **2012**, *4*, 781.
[45] P. A. Thuman-Commike, B. Greene, J. A. Malinski, J. King, W. Chiu, *Biophys. J.* **1998**, *74*, 559.
[46] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, M. J. Young, *Virology* **2000**, *277*, 450.
[47] S. Singh, A. Zlotnick, *J. Biol. Chem.* **2003**, *278*, 18249.
[48] J. Sun, C. DuFort, M. C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, C. C. Kao, B. Dragnea, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1354.
[49] M. Allen, J. W. M. Bulte, L. Liepold, G. Basu, H. A. Zywicke, J. A. Frank, M. Young, T. Douglas, *Magn. Reson. Med.* **2005**, *54*, 807.
[50] L. Liepold, S. Anderson, D. Willits, L. Oltrogge, J. A. Frank, T. Douglas, M. Young, *Magn. Reson. Med.* **2007**, *58*, 871.
[51] I. J. Minten, V. I. Claessen, K. Blank, A. E. Rowan, R. J. M. Nolte, J. J. L. M. Cornelissen, *Chem. Sci.* **2011**, *2*, 358.
[52] W. F. Rurup, F. Verbij, M. S. T. Koay, C. Blum, V. Subramaniam, J. J. L. M. Cornelissen, *Biomacromolecules* **2014**, *15*, 558.
[53] M. Uchida, M. T. Klem, M. Allen, P. Suci, M. Flenniken, E. Gillitzer, Z. Varpness, L. O. Liepold, M. Young, T. Douglas, *Adv. Mater.* **2007**, *19*, 1025.
[54] B. Wiedenheft, J. Mosolf, D. Willits, M. Yeager, K. A. Dryden, M. Young, T. Douglas, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10551.
[55] M. Allen, D. Willits, J. Mosolf, M. Young, T. Douglas, *Adv. Mater.* **2002**, *14*, 1562.
[56] S. Gider, D. D. Awschalom, T. Douglas, K. Wong, S. Mann, G. Cain, *J. Appl. Phys.* **1996**, *79*, 5324.
[57] M. Bozzi, G. Mignogna, S. Stefanini, D. Barra, C. Longhi, P. Valenti, E. Chiancone, *J. Biol. Chem.* **1997**, *272*, 3259.
[58] J. Swift, W. A. Wehbi, B. D. Kelly, X. F. Stowell, J. G. Saven, I. J. Dmochowski, *J. Am. Chem. Soc.* **2006**, *128*, 6611.
[59] Z. Yang, X. Wang, H. Diao, J. Zhang, H. Li, H. Sun, Z. Guo, *Chem. Commun.* **2007**, 3453.
[60] M. Sutter, D. Boehringer, S. Gutmann, S. Gunther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, N. Ban, *Nat. Struct. Mol. Biol.* **2008**, *15*, 939.
[61] R. Rahmanpour, T. D. H. Bugg, *FEBS J.* **2013**, *280*, 2097.
[62] E. A. Anderson, S. Isaacman, D. S. Peabody, E. Y. Wang, J. W. Canary, K. Kirshenbaum, *Nano Lett.* **2006**, *6*, 1160.
[63] M. Brasch, J. J. L. M. Cornelissen, *Chem. Commun.* **2012**, *48*, 1446.
[64] M. Brasch, A. de La Escosura, Y. Ma, C. Uetrecht, A. J. R. Heck, T. Torres, J. J. L. M. Cornelissen, *J. Am. Chem. Soc.* **2011**, *133*, 6878.
[65] T. Douglas, D. P. E. Dickson, S. Betteridge, J. Charnock, C. D. Garner, S. Mann, *Science* **1995**, *269*, 54.
[66] P. D. Garimella, A. Datta, D. W. Romanini, K. N. Raymond, M. B. Francis, *J. Am. Chem. Soc.* **2011**, *133*, 14704.
[67] H.-A. Hosein, D. R. Strongin, M. Allen, T. Douglas, *Langmuir* **2004**, *20*, 10283.
[68] F. C. Meldrum, B. R. Heywood, S. Mann, *Science* **1992**, *257*, 522.
[69] F. C. Meldrum, V. J. Wade, D. L. Nimmo, B. R. Heywood, S. Mann, *Nature* **1991**, *349*, 684.
[70] W. Wu, S. C. Hsiao, Z. M. Carrico, M. B. Francis, *Angew. Chem. Int. Ed.* **2009**, *48*, 9493; *Angew. Chem.* **2009**, *121*, 9657.
[71] T. Hermann, D. J. Patel, *Science* **2000**, *287*, 820.
[72] M. Brasch, R. M. Putri, M. S. T. Koay, J. J. L. M. Cornelissen, unpublished results.
[73] W. Earnshaw, J. King, *J. Mol. Biol.* **1978**, *126*, 721.
[74] A. O'Neil, C. Reichhardt, B. Johnson, P. E. Prevelige, T. Douglas, *Angew. Chem. Int. Ed.* **2011**, *50*, 7425; *Angew. Chem.* **2011**, *123*, 7563.
[75] A. O'Neil, P. E. Prevelige, G. Basu, T. Douglas, *Biomacromolecules* **2012**, *13*, 3902.
[76] D. P. Patterson, P. E. Prevelige, T. Douglas, *ACS Nano* **2012**, *6*, 5000.
[77] D. P. Patterson, B. Schwarz, K. El-Boubbou, J. van der Oost, P. E. Prevelige, T. Douglas, *Soft Matter* **2012**, *8*, 10158.
[78] W. Landschulz, P. Johnson, S. McKnight, *Science* **1988**, *240*, 1759.
[79] J. R. Litowski, R. S. Hodges, *J. Biol. Chem.* **2002**, *277*, 37272.
[80] I. J. Minten, L. J. A. Hendriks, R. J. M. Nolte, J. J. L. M. Cornelissen, *J. Am. Chem. Soc.* **2009**, *131*, 17771.
[81] I. J. Minten, R. J. Nolte, J. J. L. M. Cornelissen, *Macromol. Biosci.* **2010**, *10*, 539.
[82] Y. Ikezoe, S.-J. Kim, I. Yamashita, M. Hara, *Langmuir* **2009**, *25*, 4293.
[83] Y. Ikezoe, Y. Kumashiro, K. Tamada, T. Matsui, I. Yamashita, K. Shiba, M. Hara, *Langmuir* **2008**, *24*, 12836.
[84] Y. Lin, Z. Su, Z. Niu, S. Li, G. Kaur, L. A. Lee, Q. Wang, *Acta Biomater.* **2008**, *4*, 838.
[85] N. Stephanopoulos, M. Liu, G. J. Tong, Z. Li, Y. Liu, H. Yan, M. B. Francis, *Nano Lett.* **2010**, *10*, 2714.
[86] M. Uchida, S. Kang, C. Reichhardt, K. Harlen, T. Douglas, *Biochim. Biophys. Acta Gen. Subj.* **2010**, *1800*, 834.
[87] K. Uto, K. Yamamoto, N. Kishimoto, M. Muraoka, T. Aoyagi, I. Yamashita, *J. Colloid Interface Sci.* **2012**, *378*, 44.
[88] I. Yamashita, K. Iwahori, S. Kumagai, *Biochim. Biophys. Acta Gen. Subj.* **2010**, *1800*, 846.
[89] R. P. Andres, J. D. Bielefeld, J. I. Henderson, D. B. Janes, V. R. Kolagunta, C. P. Kubiak, W. J. Mahoney, R. G. Osifchin, *Science* **1996**, *273*, 1690.
[90] S.-H. Kim, J.-M. Lim, S.-K. Lee, C.-J. Heo, S.-M. Yang, *Soft Matter* **2010**, *6*, 1092.
[91] K. E. Mueggenburg, X.-M. Lin, R. H. Goldsmith, H. M. Jaeger, *Nat. Mater.* **2007**, *6*, 656.
[92] M. Okuda, Y. Kobayashi, K. Suzuki, K. Sonoda, T. Kondoh, A. Wagawa, A. Kondo, H. Yoshimura, *Nano Lett.* **2005**, *5*, 991.
[93] I. Yamashita, *Thin Solid Films* **2001**, *393*, 12.
[94] M. T. Klem, D. Willits, M. Young, T. Douglas, *J. Am. Chem. Soc.* **2003**, *125*, 10806.
[95] N. Matsukawa, K. Nishio, K. Sano, K. Shiba, I. Yamashita, *Langmuir* **2009**, *25*, 3327.
[96] V. Liljeström, J. Mikkilä, M. A. Kostiainen, *Nat. Commun.* **2014**, *5*, 4445.
[97] T. Matsui, N. Matsukawa, K. Iwahori, K.-I. Sano, K. Shiba, I. Yamashita, *Langmuir* **2007**, *23*, 1615.
[98] Strategies such as direct coating and drop-and-dry techniques have also been reported but are beyond the scope of this Minireview.[36]
[99] There are many examples of the use of cations and metal ions to arrange protein cages in higher-order arrays, however they will not be discussed here.
[100] M. A. Kostiainen, P. Hiekkataipale, A. Laiho, V. Lemieux, J. Seitsonen, J. Ruokolainen, P. Ceci, *Nat. Nanotechnol.* **2013**, *8*, 52.

Received: October 15, 2014
Published online on January 29, 2015