



Chem Soc Rev

Protein cage assembly across multiple length scales

Journal:	<i>Chemical Society Reviews</i>
Manuscript ID	CS-REV-11-2017-000818.R1
Article Type:	Review Article
Date Submitted by the Author:	07-Feb-2018
Complete List of Authors:	Aumiller, William; Indiana University , Chemistry Uchida, Masaki; Indiana University, Chemistry Douglas, Trevor; Indiana University, Department of Chemistry

SCHOLARONE™
Manuscripts

Title: Protein cage assembly across multiple length scales

Authors: William M. Aumiller, Jr.*, Masaki Uchida*, and Trevor Douglas

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

* These authors contributed equally to this work

Abstract

Within the materials science community, proteins with cage-like architectures are being developed as versatile nanoscale platforms for use in protein nanotechnology. Much effort has been focused on the functionalization of protein cages with biological and non-biological moieties to bring about new properties of not only individual protein cages, but collective bulk-scale assemblies of protein cages. In this review, we report on the current understanding of protein cage assembly, both of the cages themselves from individual subunits, and the assembly of the individual protein cages into higher order structures. We start by discussing the key properties of natural protein cages (for example: size, shape and structure) followed by a review of some of the mechanisms of protein cage assembly and the factors that influence it. We then explore the current approaches for functionalizing protein cages, on the interior or exterior surfaces of the capsids. Lastly, we explore the emerging area of higher order assemblies created from individual protein cages and their potential for new and exciting collective properties.

Contents:

1. Introduction
2. Morphology, size, and structure of protein cages
 - 2.1 Virus and Virus-like particle
 - 2.1.1 Spherical protein cages
 - 2.1.2 Tubular protein cages

- 2.2 Non-viral protein cages
 - 2.2.1 Ferritin
 - 2.2.2 Other non-viral protein cages
- 3. Assembly of monomers into protein cages
 - 3.1 Assembled structures have common protein folds
 - 3.2 What drives the assembly process?
 - 3.2.1 Empty capsid assembly mechanism
 - 3.2.2 Assembly mechanisms involving scaffold proteins
 - 3.2.3 Assembly around a nucleic acid cargo
 - 3.2.4 Assembly of other protein cages
 - 3.3 Factors influencing in vitro assembly
 - 3.3.1 Changes in pH and ionic strength can lead to different assembled structures
 - 3.4 Synthetic/computation approaches to designing protein cages
- 4. Encapsulation within protein cages
 - 4.1 Ship in a bottle like synthesis in preformed capsids
 - 4.2 Polymers inside protein cages
 - 4.3 Encapsulation via coassembly
 - 4.4 Encapsulation of enzymes
- 5. Functionalization of the exterior of Protein cages
 - 5.1 Genetic functionalization
 - 5.2 Covalent modification
 - 5.3 Noncovalent-Genetic Attachment
- 6. Higher order assembly of protein cages
 - 6.1 Protein cages as Building Blocks for Higher Order Assembly
 - 6.2 One- and Two-Dimensional Assembly of Protein Cages
 - 6.3 Three-Dimensional Assembly of Protein Cages
- 7. Applications
- 8. Conclusion

1. Introduction

Protein cages are, self-assembled, monodispersed, three dimensional structures. Many of these cages are spherical in shape, but other shapes such as rods, rings and more complex geometries are also common. The protein cage family consists of virus-like particles (VLPs),^{1,2} ferritins,^{3,4} chaperonins,⁵ and heat shock proteins,⁶ to name a few. They self-assemble from multiple copies of a single, or just a few, protein monomers into intricate supramolecular structures and possess well-defined interior, and exterior surfaces, as well as interfaces between subunit building blocks. The interior cavity is typically used to confine cargo molecules, and it is physically segregated from the external environment. In the case of viruses, nucleic acids (required for host infectivity and virus replication) are sequestered on the interior of the cage.^{2,7} Similarly, ferritin uses its interior cavity to store iron as a nanoparticle of ferric oxyhydroxide.^{8,9} The exterior surface of viruses is responsible for cellular interaction both for the purposes of infection and immunological defense responses⁷. The interface between subunits also plays a critical role in protein cage assembly and disassembly, affecting both the pathways of assembly and resulting particle morphology and also the stability of the final cage structures formed²

The vast number of protein cages and their mechanisms of assembly have intrigued biologists and virologists for decades, while simultaneously inspiring and capturing the imagination of scientists and engineers far beyond the fields of biology and virology. Viruses are by far the most abundant biological entities on earth with estimated numbers of individual particles that are difficult to imagine (e.g. an estimated 4×10^{30} viruses in ocean water alone)¹⁰, and yet just a tiny fraction of these have been well-characterized or have known structures. Because of this fact, virology is still a widely unexplored area that continues to expand. We do know, however, that the formation of virus capsids from protein subunits is an intriguing

example of molecular self-assembly¹¹ and that the process has been perfected resulting in a fine balance between structural stability and functional dynamics over millions of years of evolution. The sophisticated mechanisms of protein cage self-assembly and their delicate functionalities are beyond what current materials science can replicate, and scientists have increasingly been inspired by these protein cages for new materials. Biomimetic/bioinspired materials chemistry has used the principles behind the natural assembly of functional protein cage architectures that have been discovered to date to design and create new protein based materials.^{12, 13} In addition to using individual protein based nanoparticle structures, protein cages have been exploited as building blocks for construction of higher order bulk materials because of their size homogeneity and ability to impart a wide range of functionalities.¹⁴⁻¹⁷

Protein cages serve as a unique platform for synthetic chemistry not only because the diverse reactive groups available on naturally occurring amino acid side chains (e.g. amines, carboxylates, thiols, amides, phenols) but also the genetic capability to introduce non-native amino acids with a wide range of functionality in a site specific manner.^{18, 19} Most of protein cages mentioned in this review have near-atomic resolution structure models (either x-ray crystal structures or high-resolution Cryo-EM structural models) and based on this structural information, one can predict and select the location on the protein cage (e.g. interior, interface, or exterior) where the targeted functionality is to be introduced.

Here we focus on the assembly of protein cages across multiple length scales (i.e. from individual capsid assembly to assembly of individual capsids into higher order structures). We also cover the wide range of techniques for biofunctionalization of the cages to explore new uses and properties of the capsids and hierarchically assembled capsids. We begin with the assembly of protein cages from individual protein monomer building blocks into a cage-like structure, and

the effects of solution conditions on the final protein cage structure. A reasonably well understood mechanism of protein cage self assembly from individual subunits has been exploited to incorporate cargo molecules into the cage either *in vivo*²⁰⁻²⁴ or *in vitro*.²⁵⁻²⁷ This has proven to be a powerful approach for the development of functional protein cage nanoparticles. Additionally we review recent development of 3D bulk materials constructed via assembly of individual protein cages.^{14, 28} The ability to incorporate wide range of functionality in conjunction with the excellent homogeneity and monodispersity of protein cages make them unique building blocks to construct higher order materials. Self-assembly of individual protein cages themselves into hierarchical structures has led to new materials with the potential for emergent or collective properties.^{29, 30} Other recent reviews have covered a number of applications in which functionalized protein cages, particularly virus particles, have been used.³¹⁻³⁴ These include medicine (vaccines, imaging, gene therapy/delivery, drug delivery, tissue engineering), biotechnology (phage display, sensing, nanoreactors), and energy (catalysts, devices, battery electrodes, data storage).³¹

2. Morphology, Size, Symmetry and Structure of Protein Cages

In nature, protein cages are found in a wide range of related morphologies, sizes, and symmetries and in turn adopt many different functions. Some examples of the most familiar and widely used protein cages for materials synthesis are shown in Figure 1. Many (most) are very homogeneous in both size and shape and are highly symmetrical. This is a result of the fairly limited number of protein monomers that assemble to form morphologically distinct and precise protein cage assemblies. Typical sizes of these protein cages range from a few nanometers up to ~500 nm. VLPs are by far the largest and most diverse class of protein cages. Simplistically,

viruses consist of an assembled protein shell (capsid) with their nucleic acid cargo sequestered on the inside. VLPs assemble to form protein shells nearly identical to their native virus structures but lacking the viral genome (and other components such as membranes and accessory proteins) and they are thus non-infectious. The majority of the structurally characterized viruses consist of roughly spherical protein cages usually built on an icosahedral lattice. Rod shaped viruses are another common shape found among viruses as well as more exotic shapes discovered recently among the Archaeal viruses.³⁵ Of course there are many viruses, particularly enveloped viruses, that adopt less regular morphologies.³⁶⁻³⁸

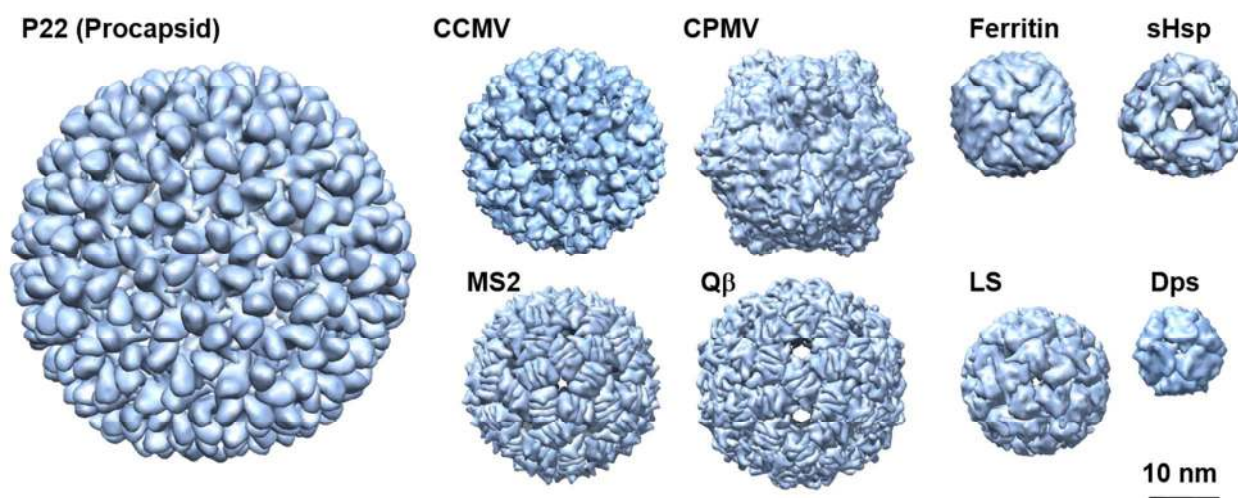


Figure 1. Space filling models of some protein cages discussed in this review. P22 procapsid (56 nm diameter, T=7) *PDB*: 3IYI, CCMV (28 nm, T=3) *PDB*: 1CWP, CPMV (30 nm, pseudo T=3) *PDB*: 1NYZ, MS2 (27 nm, T=3) *PDB*: 2MS2, Qβ (30 nm, T=3) *PDB*: 1QBE, ferritin (12 nm) *PDB*: 2FHA, sHsp (12 nm) *PDB*: 1SHS, LS (15 nm, T=1) *PDB*: 1RVV, and Dps (9 nm) *PDB*: 1QGH. These images were reproduced using UCSF Chimera (<http://www.cgl.ucsf.edu/chimera>) from the Resource for Biocomputing, Visualization, and Informatics at the University of California (supported by NIH RR-01081)

An underlying icosahedral symmetry is common among the spherical viruses. A true icosahedral virus will contain 12 5-fold, 20 3-fold, and 30 2-fold rotation axes. The smallest number of subunits needed to make a cage of this symmetry is 60, arranged so that there are 12

pentameric capsomers (capsomers are the basic building blocks from which the capsid is assembled) and no hexamers. Descriptions of viruses often invoke Triangulation number (T),³⁹ which is an allowable integer used to describe the number of nonequivalent positions in the asymmetric unit of the capsid, based on a description by Caspar and Klug.³⁹ Detailed descriptions of triangulation number can be found in the literature.³⁹⁻⁴¹ Briefly, a capsid with more than 60 subunits does not form a strict icosahedron with symmetrically equivalent subunits, but can be formed with minimal distortions using different numbers of equivalent subunits. The triangular faces on these capsids are enlarged and further subdivided into smaller triangles. The triangulation number arises from geometrical constraints where only certain numbers are allowed ($T=1, 3, 4, 7, 12, 13\dots$ etc). For an icosahedral capsid, the number of pentamers is always 12, and the number of hexamers can be calculated using the formula $10 \times (T - 1)$. For example, $T = 1$ viruses such as Adeno-associated virus (AAV) are composed of only 12 pentamers and no hexamers.⁴² Bacteriophage P22 forms a $T = 7$ capsid with 12 pentamers and 60 hexamers, for a total of 420 subunits (although in the infectious P22 phage, one pentamer is replaced with a portal complex which plays critical roles in DNA packaging as well as entry and ejection of the viral genome).⁴³

2.1 Virus and Virus-like particle protein cages

2.1.1 Spherical protein cages

Naturally occurring viruses provide a vast library of possible architectures that can be envisioned as starting materials for nanomaterials design and synthesis. The viruses most often encountered and used for materials applications are those derived from plants and bacteria,

which are well characterized in terms of structure and biophysical characteristics, and importantly can be produced reproducibly in high yield using their native host plants or heterologous expression system (typically *E. coli*). Here we briefly cover some of the key properties of the most commonly used VLPs encountered in the biomimetic and bioinspired approaches to materials chemistry.

CPMV: Cowpea mosaic virus (CPMV) is a 30 nm icosahedral plant virus that infects a number of legume species that was first reported in 1959.⁴⁴ The genome consists of 2 separate positive sense RNA strands. The outer shell is made up of 60 copies of a small (S) subunit protein and 60 copies of a large (L) subunit that consists of two distinct domains.⁴⁵ The crystal structure was reported in 1999 to 2.8 Å resolution.⁴⁶ It has become a popular platform for materials design in part because it can be produced with high yield in plants. Until recently a limitation of the CPMV system was that there was no easy method for making the CPMV cage from individual subunit building blocks and that it was difficult to reproducibly make empty virus like particles devoid of infections RNA genome.^{47, 48}

CCMV: Cowpea chlorotic mottle virus (CCMV) is an icosahedral plant virus that infects the cowpea plant creating yellow spots, hence the term “chlorotic”. The CCMV capsid assembles into a simple T=3 icosahedral capsid comprising 180 identical 20 kDa subunits.⁴⁹ It has an approximately 28 nm outer diameter and an 18 nm inner diameter. It was first isolated and characterized in 1967.⁵⁰ The virus has a tripartite genome (i.e. its genome is in 3 parts, RNA1 [3171 nt], RNA2 [2774 nt], and RNA 3 [2173 nt]) and requires 3 morphologically identical particles to pack the entire genome, RNA1 and RNA2 are packaged alone, and RNA 3 is co-packaged with a fourth RNA, RNA4 (824 nt), which is subgenomic, meaning it is synthesized from the (-) RNA3 strand.⁵¹⁻⁵³ Thus each capsid particle contains ~3000 nt.⁵³ Structural analysis

of CCMV has demonstrated that the highly basic N-termini (with 6 Arg, 3 Lys residues) are not required for capsid assembly and project into the interior of the capsid. These 180 N-termini (a total of 1620 basic amino acids) are required to package and condense the anionic RNA viral genome through complementary electrostatic interactions. The N-terminal region is disordered and therefore not observed in the crystal structure. Using the techniques of limited proteolysis combined with mass spectrometry revealed that the N-terminal region is susceptible to proteolytic digestion and is the first part of the capsid to be cleaved despite the fact that it interacts with RNA on the inside of the capsid, suggesting that this region is highly dynamic and is transiently exposed to the outside of the capsid⁵⁴. The crystal structure was reported in 1995 to 3.2 angstrom resolution.⁴⁹ Importantly CCMV has well studied *in vitro* assembly system^{53, 55, 56} and undergoes a fascinating structural transition in which pores in the structure open and close in response to changes in pH and $[Ca^{2+}]$, going from a closed form to a swollen form⁴⁹, (Figure 2). This virus can also be produced with high yield in plant and bacterial expression systems making it a very useful platform for synthetic materials applications^{57, 58}

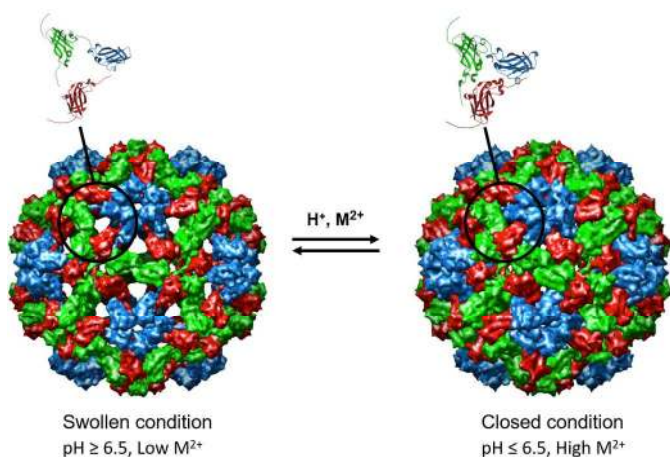


Figure 2. Cryo EM reconstruction images of the swollen and closed forms of CCMV. The swollen condition is triggered upon raising the pH to above \sim 6.5 and lowering the Ca^{2+} concentration. The swelling occurs at the quasi 3-fold axes to form \sim 2 nm pores and increases

the particle size by ~10%. Figure reproduced from Reference ⁵⁹ with permission from Wiley, copyright 2007.

BMV: Brome mosaic virus is a positive sense RNA, icosahedral virus that is 28 nm in size with $T = 3$ triangulation number that is very similar to CCMV, sharing 70% amino acid sequence identity.^{49, 52} The crystal structure was reported in 2002.⁶⁰ Like CCMV, the genome consists of 3 separate RNA strands packaged in three separate virions. RNA1 and RNA2 encode proteins 1a and 2a which are necessary for genome replication and transcription of a fourth RNA (a single subgenomic RNA or sgRNA4) from the minus strand of RNA3. RNA3 also encodes for a protein necessary for cell-to-cell movement that direct infection in the host. Coat protein is expressed from the RNA4 strand.⁶¹ The CP sequence has a higher positively charged amino acid segment that interacts with the negatively charged genome. Also, like CCMV, it undergoes a structural transition depending on solution pH and ionic strength. The native state is observed at low pH and ionic strength. At neutral pH, it undergoes a conformational change that can be stabilized by added Mg^{2+} . At higher salt and neutral pH, the capsid disassembles and the RNA genome precipitates. The capsid can be reassembled into empty $T=3$ particles and $T=1$ particles, among others, upon lowering the pH and ionic strength.^{60, 62}

Bacteriophage P22: Bacteriophage P22 is a 56 nm double stranded DNA phage that infects *Salmonella typhimurium*. It has icosahedral symmetry and forms a $T = 7$ capsid, with 415 coat proteins (CP, product of gene 5, gp5,) 60 -300 molecules of scaffold protein (SP, gp8), minor proteins (gp7, 16, and 20) and a portal protein (gp1).^{43, 63-66} The infectious P22 virion also contains other gene products known as the tail machinery, necessary for assembly and infection shown in Figure 3. The P22 system has developed into a robust platform based in part on the ease with which it self-assembles in vitro from purified scaffolding and coat protein,^{67, 68} and it

has served as a model virus for understanding dsDNA phage assembly mechanism and structures. The assembly process for the VLP, which lacks the portal complex and is comprised of 420 CP and 100-300 copies of the SP, will be described below in section 3.2. Reports of the structure of P22 have emerged over the years,^{66, 69-72} with the most recent cryo-EM structure model of P22 reported by Hyrc and coworkers to 3.3 angstrom resolution.⁷³

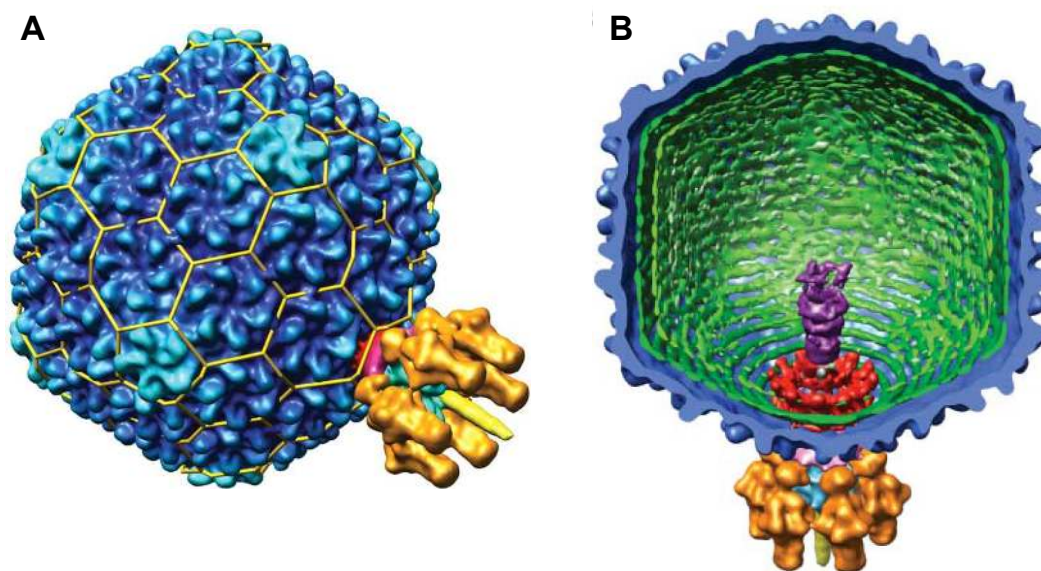


Figure 3. (A) A surface volume reconstruction image of the infectious P22 virion. The $T = 7$ organization is indicated by the yellow lattice cage. The portal complex is located at one of the 5-fold vertices. CP (gp5) is in dark blue. (B) A cutaway interior view of the infectious P22 virion. The gene products of the tail machinery are shown in different colors: gp 1 (red), gp4, (magenta), gp 7, 16, and 20 (purple), gp9 (orange), gp10 (light blue), and gp26 (yellow). Figure reproduced from reference⁴³ with permission from AAAS, copyright 2006.

Bacteriophage MS2: Bacteriophage MS2 is another bacteriophage widely used as a platform for nanomaterials synthesis and VLP-based vaccine development.^{27, 74, 75} It is an icosahedral virus that is ~ 27 nm in diameter composed of 180 subunits ($T=3$). The structure was solved to 2.8 Å resolution in the early 1990s.⁷⁶⁻⁷⁸ The genome is a positive sense single-stranded RNA that is 3569 nucleotides in length that codes for 4 proteins: the major coat protein, a

maturation protein known as protein A which mediates phage attachment to the bacterial pili, a replicase, and a lysis protein.⁷⁹ Interestingly, the fully assembled capsid consists of 3 quasi-equivalent CP structures: 30 copies of a symmetric dimer (known as a “C/C” dimer) and 60 copies of an asymmetric dimer (A/B). The primary difference for these structures is in the FG loop region. RNA hairpin structures direct the conformational change in the FG loop region to form the A/B dimers.⁷⁹⁻⁸¹

Bacteriophage Q β : Bacteriophage Q β is a 25 nm diameter icosahedral capsid that forms a T=3 structure.^{82, 83} The crystal structure was reported in 1996.⁸² The genome is made up of a single strand of positive sense RNA that is ~4200 bases long.⁸² A 29 nt hairpin in the genome has a high affinity for the CP and initiates assembly.⁸⁴⁻⁸⁶ Like MS2, its CP adopts three different conformations with 3 different dimer types. Disulfides link the individual CPs into pentamers and hexamers and result in a highly stable structure (Figure 4).⁸⁷ The dissociation temperature (thermal stability) decreases from 85-100 °C to ~40 ° upon removal of the disulfide bonds.⁸⁸

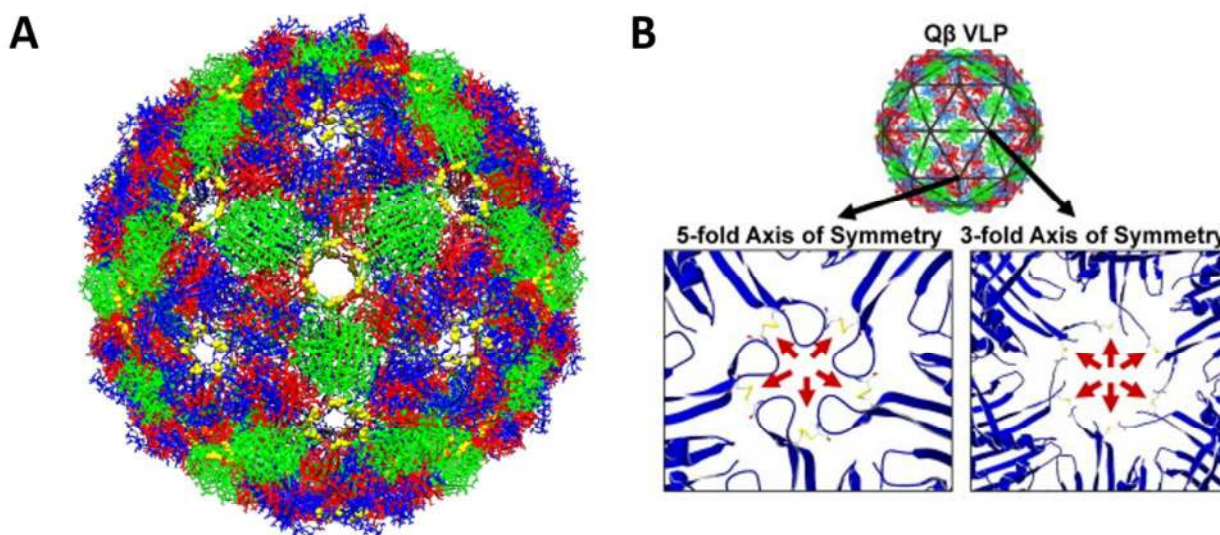


Figure 4. (A) Chimera reconstruction of bacteriophage Q β with A chains colored red, B chains colored blue, and C chains colored green. The Cysteine residues C74 and C80 that form disulfide bonds are labeled in yellow. PDB: 1QBE (B) The locations of the disulfide bonds along the FG loop are highlighted (red arrows), both along the 5-fold and 3 fold axes of symmetry. Part B reproduced from reference⁸⁸ with permission from Elsevier, copyright 2011.

2.1.2 Tubular protein structures:

TMV: Perhaps the quintessential example of a tubular rod-shaped protein structure in Nature is the tobacco mosaic virus (TMV). TMV was the first virus to be discovered over a century ago and it has been extensively studied since then.⁸⁹ TMV was also the first virus to be crystallized and the structure was reported in 1989 to 2.9 angstrom resolution.⁹⁰ Cryo-EM reconstructions⁹¹ are shown in Figure 5. TMV consists of a helical coat protein structure surrounding a single RNA strand cargo that is bound to the hollow inner channel. There are 2130 identical coat proteins formed in the right handed helix that is 300 nm in length, 18 nm in diameter and has a 4 nm inner channel with 17 coat proteins per turn.^{92, 93} TMV can be reconstituted using a double disk of the coat protein comprising 34 CPs, known as the 20S aggregate, acting as a scaffold. A specific 9nt RNA sequence acts as a recognition sequence by the TMV CPs and facilitates hydrogen bonding interactions between the RNA and the RNA binding sites on the 20S aggregates. More 20S particles are assembled in a cooperative process and the 5' end of the RNA is pulled through the central cavity.^{89, 92, 94}

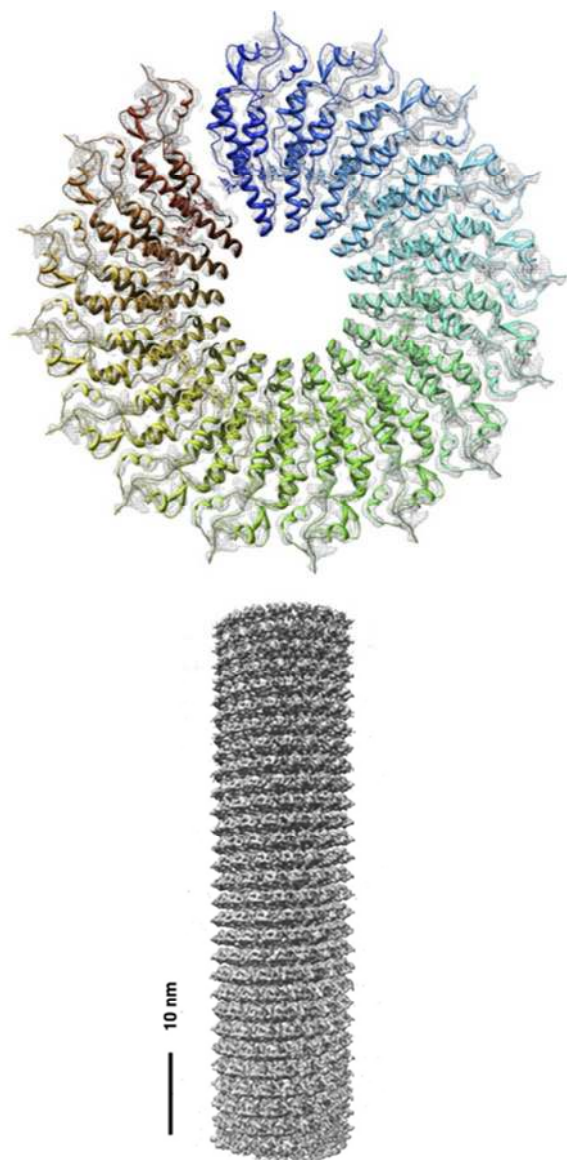


Figure 5. Density map of the TMV structure obtained by cryoEM. (a) Density of a single turn composed of 16 subunits (b) 3D construction image of a TMV rod. Figure adapted from reference⁹¹ with permission from Elsevier, copyright 2007.

Bacteriophage M13: Bacteriophage M13 is another well-characterized rod shaped filamentous virus. It is a cylindrical capsid of 880 nm in length with approximately 2700 copies of a pVIII protein along the cylinder with a width of 6.5 nm. There are 5 copies each of pIX and pVII at one end, and 5 copies of proteins pIII and pVI at the other end, which forms the tail of the phage

that infects bacteria. The genome is circular single stranded DNA, and dictates the length of the capsid.⁹⁵⁻⁹⁷ One of the most notable applications of M13 in biotechnology is perhaps its use for phage display. Here, a library of random peptide sequences is displayed at the end of M13 coat protein (typically pIII) and used for high-throughput screening to investigate protein-protein, protein-peptide, peptide-DNA, and peptide-materials interactions (see 5.1)

2.2 Non-viral protein cages

A number of non-viral protein cages have also been explored for materials design and synthesis. While the number of non-viral protein cages are far fewer than the number of viral protein cages, they have wide diversity in size and inherent functionalities, hence they have great utility as platforms for nanomaterials synthesis. A few of the most commonly encountered are discussed here.

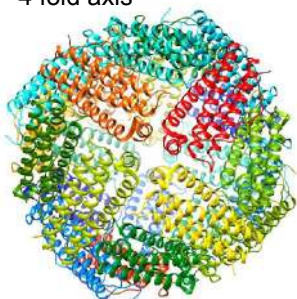
2.2.1 Ferritin

Ferritin protein cages are found in nearly all forms of life^{3,98} and function *in vivo* as iron storage containers where they encapsulate and sequester iron, usually in the form of a poorly crystalline iron oxyhydroxide.^{8,9} The general structure consists of 24 subunits with octahedral (4-fold, 3-fold, 2-fold) symmetry (Figure 6). The subunits consist of a very robust four helix bundle fold with a left-hand twist arranged in 12 antiparallel pairs. The ferritin cage has an outer diameter of ~12 nm and an inner cavity of 7-8 nm.⁹⁹ Ferritins can be further subdivided into maxi-ferritins (24 subunits), of which classical ferritins and bacterio-ferritins are members, and mini-ferritins (12 subunits), of which DNA-binding protein from starved cells (Dps) proteins are

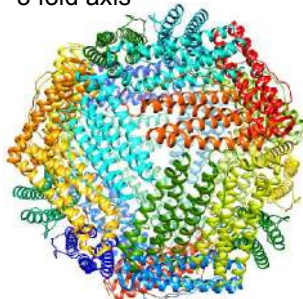
members.^{100, 101} While almost all the true ferritins have octahedral symmetry, mini-ferritins show tetrahedral symmetry (i.e. 3-fold, 2-fold).^{4, 102} Despite the quite significant differences in primary sequence among ferritins, the secondary and tertiary structures are highly conserved.^{92, 99} One notable exception to this is the ferritin isolated from the hyperthermophilic Archaeon *Archaeoglobus fulgidus* (AfFn). Despite the fact that AfFn is assembled from 24 subunits and the subunit structure has a high degree of structural similarity with other ferritins, it assembles into a cage structure not having octahedral symmetry but rather tetrahedral symmetry with four large pores (~ 4.5 nm) (Figure 6).^{103, 104} Interestingly, a double mutant of AfFn, K150A/R151A, assembles into a closed cage with octahedral symmetry, resembling the archetypal ferritin structure, suggesting a critical role for these two residues in inter-subunit interactions to form the quaternary structure¹⁰⁵. The assembly mechanism of ferritin from its constituent subunits is highly concentration dependent and starts with the unfolded monomers becoming properly folded. It is followed by rapid formation of dimers that accumulate since the dimerization is fast compared to the subsequent associations. The dimers go on to make trimers, hexamers, and dodecamers and finally forms the 24-mer fully assembled cage. The purported hexamer intermediate is the most transient of the intermediates and only detected in small amounts.^{92, 106} On the other hand, the assembly mechanism for mini-ferritins remains poorly understood.

A. Human Fn

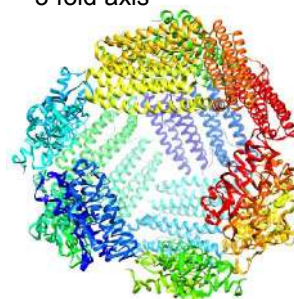
4 fold axis



3 fold axis

**B. AfFn**

3 fold axis

**C. ssDps**

3 fold axis

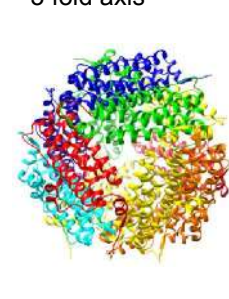


Figure 6. Ribbon diagrams of exterior surface of view of (A) human heavy-chain ferritin (PDB: 2FHA) looking down from 4 fold axis (left) and 3 fold axis, (B) ferritin from *Archaeoglobus fulgidus* (PDB: 1S3Q) looking down from 3 fold axis, and (C) Dps from *Sulfolobus solfataricus* (PDB: 2CLB) looking down from 3 fold axis. These images were reproduced using UCSF Chimera (<http://www.cgl.ucsf.edu/chimera>) from the Resource for Biocomputing, Visualization, and Informatics at the University of California (supported by NIH RR-01081)

2.2.2 Other non-viral protein cages

Lumizine synthase is an enzyme involved with riboflavin biosynthesis that can form icosahedral capsids with 60 subunits, i.e. it structurally resembles a T=1 virus capsid.¹⁰⁷⁻¹⁰⁹ Other structurally characterized icosahedral protein cages include encapsulin,⁹³ clatherin,¹¹⁰ and dihydrolipoyl acetyltransferase (E2) from *Bacillus stearothermophilus*.¹¹¹ Small heat shock protein (sHsp) from *M. jannaschii* is a 12 nm protein cage with octahedral symmetry composed of 24 subunits,⁶ meaning the overall morphology and size resemble ferritin. However, the subunit structure of sHsp consists of β -sheets instead of the α -helix bundles of the ferritin subunit. Hsp has large (3 nm) pores at the 3-fold symmetry axis and smaller (1.6 nm) pores at the 4-fold symmetry axis. Other protein cages with non-spherical morphologies have also been discovered, such as chaperonins and vault particles. The chaperonins, which are hollow cylinder-like protein-assemblies, function to encapsulate non-native state proteins within the cylindrical cavity to mediate their proper folding to the native states.⁵ The chaperonins use their ATP-dependent conformational changes to assist folding of the guest proteins. The chaperonins are categorized into two types; group I and group II based on their structure and origin. Group I chaperonins are composed of two stacked rings of supramolecular proteins each with 7-fold rotational symmetry (e.g. GroEL in bacteria, HSP60 in mitochondria) which is capped by a smaller protein (GroES in bacteria, HSP10 in mitochondria) on the top of the rings. Group II chaperonins (e.g. Thermosome in archaea) have a built-in protrusion that functions as a “lid”

structure. Vault particles adopt a barrel-like shape, with a 41 nm diameter and 73 nm length. They are termed vault particles of their structural resemblance to vaulted ceilings in Gothic cathedrals.¹¹² To date, they have been found in number of eukaryotes, however their function remains poorly understood.¹¹³

3. Assembly of monomers into protein cages

3.1 Assembled structures have common protein folds.

Since their discovery at the end of the nineteenth century, viruses have provoked discussion among scientists regarding the question of whether viruses should be considered living organisms.¹¹⁴ The origin of viruses from an evolutionary perspective also remains an active area of investigation.^{115, 116} With a number of viral structures and viral protein structures now known (>6900 entries in the PDB with the keyword virus), a few families with common protein folds have been identified and structural relationships to these have been proposed as a method for classification of viruses.¹¹⁷ Interestingly, capsids with similar architecture/folds infect hosts of all domains of life, which suggests a common ancient origin for virus families.¹¹⁷⁻¹²⁰ Four major classes have been proposed for icosahedral viruses: (1) PRD1/adenovirus-like, (2) picornavirus-like (3) HK97-like, and (4) BTV-like.¹¹⁷ Here, we focus on the well-established PRD1/adenovirus and HK-97 like viruses because their members are the most commonly encountered in the protein cages used for materials synthesis. The others are reviewed elsewhere.¹¹⁷

The capsids of the PRD1/adenovirus family consist of trimeric major capsid protein with a double β -barrel structure. The major capsid proteins form trimers, that become hexagonal

capsomers that eventually go on to make an icosahedral structure.¹²⁰ Bacteriophage PRD1 serves as the naming member of this lineage. The structure of the infectious virus consists of the protein capsid surrounding a protein-rich lipid membrane. The dsDNA genome is contained within the lipid membrane. The structure of the major capsid protein was first reported in 1999,¹²¹ followed by the entire virion structure solved by X-ray crystallography in 2004.^{122, 123} Since then, other capsid types with similar double β -barrel fold have been identified and classified as part of the lineage, some of which are given in Figure 7.

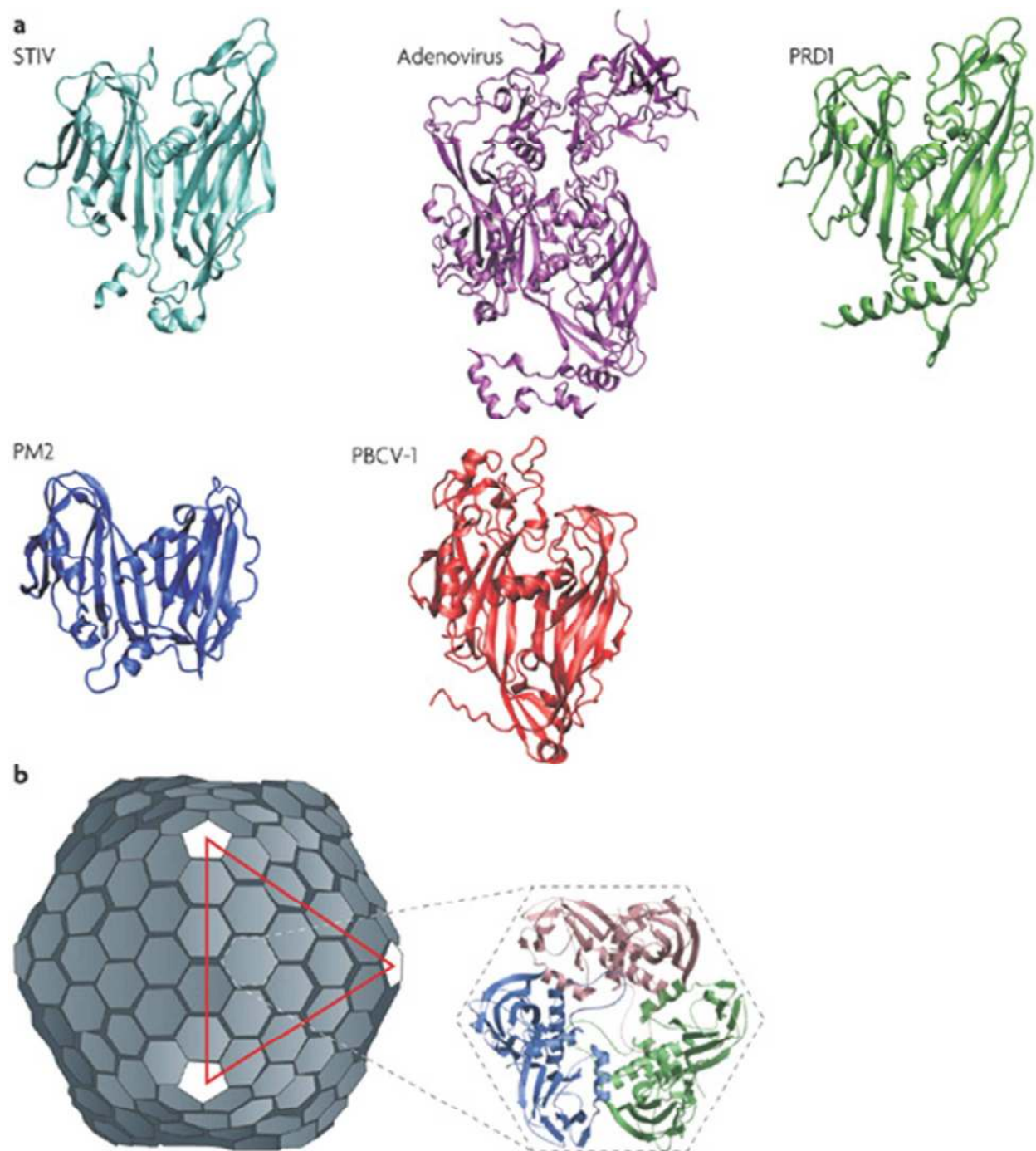


Figure 7. Structures of member capsid proteins of the PRD1-adenovirus lineage. (a) The structures are STIV (*Sulfolobus turreted icosahedral virus*, cyan), human adenovirus (purple), bacteriophage PRD1 (green), bacteriophage PM2 (blue), and *Paramecium bursaria Chlorella virus type 1* (PBCV-1, red). (b) The structure of the PRD1 virion, which forms a $T = 25$ capsid. The inset shows the trimeric capsomer with the individual monomers indicated. Figure adapted with permission from reference ¹²⁰ with permission from Nature Publishing Group, copyright 2008.

Another group that has been well characterized is the HK97 protein fold, named because HK97 was the first virus structure solved in this group. Over 40 different structures have been identified having HK97-like protein folds (Figure 8). Interestingly, only about 10-15% of the amino acid sequence is shared among the members. The structure of this protein fold is identified by a number of features that include an N-terminal region arm (N-arm) sometimes containing alpha helical portions, an extended loop of varying lengths (E-loop) that is a two-stranded anti parallel beta sheet.¹²⁴ It is also characterized by a peripheral domain (P domain) containing a “spine helix” and an unusually long beta sheet, and an axial domain (A domain) with a central beta sheet surrounded by short helices and loops.¹²⁴ The family of viruses having this protein fold structure are also characterized by the presence of an internal scaffolding protein that directs capsid assembly. Additional domains are also observed among other family members.

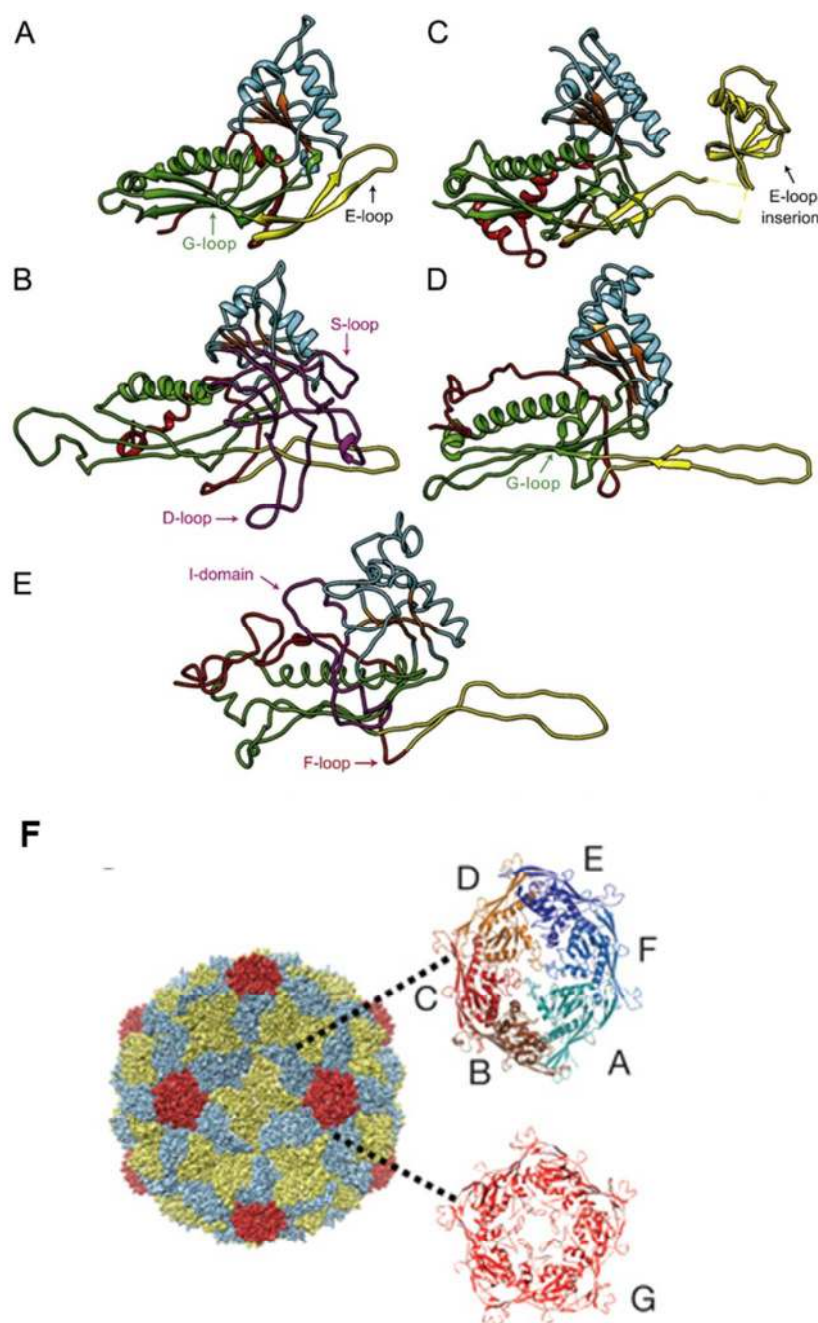


Figure 8. HK97 and related protein structures with common features highlighted. (A) HK97 in the immature state (PDB: 3E8K), (B) P22, (C) T4, (PDB: 1YUE), (D) BPP-1 (PDB: 3J4U), and (E) P-SSP7 (PDB: 2XD8). The common features are N-arm (red), E-loop (yellow), P-domain (green), A-domain (cyan), β -hinge (orange). (F) An electron density map of the HK97 capsid with capsomers (both the hexamer (6-assymmetric subunits, multicolored) and the pentamer (red) indicated). Parts A-E adapted from reference ¹²⁴ with permission from Elsevier, copyright 2015. Part F adapted from reference ¹²⁵ with permission from Nature Publishing Group, copyright 2009.

3.2 What drives the Assembly Process?

The assembly process is driven primarily by favorable interactions, typically a number of individual, weak contacts are made among capsid subunits and cargos that add up to give a globally stable structure. The primary interactions to be considered are subunit-subunit, subunit-cargo, and cargo-cargo interactions.^{126, 127} Assembly of capsids can be divided into two categories: those that can assemble without the need for a cargo molecule (empty capsid assembly) and those in which interactions between capsid subunit and cargo direct the capsid assembly. The second categories can be further subdivided into two sub-class based on type of cargo molecules: either proteinaceous molecules (called scaffold protein) or nucleic acids (typically virus genome, but can be non-genomic nucleic acid).

3.2.1 Empty Capsid Assembly

Empty capsid assembly can be understood from a model of nucleation and growth.^{127, 128} Viruses need to encapsulate their genomes inside of the capsids and the assembly of empty capsids can be modeled and understood from relatively simple models, which nevertheless provide useful insights into how subunit-subunit interactions lead to the self-assembly of the capsid architecture. In this mechanism, an assembly nucleus or seed is formed from just a few protein subunits. These clusters are transient and relatively unstable because of the low number of favorable intersubunit interactions. A critical nucleus is formed when a sufficient number of subunits form in the correct geometry that allows for further growth. Growth is rapid in comparison to the formation of nuclei. The assembly process kinetics are sigmoidal in which an

initial lag phase, where the nuclei are formed, is followed by rapid addition of subunits to complete the capsid (growth), followed by an asymptotic approach to equilibrium due to depletion of subunits (saturation). In general, increasing the subunit concentration or strengthening inter-subunit interactions will lead to more rapid assembly and the possibility increases for the formation of an excess of partially formed capsids, due to too many nucleation sites formed or kinetically trapped assemblies, that might not lead to the correct close shell capsid structure. Lowering the subunit concentration leads to large nucleation barriers, due to fewer collisions between subunits, and therefore fewer chances of forming the critical nuclei.^{127, 129}

CCMV is model virus that has been studied for decades in order to elucidate the self-assembly process of capsid formation. It has been demonstrated that under appropriate conditions, subunits of CCMV can self-assemble to form empty capsid in the absence of RNA.^{56, 130} For empty capsids, it follows the nucleation and growth model where CCMV subunit dimers associate to form a pentamer of dimers that serves as the critical nucleus for subsequent growth.¹³⁰ When the protein concentration is low, assembly continues with the addition of dimers and the formation of $T = 3$ capsids is favored. When the subunit protein concentration is high, many nucleation sites are formed which causes incomplete formation of capsids and leads to formation of pseudo $T=2$ particles.¹³⁰ As we will discuss later, CCMV capsid can also self-assemble around RNAs,⁵³ non-viral anionic polymers,¹³¹ and negatively charged gold nanoparticles.¹³² *In vitro* reconstruction of VLPs directed by negatively charged non-viral cargo molecules have been demonstrated with other viruses such as BMV and red clover necrotic mosaic virus (RCNMV)¹³³⁻¹³⁵

3.2.2 Assembly mechanisms involving scaffold proteins.

Virus particle formation may also require the use of scaffold protein (SP) to direct the self-assembly of subunit components into the correct capsid structure. One well known example of this is the bacteriophage P22, whose assembly involves association of both scaffolding proteins (SP) and coat proteins (CP) via specific interactions,¹³⁶ which results in co-assembly of 415 subunits of CP (gp 5) with about 60 – 300 subunits of SP (gp 8) (Figure 9) and CP is added along the growing edge of the shell until the capsid is formed. SP stabilizes and directs the CP into the correct geometry.¹²⁴ Intriguingly, scaffold protein can form oligomers (dimers and tetramers) as well as monomers, with the dimer being the dominant, active form in assembly.¹³⁷ This suggests a mechanism for striking a proper balance between nucleation and growth by changes in the oligomerization state of the SP.¹³⁸ The interaction between coat and scaffold protein is largely electrostatic in nature, because about 30 % of the C terminal helix-turn-helix of the SP is positively charged, with R293 on the SP and D14 on the CP being essential for proper assembly.^{124, 139, 140} Additionally, interactions between K296 (SP) and E15 (CP) are also likely to be important.¹⁴⁰ Scaffold protein is also critical for formation of T=7 capsids, rather than either smaller or larger capsids. Without scaffolding protein, aberrantly assembled particles, including small shells and spirals, that cannot package DNA are formed, and form at a slower rate¹⁴¹. Also during *in vivo* assembly, a dodecameric portal complex and ejection proteins, which are required for formation of the infectious virion, are incorporated into the procapsid at this stage (one pentameric vertex of icosahedron is occupied by the portal complex⁴³). The portal complex has been proposed to be part of the nucleation complex, which would ensure that only one portal complex is incorporated per virion.¹⁴² In the infectious virus, DNA is packaged through this portal complex after assembly of the P22 procapsid and the scaffold protein is removed after procapsid assembly during the morphological transformation of the capsid through packaging of

nucleic acid (Figure 9). Upon DNA packaging, and removal of the scaffold proteins, the spherical procapsid transforms into the mature polyhedral capsid - resulting in a change in conformation of the coat protein and an approximately 10-15% increase in radius.^{124, 143} The exact exit location of the SP is not known, but it is presumed to be through large 2.5nm pores at the center of the hexamers.¹⁴³ The scaffold protein can be recycled to take part in assembling more capsids as shown schematically in Figure 9. Most importantly and usefully for materials synthesis is the fact that assembly of the bacteriophage P22 capsid can be reproduced by expressing only CP and SP in heterologous protein expression system using *E. coli*, or mixing CP and SP under a controlled manner *in vitro*.^{67, 135, 144} In this case, the portal complexes of the infectious P22 virions are replaced with 5 CP subunits, thus the capsid is composed of 420 copies of CP with approximately 60 – 300 copies of SP. Removal of SP from the capsid and morphological transformation from the procapsid form to the matured polyhedral capsid can also be reproduced *in vitro* by treating the procapsid form of P22 with 0.5 M guanidine hydrochloride^{67, 145} and heating it at 65 °C, respectively.^{146, 147}

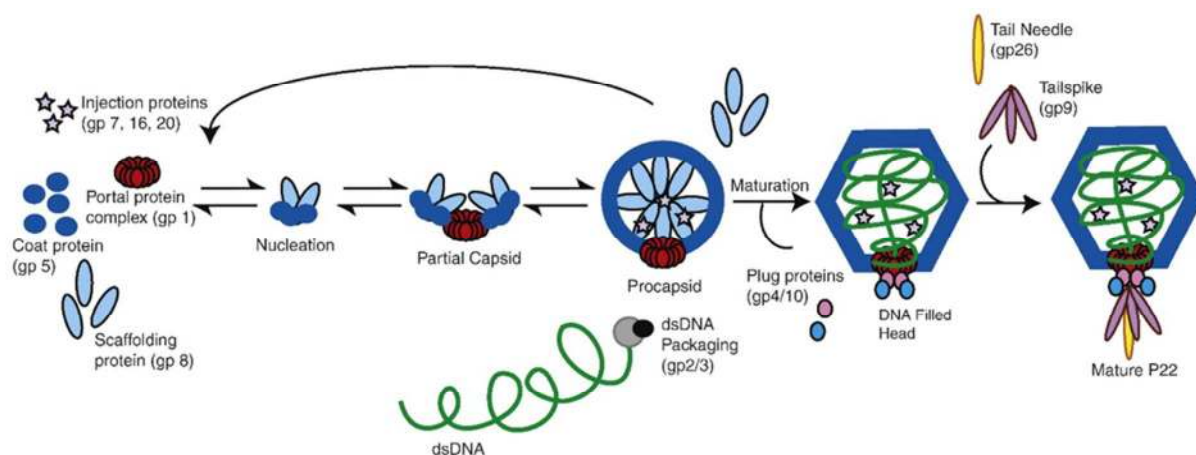


Figure 9. Scheme of P22 assembly. Figure adapted from reference ⁶⁵ with permission from Elsevier, copyright 2010.

3.2.3 Assembly around a nucleic acid cargo

Other assembly mechanisms involve assembly directly around the nucleic acid cargo as demonstrated experimentally^{53, 148-151} and suggested by simulations.¹²⁷ This mechanism generally includes most ssRNA viruses. The primary driving force for this process is electrostatic interactions between the positively charged capsid proteins and the negatively charged nucleic acid cargo, but recent work has shown that specific RNA sequences act as a signal for packaging and play a key role in the assembly of some capsids.^{151, 152} The RNA of tobacco mosaic virus initiates its own packaging and the coat protein subunits co-assemble with RNA to form the characteristic rod-shaped particle in which the nucleic acid is sequestered on the interior of the 300 nm long rod-shaped.^{153, 154} In another example, based on recent cryo-EM structures of empty VLP and fully formed CPMV, a model for assembly both with and without the RNA cargo (Figure 10) has been established. The C-terminal extension of the small subunit (S) plays an important role by not only forming favorable hydrophobic and electrostatic contacts with neighboring subunits to stabilize penton formation, but also promotes formation of an RNA binding site, and therefore can be considered a scaffolding protein or molecular chaperone.¹⁵⁵ The C-terminal extension is highly positively charged and it is cleaved during virion assembly, but stay intact longer when no RNA is present. In the virion, the interface between the two pentamers (two fold axis) has the strongest interaction with the RNA cargo and begins the encapsulation process. Many other members of the *Comoviridae* family (of which CPMV is a member) have a similar C-terminal extension that is cleaved during assembly, however the length and sequence are highly variable.

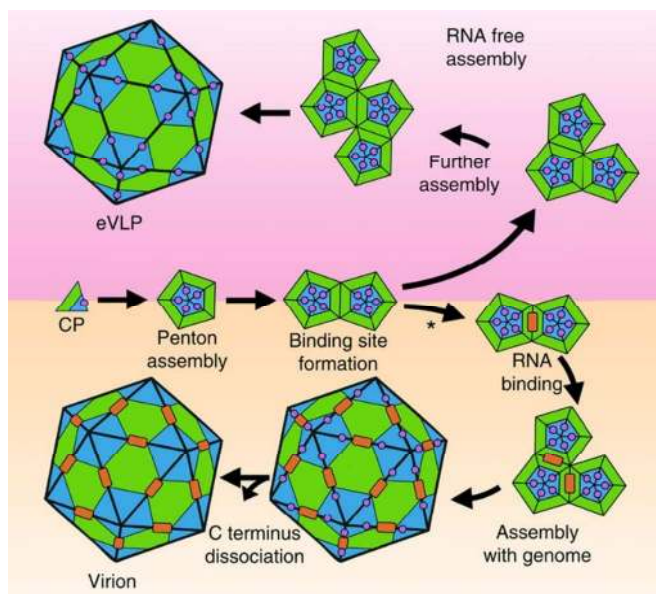


Figure 10. A model for CPMV virion and empty VLP assembly. Pentamers are made up of large (L, green) and small (S, blue) subunits. The C terminal extension is represented by the violet circle. RNA is the orange rectangle. In RNA free assembly (upper panel) C terminal cleavage is slow. In the presence of RNA (bottom panel), the C terminal extension is cleaved fast and RNA binds at the 2-fold axis. Figure reproduced from reference ¹⁵⁵ with permission from Nature Publishing Group, copyright 2015.

From these experimental observations and simulations for assembly around a polyelectrolyte (i.e. a nucleic acid cargo), two proposed mechanisms have emerged (Figure 11).^{127, 156, 157} In one mechanism the CPs assemble “en masse” in a disordered orientation and cooperatively rearrange to form the ordered capsid.¹⁵⁸ The other mechanism is similar to the nucleation and growth mechanism described above, in which the polyelectrolyte helps stabilize CP-CP interactions of an ordered nucleus, and growth continues with sequential addition of CPs. Simulations suggest that ionic strength plays a role in determining which mechanism is followed. At low ionic strength, CP-polyelectrolyte cargo interactions dominate and the en masse mechanism is observed, whereas at higher ionic strength, CP-CP interactions dominate and nucleation and growth is observed.¹⁵⁶ These observations point to the critical role pH and ionic strength play in capsid assembly, as has been observed experimentally.

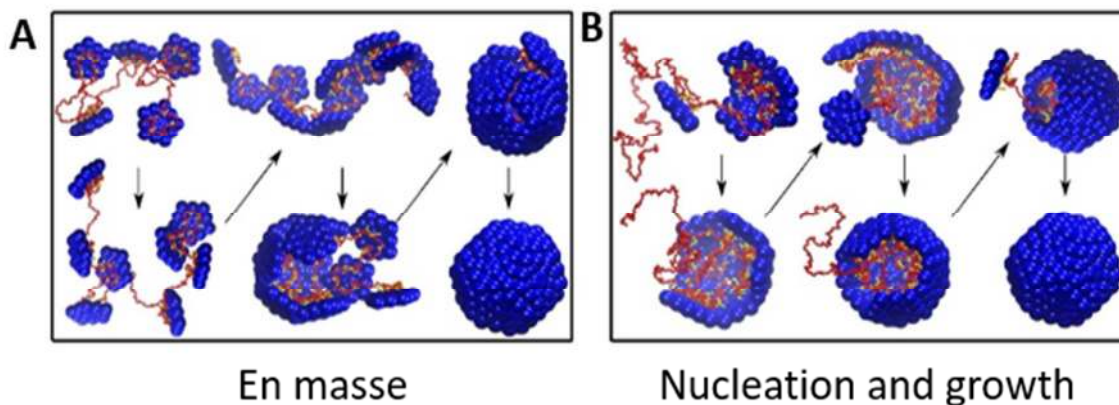
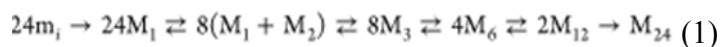


Figure 11. Assembly of a capsid around a polyelectrolyte cargo (e.g. a nucleic acid). Red is the nucleic acid cargo and blue is the coat protein. At low ionic strength (stronger coat-cargo interactions) and weak cargo-cargo interactions, an en masse mechanism is favored (A). The nucleation and growth mechanism is favored when coat-cargo interactions are weaker (higher ionic strength) and stronger coat-coat interactions. Figure adapted from reference ¹⁵⁶ with permission from Elsevier, copyright 2014.

3.2.4 Assembly of Other Protein cages

A few studies have investigated the assembly mechanism for non-viral protein cages. For example, a mechanism for the assembly of apoferritin was first put forth by Gerl and coworkers.^{106, 159} The assembly is proposed to proceed first by unstructured monomers (m_i) adopting a structured confirmation (M_1), which quickly form stable dimers. Dimers, trimers, hexamers and dodecamers intermediates were all detected. In the final step, two dodecamers assemble to make a fully formed 24-mer.⁹⁹ (Equation 1)



Recent work has used time-resolved small angle X-ray scattering (TR-SAXS) to investigate this mechanism further. The observed TR-SAXS profiles were explained by a slightly modified simple scheme given in Figure 12.¹⁶⁰ The main difference being monomers and trimers are not included as intermediates of the reaction. Like the other assembly mechanisms described previously, pH, ionic strength, and subunit concentration are important considerations. In the case of ferritin, these factors may influence not only the kinetics of the reaction, but which intermediates are observed.^{160, 161}

Another important example of non-viral protein cage assembly are the bacterial microcompartments, in which multiple copies of enzymes are enclosed inside of a single proteinaceous shell. Hagan and coworkers have adapted their previous model¹⁵⁶ for encapsulation of a polyelectrolyte cargo to study encapsulation of multiple cargos,¹⁶² like that observed with a bacterial microcompartment. Their simulation showed that assembly can occur via two distinct pathways: 1) a single-step assembly in which cargo molecules and shell subunits assemble simultaneously, or 2) a multi-step assembly in which the cargo molecules assemble first to form a dense globules, followed by adsorption and assembly of shell subunits around the globules. Relatively weak cargo-cargo interactions lead to single-step pathway, whereas stronger cargo interaction tend to follow the multi-step pathway. Their simulation study also indicates that encapsulation of high density cargo requires a balance between cargo-cargo, cargo-shell and shell-shell interactions (in the range of 5-10 $k_B T$)¹⁶², much like virus particle assembly described earlier. These observations indicate that a delicate balance of molecular interactions (a “sweet spot”) between all the subunits of a protein cage must be struck in order to form functional particles.

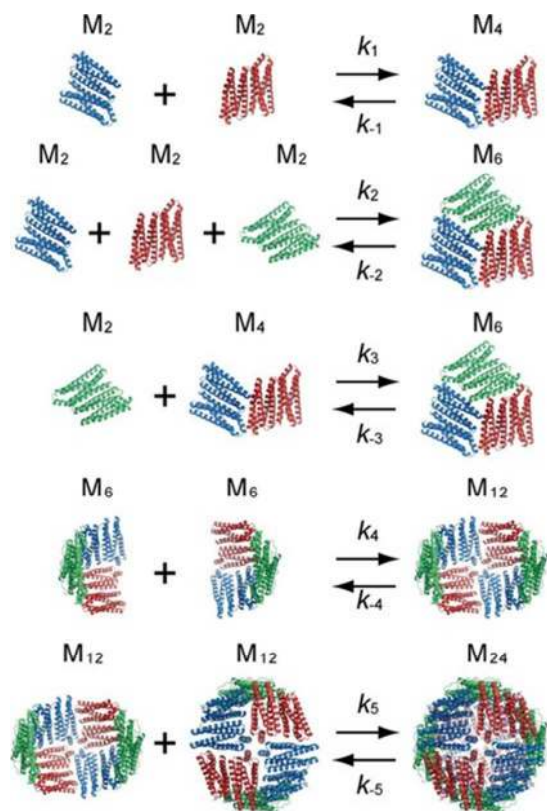


Figure 12. Proposed model for the assembly of ferritin by Sato et al. Each dimer unit is depicted as a different color. In this mechanism, hexamers (a trimer of dimers) can form directly from dimers, but also from a tetramer and a dimer. In the final step, 2 dodecamers form the 24-mer cage. Figure reproduced from reference ¹⁶⁰ with permission from the American Chemical Society, copyright 2016.

3.3 Factors influencing *in vitro* assembly

3.3.1. Changes in pH and ionic strength can lead to different assembled structures

CCMV is the first icosahedral virus that was reassembled *in vitro* to form an infectious particle.^{50, 163-166} Since then many studies have investigated the factors influencing assembly, such as changes in pH and salt concentration, coat protein concentration, and CP ratio to nucleic acid.^{130, 167-169} The CP interacts with the RNA primarily through highly basic, N-terminal arginine-rich motifs (ARM). Under appropriate conditions, CCMV capsids can self-assemble in the absence of RNA. Assembled structures range from not only single-walled capsids, but

multishell and tubular structures, depending on the pH and salt concentration (Figure 13). Multishell formation is favored at low pH (and low ionic strength) starting near pH 3.7 (the pI for CCMV CP) and up to pH 5. As pH increases, the capsid exterior becomes increasingly negatively charged, and electrostatic interactions between the exterior surface and the positively charged interior surface dominate assembly. The number of shells formed is limited to ~ 3 due to the unfavorable changes in the shell curvature as the layers grow. Multishell formation decreases and a single shell regular capsid with 28 nm becomes dominant as ionic strength increases, due to charge screening.⁵⁶

The crossover to tubular rather than spherical structures occurs near pH 6 and above. This arises due to favorable hexamer-hexamer interactions rather than hexamer-pentamer interactions. The hexagonal sheets fold to form tubular structures some of which are capped with hemispheres of hexamers and pentamers. Smaller diameter tubes are favored at higher ionic strength due to screening of the inner and outer surface mentioned above.⁵⁶

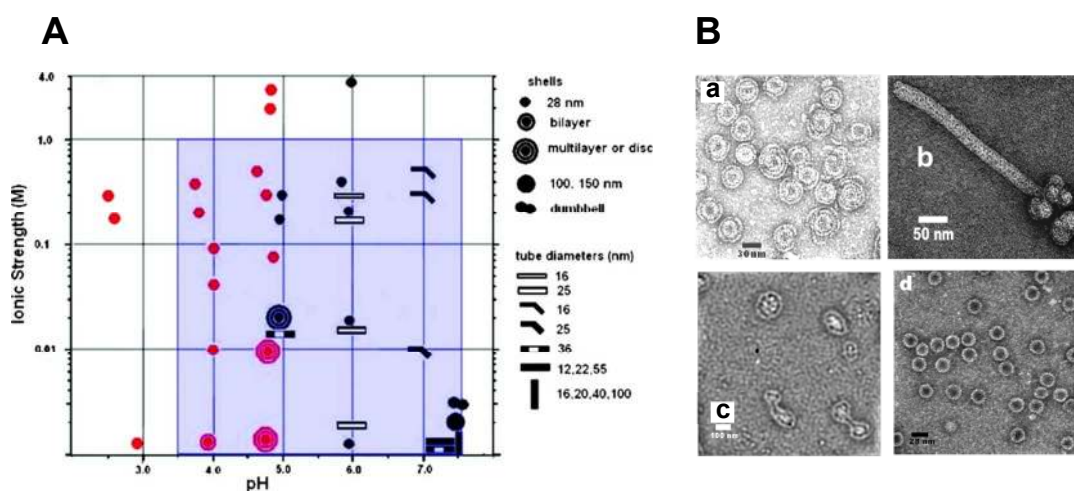


Figure 13. The range of particle morphologies possible of CCMV assembly as a function of pH and ionic strength. (A) Phase diagram of protein assembly outcome over a range of pH and ionic strength. Samples were buffered with sodium cacodylate (red) or sodium citrate (black). The blue area represents conditions of assembly from ref. ¹⁶⁸ and ¹⁶⁹. (B) Electron microscopy images

of different observed morphologies. (a) Multishelled structures are dominant at pH 4.8 and $I = 0.01$. (b) Tubular structure observed at pH 6.0 and $I = 0.01$ (c) Dumbbell shaped particles at pH 7.5, 0.001 M cacodylate buffer. (d) Single walled shells at pH 4.67 and 0.1 M acetate buffer and $I = 0.1$ Figure reproduced from reference ⁵⁶ with permission from the American Chemical Society, copyright, 2009.

pH also plays a critical role in formation of CCMV capsids around an RNA cargo. Near neutral pH, strong interactions between RNA and CP (containing the highly-positively charged arginine containing N- terminus) result in partially formed protocapsids that condense the RNA. Lowering the pH increases the strength of CP-CP interactions due to protonation of Glu81 and results in procapsid formation.⁵³ The assembly can be described as two separate steps required to obtain the properly assembled structure. First, at low ionic strength and neutral pH, the CPs undergo disordered adsorption on the RNA. When the pH is reduced, CP-CP interactions are strengthened, and the ordered capsid is formed. Fine-tuning these interactions is critical for proper in vitro assembly; if the attractions are too weak, the capsid is not formed. If they are too strong, the assembly falls into kinetic traps.¹⁴⁹ For complete packaging of the native RNA, an excess of CP is needed – specifically when there is a matched amount of charge among CP and the RNA.⁵³

3.3.2 Formation of non-native viral cages

Non-native, aberrantly assembled, virus particles that form tubular structures are known as polyheads. These have been observed in HIV,¹⁷⁰ bacteriophage T4,¹⁷¹ bacteriophage T7,¹⁷² and bacteriophage P22.^{173, 174} Generally these structures are formed as a result of mutations in the CP sequence. Remarkably, single mutation of F170 residue in the CP of P22 results in formation of polyheads up to 2 μm in length.^{173, 174} F170 is located in the β -hinge region of the CP, and causes decreased flexibility in the rest of the CP (especially in the A domain) and

consequently favors assembly of hexamers over pentamers, suggesting importance of CP flexibility to direct assembly of CPs into a proper capsid structure. The polyheads structures consist mostly of hexamers, but pentamers can be incorporated at the ends, which leads to termination of the growing structure.¹⁷³

3.4 Synthetic/computational approaches to designing protein cages

Inspired by the intricate self-assembly of native proteins into larger closed shell architectures, researchers have applied the principles learned from nature to computationally design proteins to assemble into a range of synthetic protein cages. A key concept to designing these protein cages is the underlying symmetry of the assembled cages. Highly symmetrical structures assembled from a defined number of subunit building blocks allow for a minimum number of subunit interactions to be designed, as each new interface increases the complexity of the design.¹⁷⁵ One such approach is to create fusions of known oligomeric proteins that already have intrinsic protein-protein interaction interfaces.^{12, 175-179} ¹³ For example, the design of a protein cage with cubic symmetry (432) has been successfully demonstrated by fusing two natural protein oligomers, a dimeric and a trimeric protein (Figure 14). This defines two different symmetries, two-fold and three-fold, that are necessary for assembly into the correct structure. A computationally designed linker molecule provided the correct angle needed for self-assembly of the cubic structure, but other geometries such as tetrahedral and triangular prism structures are also observed.¹³ The other approach is to design new protein-protein interfaces.^{175, 176, 180, 181} This is a new, very promising approach for precisely designing protein cages with desired size, shape, and amino acid composition for specific applications. While very specific interfaces can be designed, it comes at a cost of requiring sophisticated algorithms for design.¹⁷⁵ This can be

minimized however, by using natural oligomerization motifs as a model and then begin building new interfaces by sequence mutations.¹⁷⁶ King et al. were able to engineer a number of two component protein cages using a library of dimers and trimers.¹⁸⁰ Computationally this is accomplished by first choosing the desired architecture and symmetry docking the components (regions with large areas of contact) followed by re-designing the contact interface by changing the amino acid composition at the interface. The design feature makes use of hydrophobic protein cores surrounded by polar rims. More recent work by Baker et al showed that a 60 subunit protein cage could be computationally designed from trimeric protein building blocks.¹⁸² In another example, Jerala et al showed tetrahedral protein cages,, four-sided pyramids, and a triangular prism could be assembled from protein coiled-coil dimers.¹⁸³ *De novo* designed cages offer advantages over natural protein cages, such as customizable geometry, functionality, and increased stability. However, one disadvantage of this approach is that when computationally designed proteins are expressed *in vivo*, many sequences often result in insoluble inclusion bodies, which cannot yet be accurately predicted *a priori*.

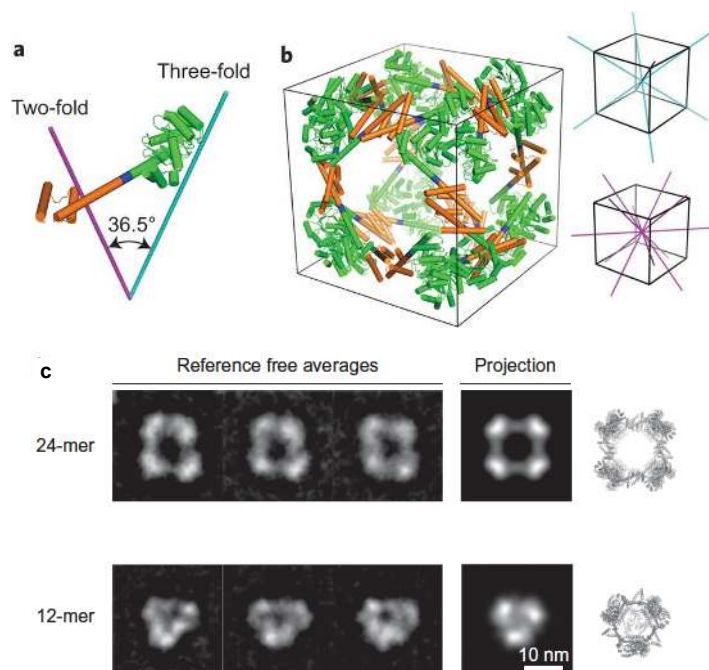


Figure 14. Model of the engineered fusion protein and the resulting octahedral structure formed. (a) The fusions are composed of a trimeric protein (KDPGal aldolase, green) and a dimeric protein (FkpA, orange) with a four residue helical linker (blue). The lines represent the three-fold (cyan) and two-fold (magenta) axes. (b) The 24-mer cage with octahedral symmetry, with the symmetry axes shown. (c) 2D class averages of the 24-mer and 12-mer obtained after aligning, clustering, and average particles from several particles (left) which are consistent with calculated projections (center). The 3D atomic models (right) are also included. Figure reproduced from reference ¹³ with permission from Nature Publishing Group, copyright 2014.

4. Encapsulation within Protein Cages

Although the biological function of the previously mentioned protein cages are different from each other, they have a relatively simple common task, i.e. sequester and protect the encapsulated cargo. The closed shell architecture of protein cages clearly defines a unique interior environment that is physically separated from the bulk, exterior, environment. Millions of years of evolution has taken advantage of this for the sequestration and protection of viral genomes within viral capsids, for the sequestration of an inert and non-toxic form of iron as iron oxide in the protein cage of ferritin and in some bacterial microcompartments the shell may act

to localize enzymes capable of transforming harmful intermediates. There are many ways to achieve sequestration of cargos, and the synthetic community has paid attention to the structural and mechanistic understanding of how biological systems have achieved this, which has provided inspiration for biomimetic approaches to materials design and synthesis.² In this section, we discuss the approaches and the inspiration behind using protein cage architectures as size constrained containers for the encapsulation and sequestration of a wide range of cargos.

Conceptually the encapsulation of a cargo on the interior of any one of these protein cages can usefully be approached in two ways: either through the initial formation of the protein cage which can then subsequently be filled with cargo or alternatively, through the entrapment of the cargo on the interior of the cage during the process of assembly from individual components into a cage. As we discussed above, examples of each of these approaches can be found in the packaging of nucleic acids within viral capsids, i.e. genomic DNA packing into a preformed procapsid as observed in bacteriophage P22,¹³⁸ or subunit proteins interacting with nucleic acids and packaging it during capsid assembly, which is observed in CCMV.⁵³

Using the preformed capsid as a reaction vessel requires that the cargo (or a cargo precursor) can traverse the shell. This usually implies the use of a small molecule or at least a cargo of size commensurate with the pore size of the capsid or that the capsid has dynamic behavior allowing the pore size to fluctuate, which could allow larger molecules to pass across the shell. Alternatively the cargo can be synthesized on the interior of the preformed capsid from small molecule precursors that can freely diffuse across the capsid shell. This is akin to a ‘ship-in-a-bottle’ synthesis/assembly and is the approach seen biologically in the biomineralization of an iron oxide particle on the interior of the iron storage protein ferritin where iron ions can

diffuse through the shell but the inorganic polymer of iron oxide is too large to leave and is therefore sequestered on the interior.^{3, 184}

Co-assembly of capsid and cargo on the other hand suggests an interaction between the capsid building blocks and the cargo. The type and strength of this interaction can vary widely from direct covalent fusion of the cargo and the capsid protein to weaker non-covalent interactions. Multivalent interactions and templating of the capsid assembly through interactions between the capsid and the cargo have biological relevance and have provided a conceptual framework that has been effectively exploited for synthetic applications.

4.1 Ship-in-a-bottle like synthesis in preformed capsids

As we pointed out previously, the ferritin family of iron storage proteins are widespread in almost all domains of life and function largely to catalyze the oxidation of ferrous ions and nucleation of a ferric oxyhydroxide nanoparticle sequestered within the cage.⁴ The protein subunits that assemble into these cage-like architectures contain catalytically active sites for oxidation of Fe(II) ion to Fe(III) (ferroxidase sites) and iron oxide nucleation which is largely through electrostatic interactions between the capsid and the accumulated ions. The cage has small well-defined pores of roughly 3 Å in diameter along the 3-fold and 4-fold symmetry axes.¹⁸⁵ Electrostatic potential calculation of ferritin suggests that the 3-fold channel is likely to be a major entrance that enables iron ions to traverse the shell and reach the interior. This indicates that ferritin has efficiently evolved this design of taking up iron from external environment into its cavity. The overall oxidation and mineralization reactions are catalyzed on the interior of the cage.¹⁸⁵⁻¹⁸⁷ This mineralization of the small diffusible precursors is an

excellent example of the ship-in-a-bottle synthesis on the interior that results in the formation of a nanoparticle product that is sequestered on the interior of the cage. In the case of ferritin this mineralization has the additional effect of continually removing soluble ions from solution thus creating a concentration gradient across the shell, which effectively drives the transport of ions into the cage and results in further accumulation on the interior. These components are all programmed into relatively simple protein subunits that can act both as the building blocks for capsid assembly and the active catalyst for the biomineralization.

Many studies have shown that a wide range of ionic and metallic nanoparticles, with no direct biological relevance, can be selectively grown inside the pre-formed protein cages derived from ferritin and related Dps proteins (Figure 15A).^{4, 188-192} The substrate flexibility in ferritins and Dps is largely due to the facile transport of cations across the capsid shell in response to an electrostatic gradient, and facilitated nucleation on the interior of the cage driven by ion accumulation at patches of high negative charge density on the interior protein interface. In addition, the protein cages themselves provide a well-defined volume and thus control and limit the size of the inorganic nanoparticles grown.

The strategy of using electrostatic interaction between protein cages and cargo molecules for cargo entrapment has been extended beyond ferritin and Dps. The observation that CCMV packages its RNA cargo primarily through electrostatic interactions between the RNA and the capsid is reminiscent of the ion-capsid interactions in ferritin, despite the fact that the charge relation between cage interior and cargo is opposite, i.e. interaction between highly positively charged N-terminus domain of capsid protein and negatively charged nucleic acid cargo. Thus the CCMV viral protein cage was initially explored as a constrained environment for a range of nanomaterial synthesis and encapsulation from negatively charged precursor molecules. The

viral capsid has pores at subunit-subunit interfaces allowing free access to small molecules through the capsid shell. The native CCMV, with its highly positively charged interior, was used to nucleate the growth of polyoxometalate crystals comprising the highly anionic $\text{H}_2\text{W}_{12}\text{O}_{42}^{10-}$ ions with NH_4^+ counterions¹⁹³. The positively charged N-terminus of the CCMV was postulated to act as a charged interface to accumulate the anions, which after nucleation could grow into the final particle filling the capsid. Using an anionic Ti ion precursor, this approach was subsequently also used to nucleate and grow TiO_2 within the CCMV capsids¹⁹⁴.

The importance of interfacial electrostatics and its utility in approaches for size constrained inorganic materials synthesis was further demonstrated in the CCMV system by means of a dramatic alteration of the charge on the interior of the capsid from positive to negative. Genetic modification of the N-terminus of the CCMV coat protein, in which all the basic residues (i.e. lysine and arginine) in the N-terminus were replaced with acidic, glutamic acid residues¹⁹⁵ did not affect the capsid assembly. However, the altered electrostatic character of the interior of the protein cage favored strong interaction with cations, which in conjunction with free molecular passage of ions across the capsid acts to promote oxidative hydrolysis leading to the formation of size-constrained iron oxide particles exclusively on the inside of the CCMV capsid; a functional mimic of the ferritin protein.

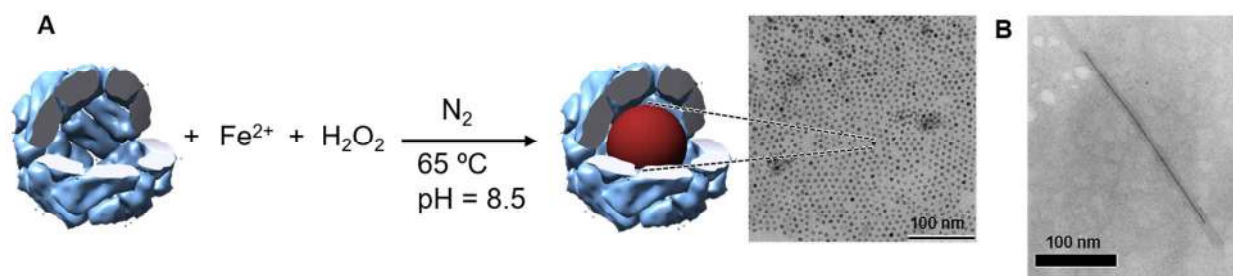


Figure 15. (A) Schematic illustration of magnetite (or maghemite) nanoparticle synthesis in ferritin. Magnetite nanoparticles can be formed and sequestered inside of ferritin even though temperature and pH used for this synthesis are far from physiological condition. TEM image on the right shows magnetite nanoparticles with approximately 6 nm in diameter synthesized inside of ferritin. (B) TEM image of a TMV containing an approximately 250 nm long nickel nanowire in the central channel. Part B reproduced from reference ¹⁹⁶ with permission from Wiley, copyright 2004.

Other protein cages such as chaperonins have been exploited to encapsulate or synthesize inorganic nanoparticles.¹⁹⁷⁻¹⁹⁹ Interestingly, ATP-dependent release of re-folded guest protein from chaperonins was successfully mimicked using preformed CdS nanoparticles and GroEL or *T.th* cpn. The CdS encapsulated inside of the chaperonin was released from the cage upon conformational changes induced by ATP hydrolysis.¹⁹⁷

The controlled synthesis of nanowires is practically important and protein cage assemblies have been used to direct these syntheses. The rod-shaped tobacco mosaic virus (TMV) has been demonstrated as a constrained template to grow inorganic nanowires (Figure 15B),^{196, 200-207} and one dimensionally aligned nanoparticle arrays.^{208, 209} The 300 nm long and 18 nm wide TMV has a 4 nm diameter interior pore in which metal nanowires such as Ni,²⁰² Co,²⁰² Cu,²⁰⁶ and Co-Fe alloy^{203, 206, 208} have been selectively grown. Although nanowires with the length of 100 – 200 nm are typically formed, fiber lengths up to 600 nm were observed, longer than a single TMV particle, suggesting some end-to-end assembly of individual capsids.²⁰²

4.2 Polymers inside protein cages.

All viruses are comprised of a capsid that surrounds and protects a viral genome that is a charged polymer of RNA or DNA. As pointed out above, in some viruses, the genome packaging takes place after the capsid has already formed. In phages, this process is energetically costly and the DNA genome is packaged at very high density with large internal pressure via ATP-driven packing process but with little or no attractive interactions between DNA and the capsid. Other viruses package their genome at much lower density and exhibit strong electrostatic interaction between the capsid and the packaged genome. As described earlier, the plant virus CCMV can be produced as an empty, nucleic acid-free capsid. Initial synthetic studies with empty CCMV demonstrated that incubation with an anionic polymer (poly(anetholesulfonic acid)) resulted in efficient encapsulation within the CCMV capsid.¹⁹³ Here, poly(anetholesulfonic acid) was incubated with the CCMV at pH 7.5 in which the CCMV took the swollen form (i.e. had larger inter-subunits pores) as discussed above, followed by lowering the pH to 4.5, well below the transition of the CCMV to the closed form. The polymer which passed across the CCMV capsid and accumulated was trapped inside of the capsid at the low pH condition. This strongly suggests that preformed capsids can be significantly more porous at the swollen form, although the size limit for free diffusion across the capsid has not been quantified.

The permeability and porous nature of the capsids suggested that site-specific extension of polymer strands anchored on the interior of the capsid could be achieved by transport of the monomers to the growing end of the polymer inside that capsid. Thus introduction of reactive amino acids, by site-directed mutagenesis, selectively on the interior of a capsid allows modified at these positions to create polymer-initiation sites from which the polymer strand can grow. Initiated polymer growth on the capsid interior enables not only the introduction of new chemical

moieties beyond what can easily be incorporated using genetic methods, but it does so at high density making efficient use of the entire capsid volume.

Organic polymer synthesis constrained inside of protein cages was first demonstrated using the copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) coupling of azide and alkyne bearing monomers. A series of polymers were synthesized through a stepwise molecular synthesis approach. In the small heat shock protein cage (sHsp) from *Methanocaldococcus jannaschii*, a branched polymer with pendant amines was synthesized using alternate addition of azide and alkyne monomers.²¹⁰ This resulted in an 8-fold increase in the number of functionalizable amine sites in comparison with wild-type sHsp, which could then be chemically labeled.²¹¹ Thus high density labeling of sHsp at amines with Gd-DTPA, a T₁ enhanced MRI contrast agent was achieved. A similar approach was taken using azide or alkyne containing metal containing compounds as the monomers to create a coordination polymer connected through the CuAAC coupling.²¹² Alternatively, a coordinatively unsaturated metal-ligand was attached selectively to the interior of the capsid lumen and served as a site for coordination polymer extension through iterative addition of a ditopic ligand and metal ion.²¹³

By far the most successful approach for high density polymer incorporation using functionalizable monomers has been with the use of “living” polymerization approaches,^{214, 215} which are fast and accommodate a wide range of monomers. Atom-transfer radical polymerization (ATRP) inside of a protein cage was first demonstrated in the bacteriophage P22 using methacrylate monomers.²¹⁶ An initiation site was introduced on the interior of the capsid through the site selective attachment of a tertiary bromide moiety on each of the 420 coat proteins. The polymerization of 2-aminoethyl methacrylate (AEMA) resulted in polymerized particles having $12,000 \pm 3000$ AEMA monomers per particle constrained within the P22 capsid.

The AEMA polymer served as a scaffold for the conjugation of a number of functional small molecules with high loading density via the pendant amine groups (Figure 16), including Gd-DTPA,²¹⁶ fluorescein isothiocyanate (FITC),²¹⁶ an isothiocyanate derivative (CoCl(dmgH)₂Pyr) and Eosin-Y Isothiocyanate.^{216,217} For example, chemical attachment of up to 10,000 Gd-DTPA to the polymer inside the P22 cavity led to development of a novel MRI contrast agent with a significantly high r_1 relaxivity. This approach was also amenable to co-polymer formation with a range of other monomers²¹⁸. Notably, polymer formation with ATRP has also been demonstrated on the interior (and exterior) of the Q β capsid where an azide-containing group was introduced and subsequently reacted with an ATRP initiating tertiary bromide via a CuAAC reaction.^{219,220}

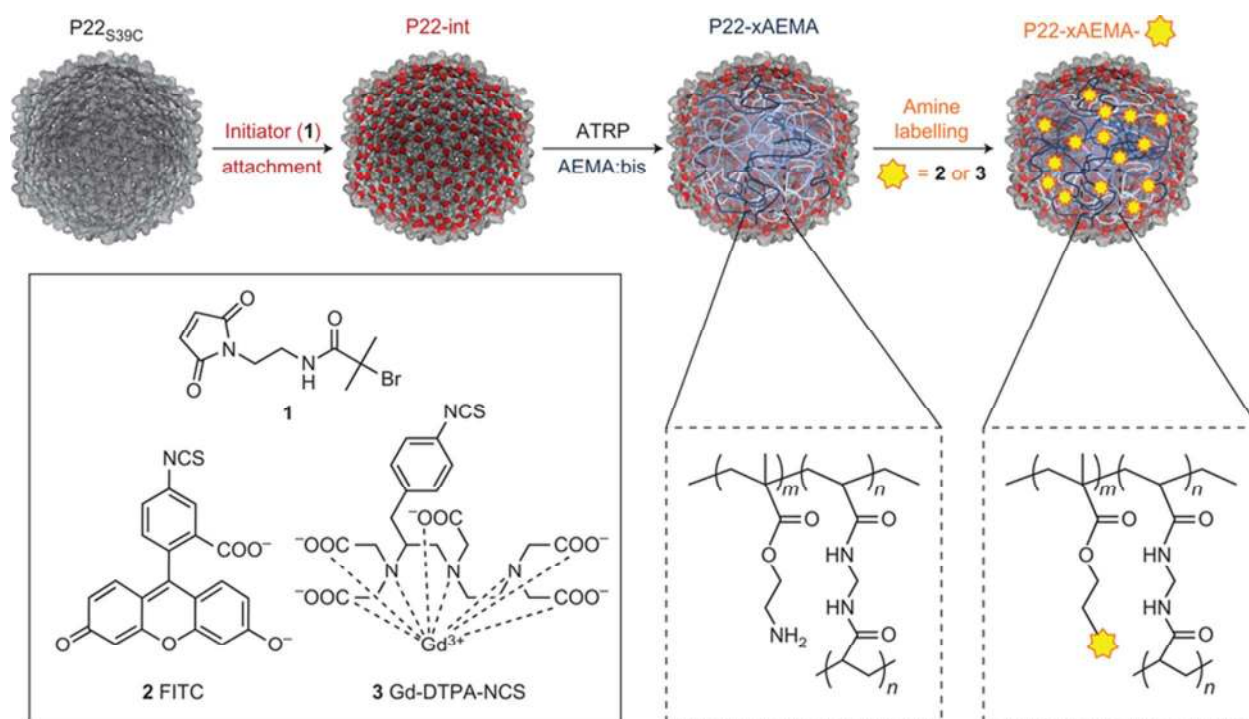


Figure 16. Atom-transfer radical polymerization (ATRP) can be used for high loading of Gd-DTPA within the P22 VLP. In the first step, an internal cysteine is labeled with a radical initiator (2-bromoiosbutyryl maleimide, **1**). Next poly-2-aminoethyl methacrylate (AEMA) is formed via ATRP. In the last step, the amino groups can be modified with FITC dye (**2**) or GD-DTPA-NCS (**3**). Figure reproduced from reference²¹⁶ with permission from Nature Publishing Group, copyright 2012.

Some protein cages have pores, which are large enough to allow macromolecules to pass through the protein shells to reach the interior cavity.²²¹⁻²²⁴ For example, a Group II chaperonin, thermosome,²²⁵ has large pores of approximately 5.4 nm diameter.²²¹ Generation four Poly(amidoamine) dendrimer, which has about 4.5 nm diameter, has successfully encapsulated inside of thermosome to use the protein cage for siRNA delivery vehicle and gold nanoparticle have been synthesized on the interior.^{222, 223} Similarly, a semiconducting polymer poly(2-methoxy-5-propyloxy sulfonate phenylene vinylene) (MPS-PPV) has been encapsulated inside of vault cage, while the MPS-PPV could not enter the cavity if vault subunit proteins were covalently cross-linked together prior to the polymer loading. This result indicates that the MPS-PPV can diffuse into the cage through inter subunit pores.²²⁶

4.3 Encapsulation via co-assembly into cage-like structures

Assembly of some virus capsids is often driven through specific directed interactions between the capsids and their endogenous cargos that facilitate the capsid assembly. As an approach it is thus amenable to synthetic manipulation and has been demonstrated in many different capsid assemblies and with many different types of cargo. This highlights the power and versatility of a bioinspired approach which harnesses sophisticated self-assembly mechanisms observed in nature for smart materials design and synthesis.

Electrostatic interactions between capsid and cargo is perhaps the simplest example of this approach. As mentioned earlier, CCMV uses electrostatic interactions between negatively charged RNA and positively charged regions of the capsid to package the genome. This strategy of self assembly of capsid subunits around nanoparticles has been applied for encapsulating pre-

formed inorganic nanoparticles (e.g. gold, iron oxides and quantum dots) inside of capsids derived from many viruses (e.g. BMV, red clover necrotic mosaic virus (RCNMV), CCMV, HIV-1 Gag protein and hepatitis B virus (HBV)).^{132-134, 227-235} In a notable example demonstrating a high yield of encapsulation (>95%), gold nanoparticles modified with carboxylic acid terminated poly(ethyleneglycol) (PEG) directed the assembly of BMV subunits (Figure 17A). Complementary electrostatics provided the interactions between the carboxyl groups and the capsid protein subunits while the PEG moiety served as a mimic for a disordered hydration layer in the RNA packaged in BMV.¹³⁴ BMV VLPs with discrete sizes, each adopting a specific T number, were obtained upon varying the size of the gold nanoparticle template, with the number of subunits in the final assembled capsid increasing with increasing template size (Figure 17B-D).²³¹ The co-assembly process of capsid and guest molecules through electrostatic interaction has also been exploited to encapsulate other cargo molecules, such as proteins and polymers.^{25,}

236-238

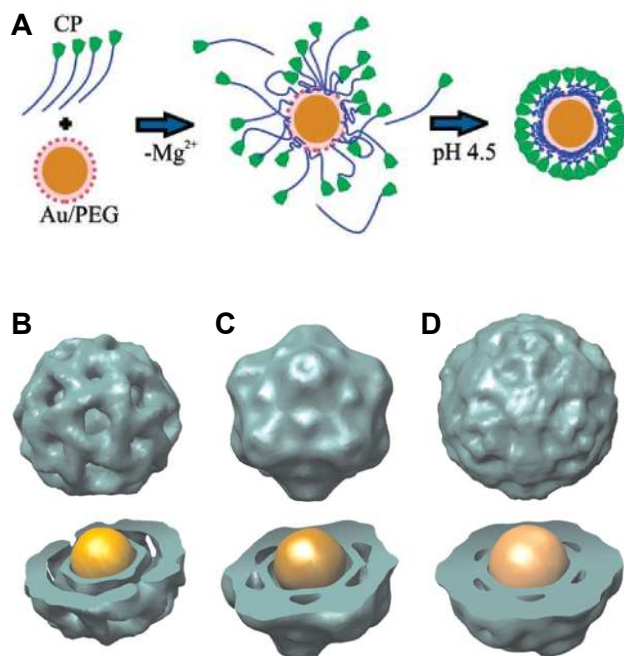


Figure 17. (A) Proposed mechanism of BMV virus-like particle (VLP) assembly around a gold nanoparticle (NP). First, electrostatic interaction leads to the formation of a disordered capsid protein (CP) - gold NP complex. The second step is a crystallization phase in which the protein-protein interactions result in the formation of a regular capsid. (B-D) Three-dimensional reconstruction (using negative stain data) of BMV VLP self-assembled around gold NPs with (B) 6 nm, (C) 9 nm, and (D) 12 nm diameter gold NPs. Structure and size of reconstituted models of (B), (C), and (D) resemble to a T = 1, a pseudo T = 2, and T = 3 models of BMV capsids, respectively. Part A reproduced from reference ¹³⁴ with permission from the American Chemical Society, copyright 2006. Part B reproduced from reference ²³¹ with permission from the National Academy of Sciences, copyright 2007.

A more elaborate level of directed cargo encapsulation has been demonstrated *via* an approach inspired by the assembly mechanism of some viruses in which specific interactions between capsid coat protein and cargo (protein or nucleic acid) play a critical role. For example, bacteriophages MS2 and Q β and RCNMV all use hairpin structures in their ssRNA to bind to the capsid proteins and thereby direct both capsid assembly and RNA encapsulation^{85, 239-241}. Similarly, as we discussed previously, bacteriophage P22 assembles initially into a procapsid form, templated by a scaffolding protein (SP), which interacts with the coat protein to direct assembly and during that process the SP is encapsulated within the procapsid⁶⁷. Co-opting these natural specific interactions has allowed for the selective sequestration of cargo with high encapsulation efficiency. Representative examples of cargo co-assembly by using specific interaction of capsids and cargos are discussed below.

4.4 Encapsulation of Enzymes

Among the many cargo molecules that have been encapsulated inside of protein cages, the encapsulation of enzymes has drawn particular interest recently. Because the interior of a viral capsid is a privileged environment, spatially separated from the bulk exterior environment by the

capsid shell, an enzyme can be sequestered inside the capsid to create well-defined nanoreactors capable of catalyzing a wide range of chemical transformations that conceptually resemble bacterial microcompartments. Encapsulation of enzymes within protein cages is a promising approach towards creating efficient catalytic materials and understanding the effects of enzyme crowding as well as multi-enzyme transformations.

Initial work in the encapsulation of active enzymes inside VLPs used the CCMV system, which undergoes reversible *in vitro* disassembly and reassembly in response to changes in solution pH as we discussed earlier. Reassembly of the capsid in the presence of the enzyme horse radish peroxidase (HRP) resulted in some fortuitous encapsulation of the enzyme within the reassembled CCMV capsid.²⁶ This passive encapsulation of an active enzyme, while not particularly efficient, paved the way for more directed approaches towards cargo sequestration and demonstrated key concepts: the capsid shell was shown to be porous towards small molecule substrates and the enzymes encapsulated within the CCMV were active but displayed some different kinetic behavior as compared to enzymes in the bulk. This observation suggests that the local environment of the capsid might affect enzyme function.

The *in vitro* assembly approach in CCMV has been refined to incorporate molecular recognition tags between the cargo and the capsid protein to achieve more efficient cargo encapsulation. Thus complementary peptides that form a heterodimeric coiled-coil were fused to the CCMV CP and the cargo protein²³⁶ allowing non-covalent coiled-coil interactions to occur prior to capsid assembly. This approach was also used to encapsulate the enzyme lipase B (PalB) from *Pseudozyma antarctica* inside CCMV and resulted in packaging of an average of 1 – 4 PalB per capsid.²⁴² The initial velocity of singly encapsulated PalB was reportedly five times faster than the velocity observed for PalB free in solution. This strategy demonstrated successful

encapsulation, but the method required incorporating unmodified CCMV CPs into the *in vitro* assembly system, together with the fusion proteins, to achieve assembly of the CCMV capsid with the PalB cargo. This approach naturally results in a distribution of species, ranging from assembled capsids containing no enzyme cargo to fully packed capsids.

There is a growing desire to directly harness the synthetic capabilities of biology and let biological systems direct the synthesis and assembly of complex materials directed only through altered genetics. Thus a goal is to establish robust methods for *in vivo* nanoreactor assembly. This requires producing all the components for assembly in the cell and introducing specific molecular recognition between the capsid protein and the cargo, with no interference from the other biomolecules within the cell. This genetically programmed assembly would take advantage of molecular level design and self-assembly, as well as metabolic engineering on the organismal level.

Directed enzyme cargo encapsulation has been shown for a number of different enzymes and capsids. Using a highly directed approach for enhanced molecular recognition between capsid and cargo, Hilvert *et al* used the protein cage derived from lumazine synthase (LS) and created mutants with an increased negative charge density on the interior of the capsid through introduction of glutamic acid residues. The fusion of a deca-arginine tag to green fluorescent protein (GFP) cargo ensured the directed encapsulation through complementary electrostatic interactions between the cargo and capsid interior.²³⁷ This approach was then demonstrated *in vivo* where the directed evolution of *Aquifex aeolicus* (AaLS) variants were screened for enhanced binding and encapsulation of HIV protease within the LS capsid. Encapsulation of the protease rendered it inaccessible to proteins in the cell and resulted in enhanced cell survival.²³⁸ The negatively charged LS mutants have been shown to be highly effective at the directed

encapsulation of a wide range of positively charge cargo proteins including positively “supercharged” GFP and ferritin variants.^{243, 244} Additionally, various enzymes such as RuBisCO and carbonic anhydrase were tethered with the supercharged proteins to direct encapsulation of the enzymes in the negatively charged LS mutants, which lead to construction of a functional carboxysome mimic.^{245, 246} Interestingly the loading of cargo in the LS mutant did not require a co-assembly but could be achieved by incubation of the assembled LS with the charged cargo. Since the cargo proteins themselves are larger than the pores of the LS cage (ferritin has a molecular weight of 5×10^6 Da) it was uncertain how the large cargos could be transported acrossn the already formed LS cages. One suggested possibility was that the LS capsid assembly could be dynamic. However, a recent cryo-EM structural model of two LS variants was determined revealing a highly unusual, cage structures with large pores (each variant exhibits tetrahedral and icosahedral symmetry, respectively) that might accommodate transport of the cargo proteins across the capsid to the interior.²⁴⁷

Directed enzyme encapsulation has also been achieved using specific interactions, through molecular recognition. Finn and coworkers have developed a unique method for directed enzyme encapsulation in the virus Q β . They used a chimeric single stranded RNA composed of 3 parts: (1) a hairpin that associates with the internal surface of the Q β CP, (2) an RNA aptamer that binds to an arginine-rich peptide (Rev), and (3) the mRNA sequence that codes for the Q β CP that also acts as spacer between the other components. Cargo proteins tagged with the Rev peptide are thus primed for directed encapsulation into the Q β capsid through the interaction between the Rev peptide and the RNA aptamer (Figure 18).²⁴ The transcribed mRNA thus serves as both as the transcript for CP expression and as the recognition element for directed encapsulation. This approach, using the co-expression of the CP/RNA mediator together with the

tagged protein cargo, has been effectively used for the directed encapsulation of other protein cargoes within the Q β system.²⁴⁸ In another example, MS2 shows sequence specific interaction between its CP and a short RNA stem-loop to trigger capsid assembly,⁸¹ hence the RNA sequence has been utilized to encapsulate cargo molecules inside of the MS2 VLPs.^{249, 250} Furthermore, encapsulation of cargo proteins has also been demonstrated by using DNA oligomers and negatively charged peptides as tags to induce capsid assembly and mediate cargo-capsid interaction.²⁷ Intriguingly, the presence of a protein stabilizing osmolyte, trimethylamine-N-oxide, has been reported to increase yield of capsid reassembly.²⁷

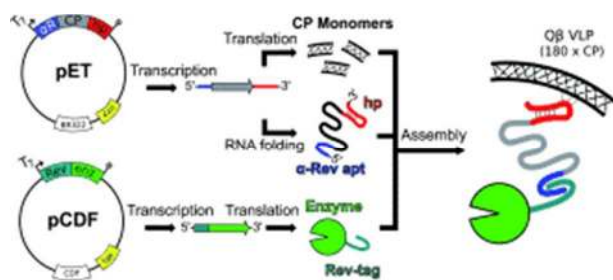


Figure 18. An example of a direct encapsulation of enzyme into a capsid by molecular recognition. A Rev-tagged enzyme is bound to the interior of the capsid via an RNA with a Rev aptamer and a Q β genome packaging hairpin that is bound to the interior of the capsid. Figure reproduced from reference²⁴ with permission from Wiley, copyright 2010.

Yet another approach for directed encapsulation involves the genetic fusion of a cargo of interest to a scaffolding protein. Of the protein cages used for directed encapsulation, the P22 bacteriophage has emerged as an extremely versatile platform for the programmed and directed encapsulation of a wide range of gene product cargoes within the *E. coli* cells used for heterologous expression. Recall that the P22 SP templates capsid assembly and that SP bound to the interior of the capsid is incorporated into the assembled P22 VLP. The wild type SP is a 303 residue long protein and it has been shown that it can be truncated to an essential C-terminal

scaffolding domain without disrupting its ability to direct assembly.^{251, 252} This truncation has been used to simultaneously direct capsid assembly and cargo encapsulation as genetic fusions to either the N-^{20, 21, 23, 253-257} or C-termini²² of the SP. Remarkably, the small helix-turn-helix domain, that is essential for directing capsid assembly, can be fused to protein cargo of up to 180kDa without affecting the fidelity of capsid assembly.²¹ By co-expressing a genetic fusion of a protein cargo and the truncated SP together with CP, a number of functional enzyme and protein cargoes have been successfully encapsulated inside the P22 VLP, which maintain their functional activity. Directed assembly of an enzyme cargo has also been demonstrated with other cages. For example, Handa and co-workers used the minor coat proteins (VP2) of the SV40 virus to direct encapsulation by genetic fusion of an enzyme cargo, cytosine deaminase within the VLP. Co-expression of the modified VP2 together with the major coat protein VP1 and the other minor coat protein VP3 resulted in assembly of SV40 VLPs with active enzymes entrapped on the interior of the capsid.²⁵⁸

The capsid morphology can have a profound effect on the enzyme cargo in terms of its local concentration, crowding effects, and access between the interior and exterior. The P22 VLP capsid undergoes a natural morphogenesis which mimics the native maturation of the infectious virus upon DNA packaging. The initially formed procapsid (PC) form of the P22 VLP expands, upon heating at 65 °C, from the spherical PC to the expanded (EX) form, which nearly doubles the interior volume of the capsid.⁷¹ Upon further heating, to 75 °C, the 12 pentameric vertices of the capsid are irreversibly lost leaving 12 pores 10 nm in size in the capsid that forms a structure affectionately referred to as wiffleball (WB).²⁵⁹ In some cases, where the cargo is thermally stable, the EX and WB morphologies of P22 can be accessed by heating the capsid and this has allowed an investigation of the kinetic behavior of encapsulated enzymes as a function of both

local enzyme concentration (crowding) and capsid porosity (substrate diffusion across the shell).^{255, 256} In the P22 encapsulation, the enzyme cargo is initially constrained to the interior capsid lumen in the PC form, but upon expansion to the EX the SP binding site is lost due to the conformational change of CP, and the cargo is free to occupy the entire volume of the capsid (Figure 19).²⁶⁰

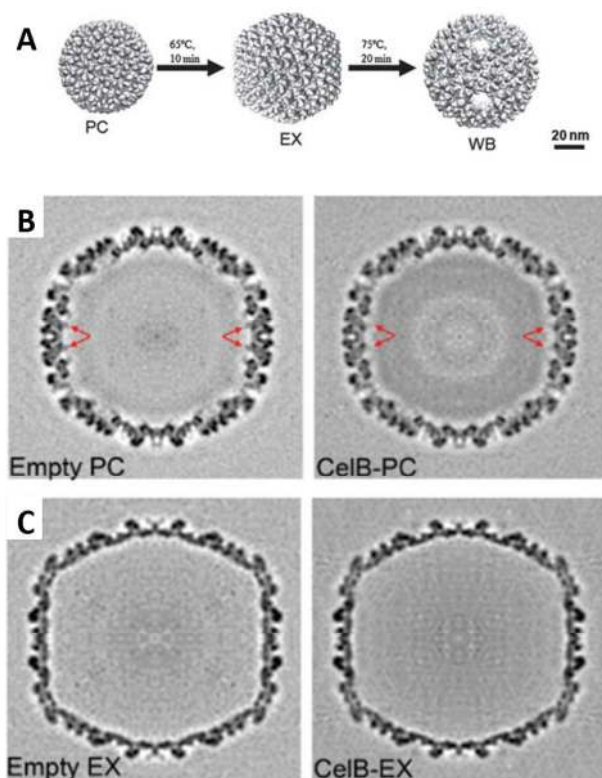


Figure 19. The morphologies of P22 and the mapping of cargo location within them. (A) P22 procapsid (PC) undergoes a morphological change to a more angular expanded (EX) from upon heating, mimicking the structure of the native virion. The internal volume increases approximately 35% and porosity of the capsid decreases. Heating further forms the wiffleball (WB) morphology, where pentamers are removed leaving large holes in the capsid wall. (B) The central sections of three-dimensional reconstructions of the empty P22 and CelB-P22. The darker region on the interior lumen of the capsid indicates higher density, caused by the presence of the bound cargos. The arrows indicate the increased density from the SP-C terminal region of the SP and the CelB-SP binding to the capsid interior. (C) Reconstructions of the EX forms with cargo. Decreased electron density reveals that the cargos are no longer bound to the capsid and can freely move within the entire capsid interior volume. Part A adapted from Ref.²⁶¹ with permission from The Royal Society of Chemistry. Part B and C reproduced from Ref.²⁶⁰ with permission from The Royal Society of Chemistry.

While a number of enzymes that have been successfully encapsulated within the P22 VLP show slightly lower k_{cat} values compared to the enzymes free in solution (likely due to the effects of molecular crowding and restricted dynamics during enzyme turnover)^{255, 256}, encapsulation of the heterodimeric NiFe hydrogenase from *E. coli*, Hyd-1, within the P22 capsid shows a dramatic enhancement of the activity as compared to the free enzyme (Figure 20).²³ The encapsulated Hyd-1 activity increases nearly 150 fold over the free enzyme and follows classic Michaelis-Menten kinetics. The enhancement is likely due to the entrapment of the enzyme at high concentration within the VLP thus shifting its equilibrium toward the more active heterodimeric form of the HyaA and HyaB subunits from the much less active monomers in bulk solution. It is also possible that the dimeric Hyd-1, presenting two copies of the SP from each subunit is favored for capsid assembly over a single copy as has been demonstrated in early studies on the effects of SP oligomerization on capsid assembly.¹³⁷ Therefore, capsid assembly itself might drive the Hyd-1 equilibrium towards the more active dimeric form and simultaneously entrap that form at high concentration within the capsid. Indeed, recent *in vitro* P22 VLP assembly study indicated that cargo proteins displaying multivalent SP preferentially encapsulated over monovalent wild type SP while assembling into VLP.²⁶²

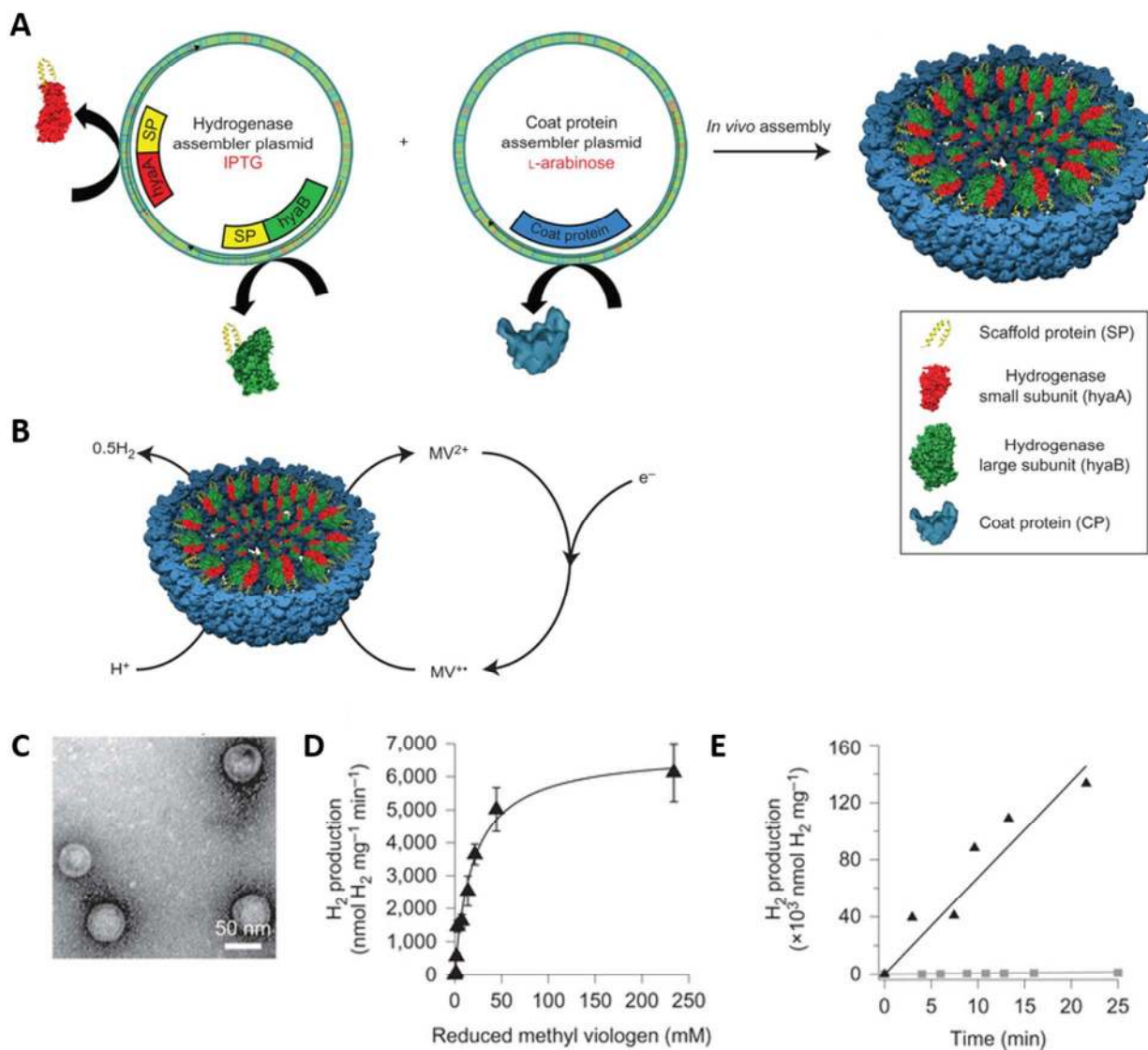


Figure 20. Encapsulation of an active hydrogenase within a P22 VLP (P22-Hyd-1). (A) Scheme showing expression of both subunits of the hydrogenase as fusions with truncated scaffold protein (hyaA-SP and hyaB-SP) under IPTG control with coat protein (CP) in a separate plasmid under L-arabinose control. The result is P22 capsid with both subunits packaged within it. Structures are PDB: 3USE (E.coli hydrogenase-1), PDB:2XYX (P22 procapsid), PDB: 2GP8 (scaffold protein binding domain) (B) Hydrogen is assayed via gas chromatography using methyl viologen (MV^+) as the electron transfer mediator and an electron source (dithionite or electrochemical reduction). (C) TEM image (negative staining) of P22-Hyd particles. (D) Specific activity of P22-Hyd at varying concentrations of methyl viologen. The enzyme activity follows Michealis-Menten kinetics. (E) Specific activity traces of P22-Hyd-1 (black triangles) and free Hyd-1 (gray squares) shows a marked increase in activity of the P22 encapsulated Hyd-

1 compared to Hyd-1 in solution. Figure adapted from reference²³ with permission from Nature Publishing Group, copyright 2015.

Unlike many other protein cages, chaperonins have relatively large pores likely due to their native function of assisting folding of guest proteins to its native state. These large pores of chaperonins allow the encapsulation of enzymes without disassembling the cage structure into subunits. For example, horseradish peroxidase was encapsulated into thermosome and used for polymerization (via ATRP) of an acrylate which occurred selectively inside of the cavity.²²¹

5. Functionalization of the exterior of Viruses and VLPs

In nature, the exterior of protein cages are in contact with the surrounding environment and must interact with the surroundings in order to function and protect the valuable cargo. The exterior surfaces of protein cages can also be rationally designed to conduct specific tasks and properties. For a virus, presentation of different functionalities on the exterior is necessary for targeting the virus to a particular location *in vivo*. Also, the exterior surfaces of ferritin has evolved to create electrostatic gradients around the three-fold channels, which provides a guidance mechanism for cations entering the protein cavity through the channels.¹⁸⁵ Because all the protein cages described here possess a highly symmetric polyvalent surface, the exterior functionalities are typically presented in a multivalent and highly symmetrical manner. For viruses, the polyvalent presentation of targeting and cell fusion moieties can enhance affinity to their host cells and infection.²⁶³ These examples found in nature inspire us to explore the exterior surfaces of protein cages to impart functionality by design. In this section, we present an overview of the variety of methods for functionalizing the exterior of a protein cage (summarized in Table 1).

Method of Modification	Examples
Genetic	<ul style="list-style-type: none"> • Amino acid modification (addition, deletion, mutation) • Phage display • Fusion proteins • Isopeptide formation (Sortase) • Split protein complementation (Spytag/Spycatcher)
Covalent (Bioconjugation)	<ul style="list-style-type: none"> • Amino acid labeling (cysteine, lysine, glutamate, aspartate, tyrosine) • Specific labeling at protein termini • Incorporation of non-natural amino acids • Atom transfer radical polymerization • PEGylation
Non-covalent genetic	<ul style="list-style-type: none"> • Decoration protein (Dec) • Biotin/Streptavidin

Table 1. Commonly used methods of exterior (and in some cases, interior) modification of protein cages discussed in this section.

5.1 Genetic functionalization

Genetic modification of the protein cage is perhaps the most widely used method for exterior functionalization. Modifications of the exterior can be achieved using molecular biology techniques to add, delete, or replace amino acid residues. Typically the protein cages are quite robust, and remain intact after minor genetic modifications.

Phage display is a powerful and widely used technique for high throughput screening of not only protein-protein and protein-peptide interactions but also interaction between peptides and abiotic materials such as inorganic materials. In phage display system, infected bacteriophage have a peptide or protein of interest “displayed” on the exterior of capsid. Some notable examples include using phage display to present a library of short peptide sequences.²⁶⁴⁻

²⁶⁶ A foreign gene is inserted within the coat protein gene and it is expressed along with the CP.

In this technique, viruses that have affinity for a particular target can be selectively amplified and characterized. Modifications to the protein cage are typically done at termini or exposed loops so that the coat protein can still fold properly.^{190, 267-272} The gene (nucleic acid) is packaged within the capsid, which allows for linkage of the phenotype (the displayed peptide) and the genotype (the encoded DNA). In a process termed biopanning, large libraries of random sequences can be screened against individual targets such as proteins, peptides, DNA, and abiotic materials. The binding phage can be isolated and its sequence determined. The filamentous phage M13 is one of the most commonly used platforms for phage display technology because of the ease of replication and commercial availability.²⁶⁵ There are a wide range of applications in molecular biology where phage display has proven useful, such as in vitro protein evolution, drug discovery, enzyme inhibitors, and finding protein partners.^{265, 273} This technique has also proven useful in the material science community. Certain peptides can be selected that show preferential binding for different inorganic surfaces/nanoparticles, based on the composition and crystallographic orientation of the material.²⁷⁴⁻²⁷⁷ In one early example, M13 bacteriophage was engineered with peptide sequences to act as a template for nucleation of ZnS or CdS nanocrystals to form nanowires.²⁷⁸

Whole proteins and protein domains of interest can also be genetically fused to capsid coat proteins.²⁷⁹⁻²⁸¹ For example, Gleiter and Lilie displayed a 6.8 kDa domain from protein Z on surface exposed loops of polyoma VLPs.²⁸² Protein Z is an antibody binding protein that was still capable of binding antibodies when attached to the VLP. In another example, Nabel and coworkers displayed trimeric antigen cargo (the ectodomain of the influenza hemagglutinin protein) at the 3-fold axes on a ferritin cage by creating a fusion at the N-terminus of the ferritin subunit to effectively create a nanoparticle based vaccine.²⁸³ Here, symmetry matching, *i.e.* the

trimeric antigen cargo was fused at a trimeric symmetry site of the platform protein cage, is an important design consideration and it lead to multivalent antigen display (eight trimeric antigen per ferritin as ferritin is composed of 24 subunit). The nanoparticle based vaccine produced an increased immunogenic response to many viral subtypes compared to the commercially available vaccine. Cell targeting peptides have also successfully displayed on the exterior of protein cages in a multi-valent manner by genetically introducing the peptides either at the end of N- or C-terminus of an individual subunit,^{190, 269, 284} or inserting into a surface exposed loop region.^{285, 286} For example, Kang and coworkers demonstrated that a non-viral protein cages, encapsulin, can also be genetically modified to display a targeting peptide ligand (Figure 21)²⁸⁵ and an encapsulated drug molecule can then be delivered to the targeted cell.

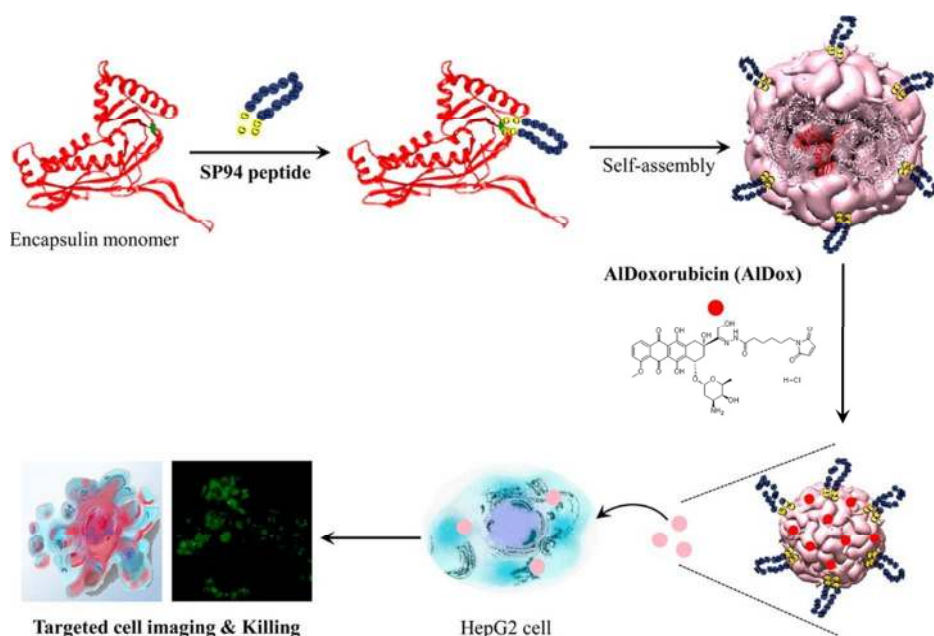


Figure 21. Functionalization and targeting of encapsulin. Encapsulin is genetically modified to contain the targeting ligand (SP94) and is assembled to form a protein cage. An anticancer drug aldoxorubicin was delivered to the HepG3 cells, resulting in cytotoxicity to the cells. Figure reproduced from reference ²⁸⁵ with permission from the American Chemical Society, copyright 2014.

Despite the successes described above the genetic fusion of functional proteins and peptides to protein cages can often result in undesired outcomes, such as the formation of insoluble fusion proteins or disruption of subunit assembly into a proper cage-like structure. To overcome this drawback, a bioconjugation technique using the enzyme sortase²⁸⁷ to mediate site specific attachment has recently been applied for functionalization of protein cages because this approach possesses great flexibility and versatility.^{288, 289} For example, GFP and influenza hemagglutinin head were successfully conjugated to the exterior surface of P22 VLP (Figure 22).²⁸⁸ The SpyTag/SpyCatcher protein ligation system²⁹⁰ shows great potential for bioconjugation to protein cages.²⁹¹ Thus, if a sortase recognized motif or a spyttag/spycatcher motif is successfully introduced onto a protein cage, nearly any protein cargo could potentially be conjugated to the protein cage in a “plug and play” modular manner using this strategy.

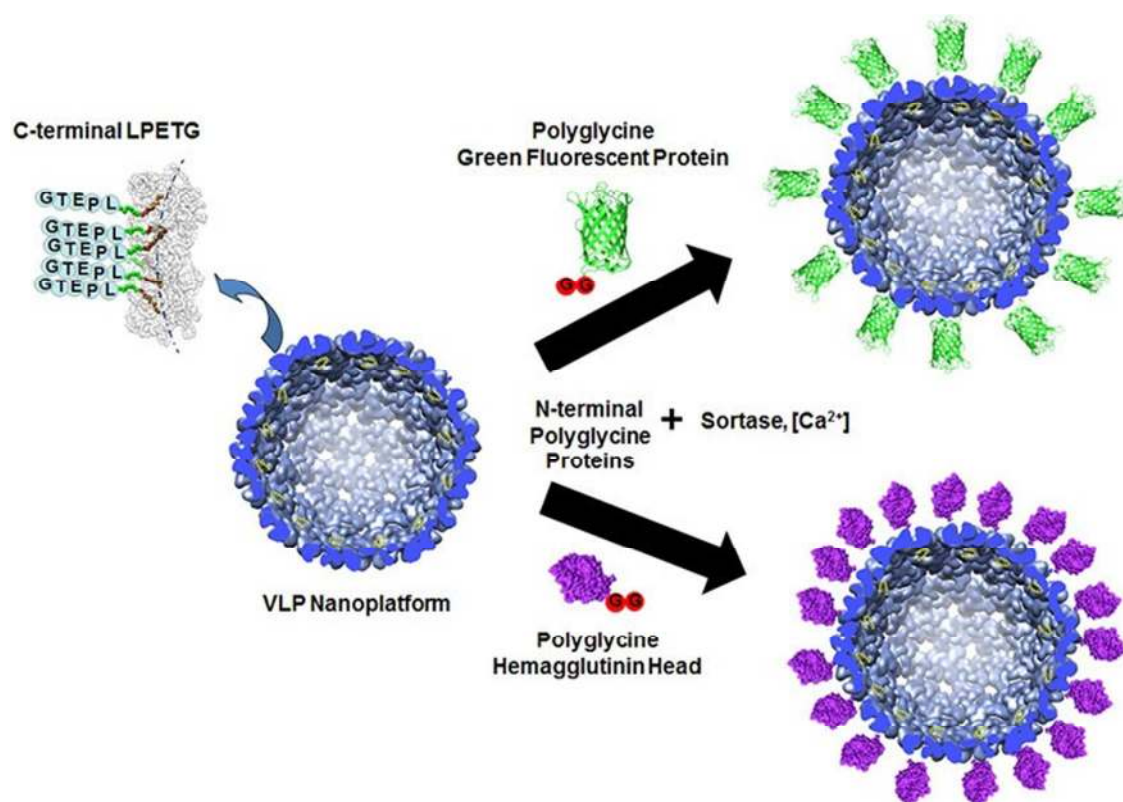


Figure 22. A modular approach to presentation of cargo on the exterior of a protein cage. The CP of P22 was genetic engineered to contain a LPETG amino acid at the the C terminal that can be covalently bound to a poly glycine sequence an N terminal of cargo (here, GFP or Hemagglutinin Head) via the enzyme sortase. This modular approach to allows for many different cargo types to be presented without the need for redesigning the modified CP sequence. Figure reproduced from reference ²⁸⁸ with permission from the American Chemical Society,. copyright 2017.

5.2 Covalent modification

Chemical conjugation is another widely used functionalization/bioconjugation approach that allows for incorporation of many different desired functional groups, both biological and chemical, into protein cages to change or enhance their function.²⁹² Because many of the protein cages utilized for nanomaterials syntheses have near-atomic resolution structure models, the number and location of functionalizable sites in a protein cage can be precisely known with knowledge of the protein sequence and structure. Covalent modification of protein cages initially utilized natural amino acids with reactive functional groups such as the sulfhydryl group (cysteine), the amino group (lysine) and the carboxyl group (glutamate and aspartate). For example, the sulfhydryl group can serve as reactive sites for thiol selective chemistry,^{293, 294} or for binding to a gold surface.²⁹⁵ High density labeling can also be achieved by labeling at amines using activated N- esters, and carboxylates can be activated into reactive esters. One major drawback of this conjugation approach is the poor selectivity of sites with the same functional group. Typically there are multiple cysteine, lysine, glutamate or aspartate residues in a protein subunit, and it is difficult to conjugate a target molecule to a specific site over others, which results in heterogeneous mixtures of modified protein cages.

In order to overcome the sometimes undesired non-selectivity of the reactions, there has been a continuing interest to develop new bioconjugation reactions with high level of site- and

chemo-selectivity under mild conditions in aqueous environments.^{292, 296, 297} For instance, the α -amine of the N-terminus is chemically unique and shows distinctive reactivity over lysines. Thus an increasing number of chemical conjugation strategies have developed to utilize the N-terminus amine as site-specific reactive group.²⁹⁷ For example, the N-terminus amine shows lower pKa (6 – 8) than other aliphatic amine (pKa \sim 10.5) as a consequence of the inductive effects of the nearby carbonyl group.^{297, 298} Hence selective acylation and alkylation of the N-terminus amine can be carried out at low to neutral pH.²⁹⁷ In an application relevant to protein cages, the TMV virus was covalently modified at both the N terminus and at an introduced cysteine residue with two different chromophores that undergo FRET to produce a “light harvesting rod.”²⁹⁹ Another emerging bioconjugation strategy is the incorporation of unnatural amino acids with bio-orthogonal reactive groups such as azido, alkyne, alanine and norbornene into a protein cage.^{18, 19} These functional groups can be uniquely modified over other natural amino acids in the protein. For example, Finn and coworkers have demonstrated functionalization of Q β mutants with azide- or alkyne-containing unnatural amino acids using Cu(I)-catalyzed [3+2] cycloaddition (known as “click chemistry”).¹⁹

Pioneering work by Francis and coworkers has developed a number of oxidative coupling reactions with electron rich aromatic species with both natural and unnatural amino acids that can be performed on proteins in aqueous solution under mild conditions.²⁹⁶ In one example, they utilize this strategy to dually functionalize both the interior and exterior of bacteriophage MS2.³⁰⁰ The interior was functionalized with a porphyrin maleimide that was introduced at a mutated C87 residue. The exterior was functionalized at the unnatural amino acid *p*-aminophenylalanine via oxidative coupling with a DNA aptamer that targets protein kinase 7 receptors on Jurkat leukemia cells. After incubation of the cells with capsids and upon light illumination, the

porphyrin generated a cytotoxic singlet oxygen species. The capsids selectively targeted and killed 76% of the Jurkat cells in the presence of erythrocytes. (Figure 23)

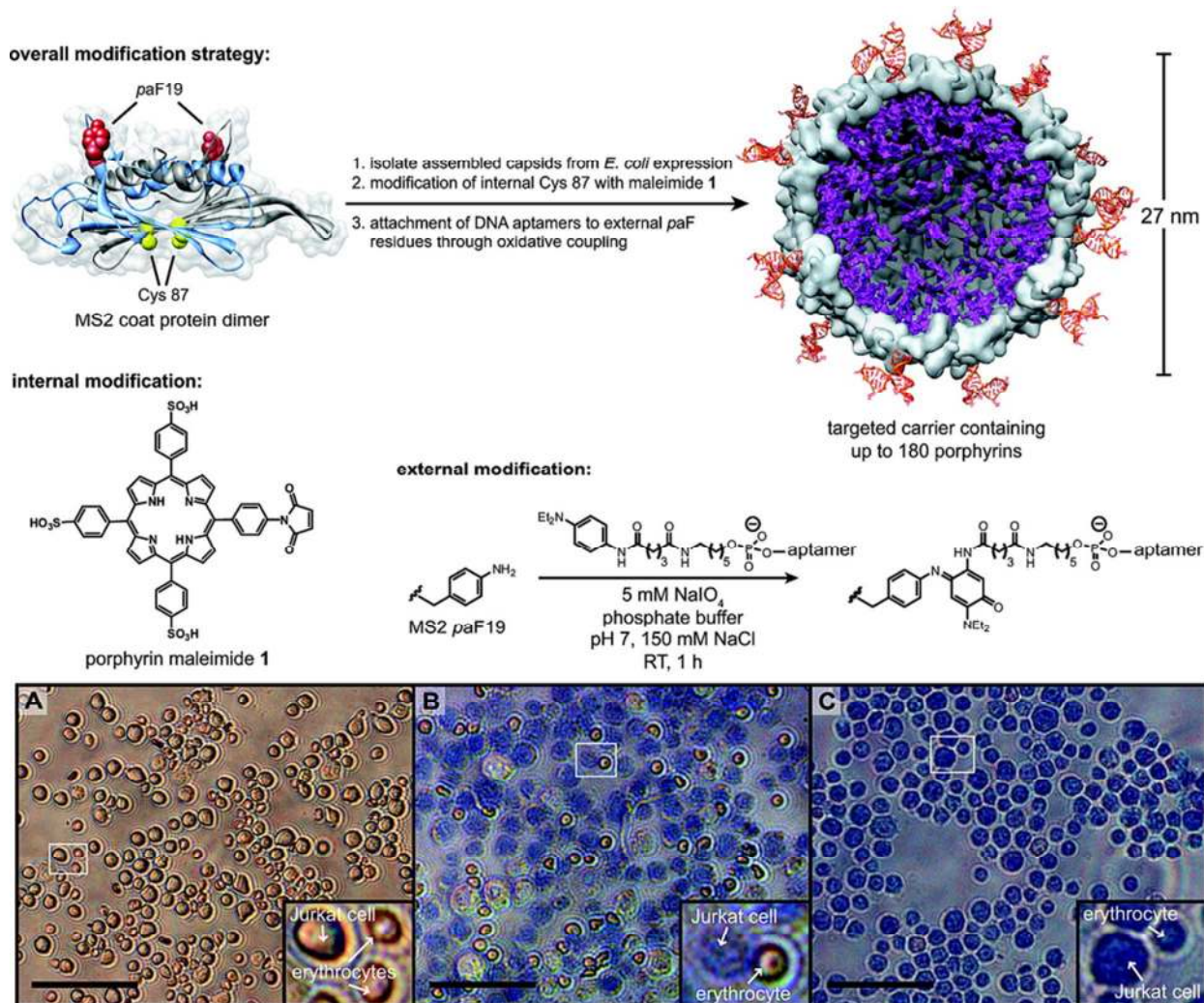


Figure 23. Overall scheme for the construction of a dually functionalized recombinant bacteriophage MS2. Interior cysteines were modified with porphyrin (purple); exterior p-aminophenylalanine were modified with DNA aptamers (~20 per particle). (A) Live control cells showing the difference in size and shape between Jurkat cells and red blood cells (RBCs). (B) Exposure of the cells to the functionalized MS2 capsids to light for 20 minutes results in selective killing of Jurkat cells as indicated by trypan blue staining. RBCs are left unharmed. (C) Postive control showing death of both cell types induced by 30% ethanol. Scale bars are 100 μ m.

Figure adapted from reference ³⁰⁰ with permission from the American Chemical Society, copyright 2010.

Atom transfer radical polymerization (ATRP) has also been used to decorate the exterior of capsids with polymers, which serve as a scaffold to conjugate imaging agents and drug molecules as demonstrated by Finn and coworkers.^{219, 220} In this study, approximately 180 ATRP initiating bromides were introduced on the exterior surface of Q β by using azide-alkyne click chemistry. Oligo(ethyleneglycol)-methacrylate with functional groups were selectively polymerized on the surface of the decorated Q β via ATRP. Direct attachment of polymer molecules instead of conducting a polymerization reaction on the capsids has also been demonstrated. The most notable example is the bioconjugation of polyethylene glycol (PEG) to proteins (so called “PEGylation”) because it is a common technique for functionalizing the proteins in an attempt to reduce renal clearance and immunogenicity, and improve pharmacokinetics. PEG with various chain lengths have successfully been conjugated on the exterior surface of protein cages such as CPMV, MS2 and ferritin.^{75, 301-303} The PEGylated protein cages exhibited a reduced interaction with a cell and an antibody.^{75, 303}

5.3 Noncovalent-Genetic Attachment

Another approach that has gained traction is the fusion of a cargo to secondary structural proteins (sometimes called auxillary proteins) with a natural affinity for a capsid exterior. Secondary proteins that bind to the exterior of some viruses are collectively known as decoration proteins. This approach offers the advantage of using genetic expression of the necessary proteins, and removes the cleanup steps typically needed for after bioconjugation to the capsid. Decoration protein functions range from being actively involved in viral infection to providing structural support to the capsid. In fact, some decoration proteins are part of the conserved structure of the virus.³⁰⁴⁻³⁰⁶ Decoration proteins were initially used as means to display proteins

or protein fragments as an alternative display platform to the M13 phage.^{307, 308} In another example, the bacteriophage T4 structure includes two secondary proteins known as Soc, which binds to local threefold interfaces between capsomers, and Hoc, which binds to the center of each capsomer. Whole proteins were functionally displayed on Soc and Hoc, and the fusions have similar binding affinities as the unmodified proteins.³⁰⁹

While P22 does not have a native decoration protein, the decoration protein “Dec” found in the highly similar bacteriophage L binds to the EX form of P22.³¹⁰ The Dec protein binds tightly as a trimer at the 60 quasi 3-fold axes and with a much lower affinity at the 20 true 3-fold sites. The N-terminus remains close to the capsid shell, while the C-terminus is projected away from the capsid³¹¹, which provides an ideal environment for presenting proteins on the exterior of the P22 capsid. For example, Schwarz *et al.* showed presentation of both a monomeric small peptide (“self peptide” derived from CD47, which inhibits phagocytosis by macrophages³¹²) and a trimeric protein cargo (the soluble region of CD40L, which is a key signal in adaptive immunity) on the exterior of P22 with Dec fusions and each of the cargo remained biologically active (Figure 24).³¹³ Dissociation constants, K_D , of the two Dec mutants to the tight binding sites on P22 were measured as 18 – 30 nM, which is nearly identical to K_D of wild type Dec (9 nM). The half-life of the tight binding site was at least 60 hours, which suggests that Dec bound P22 is a good candidate for in vivo cargo delivery.

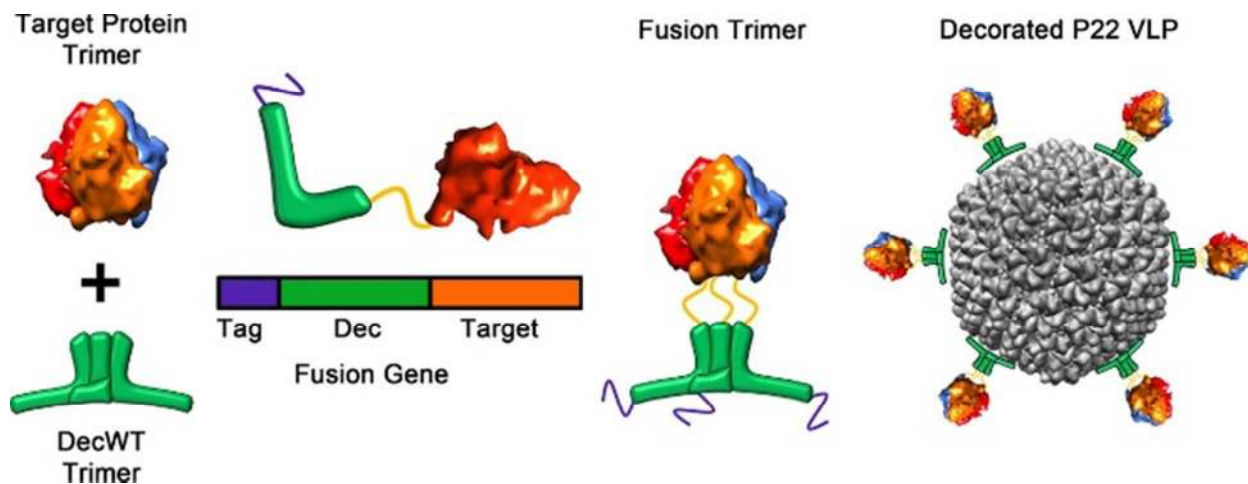


Figure 24. Presentation of a target protein trimer CD40L (PDB 1ALY) on a P22 VLP exterior. The Dec trimer is pictured in green. A polyhistidine tag (purple) is used for purification. The CD40L is genetically fused to the C terminus of Dec and presented as a trimer. Dec binds to P22 at the quasi 3-fold axes to present the CD40L on the exterior. Figure reproduced from reference ³¹³ with permission from the American Chemical Society, copyright 2015.

In other examples, entire proteins as well as small molecules can be non-covalently conjugated to the surface of platform protein cages *via* a well-characterized biotin-streptavidin interaction. For example, papillomavirus VLPs have been functionalized with protein antigens conjugated with streptavidin.^{314, 315}

6. Higher order assembly of protein cages.

Construction of higher order structure from colloidal nanoparticles in a controlled manner has drawn considerable interest in the field of materials chemistry because the assembled materials could display collective properties that are different from individual nanoparticles.^{29, 30} Exploiting the self-assembly of nanoparticle building blocks into higher order structures is a powerful, versatile and low-cost approach to designing such materials. Biological building blocks including VLPs and related cage-like proteins have several unique advantages over

synthetic colloidal particles as will be discussed below.³¹⁶ In this part, we will review recent progress towards the design of higher order structures built from protein cage building blocks via self-assembly processes.

6.1. Protein cages as Building Blocks for Higher Order Assembly

Self-assembly of biomolecular components into higher order structure with hierarchically organized arrangement over multiple length scales is one of the hallmarks of biological systems.³¹⁷ The highly organized structural arrangements of building block components could endow collective and orchestrated functionalities to the assembled structures beyond those of individual building blocks. For example, the excellent mechanical property of bone is a consequence of the highly regulated hierarchical structure of the hydroxyapatite and collagen assembly across nano to macro scales.³¹⁸ The vertebrate striated muscle is another elegant example of organized assembly of multiple biomolecular building blocks.³¹⁹ The functional properties of muscle emerge as a consequence of the hierarchical structure composed of multiple proteins including actin, myosin and titin.

Viruses are also excellent examples of nature's self-assembled constructs which are constructed from a limited number of biomolecular building blocks. They exhibit complex functionalities through their life-cycle including host-entry, replication and shedding. From a materials viewpoint, cage-like proteins are excellent building block components to construct higher order structures across multiple length scales for a number of reasons. First, proteins are genetically encoded materials, hence they are produced with near perfect uniformity. The homogeneity of a building block is critical to construct higher order structures in particular if long-range order is desired. Secondly, as we discussed earlier, a wide range of functional protein

cage building blocks can be produced by encapsulation various functional cargo molecules. Third, the established chemical and genetic modification strategies can also be used to modify the exterior surface of the protein cages, which can lead to modulation of inter-particle interactions and control the assembly of individual building blocks into higher order structures. Fourth, numerous types of protein cages, which share the same spherical morphology but with a wide range of sizes and chemical and physical properties, are available, either naturally occurring or by synthetic design. Therefore, similar construction strategies could be applied to different cages, resulting in a new class of protein-based superlattice materials. Lastly, VLPs and related cage-like proteins are in general chemically and mechanically robust, perhaps as an evolutionary consequence of protecting their genomic nucleic acids cargos from the external environment. Thus the protein shells can be used as protective coatings for encapsulated cargos including fragile enzymes.

6.2. One- and Two-dimensional Assembly of Protein Cages

One-dimensional (1D) assembly of protein cages into tube like structures has been demonstrated through a head-to-head assembly of individual cages. Aida *et al.* have demonstrated fabrication of a micrometer long nanotube self-assembled from a chemically modified GroEL chaperonin.³²⁰⁻³²² The apical domain of GroEL was modified with merocyanine (MC). The MC decorated GroEL “polymerized” into a tube structure in the presence of Mg^{2+} through a head-to-head assembly of GroEL due to coordination between MC and Mg^{2+} . A 1D tube of GroEL encapsulating superparamagnetic iron oxide nanoparticles within the cavity was shown to form a bundle of tubes under the influence of an external magnetic field.³²³

Ordered two-dimensional (2D) arrays of proteins have been studied for decades as they are of interest in electron crystallography and nanotechnology. Such arrays can be developed at flat surfaces including air-liquid and liquid-liquid interfaces as substrate to support the growth of the arrays.³²⁴⁻³³¹ For example, some of the earliest examples of self-assembled 2D arrays used the protein cage ferritin, and were prepared on lipid monolayers at the air-water interface,³²⁴⁻³²⁶ where interaction of protein cages with polar head groups of lipids was key to manipulate the array formation. This approach is versatile and close-packed 2D arrays have also been developed from larger protein cages including CPMV,³²⁷ TYMV,³²⁸ and P22.³²⁹ It is important to note that arrangement of protein cages into ordered arrays is primarily dependent on the properties of the exterior surfaces of protein cages. This allows for the development of the same array structures regardless of the cargo molecules encapsulated within protein cages. For example, ferritin proteins encapsulating iron oxide or indium oxide nanoparticles within the interior cavity have been successfully assembled into near perfect hexagonally packed 2D arrays under the same conditions that lead to the empty apoferritin array assembly.^{332, 333}

Assembly of protein cages at the air-liquid and liquid-liquid interfaces is typically a dynamic process, thus the arrays formed at the interfaces are usually transferred to solid substrates, or an inter-particle chemical cross-linkage is introduced to obtain stable 2D monolayers.^{330, 331} Alternatively, direct deposition of protein cages on the surfaces of solid substrates has also been investigated. The deposition can be achieved either through the formation of covalent bonds between the protein cage and the substrate or through weak physical interactions between them.^{334, 335} For example, an engineered CCMV Janus particle, which possessed a reactive thiol group only on one side of its exterior surface, was shown to bind selectively to a gold substrate via a covalent linkage and formed a 2D array of the CCMV.²⁹⁵

Controlled deposition of CCMV was also achieved by manipulating electrostatic interactions between the overall negatively charged CCMV and the positively charged substrate.³³⁶

The bottom-up approaches for constructing 2D arrays of protein cages can be readily combined with top-down techniques such as lithography, which lead to controlled placement of protein cages into a designed pattern. Specific interaction between protein cages and substrates are required to accomplish site-specific immobilization of the cages. For example, De Yoreo et al. demonstrated alignment of virus capsids in a line (one capsid wide) through a combination of exterior surface engineering of CPMV and a dip-pen lithography patterning of a substrate^{334, 335}. They genetically engineered a CPMV mutant decorated with six contiguous histidine residues (His-CPMV). The His-CPMV was successfully placed as lines with 30 nm width on nitrilotriacetic acid (NTA)-terminated patterned on a gold substrate (Figure 25) *via* a bond formation with Ni-NTA. The line width corresponds to the 1D alignment of single capsids.

Modification of the exterior surface of a protein cage with an additional peptide sequence is another approach that has been used to immobilize a protein cage in site-specific manner. Peptide sequences which exhibit high affinity to specific metal, inorganic and polymer materials have been identified by several techniques such as phage display library²⁷⁷ and *de novo* design.³³⁷ These peptides can be readily introduced on the exterior surfaces of protein cages *via* genetic or chemical modification and are utilized to immobilize protein cages on a targeted substrate^{338, 339}. For example, a titanium binding peptide (RKLPDA) was introduced onto the exterior surface of ferritin allowing the ferritin to be localized to titanium patterns on a silica substrate.³⁴⁰ This result illustrates that engineering of protein cages in combination with top-down patterning techniques will lead to further development of fabricating protein cage based devices.

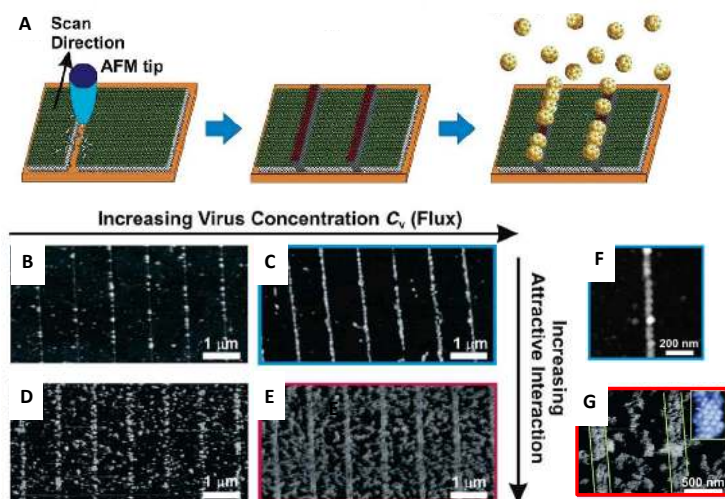


Figure 25. (A) Schematic of nanografting and virus deposition process. (A) A Ni-NTA terminated surface is created at specific locations on a gold substrate (red lines) to allow for deposition of His-tagged CPMV (yellow spheres) via Ni(II) chelation. (B-E) AFM images showing the coverage of the CPMV on the substrate as well as the increasing order observed by increases in virus concentration (flux) and increasing inter-viral attractive interaction (increasing PEG concentration). (B) At a condition of low flux and weak inter-viral attractive interaction (0 wt% PEG), His-CPMV attaches almost exclusively to the Ni-NTA line, but the coverage is low. (C) At higher flux with weak interaction, His-CPMV still exclusively attaches to the line, but with increased coverage. (D, E) As the inter-viral interaction is increased (by addition of 1 wt% PEG), virus assembly spreads outward from the lines, and clusters of viruses lie between the lines. (F) Higher resolution AFM image of condition (C), showing His-CPMV alignment at a single capsid width. (G) Higher resolution image of condition (E), showing ordered packing of virus capsid on the Ni-NTA line. Figure reproduced from Reference³³⁵ with permission from the American Chemical Society, copyright 2006.

6.3. Three-Dimensional Assembly of Protein Cages

Construction of three-dimensional (3D) superlattice from functional nanoparticle building blocks in a controlled manner is of significant interest because they have the potential for making a new class of materials that exploit collective behavior and properties arising from those of the individual particles³⁴¹⁻³⁴⁴. Although synthetic particles such as metallic, inorganic and polymer nanoparticles have been extensively studied as building blocks, protein cages are appealing multifunctional nanoparticles as has been described in the previous sections. In the assembly of

nanoparticles, the formation of amorphous aggregates is common^{28, 345-347} but the successful formation of long-range ordered assemblies of nanoparticles, i.e. superlattices, has been demonstrated by optimizing interactions between building block components^{16, 343, 348-350}. A range of methods have been explored to construct 3D assemblies of protein cages and in many cases linker molecules have been utilized to direct and control assembly. Electrostatic interaction is one of the most well explored interactions to mediate assembly of protein cages into bulk materials.^{14, 15, 17, 351-353} For example, Kostianen *et al.* have demonstrated the formation of protein cage based superlattice materials through electrostatic interactions between the protein cages and oppositely charged mediator particles.^{14, 15} Chemically modified gold nanoparticles bearing positive surface charge were used to mediate 3D assembly of CCMV, which has a negatively charged exterior surface. Inter-particle interactions were manipulated by changing ionic strength and pH of the solution (Figure 26A). When the interparticle interactions were too strong, an amorphous aggregate was formed, whereas an ordered array with a face-centered-cubic (FCC) arrangement of CCMV was obtained under an optimal range of interaction (Figure 26B).¹⁴ They have further explored the approach and demonstrated that superlattices with not only FCC structures, but also hexagonal-close-packed (HCP) structures, and body-centered-cubic (BCC) structures have been realized by changing the type and size of the mediator (Figure 26C).^{15, 353}

Catalytically active VLP superlattices were recently developed using a similar assembly approach combined with genetic modification of VLP exterior surface to modulate interparticle interactions.¹⁷ In this study, two enzymes, which catalyze a sequential two-step reaction to form isobutanol, were encapsulated individually into the P22 VLPs. The superlattices constructed from the two populations of P22 VLPs retained the coupled two-step catalytic activity, despite

the fact that this was now a bulk material with the enzymes immobilized within it. This result indicates that the assembled lattice is porous and the diffusion of small molecules substrates and products through the lattice is not the rate-limiting step in the overall reaction under the conditions it was investigated. As it is a bulk material, the catalytically active superlattice is readily condensed, recycled and reused.

Higher order assembly mediated by electrostatic interaction has been applied for developing protein cage templated binary arrays of encapsulated inorganic nanoparticles. Beck et al., used two types of ferritin cages, one with a positively charged exterior surface and the other having a negatively charged exterior surface.³⁵⁴ The oppositely charged ferritin mutants were co-crystallized with a tetragonal lattice structure. By synthesizing cerium oxide and cobalt oxide inside of negatively charged and positively charged ferritins independently prior to the co-crystallization process, the protein cages served as a template to construct a binary crystalline array of cerium oxide and cobalt oxide nanoparticles.

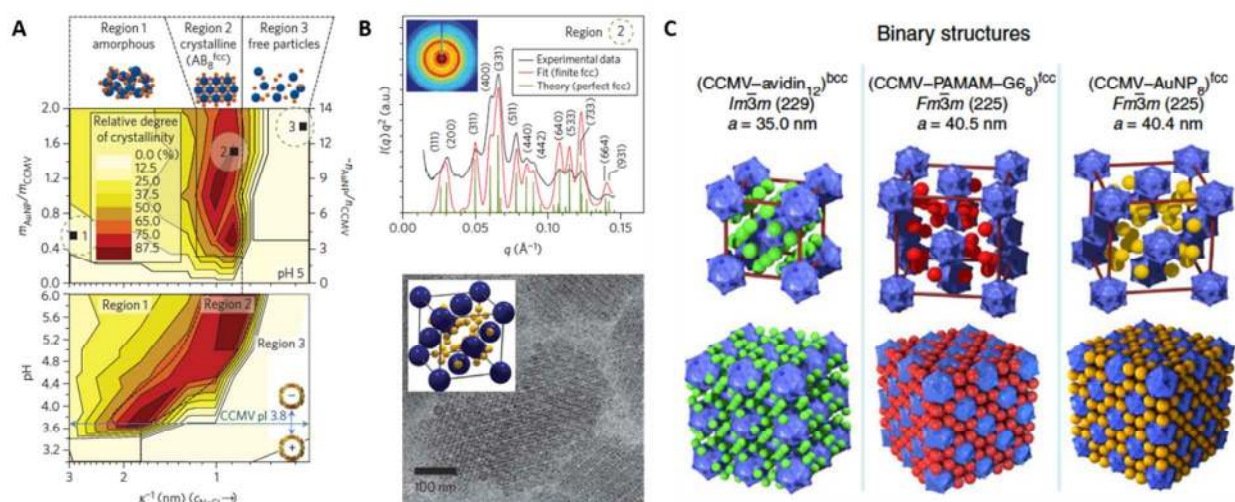


Figure 26. (A) Superlattice formation of positively charged gold nanoparticles (AuNP) and negatively charged CCMV at different Debye screening lengths, κ^{-1} , (adjusted by NaCl concentration), AuNP/CCMV mass ratios (top) and pH (bottom) determined by small angle x-ray scattering (SAXS). At a given pH, the superlattice is formed only in a narrow NaCl concentration window (Region 2), which is shifted to smaller κ^{-1} (i.e. higher NaCl concentration) when the pH is increased. Superlattice formation is possible only when the pH is higher than pI of CCMV (pH = 3.8). High NaCl concentration screens the electrostatic interaction between CCMV and AuNP, resulting in non-assembly of the particles (Region 3). (B) Two-dimensional (inset) and integrated one-dimensional SAXS data obtained from a CCMV superlattice (sample 2 in (A)) (top). The experimental scattering pattern (black trace) matches well with the respective theoretical scattering pattern of an AB_8^{FCC} -type structure (red and green traces). Crystalline superlattices are also observed with cryo-TEM (bottom). The particles arrange into an AB_8^{FCC} type lattice, where CCMV (blue) adopts an FCC structure and the voids between the CCMVs are filled with eight AuNPs (yellow) as shown in inset. (C) Crystal structure models and lattice parameters of CCMV-avidin, CCMV-G6 PAMAM dendrimer and CCMV-Au nanoparticle superlattice. Superlattices with different structures are yielded depending on linker molecules, which mediate assembly of CCMV. Parts A and B reproduced from reference ¹⁴ with permission from Nature Publishing Group, copyright 2013. Part C adapted from reference ¹⁵ with permission from Nature Publishing Group, copyright 2014.

Three-dimensional assembly of protein cages do occur naturally *in vivo*. Iridovirus such as Wiseana iridescent virus (WIV) is known to display a “rainbow-like” opalescent hue in heavily infected host insects.^{355, 356} This is because paracrystalline arrays of viral particles are formed within the cell cytoplasm of the infected hosts and a weak optical interference effect due to Bragg scattering from the assembled arrays is observed.³⁵⁷ Vaia *et al.*, have successfully reproduced the assembly of WIV particles *in vitro* by various sedimentation or thin-film assembly techniques and demonstrated that optical iridescence of the *in vitro* assembled materials is much stronger than that commonly observed *in vivo*.³⁵⁷ The inter-particle interaction is characterized by weak long-range electrostatic repulsion and short-range attractive interaction due to depletion and van der Waals forces.

Complementary pairs of oligonucleotides have been another widely used linker molecule to mediate assembly of nanoparticles into higher order structures.^{342-345, 348, 349} Various physical parameters of the oligonucleotide linkers including strength and specificity of the

complementary strand interactions, and the length and shapes of the linkers can be finely tuned by rationally designed sequence of the linkers. Thus a wide range of ordered arrays of not only close packed structures, but also loosely packed structures such as a diamond lattice have been realized via this approach.^{16, 343, 350} Finn *et al.*, applied oligonucleotide linkers for directing assembly of protein cages. They demonstrated temperature dependent assembly and disassembly of two populations of virus capsids, each conjugated with one of the complementary pair of oligonucleotide linkers. They have also reported construction of a binary superlattice with NaTl (B32)-type crystalline structures assembled from oligonucleotide decorated virus capsid and gold nanoparticle.¹⁶ In principle, complementary pairs of peptide motifs such as coiled-coil peptides have the potential to mediate assembly of protein cages in a sequences specific manner similar to the oligonucleotide-based linker.³⁵⁸ Although they have not yet been explored as assembly mediators to nearly the same extent that oligonucleotide-based linkers have, polypeptide linkers could be significantly more diverse than oligonucleotide linkers and have great potential to direct and control assembly of protein cage building blocks into complex structures.³⁵⁹⁻³⁶²

Two-dimensional arrays of protein cages discussed earlier are readily expanded to construction of 3D architectures via layer-by-layer deposition.^{28, 338, 363-371} For instance, Sano *et al.* fabricated 3D layer of a ferritin mutant (T1-LF) which possesses a titanium binding peptide on its exterior surface.³³⁸ The T1-LF exhibit two unique functions, specific binding to Ti and capability to induce titania mineralization around the cage. In the fabrication process, the first layer of the T1-LF was selectively adsorbed on a Ti pattern formed on a Pt substrate due to specific affinity of the Ti binding peptide. Next, the substrate was soaked in a solution containing titanium bis(ammonium lacto) dihydroxide (TiBALDH), a precursor of titania and a thin titania layer was formed on the top of T1-LF layer. The titania layer worked as a binding target of the

second T1-LF layer. Hence stacking layers of ferritin and titania were successfully fabricated by utilizing binding and mineralization capability of the T1-LF.

Anisotropic shaped filamentous and tubular cages such as TMV and bacteriophage M13 have also been utilized to fabricate 2D film and 3D assemblies of cages, because they could exhibit liquid-crystalline ordering under appropriate conditions.³⁷²⁻³⁷⁶ For example, Lee *et al.* demonstrated formation of a film of M13 filaments with long-range order on a substrate through simply dipping and pulling the substrate vertically from a suspension of the M13 phage at precisely controlled speeds.³⁷⁶ Importantly, the helical structure of individual M13 phage was mirrored in the hierarchical assembly and a helical twist to the organization of the assembly was observed when fabricated under appropriate conditions. TMV, whose surface is negatively charged, has been assembled into a bundle of superlattice wires mediated by an oppositely charged metal nanoparticle. Similar to the example of M13 film, the assembled bundle shows right-handed helical twisting because of the right-handed structure of the individual TMV building units.³⁷³

One of the unique features of protein cages over other nanoparticles is their ordered and highly symmetric arrangement of subunits (as discussed in previous sections), thus they present well-defined symmetry specific sites. By exploiting symmetry specific sites on the protein cages as binding sites for linker molecules, protein cages can serve as coordination nodes for the construction of three dimensional network structures, which conceptually resemble the metal ions in metal-organic frameworks (MOFs). For example, Tezcan *et al.* have exploited the 3-fold symmetry site of the ferritin cage to direct its assembly into a crystalline material.³⁷⁷ They substituted threonine 122, located at the 3-fold symmetry site of ferritin, with histidine. The three histidine residues at the 3-fold symmetry site of the cage (i.e. one from each subunit) coordinate

with Zn^{2+} with tetrahedral coordination geometry, and an additional ditopic ligand molecule can occupy the last coordination site of Zn^{2+} and serve as a linker to mediate assembly between ferritin cages. As a consequence, the engineered ferritin formed a BCC type crystal in the presence of Zn^{2+} and a ditopic linker, instead of an FCC type crystal which wild type ferritin typically forms (Figure 27). Each ferritin cage possesses eight 3-fold symmetry axes, hence the engineered ferritin crystallized into a BCC structure that has eight nearest neighbor positions.

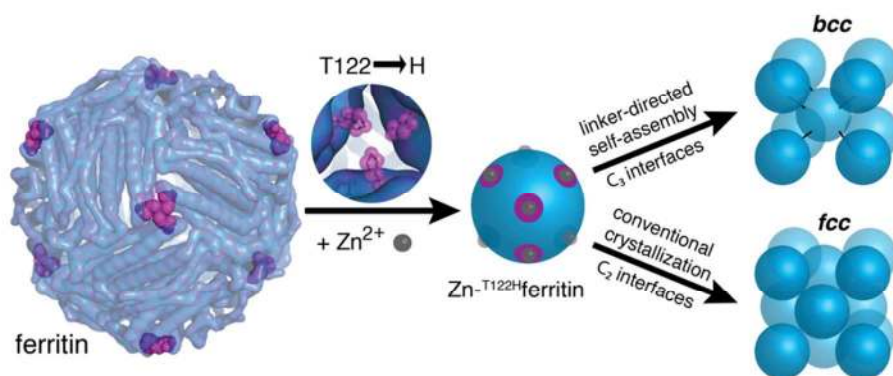


Figure 27. Schematic illustration for metal/linker-directed self-assembly of ferritin into three-dimensional crystals. Surface exposed Zn^{2+} binding sites are engineered at 3-fold symmetry (C_3) sites of ferritin by mutating threonine 122 to histidine (T122H). In the presence of ditopic organic linkers, the T122H ferritin mutant forms a BCC lattice through coordination of ferritins at the C_3 sites. On the other hand, in the absence of such ditopic linkers, the mutant forms a FCC lattice, which ferritin typically crystallized into, through coordination at 2-fold symmetry (C_2) sites. Figure reproduced from reference ³⁷⁷ with permission from the American Chemical Society, copyright 2015.

An engineered protein can also serve as a linker molecule to direct self-assembly of protein cages into higher order structures. For example, a decoration protein (Dec), which binds at symmetry specific sites on the expanded (mature) form of P22 was used to construct protein-based linkers to mediate controlled assembly of the P22 VLP.²⁸ A linear ditopic linker and a tetrahedral tetratopic linker were genetically engineered through either a point mutation of Dec or fusion of Dec to the exposed termini on the smaller Dps protein cage respectively. Bulk

assembly and layer-by-layer deposition of P22 VLP in solution was successfully achieved using both linker molecules (Figure 28). Although the obtained bulk assemblies of P22 VLPs in this study did not show a long-range ordered structure, optimal manipulation of interaction between protein cages and linker proteins could lead to construction of protein only superlattice materials because Dec binds to symmetry specific sites of P22.^{310,311}

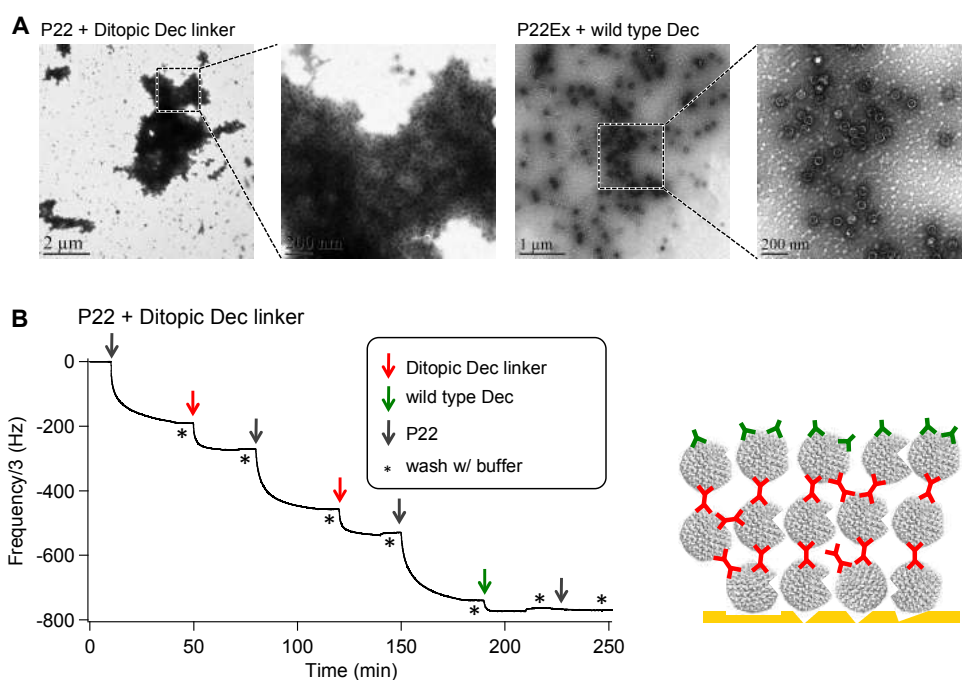


Figure 28. (A) TEM images of P22 mixed with a ditopic linker derived from Dec (left) or wild type (wt) Dec (right). In the case of P22 with ditopic Dec linker, large clusters of micrometer to tens of micrometers in size, which composed of P22 VLP, were observed. In the case of P22 with wtDec, each VLP were well dispersed and no cluster of VLP was observed. (B) Quartz crystal microbalance (QCM) resonance frequency changes upon alternate injection of P22, ditopic Dec linker and wtDec on a gold coated QCM sensor. The decrease in resonance frequency corresponds to the increase in mass on the sensor due to the deposition of the proteins. Layer-by-layer deposition was demonstrated through the alternative addition of P22 and ditopic Dec linker. Addition of wtDec caps the deposited layer and passivates from further attachment of P22 because it has only one binding site for P22. The cartoon on the right depicts observed layer-by-layer deposition. Figure adapted from reference²⁸ with permission from Wiley, copyright 2015 .

7. Conclusion

It is evident that the natural world has a seemingly infinite number of protein cages, specifically viral protein cages waiting to be discovered and exploited. Even the small number of cages that have been thoroughly characterized to date have provided us with a vast library that can be used for further design and manipulation for materials synthesis. Fundamental studies into the mechanisms of viral capsid assembly and protein-protein interactions are being harnessed for development of novel nanomaterials and armed with the knowledge of Nature's synthetic methods, novel designer cages can be constructed to have specific features (such as stability, size, rigidity, and porosity). With increasing synthetic precision a wide range of desired moieties can be presented on either the interior or exterior of these cages taking advantage of the high symmetry and associated multivalent presentation of these platforms. Thus these closed shell protein architectures provide interior and exterior surfaces for presentation and cargo encapsulation as well as intersubunit interfaces that allow for fine tuning of the structural properties and inspiration for *de novo* design. The diversity and synthetic versatility of these platforms has made them attractive materials for an increasing number of applications in materials science, chemistry, biology, and medicine. These applications have grown extensively in the last few decades, and will continue to grow as the library of cages, our foundational understanding of individual protein cages, and the generality of new approaches develops. New advances in directing the interactions between functionally active individual particles also holds the promise of future applications where the collective properties of an ensemble of particles can be harnessed. Viruses no longer need to be understood only as hostile pathogens but rather as a

source of active nanomaterial inspiration and the raw materials for a new and exciting materials future.

Author Information

*Email:

Corresponding Author

Email: trevdoug@indiana.edu

Notes:

The authors declare no competing financial interests

Biographies

William M Aumiller Jr received a B.S degree in chemistry from Juniata College in 2008. In 2015 he received his Ph.D. in chemistry from The Pennsylvania State University under the direction of Prof. Christine D. Keating. He developed model systems based on aqueous phase separation of polymers for the purpose of mimicking intracellular compartmentalization and phase separation. He is currently a postdoctoral research associate in Trevor Douglas's group at Indiana University where he is studying higher order assembly of virus particles.

Masaki Uchida is an Associate Scientist in the Douglas Lab at Indiana University. His current research interests are in bioinspired materials synthesis and assembly of protein cages into higher order structures. He received his PhD degree in Materials Chemistry in 2002 from Kyoto

University, Japan under the supervision of Prof. Tadashi Kokubo. He did his post-doctoral training at the National Institute of Advanced Science and Technology, Japan and at Montana State University.

Trevor Douglas is the Earl Blough Professor of Chemistry at Indiana University. His research has focused on the development of biomimetic approaches to construction of new functional materials. He received his PhD in 1991 from Cornell University working with Professor Klaus Theopold and did his post-doctoral training at Bath University with Professor Stephen Mann, in the area of biomineralization and biomimetic materials.

Conflicts of Interest

There are no conflicts of interest to declare.

Acknowledgments

The authors would like to acknowledge funding support from the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering (DE-SC0016155), the National Institutes of Health (R01 AI104905), and the National Science Foundation (NSF-BMAT DMR- 1507282).

References

1. J. E. Johnson and W. Chiu, *Curr. Opin. Struct. Biol.*, 2000, **10**, 229-235.
2. T. Douglas and M. Young, *Science*, 2006, **312**, 873-875.
3. P. M. Harrison and P. Arosio, *Biochim Biophys Acta-Bioenerg*, 1996, **1275**, 161-203.
4. M. Uchida, S. Kang, C. Reichhardt, K. Harlen and T. Douglas, *Biochim. Biophys. Acta*, 2010, **1800**, 834-845.
5. A. L. Horwich, W. A. Fenton, E. Chapman and G. W. Farr, *Annu. Rev. Cell Dev. Biol.*, 2007, **23**, 115-145.
6. K. K. Kim, R. Kim and S. H. Kim, *Nature*, 1998, **394**, 595-599.

7. H. F. Lodish, *Molecular cell biology*, W.H. Freeman-Macmillan Learning, New York, Eighth edition. edn., 2016.
8. E. C. Theil and D. J. Goss, *Chem. Rev.*, 2009, **109**, 4568-4579.
9. P. Arosio and S. Levi, *Ferritin, iron homeostasis, and oxidative damage*, Free Radic Biol Med, 2002.
10. C. A. Suttle, *Nature*, 2005, **437**, 356-361.
11. M. G. Mateu, *Arch. Biochem. Biophys.*, 2013, **531**, 65-79.
12. Y. T. Lai, D. Cascio and T. O. Yeates, *Science*, 2012, **336**, 1129-1129.
13. Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Laganowsky, F. J. Asturias, J. A. Tainer, C. V. Robinson and T. O. Yeates, *Nat. Chem.*, 2014, **6**, 1065-1071.
14. M. A. Kostianen, P. Hiekkataipale, A. Laiho, V. Lemieux, J. Seitsonen, J. Ruokolainen and P. Ceci, *Nat. Nanotechnol.*, 2013, **8**, 52-56.
15. V. Liljestrom, J. Mikkila and M. A. Kostianen, *Nat Commun*, 2014, **5**, 4445.
16. P. Cigler, A. K. R. Lytton-Jean, D. G. Anderson, M. G. Finn and S. Y. Park, *Nat. Mater.*, 2010, **9**, 918-922.
17. M. Uchida, K. McCoy, M. Fukuto, L. Yang, H. Yoshimura, H. M. Miettinen, B. LaFrance, D. P. Patterson, B. Schwarz, J. A. Karty, P. E. Prevelige Jr, B. Lee and T. Douglas, *ACS Nano*, In press.
18. Z. M. Carrico, D. W. Romanini, R. A. Mehl and M. B. Francis, *Chem. Comm.*, 2008, **369**, 1205-1203.
19. E. Strable, D. E. Prasuhn, A. K. Udit, S. Brown, A. J. Link, J. T. Ngo, G. Lander, J. Quispe, C. S. Potter, B. Carragher, D. A. Tirrell and M. G. Finn, *Bioconj. Chem.*, 2008, **19**, 866-875.
20. A. O'Neil, P. E. Prevelige, G. Basu and T. Douglas, *Biomacromolecules*, 2012, **13**, 3902-3907.
21. D. P. Patterson, B. Schwarz, R. S. Waters, T. Gedeon and T. Douglas, *ACS Chemical Biology*, 2014, **9**, 359-365.
22. S. Qazi, H. M. Miettinen, R. A. Wilkinson, K. McCoy, T. Douglas and B. Wiedenheft, *Mol. Pharm.*, 2016, **13**, 1191-1196.
23. P. C. Jordan, D. P. Patterson, K. N. Saboda, E. J. Edwards, H. M. Miettinen, G. Basu, M. C. Thielges and T. Douglas, *Nat. Chem.*, 2016, **8**, 179-185.
24. J. D. Fiedler, S. D. Brown, J. L. Lau and M. G. Finn, *Angew. Chem. Int. Ed.*, 2010, **49**, 9648-9651.
25. B. Wörsdörfer, Z. Pianowski and D. Hilvert, *J. Am. Chem. Soc.*, 2012, **134**, 909-911.
26. M. Comellas-Aragones, H. Engelkamp, V. I. Claessen, N. A. J. M. Sommerdijk, A. E. Rowan, P. C. M. Christianen, J. C. Maan, B. J. M. Verduin, J. J. L. M. Cornelissen and R. J. M. Nolte, *Nat Nano*, 2007, **2**, 635-639.
27. J. E. Glasgow, S. L. Capehart, M. B. Francis and D. Tullman-Ercek, *ACS Nano*, 2012, **6**, 8658-8664.
28. M. Uchida, B. LaFrance, C. C. Broomell, P. E. Prevelige, Jr. and T. Douglas, *Small*, 2015, **11**, 1562-1570.
29. A. Böker, J. He, T. Emrick and T. P. Russell, *Soft Matter*, 2007, **3**, 1231.
30. Z. Nie, A. Petukhova and E. Kumacheva, *Nat Nanotechnol*, 2010, **5**, 15-25.
31. A. M. Wen and N. F. Steinmetz, *Chem. Soc. Rev.*, 2016, **45**, 4074-4126.
32. J. N. Culver, A. D. Brown, F. H. Zang, M. Gnerlich, K. Gerasopoulos and R. Ghodssi, *Virology*, 2015, **479**, 200-212.

33. M. Rother, M. G. Nussbaumer, K. Renggli and N. Bruns, *Chem. Soc. Rev.*, 2016, **45**, 6213-6249.
34. L. Schoonen and J. C. M. van Hest, *Nanoscale*, 2014, **6**, 7124-7141.
35. D. Prangishvili, *Annu. Rev. Microbiol.*, 2013, **67**, 565-585.
36. T. Noda, *Frontiers in Microbiology*, 2012, **3**.
37. G. P. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Y. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken and P. J. Zhang, *Nature*, 2013, **497**, 643-646.
38. P. Ge, J. Tsao, S. Schein, T. J. Green, M. Luo and Z. H. Zhou, *Science*, 2010, **327**, 689-693.
39. D. L. D. Caspar and A. Klug, *Cold Spring Harbor Symp. Quant. Biol.*, 1962, **27**, 1-24.
40. B. V. Prasad and M. F. Schmid, *Adv. Exp. Med. Biol.*, 2012, **726**, 17-47.
41. M. Carrillo-Tripp, C. M. Shepherd, I. A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. Brooks, 3rd and V. S. Reddy, *Nucleic Acids Res.*, 2009, **37**, D436-442.
42. S. Daya and K. I. Berns, *Clin. Microbiol. Rev.*, 2008, **21**, 583-593.
43. G. C. Lander, L. Tang, S. R. Casjens, E. B. Gilcrease, P. Prevelige, A. Poliakov, C. S. Potter, B. Carragher and J. E. Johnson, *Science*, 2006, **312**, 1791-1795.
44. S. R. Chant, *Ann. Appl. Biol.*, 1959, **47**, 565-+.
45. F. Sainsbury, M. C. Canizares and G. P. Lomonossoff, *Annual Review of Phytopathology*, Vol 48, 2010, **48**, 437-455.
46. T. W. Lin, Z. G. Chen, R. Usha, C. V. Stauffacher, J. B. Dai, T. Schmidt and J. E. Johnson, *Virology*, 1999, **265**, 20-34.
47. F. Sainsbury, P. Saxena, A. A. A. Aljabali, K. Saunders, D. J. Evans and G. P. Lomonossoff, *Virus Hybrids as Nanomaterials: Methods and Protocols*, 2014, **1108**, 139-153.
48. K. Saunders, F. Sainsbury and G. P. Lomonossoff, *Virology*, 2009, **393**, 329-337.
49. J. A. Speir, S. Munshi, G. Wang, T. S. Baker and J. E. Johnson, *Structure*, 1995, **3**, 63-78.
50. J. B. Bancroft and E. Hiebert, *Virology*, 1967, **32**, 354-&.
51. P. Annamalai and A. L. N. Rao, *Virology*, 2005, **332**, 650-658.
52. A. L. N. Rao, *Annu. Rev. Phytopathol.*, 2006, **44**, 61-87.
53. R. F. Garmann, M. Comas-Garcia, C. M. Knobler and W. M. Gelbart, *Acc. Chem. Res.*, 2016, **49**, 48-55.
54. J. A. Speir, B. Bothner, C. Qu, D. A. Willits, M. J. Young and J. E. Johnson, *J. Virol.*, 2006, **80**, 3582-3591.
55. J. B. Bancroft, G. J. Hills and R. Markham, *Virology*, 1967, **31**, 354-379.
56. L. Lavelle, M. Gingery, M. Phillips, W. M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres and J. Ruiz-Garcia, *J. Phys. Chem. B*, 2009, **113**, 3813-3819.
57. L. Lavelle, J. P. Michel and M. Gingery, *J. Virol. Methods*, 2007, **146**, 311-316.
58. J. P. Michel, M. Gingery and L. Lavelle, *J. Virol. Methods*, 2004, **122**, 195-198.
59. M. Uchida, M. T. Klem, M. Allen, P. Suci, M. Flenniken, E. Gillitzer, Z. Varpness, L. O. Liepold, M. Young and T. Douglas, *Adv. Mater.*, 2007, **19**, 1025-1042.
60. R. W. Lucas, S. B. Larson and A. McPherson, *J. Mol. Biol.*, 2002, **317**, 95-108.
61. J. Diez, M. Ishikawa, M. Kaido and P. Ahlquist, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 3913-3918.

62. P. Pfeiffer and L. Hirth, *Virology*, 1974, **61**, 160-167.
63. J. King, D. Botstein, S. Casjens, W. Earnshaw, S. Harrison and E. Lenk, *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.*, 1976, **276**, 37-&.
64. P. E. Prevelige and J. King, *Progress in Medical Virology*, 1993, **40**, 206-221.
65. C. M. Teschke and K. N. Parent, *Virology*, 2010, **401**, 119-130.
66. G. C. Lander, R. Khayat, R. Li, P. E. Prevelige, C. S. Potter, B. Carragher and J. E. Johnson, *Structure*, 2009, **17**, 789-799.
67. P. E. Prevelige, D. Thomas and J. King, *J. Mol. Biol.*, 1988, **202**, 743-757.
68. M. T. Fuller and J. King, *J. Mol. Biol.*, 1982, **156**, 633-665.
69. P. A. ThumanCommike, B. Greene, J. Jakana, B. V. V. Prasad, J. King, P. E. Prevelige and W. Chiu, *J. Mol. Biol.*, 1996, **260**, 85-98.
70. J. Chang, P. Weigele, J. King, W. Chiu and W. Jiang, *Structure*, 2006, **14**, 1073-1082.
71. K. N. Parent, R. Khayat, L. H. Tu, M. M. Suhanovsky, J. R. Cortines, C. M. Teschke, J. E. Johnson and T. S. Baker, *Structure*, 2010, **18**, 390-401.
72. D. H. Chen, M. L. Baker, C. F. Hryc, F. DiMaio, J. Jakana, W. Wu, M. Dougherty, C. Haase-Pettingell, M. F. Schmid, W. Jiang, D. Baker, J. A. King and W. Chiu, *Proc Natl Acad Sci U S A*, 2011, **108**, 1355-1360.
73. C. F. Hryc, D. H. Chen, P. V. Afonine, J. Jakana, Z. Wang, C. Haase-Pettingell, W. Jiang, P. D. Adams, J. A. King, M. F. Schmid and W. Chiu, *Proc Natl Acad Sci U S A*, 2017, DOI: 10.1073/pnas.1621152114.
74. J. M. Hooker, E. W. Kovacs and M. B. Francis, *J. Am. Chem. Soc.*, 2004, **126**, 3718-3719.
75. E. W. Kovacs, J. M. Hooker, D. W. Romanini, P. G. Holder, K. E. Berry and M. B. Francis, *Bioconj. Chem.*, 2007, **18**, 1140-1147.
76. R. Golmohammadi, K. Valegard, K. Fridborg and L. Liljas, *J. Mol. Biol.*, 1993, **234**, 620-639.
77. K. Valegard, L. Liljas, K. Fridborg and T. Unge, *Acta Crystallogr. Sect. B: Struct. Sci.*, 1991, **47**, 949-960.
78. K. Valegard, L. Liljas, K. Fridborg and T. Unge, *Nature*, 1990, **345**, 36-41.
79. Y. Fu and J. M. Li, *Virus Res.*, 2016, **211**, 9-16.
80. E. C. Dykeman, P. G. Stockley and R. Twarock, *J. Mol. Biol.*, 2010, **395**, 916-923.
81. P. G. Stockley, O. Rolfsson, G. S. Thompson, G. Basnak, S. Francese, N. J. Stonehouse, S. W. Homans and A. E. Ashcroft, *J. Mol. Biol.*, 2007, **369**, 541-552.
82. R. Golmohammadi, K. Fridborg, M. Bundule, K. Valegard and L. Liljas, *Structure*, 1996, **4**, 543-554.
83. S. D. Brown, J. D. Fiedler and M. G. Finn, *Biochemistry*, 2009, **48**, 11155-11157.
84. P. Y. Fang, L. M. G. Ramos, S. Y. Holguin, C. Hsiao, J. C. Bowman, H. W. Yang and L. D. Williams, *Nucleic Acids Res.*, 2017, **45**, 3519-3527.
85. G. W. Witherell and O. C. Uhlenbeck, *Biochemistry*, 1989, **28**, 71-76.
86. A. E. Ashcroft, H. Lago, J. M. B. Macedo, W. T. Horn, N. J. Stonehouse and P. G. Stockley, *J. Nano. Nanotechnol.*, 2005, **5**, 2034-2041.
87. H. Takamatsu and K. Iso, *Nature*, 1982, **298**, 819-824.
88. B. C. Bundy and J. R. Swartz, *J. Biotechnol.*, 2011, **154**, 230-239.
89. J. M. Alonso, M. L. Gorzny and A. M. Bittner, *Trends Biotechnol.*, 2013, **31**, 530-538.
90. K. Namba, R. Pattanayek and G. Stubbs, *J. Mol. Biol.*, 1989, **208**, 307-325.

91. C. Sachse, J. Z. Chen, P. D. Coureux, M. E. Stroupe, M. Fandrich and N. Grigorieff, *J. Mol. Biol.*, 2007, **371**, 812-835.
92. B. J. G. E. Pieters, M. B. van Eldijk, R. J. M. Nolte and J. Mecinovic, *Chem. Soc. Rev.*, 2016, **45**, 24-39.
93. M. Sutter, D. Boehringer, S. Gutmann, S. Gunther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban and N. Ban, *Nat. Struct. Mol. Biol.*, 2008, **15**, 939-947.
94. P. J. G. Butler, *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.*, 1999, **354**, 537-550.
95. S. Y. Lee, J. S. Lim and M. T. Harris, *Biotechnol. Bioeng.*, 2012, **109**, 16-30.
96. J. W. Kehoe and B. K. Kay, *Chem. Rev.*, 2005, **105**, 4056-4072.
97. L. Makowski, *Curr. Opin. Struct. Biol.*, 1994, **4**, 225-230.
98. A. Lewin, G. R. Moore and N. E. Le Brun, *Dalton Trans.*, 2005, 3597-3610.
99. Y. Zhang and B. P. Orner, *International Journal of Molecular Sciences*, 2011, **12**, 5406-5421.
100. L. E. Bevers and E. C. Theil, *Prog Mol Subcell Biol*, 2011, **52**, 29-47.
101. M. Almiron, A. J. Link, D. Furlong and R. Kolter, *Genes Dev.*, 1992, **6**, 2646-2654.
102. A. Ilari, C. Savino, S. Stefanini, E. Chiancone and D. Tsernoglou, *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 1999, **55**, 552-553.
103. J. Swift, C. A. Butts, J. Cheung-Lau, V. Yerubandi and I. J. Dmochowski, *Langmuir*, 2009, **25**, 5219-5225.
104. E. Johnson, D. Cascio, M. R. Sawaya, M. Gingery and I. Schröder, *Structure*, 2005, **13**, 637-648.
105. B. Sana, E. Johnson, P. Le Magueres, A. Criswell, D. Cascio and S. Lim, *J. Biol. Chem.*, 2013, **288**, 32663-32672.
106. M. Gerl and R. Jaenicke, *Eur. Biophys. J.*, 1987, **15**, 103-109.
107. K. Schott, R. Ladenstein, A. König and A. Bacher, *J. Biol. Chem.*, 1990, **265**, 12686-12689.
108. X. F. Zhang, P. V. Konarev, M. V. Petoukhov, D. I. Svergun, L. Xing, R. H. Cheng, I. Haase, M. Fischer, A. Bacher, R. Ladenstein and W. Meining, *J. Mol. Biol.*, 2006, **362**, 753-770.
109. R. Ladenstein, M. Fischer and A. Bacher, *FEBS J.*, 2013, **280**, 2537-2563.
110. Y. Cheng, W. Boll, T. Kirchhausen, S. C. Harrison and T. Walz, *J. Mol. Biol.*, 2007, **365**, 892-899.
111. J. L. Milne, X. Wu, M. J. Borgnia, J. S. Lengyel, B. R. Brooks, D. Shi, R. N. Perham and S. Subramaniam, *J. Biol. Chem.*, 2006, **281**, 4364-4370.
112. L. H. Rome and V. A. Kickhoefer, *ACS Nano*, 2013, **7**, 889-902.
113. H. Tanaka, K. Kato, E. Yamashita, T. Sumizawa, Y. Zhou, M. Yao, K. Iwasaki, M. Yoshimura and T. Tsukihara, *Science*, 2009, **323**, 384-388.
114. D. Moreira and P. Lopez-Garcia, *Nat. Rev. Microbiol.*, 2009, **7**, 306-311.
115. P. Forterre and D. Prangishvili, *Res. Microbiol.*, 2009, **160**, 466-472.
116. E. V. Koonin, V. V. Dolja and M. Krupovic, *Virology*, 2015, **479**, 2-25.
117. N. G. A. Abrescia, D. H. Bamford, J. M. Grimes and D. I. Stuart, *Annu. Rev. Biochem.*, 2012, **81**, 795-822.
118. D. H. Bamford, J. M. Grimes and D. I. Stuart, *Curr. Opin. Struct. Biol.*, 2005, **15**, 655-663.
119. S. D. Benson, J. K. H. Bamford, D. H. Bamford and R. M. Burnett, *Mol. Cell*, 2004, **16**, 673-685.

120. M. Krupovic and D. H. Bamford, *Nat. Rev. Microbiol.*, 2008, **6**, 941-948.
121. S. D. Benson, J. K. H. Bamford, D. H. Bamford and R. M. Burnett, *Cell*, 1999, **98**, 825-833.
122. J. J. B. Cockburn, N. G. A. Abrescia, J. M. Grimes, G. C. Sutton, J. M. Diprose, J. M. Benevides, G. J. Thomas, J. K. H. Bamford, D. H. Bamford and D. I. Stuart, *Nature*, 2004, **432**, 122-125.
123. N. G. A. Abrescia, J. J. B. Cockburn, J. M. Grimes, G. C. Sutton, J. M. Diprose, S. J. Butcher, S. D. Fuller, C. S. Martin, R. M. Burnett, D. I. Stuart, D. H. Bamford and J. K. H. Bamford, *Nature*, 2004, **432**, 68-74.
124. M. M. Suhanovsky and C. M. Teschke, *Virology*, 2015, **479**, 487-497.
125. I. Gertsman, L. Gan, M. Guttman, K. Lee, J. A. Speir, R. L. Duda, R. W. Hendrix, E. A. Komives and J. E. Johnson, *Nature*, 2009, **458**, 646-650.
126. W. K. Kegel and P. van der Schoot, *Biophys. J.*, 2006, **91**, 1501-1512.
127. J. D. Perlmutter and M. F. Hagan, *Annu. Rev. Phys. Chem.*, 2015, **66**, 217-239.
128. P. E. Prevelige, D. Thomas and J. King, *Biophys. J.*, 1993, **64**, 824-835.
129. H. D. Nguyen, V. S. Reddy and C. L. Brooks, *Nano Lett.*, 2007, **7**, 338-344.
130. A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres and M. J. Young, *Virology*, 2000, **277**, 450-456.
131. Y. F. Hu, R. Zandi, A. Anavitarte, C. M. Knobler and W. M. Gelbart, *Biophys. J.*, 2008, **94**, 1428-1436.
132. S. E. Aniagyei, C. J. Kennedy, B. Stein, D. A. Willits, T. Douglas, M. J. Young, M. De, V. M. Rotello, D. Srisathiyarayanan, C. C. Kao and B. Dragnea, *Nano Lett.*, 2009, **9**, 393-398.
133. X. Huang, L. M. Bronstein, J. Retrum, C. Dufort, I. Tsvetkova, S. Aniagyei, B. Stein, G. Stucky, B. McKenna, N. Remmes, D. Baxter, C. C. Kao and B. Dragnea, *Nano Lett.*, 2007, **7**, 2407-2416.
134. C. Chen, M.-C. Daniel, Z. T. Quinkert, M. De, B. Stein, V. D. Bowman, P. R. Chipman, V. M. Rotello, C. C. Kao and B. Dragnea, *Nano Lett.*, 2006, **6**, 611-615.
135. K. N. Parent, A. Zlotnick and C. M. Teschke, *J. Mol. Biol.*, 2006, **359**, 1097-1106.
136. P. A. Thuman-Commike, B. Greene, J. A. Malinski, M. Burbea, A. McGough, W. Chiu and P. E. Prevelige, *Biophys. J.*, 1999, **76**, 3267-3277.
137. M. H. Parker, W. F. Stafford and P. E. Prevelige, *J. Mol. Biol.*, 1997, **268**, 655-665.
138. P. E. Prevelige and B. A. Fane, *Adv. Exp. Med. Biol.*, 2012, **726**, 325-350.
139. J. R. Cortines, P. R. Weigele, E. B. Gilcrease, S. R. Casjens and C. M. Teschke, *Virology*, 2011, **421**, 1-11.
140. J. R. Cortines, T. Motwani, A. A. Vyas and C. M. Teschke, *J. Virol.*, 2014, **88**, 5287-5297.
141. P. A. Thuman-Commike, B. Greene, J. A. Malinski, J. King and W. Chiu, *Biophys. J.*, 1998, **74**, 559-568.
142. S. D. Moore and P. E. Prevelige, Jr., *J. Mol. Biol.*, 2002, **315**, 975-994.
143. B. V. Prasad, P. E. Prevelige, E. Marietta, R. O. Chen, D. Thomas, J. King and W. Chiu, *J. Mol. Biol.*, 1993, **231**, 65-74.
144. M. T. Fuller and J. King, *Biophys. J.*, 1980, **32**, 381-401.
145. M. T. Fuller and J. King, *Virology*, 1981, **112**, 529-547.
146. M. L. Galisteo and J. King, *Biophys. J.*, 1993, **65**, 227-235.

147. C. M. Teschke, A. McGough and P. A. Thuman-Commike, *Biophys. J.*, 2003, **84**, 2585-2592.
148. C. Chen, J. C. Y. Wang and A. Zlotnick, *PLoS Path.*, 2011, **7**.
149. R. F. Garmann, M. Comas-Garcia, A. Gopal, C. M. Knobler and W. M. Gelbart, *J. Mol. Biol.*, 2014, **426**, 1050-1060.
150. R. F. Garmann, R. Sportsman, C. Beren, V. N. Manoharan, C. M. Knobler and W. M. Gelbart, *J. Am. Chem. Soc.*, 2015, **137**, 7584-7587.
151. E. C. Dykeman, P. G. Stockley and R. Twarock, *J. Mol. Biol.*, 2013, **425**, 3235-3249.
152. P. G. Stockley, R. Twarock, S. E. Bakker, A. M. Barker, A. Borodavka, E. Dykeman, R. J. Ford, A. R. Pearson, S. E. V. Phillips, N. A. Ranson and R. Tuma, *J. Biol. Phys.*, 2013, **39**, 277-287.
153. G. Stubbs, *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.*, 1999, **354**, 551-557.
154. A. Klug, *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.*, 1999, **354**, 531-535.
155. E. L. Hesketh, Y. Meshcheriakova, K. C. Dent, P. Saxena, R. F. Thompson, J. J. Cockburn, G. P. Lomonosoff and N. A. Ranson, *Nat. Commun.*, 2015, **6**.
156. J. D. Perlmutter, M. R. Perkett and M. F. Hagan, *J. Mol. Biol.*, 2014, **426**, 3148-3165.
157. O. M. Elrad and M. F. Hagan, *Phys. Biol.*, 2010, **7**.
158. A. McPherson, *Bioessays*, 2005, **27**, 447-458.
159. M. Gerl, R. Jaenicke, J. M. A. Smith and P. M. Harrison, *Biochemistry*, 1988, **27**, 4089-4096.
160. D. Sato, H. Ohtomo, Y. Yamada, T. Hikima, A. Kurobe, K. Fujiwara and M. Ikeguchi, *Biochemistry*, 2016, **55**, 287-293.
161. D. Sato, S. Takebe, A. Kurobe, H. Ohtomo, K. Fujiwara and M. Ikeguchi, *Biochemistry*, 2016, **55**, 482-488.
162. J. D. Perlmutter, F. Mohajerani and M. F. Hagan, *Elife*, 2016, **5**.
163. E. Hiebert, J. B. Bancroft and C. E. Bracker, *Virology*, 1968, **34**, 492-508.
164. J. B. Bancroft, G. W. Wagner and C. E. Bracker, *Virology*, 1968, **36**, 146-149.
165. J. B. Bancroft, C. E. Bracker and G. W. Wagner, *Virology*, 1969, **38**, 324-335.
166. E. Hiebert and J. B. Bancroft, *Virology*, 1969, **39**, 296-311.
167. X. X. Zhao, J. M. Fox, N. H. Olson, T. S. Baker and M. J. Young, *Virology*, 1995, **207**, 486-494.
168. K. W. Adolph and P. J. Butler, *J. Mol. Biol.*, 1974, **88**, 327-341.
169. J. B. Bancroft, *Adv. Virus Res.*, 1970, **16**, 99-134.
170. I. J. L. Byeon, X. Meng, J. W. Jung, G. P. Zhao, R. F. Yang, J. W. Ahn, J. Shi, J. Concel, C. Aiken, P. J. Zhang and A. M. Gronenborn, *Cell*, 2009, **139**, 780-790.
171. A. C. Steven, E. Couture, U. Aebi and M. K. Showe, *J. Mol. Biol.*, 1976, **106**, 187-221.
172. A. C. Steven, P. Serwer, M. E. Bisher and B. L. Trus, *Virology*, 1983, **124**, 109-120.
173. M. M. Suhanovsky, K. N. Parent, S. E. Dunn, T. S. Baker and C. M. Teschke, *Mol. Microbiol.*, 2010, **77**, 1568-1582.
174. K. N. Parent, R. S. Sinkovits, M. M. Suhanovsky, C. M. Teschke, E. H. Egelman and T. S. Baker, *Phys. Biol.*, 2010, **7**.
175. N. P. King and Y. T. Lai, *Curr. Opin. Struct. Biol.*, 2013, **23**, 632-638.
176. Y. T. Lai, N. P. King and T. O. Yeates, *Trends Cell Biol.*, 2012, **22**, 653-661.
177. J. E. Padilla, C. Colovos and T. O. Yeates, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 2217-2221.

178. J. C. Sinclair, K. M. Davies, C. Venien-Bryan and M. E. M. Noble, *Nat. Nanotechnol.*, 2011, **6**, 558-562.
179. Y. T. Lai, K. L. Tsai, M. R. Sawaya, F. J. Asturias and T. O. Yeates, *J. Am. Chem. Soc.*, 2013, **135**, 7738-7743.
180. N. P. King, J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates and D. Baker, *Nature*, 2014, **510**, 103-108.
181. J. B. Bale, S. Gonen, Y. Liu, W. Sheffler, D. Ellis, C. Thomas, D. Cascio, T. O. Yeates, T. Gonen, N. P. King and D. Baker, *Science (New York, NY)*, 2016, **353**, 389-394.
182. Y. Hsia, J. B. Bale, S. Gonen, D. Shi, W. Sheffler, K. K. Fong, U. Nattermann, C. Xu, P. S. Huang, R. Ravichandran, S. Yi, T. N. Davis, T. Gonen, N. P. King and D. Baker, *Nature*, 2016, **535**, 136-139.
183. A. Ljubetic, F. Lapenta, H. Gradisar, I. Drobnak, J. Aupic, Z. Strmsek, D. Lainscek, I. Hafner-Bratkovic, A. Majerle, N. Krivec, M. Bencina, T. Pisanski, T. C. Velickovic, A. Round, J. M. Carazo, R. Melero and R. Jerala, *Nat. Biotechnol.*, 2017, **35**, 1094-1101.
184. E. C. Theil, *Curr. Opin. Chem. Biol.*, 2011, **15**, 304-311.
185. T. Douglas and D. R. Ripoll, *Protein Sci.*, 1998, **7**, 1083-1091.
186. T. Tosha, H.-L. Ng, O. Bhattasali, T. Alber and E. C. Theil, *J. Am. Chem. Soc.*, 2010, **132**, 14562-14569.
187. T. Tosha, M. R. Hasan and E. C. Theil, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 18182-18187.
188. G. Jutz, P. van Rijn, B. Santos Miranda and A. Boker, *Chem. Rev.*, 2015, **115**, 1653-1701.
189. F. C. Meldrum, V. J. Wade, D. L. Nimmo, B. R. Heywood and S. Mann, *Nature*, 1991, **349**, 684-687.
190. M. Uchida, M. L. Flenniken, M. Allen, D. A. Willits, B. E. Crowley, S. Brumfield, A. F. Willis, L. Jackiw, M. Jutila, M. J. Young and T. Douglas, *J. Am. Chem. Soc.*, 2006, **128**, 16626-16633.
191. M. Allen, D. Willits, M. Young and T. Douglas, *Inorg. Chem.*, 2003, **42**, 6300-6305.
192. S. Kang, J. Lucon, Z. B. Varpness, L. Liepold, M. Uchida, D. Willits, M. Young and T. Douglas, *Angewandte Chemie-International Edition*, 2008, **47**, 7845-7848.
193. T. Douglas and M. Young, *Nature*, 1998, **393**, 152-155.
194. C. Jolley, M. Klem, R. Harrington, J. Parise and T. Douglas, *Nanoscale*, 2011, **3**, 1004-1007.
195. T. Douglas, E. Strable, D. Willits, A. Aitouchen, M. Libera and M. Young, *Adv. Mater.*, 2002, **14**, 415-418.
196. M. Knez, M. Sumser, A. M. Bittner, C. Wege, H. Jeske, T. P. Martin and K. Kern, *Adv. Funct. Mater.*, 2004, **14**, 116-124.
197. D. Ishii, K. Kinbara, Y. Ishida, N. Ishii, M. Okochi, M. Yohda and T. Aida, *Nature*, 2003, **423**, 628-632.
198. R. A. McMillan, C. D. Paavola, J. Howard, S. L. Chan, N. J. Zaluzec and J. D. Trent, *Nat. Mater.*, 2002, **1**, 247-252.
199. R. A. McMillan, J. Howard, N. J. Zaluzec, H. K. Kagawa, R. Mogul, Y.-F. Li, C. D. Paavola and J. D. Trent, *J. Am. Chem. Soc.*, 2005, **127**, 2800-2801.
200. K. M. Bromley, A. J. Patil, A. W. Perriman, G. Stubbs and S. Mann, *J. Mater. Chem.*, 2008, **18**, 4796-4796.

201. J. C. Zhou, C. M. Soto, M.-S. Chen, M. A. Bruckman, M. H. Moore, E. Barry, B. R. Ratna, P. E. Pehrsson, B. R. Spies and T. S. Confer, *J. Nanobiotechnol.*, 2012, **10**, 18.
202. M. Knez, A. M. Bittner, F. Boes, C. Wege, H. Jeske, E. Maiß and K. Kern, *Nano Lett.*, 2003, **3**, 1079-1082.
203. S. Balci, K. Hahn, P. Kopold, A. Kadri, C. Wege, K. Kern and A. M. Bittner, *Nanotechnology*, 2012, **23**, 045603.
204. M. Kobayashi, K. Onodera, Y. Watanabe and I. Yamashita, *Chem. Lett.*, 2010, **39**, 616-618.
205. X. Z. Fan, E. Pomerantseva, M. Gnerlich, A. Brown, K. Gerasopoulos, M. McCarthy, J. Culver and R. Ghodssi, *J. Vac. Sci. Technol.*, 2013, **31**, 050815.
206. S. Balci, A. M. Bittner, K. Hahn, C. Scheu, M. Knez, A. Kadri, C. Wege, H. Jeske and K. Kern, *Electrochim. Acta*, 2006, **51**, 6251-6257.
207. R. Tsukamoto, M. Muraoka, M. Seki, H. Tabata and I. Yamashita, *Chem. Mater.*, 2007, **19**, 2389-2391.
208. M. Kobayashi, M. Seki, H. Tabata, Y. Watanabe and I. Yamashita, *Nano Lett.*, 2010, **10**, 773-776.
209. E. Dujardin, C. Peet, G. Stubbs, J. N. Culver and S. Mann, *Nano Lett.*, 2003, **3**, 413-417.
210. M. J. Abedin, L. Liepold, P. Suci, M. Young and T. Douglas, *J. Am. Chem. Soc.*, 2009, **131**, 4346-4354.
211. L. O. Liepold, M. J. Abedin, E. D. Buckhouse, J. A. Frank, M. J. Young and T. Douglas, *Nano Lett.*, 2009, **9**, 4520-4526.
212. J. Lucon, M. J. Abedin, M. Uchida, L. Liepold, C. C. Jolley, M. Young and T. Douglas, *Chem. Comm.*, 2010, **46**, 264-266.
213. M. Uchida, D. S. Morris, S. Kang, C. C. Jolley, J. Lucon, L. O. Liepold, B. LaFrance, P. E. Prevelige and T. Douglas, *Langmuir*, 2012, **28**, 1998-2006.
214. J.-S. Wang and K. Matyjaszewski, *J. Am. Chem. Soc.*, 1995, **117**, 5614-5615.
215. M. Kato, M. Kamigaito and M. Sawamoto, *Macromolecules*, Macromolecules, 1995.
216. J. Lucon, S. Qazi, M. Uchida, G. J. Bedwell, B. LaFrance, P. E. Prevelige Jr and T. Douglas, *Nat. Chem.*, 2012, **4**, 781-788.
217. E. Edwards, R. Roychoudhury, B. Schwarz, P. Jordan, J. Lisher, M. Uchida and T. Douglas, *J. Mater. Chem. B*, 2016, **4**, 5375-5384.
218. J. Lucon, E. Edwards, S. Qazi, M. Uchida and T. Douglas, *Eur. Polym. J.*, 2013, **49**, 2976-2985.
219. J. K. Pokorski, K. Breitenkamp, L. O. Liepold, S. Qazi and M. G. Finn, *J. Am. Chem. Soc.*, 2011, **133**, 9242-9245.
220. M. L. Hovlid, J. L. Lau, K. Breitenkamp, C. J. Higginson, B. Laufer, M. Manchester and M. G. Finn, *ACS Nano*, 2014, **8**, 8003-8014.
221. K. Renggli, N. Sauter, M. Rother, M. G. Nussbaumer, R. Urbani, T. Pfohl and N. Bruns, *Polymer Chemistry*, 2017, **8**, 2133-2136.
222. M. G. Nussbaumer, J. T. Duskey, M. Rother, K. Renggli, M. Chami and N. Bruns, *Adv Sci*, 2016, **3**, 1600046.
223. M. G. Nussbaumer, C. Bisig and N. Bruns, *Chem. Commun.*, 2016, **52**, 10537-10539.
224. X. Wang, C. Wang, M. Pan, J. Wei, F. Jiang, R. Lu, X. Liu, Y. Huang and F. Huang, *ACS Appl Mater Interfaces*, 2017, **9**, 25387-25396.
225. L. Ditzel, J. Lowe, D. Stock, K. O. Stetter, H. Huber, R. Huber and S. Steinbacher, *Cell*, 1998, **93**, 125-138.

226. B. C. Ng, M. Yu, A. Gopal, L. H. Rome, H. G. Monbouquette and S. H. Tolbert, *Nano Lett.*, 2008, **8**, 3503-3509.
227. S. K. Dixit, N. L. Goicochea, M.-C. Daniel, A. Murali, L. Bronstein, M. De, B. Stein, V. M. Rotello, C. C. Kao and B. Dragnea, *Nano Lett.*, 2006, **6**, 1993-1999.
228. S. E. Aniagyei, C. DuFort, C. C. Kao and B. Dragnea, *J. Mater. Chem.*, 2008, **18**, 3763-3712.
229. A. G. Malyutin, H. Cheng, O. R. Sanchez-Felix, K. Carlson, B. D. Stein, P. V. Konarev, D. I. Svergun, B. Dragnea and L. M. Bronstein, *ACS Appl. Mat. Inter.*, 2015, **7**, 12089-12098.
230. I. Tsvetkova, C. Chen, S. Rana, C. C. Kao, V. M. Rotello and B. Dragnea, *Soft Matter*, 2012, **8**, 4571.
231. 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 and B. Dragnea, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 1354-1359.
232. X. Huang, B. D. Stein, H. Cheng, A. Malyutin, I. B. Tsvetkova, D. V. Baxter, N. B. Remmes, J. Verchot, C. Kao, L. M. Bronstein and B. Dragnea, *ACS Nano*, 2011, **5**, -4045.
233. L. Loo, R. H. Guenther, S. A. Lommel and S. Franzen, *J. Am. Chem. Soc.*, 2007, **129**, 11111-11117.
234. L. Loo, R. H. Guenther, V. R. Basnayake, S. A. Lommel and S. Franzen, *J. Am. Chem. Soc.*, 2006, **128**, 4502-4503.
235. M. Vieweger, N. Goicochea, E. S. Koh and B. Dragnea, *ACS Nano*, 2011, **5**, 7324-7333.
236. I. J. Minten, L. J. A. Hendriks, R. J. M. Nolte and J. J. L. M. Cornelissen, *J. Am. Chem. Soc.*, 2009, **131**, 17771-17773.
237. F. P. Seebeck, K. J. Woycechowsky, W. Zhuang, J. P. Rabe and D. Hilvert, *J. Am. Chem. Soc.*, 2006, **128**, 4516-4517.
238. B. Wörsdörfer, K. J. Woycechowsky and D. Hilvert, *Science*, 2011, **331**, 589-592.
239. D. S. Peabody, *EMBO J.*, 1993, **12**, 595-600.
240. F. Lim, M. Spingola and D. S. Peabody, *J. Biol. Chem.*, 1996, **271**, 31839-31845.
241. A. M. Parrott, H. Lago, C. J. Adams, A. E. Ashcroft, N. J. Stonehouse and P. G. Stockley, *Nucleic Acids Res.*, 2000, **28**, 489-497.
242. I. J. Minten, V. I. Claessen, K. Blank, A. E. Rowan, R. J. M. Nolte and J. J. L. M. Cornelissen, *Chemical Science*, 2011, **2**, 358-362.
243. T. Beck, S. Tetter, M. Kunzle and D. Hilvert, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 937-940.
244. E. Sasaki and D. Hilvert, *J. Phys. Chem. B*, 2016, **120**, 6089-6095.
245. Y. Azuma, R. Zschoche, M. Tinzl and D. Hilvert, *Angew. Chem. Int. Ed. Engl.*, 2016, **55**, 1531-1534.
246. R. Frey, S. Mantri, M. Rocca and D. Hilvert, *J. Am. Chem. Soc.*, 2016, **138**, 10072-10075.
247. E. Sasaki, D. Bohringer, M. van de Waterbeemd, M. Leibundgut, R. Zschoche, A. J. Heck, N. Ban and D. Hilvert, *Nat Commun*, 2017, **8**, 14663.
248. J. K. Rhee, M. Hovlid, J. D. Fiedler, S. D. Brown, F. Manzenrieder, H. Kitagishi, C. Nycholat, J. C. Paulson and M. G. Finn, *Biomacromolecules*, 2011, **12**, 3977-3981.
249. M. Wu, W. L. Brown and P. G. Stockley, *Bioconj. Chem.*, 1995, **6**, 587-595.

250. 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, J. d. C. Caldeira, B. Chackerian, W. Wharton and D. S. Peabody, *ACS Nano*, 2011, **5**, 5729-5745.
251. M. H. Parker, S. Casjens and P. E. Prevelige Jr, *J. Mol. Biol.*, 1998, **281**, 69-79.
252. P. R. Weigele, L. Sampson, D. Winn-Stapley and S. R. Casjens, *J. Mol. Biol.*, 2005, **348**, 831-844.
253. A. O'Neil, C. Reichhardt, B. Johnson, P. E. Prevelige and T. Douglas, *Angew. Chem. Int. Ed.*, 2011, **50**, 7425-7428.
254. D. P. Patterson, K. McCoy, C. Fijen and T. Douglas, *J. Mater. Chem. B*, 2014, **2**, 5948-5951.
255. D. P. Patterson, B. LaFrance and T. Douglas, *Chem. Commun.*, 2013, **49**, 10412.
256. D. P. Patterson, P. E. Prevelige and T. Douglas, *ACS Nano*, 2012, **6**, 5000-5009.
257. A. O'Neil, P. E. Prevelige and T. Douglas, *Biometer. Sci.*, 2013, **1**, 881.
258. T. Inoue, M.-a. Kawano, R.-u. Takahashi, H. Tsukamoto, T. Enomoto, T. Imai, K. Kataoka and H. Handa, *J. Biotechnol.*, 2008, **134**, 181-192.
259. C. M. Teschke, A. McGough and P. A. Thuman-Commike, *Biophys. J.*, 2003, **84**, 2585-2592.
260. A. Llauro, D. Luque, E. Edwards, B. L. Trus, J. Avera, D. Reguera, T. Douglas, P. J. Pablo and J. R. Caston, *Nanoscale*, 2016, **8**, 9328-9336.
261. D. P. Patterson, B. Schwarz, K. El-Boubbou, J. van der Oost, P. E. Prevelige and T. Douglas, *Soft Matter*, 2012, **8**, 10158.
262. J. Sharma, M. Uchida, H. M. Miettinen and T. Douglas, *Nanoscale*, 2017, **9**, 10420-10430.
263. M. Mammen, S. K. Choi and G. M. Whitesides, *Angew. Chem.*, 1998, **37**, 2754-2794.
264. S. H. Yang, W. J. Chung, S. McFarland and S. W. Lee, *Chem Rec*, 2013, **13**, 43-59.
265. B. P. Gray and K. C. Brown, *Chem. Rev.*, 2014, **114**, 1020-1081.
266. G. P. Smith and V. A. Petrenko, *Chem. Rev.*, 1997, **97**, 391-410.
267. M. Uchida, A. D. Willits, K. Muller, A. F. Willis, L. Jackiw, M. Jutila, M. Young, A. E. Porter and T. Douglas, *Adv. Mater.*, 2009, **21**, 458-462.
268. M. Uchida, H. Kosuge, M. Terashima, D. A. Willits, L. O. Liepold, M. J. Young, M. V. McConnell and T. Douglas, *ACS Nano*, 2011, **5**, 2493-2502.
269. M. L. Flenniken, D. A. Willits, A. L. Harmsen, L. O. Liepold, A. G. Harmsen, M. J. Young and T. Douglas, *Chem. Biol.*, 2006, **13**, 161-170.
270. A. Servid, P. Jordan, A. O'Neil, P. Prevelige and T. Douglas, *Biomacromolecules*, 2013, **14**, 2989-2995.
271. J. P. Phelps, N. Dang and L. Rasochova, *J. Virol. Methods*, 2007, **141**, 146-153.
272. A. C. Tissot, R. Renhofa, N. Schmitz, I. Cielens, E. Meijerink, V. Ose, G. T. Jennings, P. Saudan, P. Pumpens and M. F. Bachmann, *PLoS One*, 2010, **5**, e9809.
273. C. H. Wu, I. J. Liu, R. M. Lu and H. C. Wu, *J. Biomed. Sci.*, 2016, **23**.
274. S. R. Whaley, D. S. English, E. L. Hu, P. F. Barbara and A. M. Belcher, *Nature*, 2000, **405**, 665-668.
275. R. R. Naik, S. J. Stringer, G. Agarwal, S. E. Jones and M. O. Stone, *Nat. Mater.*, 2002, **1**, 169-172.
276. H. B. Chen, X. D. Su, K. G. Neoh and W. S. Choe, *Anal. Chem.*, 2006, **78**, 4872-4879.
277. M. Sarikaya, C. Tamerler, A. Jen, K. Schulten and F. Baneyx, *Nat. Mater.*, 2003, **2**, 577-585.

278. C. Mao, C. E. Flynn, A. Hayhurst, R. Sweeney, J. Qi, G. Georgiou, B. Iverson and A. M. Belcher, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 6946-6951.
279. H. R. Hoogenboom, A. D. Griffiths, K. S. Johnson, D. J. Chiswell, P. Hudson and G. Winter, *Nucleic Acids Res.*, 1991, **19**, 4133-4137.
280. P. A. Kratz, B. Böttcher and M. Nassal, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 1915-1920.
281. A. Schneemann, J. A. Speir, G. S. Tan, R. Khayat, D. C. Ekiert, Y. Matsuoka and I. A. Wilson, *J. Virol.*, 2012, **86**, 11686-11697.
282. S. Gleiter and H. Lilie, *Protein Sci.*, 2001, **10**, 434-444.
283. M. Kanekiyo, C.-J. Wei, H. M. Yassine, P. M. McTamney, J. C. Boyington, J. R. R. Whittle, S. S. Rao, W.-P. Kong, L. Wang and G. J. Nabel, *Nature*, 2013, **499**, 102-106.
284. J. K. Pokorski, M. L. Hovlid and M. G. Finn, *ChemBioChem*, 2011, **12**, 2441-2447.
285. H. Moon, J. Lee, J. Min and S. Kang, *Biomacromolecules*, 2014, **15**, 3794-3801.
286. B. K. Culpepper, D. S. Morris, P. E. Prevelige and S. L. Bellis, *Biomaterials*, 2013, **34**, 2455-2462.
287. S. Tsukiji and T. Nagamune, *ChemBioChem*, 2009, **10**, 787-798.
288. D. Patterson, B. Schwarz, J. Avera, B. Western, M. Hicks, P. Krugler, M. Terra, M. Uchida, K. McCoy and T. Douglas, *Bioconjug Chem*, 2017, DOI: 10.1021/acs.bioconjchem.7b00296.
289. L. Schoonen, J. Pille, A. Borrmann, R. J. Nolte and J. C. van Hest, *Bioconjug Chem*, 2015, **26**, 2429-2434.
290. S. C. Reddington and M. Howarth, *Curr. Opin. Chem. Biol.*, 2015, **29**, 94-99.
291. H. Moon, Y. Bae, H. Kim and S. Kang, *Chem. Commun.*, 2016, **52**, 14051-14054.
292. N. Stephanopoulos and M. B. Francis, *Nat. Chem. Biol.*, 2011, **7**, 876-884.
293. M. L. Flenniken, D. A. Willits, S. Brumfield, M. J. Young and T. Douglas, *Nano Lett.*, 2003, **3**, 1573-1576.
294. R. A. Miller, A. D. Presley and M. B. Francis, *J. Am. Chem. Soc.*, 2007, **129**, 3104-3109.
295. M. T. Klem, D. Willits, M. Young and T. Douglas, *J. Am. Chem. Soc.*, 2003, **125**, 10806-10807.
296. A. M. ElSohly and M. B. Francis, *Acc. Chem. Res.*, 2015, **48**, 1971-1978.
297. C. B. Rosen and M. B. Francis, *Nat. Chem. Biol.*, 2017, **13**, 697-705.
298. T. J. Sereda, C. T. Mant, A. M. Quinn and R. S. Hodges, *J. Chromatogr.*, 1993, **646**, 17-30.
299. M. T. Dedeo, K. E. Duderstadt, J. M. Berger and M. B. Francis, *Nano Lett.*, 2010, **10**, 181-186.
300. N. Stephanopoulos, G. J. Tong, S. C. Hsiao and M. B. Francis, *Acs Nano*, 2010, **4**, 6014-6020.
301. Y. Kulmashiro, Y. Ikezoe, K. Tamada and M. Hara, *J. Phys. Chem. B*, 2008, **112**, 8291-8297.
302. K. S. Raja, Q. Wang, M. J. Gonzalez, M. Manchester, J. E. Johnson and M. G. Finn, *Biomacromolecules*, 2003, **4**, 472-476.
303. N. F. Steinmetz and M. Manchester, *Biomacromolecules*, 2009, **10**, 784-792.
304. G. C. Lander, A. Evilevitch, M. Jeembaeva, C. S. Potter, B. Carragher and J. E. Johnson, *Structure*, 2008, **16**, 1399-1406.
305. L. Qin, A. Fokine, E. O'Donnell, V. B. Rao and M. G. Rossmann, *J. Mol. Biol.*, 2010, **395**, 728-741.
306. Y. Xiang and M. G. Rossmann, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 4806-4810.

307. A. Gupta, M. Onda, I. Pastan, S. Adhya and V. K. Chaudhary, *J. Mol. Biol.*, 2003, **334**, 241-254.
308. F. Yang, P. Forrer, Z. Dauter, J. F. Conway, N. Cheng, M. E. Cerritelli, A. C. Steven, A. Plückthun and A. Wlodawer, *Nat. Struct. Mol. Biol.*, 2000, **7**, 230-237.
309. Z.-j. Ren and L. W. Black, *Gene*, 1998, **215**, 439-444.
310. L. Tang, E. B. Gilcrease, S. R. Casjens and J. E. Johnson, *Structure*, 2006, **14**, 837-845.
311. K. N. Parent, C. T. Deedas, E. H. Egelman, S. R. Casjens, T. S. Baker and C. M. Teschke, *Biomaterials*, 2012, **33**, 5628-5637.
312. P. L. Rodriguez, T. Harada, D. A. Christian, D. A. Pantano, R. K. Tsai and D. E. Discher, *Science*, 2013, **339**, 971-975.
313. B. Schwarz, P. Madden, J. Avera, B. Gordon, K. Larson, H. M. Miettinen, M. Uchida, B. LaFrance, G. Basu and A. Rynda-Applé, *ACS nano*, 2015, **9**, 9134-9147.
314. B. Chackerian, P. Lenz, D. R. Lowy and J. T. Schiller, *J. Immunol.*, 2002, **169**, 6120-6126.
315. B. Chackerian, D. R. Lowy and J. T. Schiller, *J. Clin. Investig.*, 2001, **108**, 415-423.
316. Z. Liu, J. Qiao, Z. Niu and Q. Wang, *Chem. Soc. Rev.*, 2012, **41**, 6178-6194.
317. G. M. Whitesides and B. Grzybowski, *Science*, 2002, **295**, 2418-2421.
318. U. G. Wegst, H. Bai, E. Saiz, A. P. Tomsia and R. O. Ritchie, *Nat. Mater.*, 2015, **14**, 23-36.
319. L. Tskhovrebova and J. Trinick, *Nat. Rev. Mol. Cell Biol.*, 2003, **4**, 679-689.
320. S. Biswas, K. Kinbara, T. Niwa, H. Taguchi, N. Ishii, S. Watanabe, K. Miyata, K. Kataoka and T. Aida, *Nat. Chem.*, 2013, **5**, 613-620.
321. T. Sendai, S. Biswas and T. Aida, *J. Am. Chem. Soc.*, 2013, **135**, 11509-11512.
322. S. Biswas, K. Kinbara, N. Oya, N. Ishii, H. Taguchi and T. Aida, *J. Am. Chem. Soc.*, 2009, **131**, 7556-+.
323. S. Sim, D. Miyajima, T. Niwa, H. Taguchi and T. Aida, *J. Am. Chem. Soc.*, 2015, **137**, 4658-4661.
324. P. Fromherz, *Nature*, 1971, **231**, 267-268.
325. E. E. Uzgiris and R. D. Kornberg, *Nature*, 1983, **301**, 125-129.
326. H. Yoshimura, T. Scheybani, W. Baumeister and K. Nagayama, *Langmuir*, 1994, **10**, 3290-3295.
327. S. Kewalramani, S. Wang, Y. Lin, H. G. Nguyen, Q. Wang, M. Fukuto and L. Yang, *Soft Matter*, 2011, **7**, 939-945.
328. M. Fukuto, Q. L. Nguyen, O. Vasilyev, N. Mank, C. L. Washington-Hughes, I. Kuzmenko, A. Checco, Y. Mao, Q. Wang and L. Yang, *Soft Matter*, 2013, **9**, 9633.
329. H. Yoshimura, E. Edwards, M. Uchida, K. McCoy, R. Roychoudhury, B. Schwarz, D. Patterson and T. Douglas, *J. Phys. Chem. B*, 2016, DOI: 10.1021/acs.jpcc.6b01425.
330. G. Kaur, J. He, J. Xu, S. Pingali, G. Jutz, A. Boker, Z. Niu, T. Li, D. Rawlinson, T. Emrick, B. Lee, P. Thiyagarajan, T. P. Russell and Q. Wang, *Langmuir*, 2009, **25**, 5168-5176.
331. J. T. Russell, Y. Lin, A. Boker, L. Su, P. Carl, H. Zettl, J. He, K. Sill, R. Tangirala, T. Emrick, K. Littrell, P. Thiyagarajan, D. Cookson, A. Fery, Q. Wang and T. P. Russell, *Angew. Chem. Int. Ed. Engl.*, 2005, **44**, 2420-2426.
332. I. Yamashita, *Thin Solid Films*, 2001, **393**, 12-18.
333. M. Okuda, Y. Kobayashi, K. Suzuki, K. Sonoda, T. Kondoh, A. Wagawa, A. Kondo and H. Yoshimura, *Nano Lett.*, 2005, **5**, 991-993.

334. C. L. Cheung, J. A. Camarero, B. W. Woods, T. Lin, J. E. Johnson and J. J. De Yoreo, *J. Am. Chem. Soc.*, 2003, **125**, 6848-6849.
335. C. L. Cheung, S.-W. Chung, A. Chatterji, T. Lin, J. E. Johnson, S. Hok, J. Perkins and J. J. De Yoreo, *J. Am. Chem. Soc.*, 2006, **128**, 10801-10807.
336. P. A. Suci, M. T. Klem, T. Douglas and M. Young, *Langmuir*, 2005, **21**, 8686-8693.
337. M. Okochi, T. Sugita, S. Furusawa, M. Umetsu, T. Adschiri and H. Honda, *Biotechnol. Bioeng.*, 2010, **106**, 845-851.
338. K. Sano, S. Yoshii, I. Yamashita and K. Shiba, *Nano Lett.*, 2007, **7**, 3200-3202.
339. T. Hayashi, K. Sano, K. Shiba, Y. Kumashiro, K. Iwahori, I. Yamashita and M. Hara, *Nano Lett.*, 2006, **6**, 515-519.
340. T. Hashimoto, K. Gamo, M. Fukuta, B. Zheng, N. Zettsu, I. Yamashita, Y. Uraoka and H. Watanabe, *Appl. Phys. Lett.*, 2011, **99**, 263701.
341. C. B. Murray, C. R. Kagan and M. G. Bawendi, *Annu. Rev. Mater. Sci.*, 2000, **30**, 545-610.
342. W. B. Rogers, W. M. Shih and V. N. Manoharan, *Nat. Rev. Mater.*, 2016, **1**, 16008.
343. R. J. Macfarlane, B. Lee, M. R. Jones, N. Harris, G. C. Schatz and C. A. Mirkin, *Science*, 2011, **334**, 204-208.
344. Y. Zhang, F. Lu, K. G. Yager, D. van der Lelie and O. Gang, *Nat. Nanotechnol.*, 2013, **8**, 865-872.
345. C. A. Mirkin, R. L. Letsinger, R. C. Mucic and J. J. Storhoff, *Nature*, 1996, **382**, 607-609.
346. E. Strable, J. E. Johnson and M. G. Finn, *Nano Lett.*, 2004, **4**, 1385-1389.
347. C. C. Broomell, H. Birkedal, C. L. P. Oliveira, J. S. Pedersen, J.-A. Gertenbach, M. Young and T. Douglas, *Soft Matter*, 2010, **6**, 3167.
348. S. Y. Park, A. K. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz and C. A. Mirkin, *Nature*, 2008, **451**, 553-556.
349. D. Nykypanchuk, M. M. Maye, D. van der Lelie and O. Gang, *Nature*, 2008, **451**, 549-552.
350. W. Liu, M. Tagawa, H. L. Xin, T. Wang, H. Emamy, H. Li, K. G. Yager, F. W. Starr, A. V. Tkachenko and O. Gang, *Science*, 2016, **351**, 582-586.
351. M. A. Kostiainen, O. Kasyutich, J. J. L. M. Cornelissen and R. J. M. Nolte, *Nat. Chem.*, 2010, **2**, 394-399.
352. M. A. Kostiainen, P. Ceci, M. Fornara, P. Hiekkataipale, O. Kasyutich, R. J. M. Nolte, J. J. L. M. Cornelissen, R. D. Desautels and J. van Lierop, *ACS Nano*, 2011, **5**, 6394-6402.
353. V. Liljestrom, J. Seitsonen and M. A. Kostiainen, *ACS Nano*, 2015, **9**, 11278-11285.
354. M. Kunzle, T. Eckert and T. Beck, *J. Am. Chem. Soc.*, 2016, **138**, 12731-12734.
355. T. Williams, *Neotrop. Entomol.*, 2008, **37**, 615-632.
356. M. Folwer and J. S. Robertson, *J. Invertebr. Pathol.*, 1972, **19**, 154-155.
357. S. B. Juhl, E. P. Chan, Y. H. Ha, M. Maldovan, J. Brunton, V. Ward, T. Dokland, J. Kalmakoff, B. Farmer, E. L. Thomas and R. A. Vaia, *Adv. Funct. Mater.*, 2006, **16**, 1086-1094.
358. A. Servid, P. Jordan, A. O'Neil, P. Prevelige and T. Douglas, *Biomacromolecules*, 2013, **14**, 2989-2995.
359. B. Apostolovic, M. Danial and H. A. Klok, *Chem. Soc. Rev.*, 2010, **39**, 3541-3575.

360. J. M. Fletcher, R. L. Harniman, F. R. Barnes, A. L. Boyle, A. Collins, J. Mantell, T. H. Sharp, M. Antognozzi, P. J. Booth, N. Linden, M. J. Miles, R. B. Sessions, P. Verkade and D. N. Woolfson, *Science*, 2013, **340**, 595-599.
361. D. N. Woolfson, G. J. Bartlett, M. Bruning and A. R. Thomson, *Curr. Opin. Struct. Biol.*, 2012, **22**, 432-441.
362. H. Gradisar, S. Bozic, T. Doles, D. Vengust, I. Hafner-Bratkovic, A. Mertelj, B. Webb, A. Sali, S. Klavzar and R. Jerala, *Nat. Chem. Biol.*, 2013, **9**, 362-366.
363. S. Kang, P. A. Suci, C. C. Broomell, K. Iwahori, M. Kobayashi, I. Yamashita, M. Young and T. Douglas, *Nano Lett.*, 2009, **9**, 2360-2366.
364. Y. J. Kang, M. Uchida, H.-H. Shin, T. Douglas and S. Kang, *Soft Matter*, 2011, **7**, 11078.
365. N. F. Steinmetz, E. Bock, R. P. Richter, J. P. Spatz, G. P. Lomonosoff and D. J. Evans, *Biomacromolecules*, 2008, **9**, 456-462.
366. N. F. Steinmetz, G. Calder, G. P. Lomonosoff and D. J. Evans, *Langmuir* 2006, **22**, 10032-10037.
367. B. D. Tiu, D. L. Kernan, S. B. Tiu, A. M. Wen, Y. Zheng, J. K. Pokorski, R. C. Advincula and N. F. Steinmetz, *Nanoscale*, 2017, **9**, 1580-1590.
368. P. A. Suci, M. T. Klem, F. T. Arce, T. Douglas and M. Young, *Langmuir* 2006, **22**, 8891-8896.
369. P. Suci, M. T. Klem, M. Young and T. Douglas, *Soft Matter* 2008, **4**, 2519-2523.
370. P. J. Yoo, K. T. Nam, A. M. Belchert and P. T. Hammond, *Nano Lett.*, 2008, **8**, 1081-1089.
371. P. J. Yoo, K. T. Nam, J. F. Qi, S. K. Lee, J. Park, A. M. Belcher and P. T. Hammond, *Nat. Mater.*, 2006, **5**, 234-240.
372. T. Li, R. E. Winans and B. Lee, *Langmuir*, 2011, **27**, 10929-10937.
373. V. Liljestrom, A. Ora, J. Hassinen, H. T. Rekola, Nonappa, M. Heilala, V. Hynninen, J. J. Joensuu, R. H. A. Ras, P. Torma, O. Ikkala and M. A. Kostianen, *Nat Commun*, 2017, **8**, 671.
374. T. Li, X. Zan, R. E. Winans, Q. Wang and B. Lee, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 6638-6642.
375. P. J. Yoo, K. T. Nam, J. F. Qi, S. K. Lee, J. Park, A. M. Belcher and P. T. Hammond, *Nat. Mater.*, 2006, **5**, 234-240.
376. W. J. Chung, J. W. Oh, K. Kwak, B. Y. Lee, J. Meyer, E. Wang, A. Hexemer and S. W. Lee, *Nature*, 2011, **478**, 364-368.
377. P. A. Sontz, J. B. Bailey, S. Ahn and F. A. Tezcan, *J. Am. Chem. Soc.*, 2015, **137**, 11598-11601.

TOC sentence

Assembly of protein cages across multiple length scales is described, with an emphasis on their use as functional biomaterials.

TOC image

