

Protein Assembly by Design

Jie Zhu,[‡] Nicole Avakyan,[‡] Albert A. Kakkis, Alexander M. Hoffnagle, Kenneth Han, Yiying Li, Zhiyin Zhang, Tae Su Choi, Youjeong Na, Chung-Jui Yu and F. Akif Tezcan^{*}

Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0340, United States

[‡] These authors contributed equally.

^{*} Corresponding author: tezcan@ucsd.edu

Abstract

Proteins are nature's primary building blocks for the construction of sophisticated molecular machines and dynamic materials, ranging from protein complexes such as photosystem II and nitrogenase that drive biogeochemical cycles to cytoskeletal assemblies and muscle fibers for motion. Such natural systems have inspired extensive efforts in the rational design of artificial protein assemblies in the last two decades. As molecular building blocks, proteins are highly complex, both in terms of their three-dimensional structures and chemical compositions. To enable control over the self-assembly of such complex molecules, scientists have devised many creative strategies by combining tools and principles of experimental and computational biophysics, supramolecular chemistry, inorganic chemistry, materials science, and polymer chemistry, among others. Owing to these innovative strategies, what started as a purely structure-building exercise two decades ago has, in short order, led to artificial protein assemblies with unprecedented structures and functions and protein-based materials with unusual properties. Our goal in the Review is to give an overview of this exciting and highly interdisciplinary area of research, first outlining the design strategies and tools that have been devised for controlling protein self-assembly, then describing the diverse structures of artificial protein assemblies, and finally highlighting the emergent properties and functions of these assemblies.

Table of contents

1. Introduction	4
2. Design principles and tools for protein assembly	8
2.1 Construction principles of natural protein assemblies	8
2.2 Tools and strategies for the design of artificial protein assemblies	10
2.2.1 <i>Symmetric protein fusion</i>	12
2.2.2 <i>Computational interface design</i>	13
2.2.3 <i>Metal coordination</i>	15
2.2.4 <i>Covalent bonding</i>	16
2.2.5 <i>Host-guest interactions</i>	17
2.2.6 <i>Electrostatic interactions</i>	18
2.2.7 <i>Hybridization with non-proteinaceous components</i>	19
3. Finite and extended protein assemblies	20
3.1 Dimers	20
3.1.1 <i>Dimerization by domain swapping</i>	20
3.1.2 <i>Dimerization through computational interface design</i>	23
3.1.3 <i>Dimerization through metal-coordination</i>	27
3.2 Small oligomers	30
3.2.1 <i>Oligomerization through computational interface design</i>	30
3.2.2 <i>Oligomerization through electrostatic interactions</i>	37
3.2.3 <i>Oligomerization by genetic fusion and domain swapping</i>	38
3.2.4 <i>Metal-mediated oligomerization</i>	40
3.3 Protein cages	45
3.3.1 <i>Protein cages through genetic fusion</i>	45
3.3.2 <i>Computationally designed protein cages</i>	50
3.3.3 <i>Metal-directed protein cages</i>	54
3.3.4 <i>Reengineering of natural cages</i>	57
3.4 Extended 1D assemblies	61
3.4.1 <i>1D assemblies mediated by host-guest and receptor-ligand interactions</i>	62
3.4.2 <i>1D assemblies through covalent bonding</i>	64
3.4.3 <i>Metal-directed 1D assemblies</i>	67
3.4.4 <i>Electrostatically directed 1D assemblies</i>	72
3.4.5 <i>Computationally designed 1D assemblies</i>	73
3.5 Extended 2D assemblies	75
3.5.1 <i>2D protein assemblies through genetic fusion</i>	76
3.5.2 <i>2D assemblies through covalent bonding</i>	77
3.5.3 <i>Metal-mediated 2D assemblies</i>	81
3.5.4 <i>2D assemblies mediated by protein-ligand interactions</i>	84

3.5.5	<i>Computationally designed 2D assemblies</i>	87
3.5.6	<i>2D protein assembly at interfaces</i>	89
3.6	Extended 3D protein assemblies	92
3.6.1	<i>3D protein assemblies through rational design of protein surfaces</i>	92
3.6.2	<i>Effector-mediated 3D protein assembly</i>	97
3.6.2.1	Metal-directed 3D assemblies	97
3.6.2.2	3D assemblies through receptor-ligand interactions	99
3.6.3	<i>Electrostatically directed 3D protein assembly</i>	100
3.6.3.1	3D assemblies mediated by small molecules	100
3.6.3.2	3D assemblies through interactions with nanoparticles	102
3.6.4	<i>3D protein-DNA hybrid lattices</i>	104
4.	Properties, Functions and Applications of Designed Protein Assemblies	107
4.1	Stability of protein assemblies	108
4.2	Stimuli-responsive and reconfigurable protein assemblies	110
4.2.1	<i>Stimuli-responsive protein assemblies</i>	111
4.2.2	<i>Reconfigurable protein assemblies</i>	115
4.3	Encapsulation, scaffolding and structural organization by designed protein assemblies	118
4.3.1	<i>Encapsulation</i>	119
4.3.2	<i>Scaffolding for applications in structural biology</i>	121
4.3.3	<i>Scaffolding of biological molecules</i>	123
4.3.4	<i>Scaffolding of inorganic and synthetic components</i>	126
4.4	Mechanical properties of designed protein assemblies	131
4.5	Biochemical functions of designed protein assemblies	135
4.5.1	<i>Binding and recognition</i>	135
4.5.2	<i>Membrane-related functions</i>	137
4.5.3	<i>Catalysis</i>	139
5.	Future directions and challenges	142
6.	List of abbreviations	144
7.	Author biographies	146
8.	Acknowledgments	148
9.	References	149

1. Introduction

Life likely emerged through the self-assembly and organization of a primordial mixture of ions, minerals and small, ångström-scale organic molecules.¹⁻⁴ Yet, the complexity of living systems we know them today largely rests on the existence of extended biological polymers such as polypeptides, polynucleotides and polysaccharides, in which molecular components are covalently linked into linear arrays in a modular, genetically encoded fashion.^{5,6} The stable, covalent pre-arrangement of their components allows biopolymers to arrange into elaborate nano- and microscale architectures, endowing them with the ability to act as recognition elements and store energy (polysaccharides),⁷⁻¹¹ maintain and transmit genetic information with high fidelity (polynucleotides),¹²⁻¹⁵ and record and transduce chemical, physical and mechanical information (polypeptides).¹⁶⁻²⁷

From the point of chemical versatility, polypeptides reign supreme among biopolymers, as they are composed of twenty distinct amino acid components. The compositional complexity of these linear polymers with twenty different building blocks creates an incalculable number of chemical interactions between the building blocks and the environment surrounding them.^{28,29} Despite this, polypeptides have the ability to fold into discrete and often singular architectures with nanometer dimensions, namely proteins.³⁰⁻³² Such a well-defined spatial organization of twenty functionalities in three-dimensional space enables proteins to interact with and act upon almost any other type of matter (organic or inorganic, biological or abiological) or external stimuli with high precision.¹⁶⁻²⁷ Importantly, proteins can also associate with copies of themselves or of other proteins in a specific fashion to form larger complexes and assemblies.³³⁻³⁷ Protein-protein interactions are not only critical for the high-fidelity transmission of chemical information in a cell, but also for the construction of large (several nm's to μ m's) protein assemblies that execute complex, multi-step biochemical processes or form structural materials that shape the cell and allow it to dynamically interact with the environment (**Figure 1**).³⁸⁻⁴⁹

The scaling of structural and functional diversity with the increasing hierarchical organization and dimensions of proteins is illustrated in **Figure 2**. While the *folding* of polypeptide chains into tertiary protein structures engenders the essential functions of an autonomous cell (e.g., catalysis, recognition and binding, signaling, electron transport, etc.),⁵⁰⁻⁵³ it is the *assembly* of proteins into supramolecular and extended structures that provides organisms with the necessary functional complexity.^{41,45,46,49,54,55} Inspired by these sophisticated machines and materials, there has been great interest in the control of protein self-assembly by design. As we aim to capture in this Review, this exciting new field has witnessed tremendous advances in just over a decade, progressing from the construction of simple protein dimers to the design of 3D protein crystals, megadalton-scale protein cages, *in vivo* active enzymes and stimuli-responsive materials with unprecedented properties and functions.

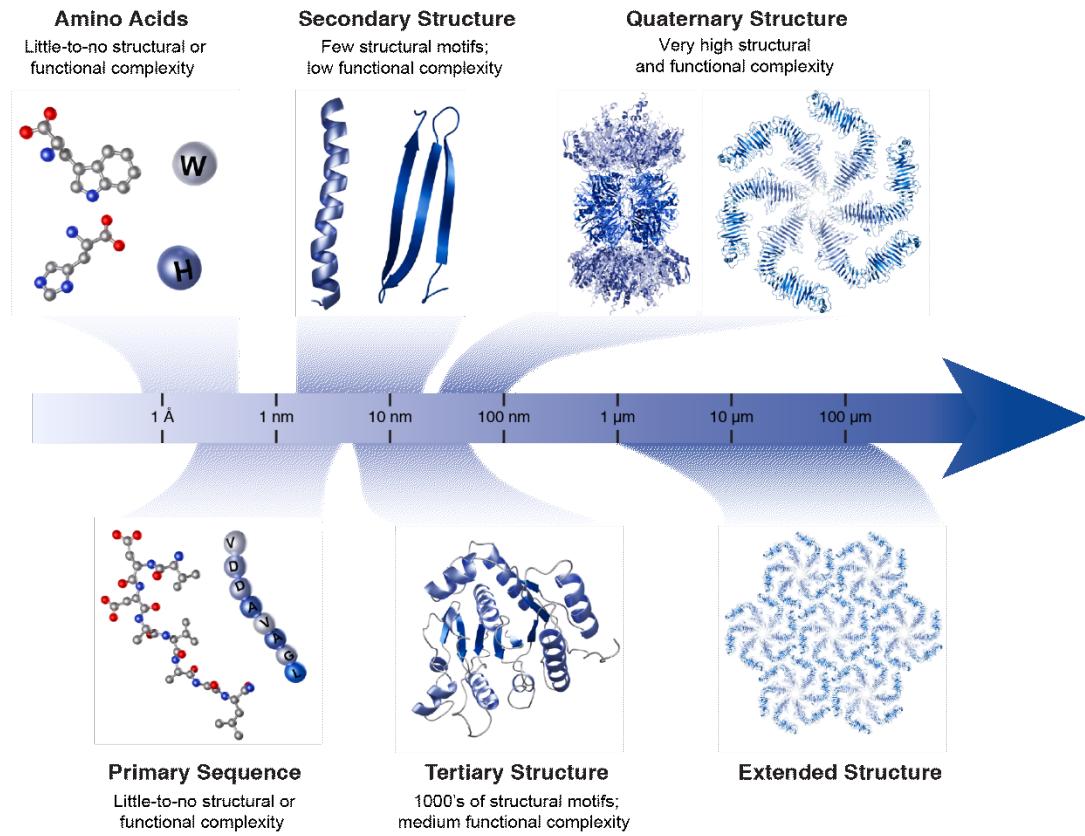


Figure 1. Complexification of protein assembly, from ångström-scale amino acids to extended, micron-scale protein structures. Individual amino acids are first covalently linked to form unstructured polymers, termed polypeptides. Polypeptides adopt secondary structure motifs, such as α -helices and β -strands, that combine to determine the tertiary structures of proteins. Discrete, folded polypeptides with either identical or distinct tertiary structures can assemble via non-covalent interactions into supramolecular complexes, termed quaternary structures. Further non-covalent interactions at the interfaces of symmetric quaternary structures can give rise to extended structures, exemplified here by a bacterial S-layer structure.

At a first glance, protein self-assembly and protein folding may appear quite similar to one another in that they are driven by the same intermolecular forces: van der Waals, hydrophobic, electrostatic, hydrogen bonding, metal coordination, disulfide, solvation/desolvation, solvent and configurational entropy.^{28,29} Thus, one may be tempted to think that the task of rationally designing or predicting protein self-assembly is also similar to designing and predicting protein folding. However, the two processes are quite dissimilar. Folding of a polypeptide into a tertiary structure is a self-contained, intramolecular process that is largely independent of environmental parameters (at or near ambient conditions) or concentration (at low volume fractions). The process is self-specific (there is little cross-talk with other species in solution), almost always proceeds under thermodynamic control, and is dominated by rather stringent steric/dihedral constraints of the polypeptide backbone.^{31,32,56} These constraints, coupled with appropriate amino acid sidechain interactions, enable most polypeptide sequences to spontaneously fold into singular 3D structures, as postulated by Anfinsen.^{57,58} Indeed, the sequence-folding patterns contained within the immense repository of experimentally determined

protein structures have been used to develop knowledge-based tools and deep-learning methods to predict 3D protein structures *a priori* from amino acid sequences with atomic-level accuracy.⁵⁹⁻⁶⁴ In parallel, it has also been possible to develop computational platforms based on the same empirical parameters to design novel protein folds from scratch.⁶⁵⁻⁷⁰

The challenge of designing protein assembly is an altogether different matter. Unlike folding, the self-assembly of a protein is both environment- and concentration-dependent, can be complicated by crosstalk with other species in solution, is not subject to any prescribed steric constraints and does not always operate under thermodynamic control, meaning that it can be pathway-dependent and lead to different structural outcomes under different environmental conditions.^{54,71-74} It is well-appreciated in nanoscience that most nanoparticles (such as proteins) that interact via strong, short-range attractive forces tend to form amorphous aggregations rather than ordered structures.⁷⁵ Indeed, as anyone who has dabbled in protein crystallization can attest to, the most probable outcome of protein self-assembly is heterogeneous aggregation. At the same time, it is also frequently observed that a single protein can crystallize in several different space group symmetries, featuring various protein-protein contacts that can be hard to rationalize in terms of their thermodynamic favorability even *a posteriori*. This is because the free-energy landscape for the self-assembly of any protein is multidimensional, shallow and marked by many energy minima whose magnitudes are readily altered by external perturbations.

While such a complex energy landscape makes the prediction and design of protein self-assembly difficult, this complication is not unique to proteins. It also applies to the self-assembly and crystallization of small molecules, inorganic complexes, nanoparticles, and even large colloids, which have been extensively investigated over the last several decades.⁷⁶⁻⁸⁰ Consequently, there exists a knowledge base and a good understanding of how self-assembly can be controlled across different length scales through the manipulation of the intrinsic properties of objects (e.g., shape, charge, size), through the design of chemical/physical interactions between them, and through the use of concepts such as symmetry and templating. Indeed, these concepts of molecular/nanoscale self-assembly and supramolecular chemistry have been combined with the tools of protein design and engineering in many creative ways, fueling the rapid progress of the protein self-assembly by design.

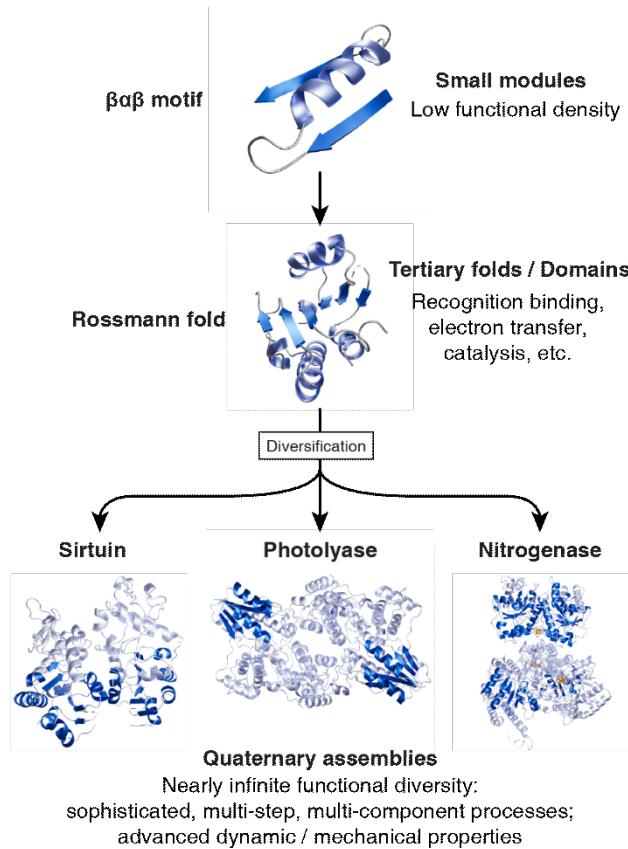


Figure 2. Protein evolution entails the incorporation of a small structural module into a larger, functional tertiary fold, which can then associate with other functional domains to form quaternary assemblies with diverse, multi-component functions. This process is exemplified here with the $\beta\alpha\beta$ motif-containing Rossmann fold, which is part of protein complexes with functions spanning histone deacetylation (sirtuin), DNA repair (photolyase), and dinitrogen reduction (nitrogenase). In all three quaternary assemblies, the Rossmann fold functions as the nucleotide-binding domain.

There have been excellent reviews that have covered the topic of protein self-assembly and highlighted its broad reach across protein design and engineering, chemical and structural biology, bio- and nanotechnology and materials science.^{49,81-101} Our primary goal is not only to update those reviews with the most recent examples from literature, but to provide a logical framework that we believe aptly describes the progression of the field from structure-building to property- and function-building. Although we acknowledge the key importance of disordered or heterogeneous protein ensembles, our focus here will be entirely on the design and construction of structurally well-defined, compositionally uniform protein assemblies, which lend themselves more readily to establishing design-structure-property-function relationships. This review will also not include the topic of peptide engineering and assembly, which is a diverse field in its own right and has been extensively covered in many reviews, including one in this issue.¹⁰²⁻¹⁰⁹ Section 2 will start with a summary of modes of protein self-assembly in natural systems, followed by a description of different tools and strategies that have been employed to design artificial protein architectures. Section 3 will focus on the design and construction of different classes of structures

using the tools described in Section 2. Pivoting from structure to property and function, Section 4 will focus on artificial protein assemblies with emergent physical, material and functional properties and dynamic behavior. We will conclude by giving an overview of the field of protein assembly by design and directions for future exploration.

2. Design principles and tools for protein assembly

2.1 Construction principles of natural protein assemblies

Natural evolution has created countless examples of functional protein assemblies that have inspired the design efforts covered in this Review. Before describing these examples, we will briefly summarize the general design parameters of natural protein assemblies, namely their shape/structure/dimensionality, symmetry, compositions, and connectivity, and how these relate to biological functions. We also note that by “protein assemblies” we refer exclusively to protein complexes that act as a unit and are sufficiently long-lived to be structurally characterized by conventional tools such as crystallography, electron microscopy, nuclear magnetic resonance spectroscopy (NMR), atomic force microscopy (AFM), small angle X-ray scattering (SAXS), etc.

Most generally, protein assemblies can be structurally characterized as closed/finite or open/extended.¹¹⁰ Finite protein assemblies (**Figure 3a**) are dimeric or oligomeric (i.e., consisting of three or more protein monomers) and physically bounded. Although the functions of finite protein assemblies are quite diverse, they are typically involved in the execution of biochemical processes such as signaling, catalysis and binding/recognition.¹¹⁰ Importantly, most of these biological functions derive explicitly from the formation of larger protein assemblies; in other words, they cannot be performed by the monomeric components in isolation. A classic example is hemoglobin (**Figure 3a-ii**), whose ability to bind O₂ with positive cooperativity and in an allosterically controllable fashion (by pH, CO₂ or 2,3-bisphosphoglycerate (BPG)) is wholly dependent on its tetrameric assembly state and central to its biological role as a regulable O₂ transporter.¹¹¹ In contrast, the monomeric myoglobin is incapable of cooperativity and allosteric control and acts in O₂ storage and delivery (rather than transport).¹¹² Another inherent benefit of protein dimerization or oligomerization is the stabilization of the protein subunits through the formation of intermolecular bonds and the reduction of exposed protein surfaces.¹¹⁰

Extended protein assemblies (**Figure 3b**) are polymeric (i.e., consisting of many monomers), usually possess crystalline order and are characterized by their dimensionalities, 1D, 2D or 3D, that span nm-to-μm length scales. Rather than performing biochemical tasks, extended protein assemblies serve as mechanical/architectural elements and in scaffolding/transport of other cellular components, commensurate with their dimensionalities. For example, 1D cytoskeletal filaments (e.g., actin (**Figure 3b-vi**), and microtubules)¹¹³⁻¹¹⁵ are involved in controlling cellular shape, movement and intracellular transport; 2D protein arrays (e.g., bacterial S-layers, **Figure 3b-vii**)^{116,117} act as protective layers or selective membranes; 3D protein lattices (e.g., crystals of insulin in pancreatic β-cells, cypovirus polyhedrin, **Figure 3b-viii**, peroxisomal alcohol oxidase)¹¹⁸⁻¹²¹ serve as stable platforms for storage, scaffolding and catalysis. A particular class of finite protein assemblies are 0D, cage-like architectures (e.g.,

ferritin (**Figure 3a-iv**), virus capsids).^{122,123} Although such assemblies are typically more “polymeric” rather than “oligomeric” in terms of component quantity and also primarily serve in scaffolding and encapsulation, they are physically bounded and thus categorized structurally as finite or closed.

Protein assemblies can also be classified according to whether they possess symmetry or not. Symmetry is a powerful design element that is ubiquitous in natural protein assemblies.¹¹⁰ It allows the formation of large protein complexes from a small number of building blocks while also minimizing the numbers of distinct contact types and associative surfaces between proteins.^{110,124} These advantages of symmetry are not only important for the natural evolution but also for the rational, bottom-up construction of protein assemblies, as they minimize the genetic and therefore the design burden. Consequently, the majority of natural protein assemblies are symmetric, with cyclic, dihedral or cubic point group symmetries in the case of closed architectures, helical or cyclic + translation symmetries in 1D structures, plane group symmetries in 2D assemblies and space group symmetries in 3D arrays.

Protein assemblies can be homomeric (i.e., composed of one type of protein monomer) (**Figure 3a-i, iii and iv**) or heteromeric (i.e., composed of two or more different types of protein monomers) (**Figure 3a-ii and v**). A large fraction of symmetrical protein assemblies is homomeric, while a much smaller fraction of symmetrical protein assemblies is heteromeric. The functional complexity of protein assemblies generally scales with their heteromeric composition, in that the different types of protein subunits within an assembly perform different functions. For example, photosystem II (PSII, itself a C_2 symmetric heteromer, **Figure 3a-v**) consists of ~20 different protein subunits,¹²⁵ enabling PSII to perform many coupled tasks (light-harvesting, long-distance energy and electron transfer, charge separation, water oxidation, generation of proton-motive force), which would be impossible to accomplish with a smaller or a homomeric protein assembly.

The key determinant of the structures, properties, and functions of protein assemblies are the connections between protein monomers. Along with the protein subunits themselves, protein-protein interfaces dictate the geometry/shape of an assembly, its rigidity or flexibility, if it can associate/dissociate or change its structure in response to a stimulus, or whether it contains a functional cofactor or an active site. Protein-protein interfaces (**Figure 3c**) in natural assemblies are mediated by the same non-covalent and covalent interactions that stabilize tertiary folds (e.g., electrostatic/polar, disulfide bonds (**Figure 3c-ix**), metal coordination (**Figure 3c-x**), hydrophobic (**Figure 3c-xi**), solvation/configurational entropy), but can vary widely in shape, composition and size.¹²⁶ Although most interfaces tend to be flat (**Figure 3c-xi**), they can also have quite irregular, non-uniform shapes (**Figure 3c-ix**). Hydrophobic hot spots or extended patches are common features of stable protein-protein interfaces. Most protein-protein interfaces have buried surface areas of 1000-2000 Å², yet some can be smaller (e.g., ~600 Å² in the Zn-mediated dimer interface of Rad50), or much larger, with many dimeric interfaces burying more than 5000 Å².³⁷

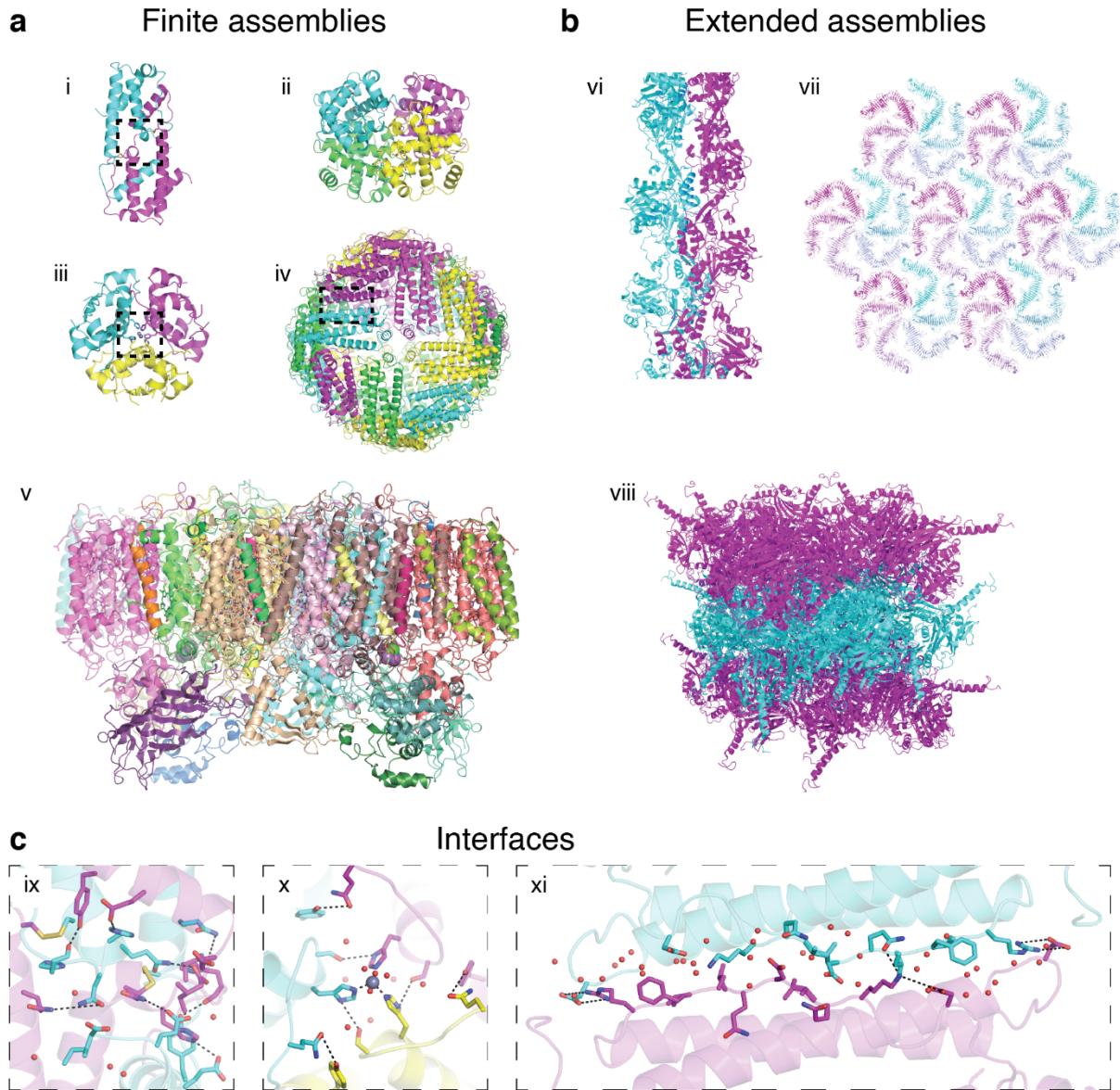


Figure 3. Finite and extended natural protein assemblies. **a)** Examples of finite protein assemblies: Interleukin-5 (i, PDB ID: 1HUL), hemoglobin (ii, PDB ID: 1HHO), insulin (iii, PDB ID: 1ZNI), human heavy chain ferritin (iv, PDB ID: 6B8F), and photosystem II (v, PDB ID: 1AXT). **b)** Examples of extended protein assemblies: 1D actin filament (vi, PDB ID: 6BNO), 2D S-layer (vii, PDB ID: 5N8P) and 3D infectious cytoplasmic polyhedrosis virus protein crystal (viii, PDB ID: 2OH5). **c)** Close-up views of protein-protein interfaces of interleukin-5 (ix), insulin (x), and ferritin (xi) highlighting non-covalent, covalent, and metal-mediated interactions. In addition to these enthalpic contributions, the expulsion of waters upon the burial of interfacial amino acid residues (i.e., the “hydrophobic effect”) represents an important entropic contribution to interface stability.

2.2 Tools and strategies for the design of artificial protein assemblies

In a traditional engineering-based approach, a design task ideally starts with the questions “what function should the designed object serve? what properties should it have?”,

followed by “what form should the object possess to fulfill the desired function or properties” and “what are the available building blocks and how should they be put together”? A look at the chemical and structural sophistication of natural protein assemblies (**Figure 3**) quickly reveals the challenges in posing the design questions in this order. First, the molecular interfaces that connect the protein subunits in natural assemblies are often too extensive and heterogeneous to be routinely designed from scratch. Second, we have a limited understanding of structure-property or structure-function relationships in proteins, though we have certainly learned that a static 3D picture of a protein or a protein assembly (even at atomic resolution) is hardly sufficient for predicting its properties and functions with great accuracy. Given these two ability/knowledge gaps, the majority of efforts in the area of designing protein assemblies have been directed at obtaining target structures or shapes (a great challenge in its own right) through the development of bottom-up construction strategies. Importantly, these strategies have not only yielded numerous examples of novel protein architectures (Section 3), but also lent themselves well to generating and discovering new functions and properties (Section 4). The latter point emphasizes the value of structure-building tools regardless of functional intent.

In this section, we will provide an overview of the different strategies that have been developed over the last two decades to control protein self-assembly and to construct supramolecular or extended protein architectures (**Figure 4**). Further details on these strategies will be provided when discussing specific examples in Section 3. There are two take-away points: 1) There has been tremendous progress in the rational and computational design of protein structures and protein-protein interfaces. 2) Rational design of protein assembly in the laboratory is not limited by the biochemical constraints of the cellular environment and an adherence to the construction strategies that nature uses. Thus, scientists have been able to adapt tools and materials from various disciplines (supramolecular and inorganic chemistry, reticular chemistry, inorganic and DNA nanotechnology, and polymer chemistry) to devise many innovative design approaches for protein assembly.

Regardless of the approach used, the concept of symmetry has featured prominently in the design of artificial protein assemblies, from protein cages to 1D, 2D and 3D protein crystalline arrays.¹²⁷⁻¹³⁰ Recently, Laniado and Yeates compiled a rule set for designing symmetric protein assemblies from oligomeric components and identified 124 distinct types of symmetry-combination materials.¹³¹ The readers are referred to this article and the many excellent reviews that articulate the importance of symmetry considerations in the design of ordered protein assemblies^{45,83,87,132-134} as well as in the synthesis of supramolecular and reticular materials.¹³⁵⁻¹³⁹

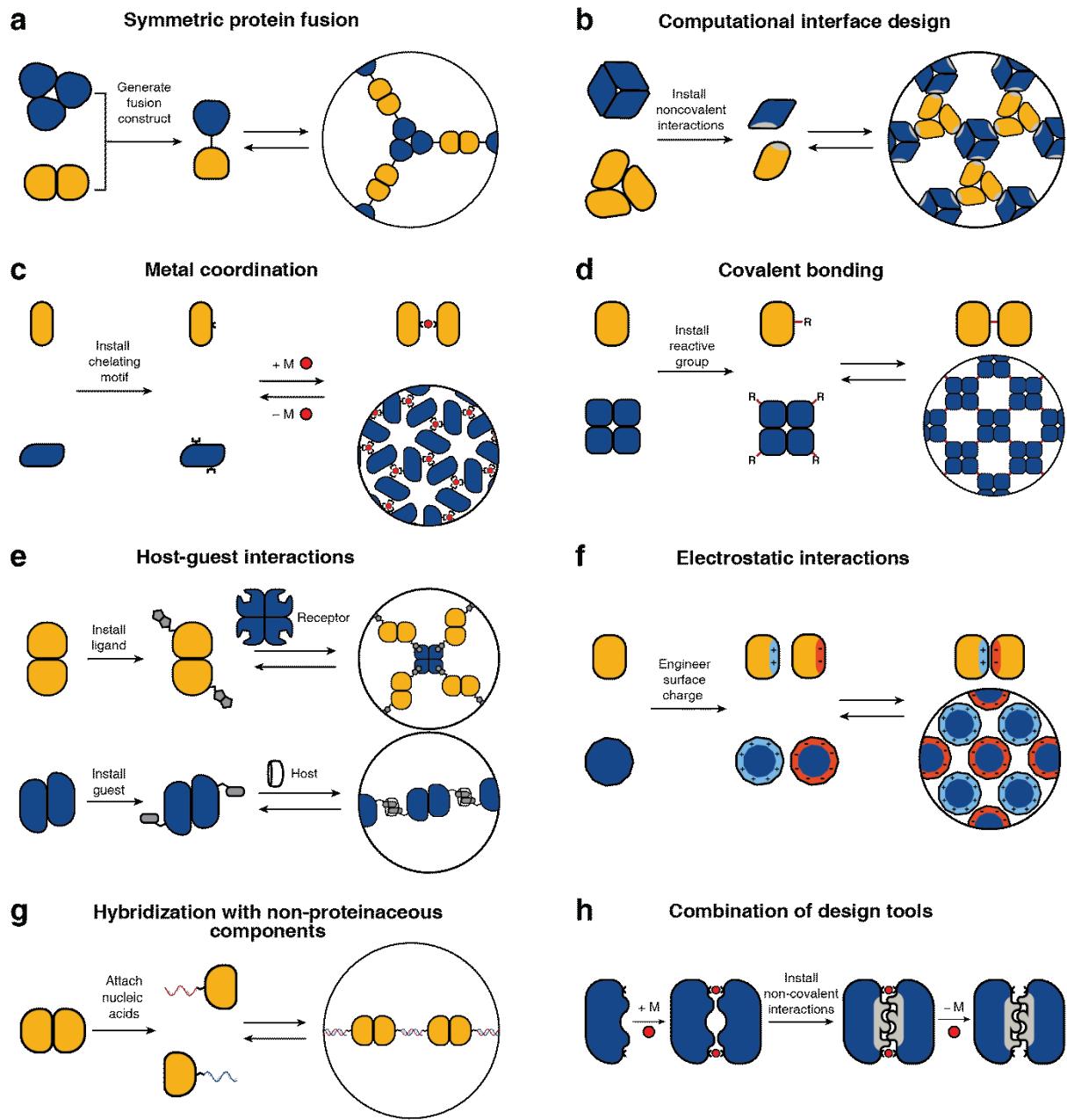


Figure 4. Schematic representation of the design tools and strategies for protein self-assembly, which encompass non-covalent, covalent, and metal-mediated interactions. Each strategy has been successfully used to construct both finite and extended assemblies with a wide array of protein building blocks. Selection of the design strategy is largely dictated by the desired stability, selectivity, and reversibility of the target assembly.

2.2.1 Symmetric protein fusion

Most approaches for constructing protein assemblies directly involve the design or modification of a protein-protein interface (*vide infra*). One of the earliest studies, wherein the rational design of protein assemblies was explicitly articulated, used an alternative approach, which we refer to as symmetric protein fusion. This method was introduced by the Yeates Group

and exploits the natural, self-associative interfaces of oligomeric proteins to drive self-assembly. It is based on the principle that most symmetric objects and extended materials can be generated from the proper combination of two symmetry elements, as summarized by Padilla, Laniado and Yeates,^{127,131} and earlier in the supramolecular chemistry literature.¹³⁵⁻¹³⁹ In the symmetric protein fusion method, this combination is achieved by genetically fusing the monomeric subunits of natively oligomeric proteins that possess appropriate symmetries (**Figure 4a**). The proper orientation of subunits with respect to one another is central to obtaining the desired assembly geometry and requires a careful integration of a rigid peptide linker domain.

In the first demonstration of this strategy, Padilla et al. fused the monomeric components of C_2 and C_3 symmetric protein oligomers with an α -helical linker. This generated a construct that assembled into a tetrahedral cage, enabled by the fusion-enforced orientation of the resulting C_2 and C_3 symmetric interfaces at the desired angle ($\sim 54.7^\circ$).¹²⁷ A second fusion constructed from the subunits of two C_2 symmetric dimers (with non-intersecting symmetry axes) led to the formation of 1D protein filaments, further providing proof-of-principle.¹²⁷ Using a similar strategy, Sinclair et al. combined protein building blocks with matching rotational symmetries to construct 1D and 2D crystalline arrays.¹²⁸ Linker length and rigidity played central roles in determining whether or not fusion constructs would form assemblies with long range order.¹²⁸ To make genetic protein fusion a more versatile tool for self-assembly, an increasing diversity of domains have been used as building blocks, including *de novo* designed coiled coils.¹⁴⁰⁻¹⁴² Cannon et al. recently assembled an icosahedral cage using a doubly fused, three-component construct comprised of a dimer, pentamer, and trimer.¹⁴³

An inherent challenge of the fusion approach is the linker design and the requirement to choose natively oligomeric protein building blocks with appropriate symmetries/topologies. To enable geometrically specific assembly, the linker must be of optimal length/rigidity and positioned accurately while the stability of the monomers is maintained. The high association constants of the native protein-protein interfaces may also increase the likelihood of kinetically trapped aggregates. The use of externally tunable (i.e., non-obligate) protein-protein interfaces along with algorithms that accurately model linker lengths/placement could greatly increase the scope of the symmetric protein fusion approach.

2.2.2 Computational interface design

The computational design of self-assembling supramolecular or extended protein architectures involves the modeling of the geometric arrangement of multiple protein units to form a desired architecture, followed by the design of energetically favorable protein-protein interfaces to stabilize that architecture (**Figure 4b**). All computational protein design calculations contain an energy function with which to evaluate a protein structure.^{144,145} Energy functions used in protein modeling can be molecular mechanics-based functions similar to those used in molecular dynamics simulations,¹⁴⁶⁻¹⁵¹ or they can be statistical functions derived from rotamer configurations and sequence patterns obtained from a database of protein structures.¹⁵²⁻¹⁵⁴ The former is often too computationally expensive to reproducibly and accurately evaluate in protein design calculations, whereas the latter cannot effectively model unfamiliar structures that do not frequently appear in the structural database.¹⁴⁴ As a result, energy functions used in protein

design frequently use a combination of physical terms, such as electrostatic and van der Waals' interactions, and statistical terms, such as torsion angle probabilities.^{155,156}

In principle, an accurate energy function, combined with a method for traversing the energy landscape, is sufficient to predict a protein's folded structure given its sequence. However, the protein design problem is in some ways the inverse of the protein folding problem.¹⁵⁷ Rather than attempting to predict the lowest energy structure of a protein given its sequence, the aim of protein design is to predict a sequence that will stabilize a desired structure. Complicating matters is the fact that many amino acid side chains can adopt multiple conformations, or rotamers, and therefore the number of rotamers that must be sampled is much larger than just the number of natural amino acids. While most protein design calculations use discrete rotamer libraries that help limit the search space,¹⁵⁸⁻¹⁶⁴ even the smallest rotamer libraries are impossible to exhaustively sample except in the case of calculations involving only a few positions. There are several ways to address this problem.¹⁴⁵ Most commonly, Monte Carlo methods are employed to randomly sample the design space and attempt to converge on a favorable sequence.¹⁶⁵⁻¹⁷⁰ Because of the inherent stochasticity of these searches, multiple iterations are frequently required to obtain the best sequence or sequences. Alternatives to Monte Carlo searches include dead-end elimination,^{171,172} in which physically incompatible rotamers are first identified and excluded and an exhaustive search is then performed with the remaining rotamers, and mean-field calculations,^{173,174} in which sequences are evaluated by considering the average positions of all possible rotamers of an amino acid. Dead-end elimination methods have been successfully used for small proteins,^{66,156,175,176} but are usually too computationally expensive for larger proteins, whereas mean-field calculations perform well with hydrophobic core residues but are less effective with surface residues.^{145,169}

A protein's sequence and tertiary structure can be designed with a suitable energy function and a method to explore the search space of possible rotamers. However, the design of supramolecular protein assemblies also requires the optimization of the rigid body orientations of the protein subunits. This can be accomplished by protein docking calculations, which attempt to predict the quaternary structure of two or more protein domains.¹⁷⁷ Docking calculations typically start with a global search of the degrees of freedom of the system using simplified representations of the proteins, such as by excluding amino acid side chains. This low-resolution docking step is then followed by an all-atom, high resolution docking stage with finer perturbations.¹⁷⁷ As with protein sequence design, protein docking calculations can be performed with a variety of energy functions and approaches to sample the degrees of freedom of the system. Notable examples include fast Fourier transform (FFT) docking,¹⁷⁸ which largely assesses surface complementarity, docking with Monte Carlo methods,¹⁷⁹ and docking based on biochemical interaction data.¹⁸⁰ In addition, several servers that integrate multiple docking calculations to predict multimeric protein structures have been developed.^{181,182}

As discussed above, multiple tools have been developed for each part of the computational protein design process, and different calculations have been combined to design protein oligomers.¹⁷⁶ However, the most frequently used software for the computational design of protein assemblies is currently Rosetta¹⁸³ which has been used to construct dimers,¹⁸⁴ small oligomers,¹⁸⁵ protein cages,¹⁸⁶ 1D helical filaments,¹⁸⁷ and 2D protein arrays,¹⁸⁸ which will be discussed in more detail in Section 3.

As evidenced by this large array of artificial protein structures, computational design has emerged as a powerful tool for designing and optimizing non-covalent protein-protein interactions. However, this approach still requires a considerable amount of trial-and-error and is inherently geared toward finding deep-energy minima primarily through hydrophobic/packing interactions, which is suitable for optimizing association energies and specificities, but not for dynamics. It is important to note that there has been steady progress in the computational design of polar/H-bonding interactions.^{189,190} With additional improvements in scoring functions and protocols to effectively model polar interactions, protein solvation and metal coordination, it will increasingly be possible to computationally design protein assemblies not only with desired structures but also with complex dynamic behavior and functions.

2.2.3 Metal coordination

Although natural protein-protein interfaces are primarily formed through non-covalent interactions, it is estimated that 5-10% of oligomeric proteins contain interfacial metal ions or metallocofactors.¹⁹¹ Such naturally occurring interfacial metal centers drive protein self-assembly or stabilize quaternary structures, mediate transient protein-protein interactions, and act as catalytic centers.¹⁹² From a structure-building perspective, metal coordination bonds are highly appealing, as they are considerably stronger than non-covalent interactions but reversible, enabling the formation of protein-protein interfaces on a small design footprint and under thermodynamic control.¹⁹³ These criteria are satisfactorily met by mid-to-late first-row transition metal ions (Mn²⁺ to Zn²⁺), which are labile for ligand substitution, yet can form thermodynamically stable complexes. Metal-ligand bonds are highly directional, meaning that the stereochemical preferences of metal ions can dictate the symmetry and structures of protein assemblies.¹⁹⁴ Furthermore, metals have inherent reactivities and their coordination bonds are inherently environment-sensitive (e.g., to solution pH, redox potential or the presence of extrinsic chelators), thus providing a facile means to control the thermodynamics and kinetics of protein self-assembly and to construct stimuli-responsive or reactive protein assemblies.^{195,196} Finally, metal-mediated protein-protein interactions can be designed using a variety of natural amino acids (predominantly, histidine (His), aspartic acid (Asp), glutamic acid (Glu), cysteine (Cys)) as well as synthetic, non-biological ligands (e.g., phenanthroline, hydroxyquinoline, bi- or terpyridine) to access diverse modalities of protein self-assembly (**Figure 4c**).^{197,198}

The advantages of metal coordination have been widely exploited in molecular self-assembly,¹³⁵⁻¹³⁹ and were first adapted explicitly for the design of supramolecular protein assemblies in 2007 by the Tezcan Group.¹⁹⁹ An obvious challenge in using metal coordination for controlling protein-protein interactions is the potential lack of selectivity that stems from the fact that the surface of any protein is replete with metal-binding amino acid sidechains. To circumvent this challenge, two pairs of metal-chelating, *i/i+4* bis-His motifs were incorporated on the surface of cytochrome *cb*₅₆₂ (cyt *cb*₅₆₂), a monomeric, four-helix-bundle protein. Despite its minimal design footprint, the resulting construct (termed MBPC1) formed a discrete *D*₂ symmetric tetramer upon binding four Zn²⁺ ions in a tetrahedral coordination geometry.¹⁹⁹ Notably, the same building block self-assembled into two other discrete oligomers upon binding Ni²⁺ (a C₃-symmetric trimer) and Cu²⁺ (a C₂-symmetric dimer), as dictated by the stereochemical preferences of the metal ions¹⁹⁴, thus showing the unique versatility of metal coordination in

directing protein-protein interactions. Building up on these proof-of-principle experiments, the approach of Metal-Directed Protein Self-Assembly (MDPSA) has been broadly adopted to construct protein assemblies with a variety of structures and dimensionalities,^{129,200-202} as well as dynamic,²⁰²⁻²⁰⁴ stimuli-responsive^{195,204} and functional architectures,²⁰⁵⁻²⁰⁷ using both natural^{199,202} and non-natural metal-binding motifs^{208,209} installed on protein surfaces. These examples are discussed in detail in Sections 3 and 4.

Despite these advantages, the lack of selectivity in metal-protein interactions still represents an important hurdle in MDPSA, often requiring trial-and-error as part of the design process. This selectivity issue can be particularly exacerbated in the crowded and heterogeneous cellular environment (i.e., for *in vivo* applications) wherein metal concentrations are tightly controlled. These challenges can in part be ameliorated by combining the design of metal coordination on protein surfaces with the computational design of protein-protein interfaces. This combined approach has indeed been shown to yield metalloprotein oligomers that efficiently form in bacterial cells by selective binding of metal ions and catalyze enzymatic reactions *in vivo*.^{205,210} Another area of improvement for MDPSA is a better understanding of the energetics of metal-protein interactions. Compared to the relatively well-described hydrophobic packing interactions and, increasingly, polar interactions,²¹¹ there are still no accurate energy or scoring functions for the computational modeling of metal coordination in proteins,²¹² even for closed-shell metal ions such as Zn²⁺.²¹³ Once such energy/scoring functions are established, MDPSA and computational design may provide a particularly powerful combination in the design of functional protein assemblies.

2.2.4 Covalent bonding

As in the case of metal coordination, covalent bonding offers a powerful tool for structure building, primarily in the form of disulfide bonds (**Figure 4d**). With bond dissociation energies of up to 60 kcal/mol,²¹⁴ disulfide bonds are more than an order of magnitude stronger than typical non-covalent bonds, yet they are reversible through a two-electron redox equilibrium. This allows disulfides to stabilize protein structures on very small interaction footprints, while also enabling responsiveness to external stimuli (solution pH, redox potential/state) and enzymatic regulation.²¹⁵ Although commonly found in cytosolic proteins for intrachain crosslinking of tertiary folds, disulfide bonds also frequently serve to stabilize quaternary or extended protein structures through interchain bonding (e.g., in antibodies and extracellular proteins).²¹⁶⁻²¹⁹

In analogy to metal coordination, the combination of high stability, reversibility, and stimuli-responsiveness of the disulfide bond has made it a popular tool in the self-assembly of synthetic supramolecular complexes and generation of dynamic covalent libraries.²²⁰⁻²²² In terms of protein self-assembly, an important feature of disulfide bonding is its self-selectivity: a Cys sidechain will only couple with another Cys sidechain. Indeed, this can be exploited to readily generate disulfide-bonded protein homodimers. In one case, Banatao et al. prepared homodimers of three different single-Cys variants of lysozyme and demonstrated that these variants formed 3D crystals with morphologies inaccessible with the monomeric protein.²²³

Unlike homodimerization, however, the self-assembly of most oligomeric or extended protein architectures would necessitate the formation of more than a single disulfide bond, in turn

requiring the building blocks to possess more than one surface Cys residue. In such cases, it is imperative that the multiple Cys residues be placed in precise positions to pair correctly and yield the desired assembly structures. This is relatively easily accomplished with toroidal building blocks, such as the homohexameric Hemolysin-coregulated protein 1 (Hcp1), whose top and bottom faces can be tailored with symmetry-related Cys residues to yield 1D nanotubes continuously linked by inter-hexamer disulfide bonds (see Section 3.4.2).²²⁴ Alternatively, building blocks that possess 2D or 3D symmetry, such as the C_4 symmetric L-rhamnulose-1-phosphate aldolase (RhuA) and the octahedral ferritin, can be engineered in their vertices with Cys residues to self-assemble into disulfide-mediated 2D lattices (see Sections 3.5.2).²²⁵⁻²²⁸ Particularly, the 2D C^{98} RhuA lattices, described in more detail in Section 3.5.2, illustrate the key importance of the reversibility of disulfide bonds in the formation of defect-free lattices, while also highlighting how the flexibility of these bonds can give rise to coherent lattice dynamics (Section 4.2.2).²²⁵⁻²²⁷ Other types of covalent linkages have also been used to drive protein self-assembly, such as native chemical ligation,²²⁹ Tyr dimerization,²³⁰ and sulfo-NHS/EDC coupling,²³¹ however, the irreversibility of these linkages generally translates into a lack of order and structural homogeneity.

It is apparent that disulfide bonds provide unique advantages toward designing protein assemblies, but they also suffer from a lack of selectivity and a reliance on accurate geometric alignment. For example, in the case of non-symmetric protein building blocks, ensuring the geometrically specific formation of multiple disulfide bonds during protein self-assembly would require additional design elements. At least in one case (cyt cb_{562}), such specificity was achieved through the simultaneous incorporation of metal-coordination and computationally designed non-covalent interactions, which yielded the formation of tetrameric architectures containing up to six interfacial disulfide bonds.^{210,232,233} Further specificity could potentially be attained through the implementation of abiological covalent bonds such as boronate esters, imines, and triazines,²³⁴⁻²³⁷ particularly if the relative instabilities of these bonds in aqueous media can be ameliorated.

2.2.5 Host-guest interactions

As amply demonstrated in natural systems and in supramolecular chemistry,²³⁸⁻²⁴¹ “host-guest” interactions can be readily applied to drive protein self-assembly in a modular fashion. Generally, a “host” is defined as a large molecule or a macromolecule (even a protein) that specifically recognizes a smaller “guest” molecule through a synergy of non-covalent interactions.²⁴²⁻²⁴⁴ The high specificity and reversibility of host-guest interactions allow for self-assembly with high fidelity and readily lend themselves to the generation of dynamic, switchable, and stimuli-responsive complexes.²⁴⁵⁻²⁴⁹

Depending on the nature of chemical components, the host-guest pairs can be either *biological* or *synthetic*. The most common biological host-guest pairs for protein assembly are streptavidin-biotin and cytochrome/myoglobin-heme pairs, featuring high-affinity interactions between the host protein and cognate cofactor.²⁵⁰⁻²⁵⁶ For example, Ringler and Schulz generated a quadratic network by mixing the C_4 -symmetric enzyme RhuA labeled with biotin with the D_2 -symmetric streptavidin.¹²⁸ Similarly, the Hayashi Lab exploited heme-apocytochrome/apomyoglobin interactions to build various supramolecular protein assemblies.^{250,251} Synthetic host-guest pairs used for protein self-assembly often comprise

macrocyclic hosts (e.g., cyclodextrins, cucurbiturils, calixarenes) and various small molecule guests, which are widely employed in the self-assembly of synthetic supramolecular complexes.^{257,258} The Brunsved Group developed synthetic host-guest pairs for protein assembly, obtaining protein heterodimers stabilized exclusively by molecular recognition between cyclodextrin and lithocholic acid.²⁵⁹ Follow-up studies have expanded this strategy to obtain both discrete and extended assemblies driven by selective recognition between synthetic macrocycles and small molecule ligands.²⁶⁰⁻²⁶⁸

Host-guest interactions represent a versatile tool to drive protein self-assembly that does not require extensive interface design and often incorporates non-proteinaceous building blocks, yielding multi-component assemblies with relative ease (**Figure 4e**). One potential drawback of using this tool is its dependence on chemical modification of the protein with the host/guest molecules. The linkers that connect host/guest molecules to the protein building blocks are often long and flexible, which may preclude the formation of structurally ordered protein assemblies. To address these challenges, future work in host-guest recognition driven protein self-assembly could involve more widespread utilization of host-guest interactions between a synthetic host and native amino acid residues of a protein building block (e.g., sulfocalixarene-lysine), which would eliminate the need for bioconjugation.^{257,266-269} A particular advantage of synthetic host-guest pairs is their modularity, which can potentially allow the interprotein complexation affinities and specificities to be tuned by synthetic modifications to the guest or the host.²⁷⁰

2.2.6 Electrostatic interactions

Electrostatic interactions are widely utilized in both natural and designed protein assemblies,²⁷¹⁻²⁷⁴ as they can be implemented both locally through salt bridges and globally between oppositely charged domains (**Figure 4f**). To drive assembly through *local* electrostatic interactions, researchers leverage amino acid residues with ionic side chains (i.e., Glu, Asp, arginine (Arg), and lysine (Lys)) to generate charge anisotropy across specific interfaces of the protein building block.²⁷⁵ To drive assembly through *global* electrostatic interactions, one can extend charge anisotropy across an entire protein surface or use charged protein spheres as macromolecular point charges.^{276,277} The Debye lengths of typical electrolyte solutions can extend well over several nm's (i.e., on the same length scale as most protein building blocks),²⁷² meaning that electrostatic interactions can exert a sizeable energetic influence on self-assembly thermodynamics and kinetics at long distances. In both a local and global context, the overall charge of the protein interface/surface (characterized by the isoelectric point) varies with pH, giving rise to assemblies whose affinity and association kinetics can be readily controlled by solution conditions.²⁷⁶ Researchers have generated protein assemblies of a wide structural variety, from discrete oligomers to superlattices, via local and global electrostatic interactions.^{272,276-281} These assemblies have incorporated multiple different protein building blocks and even non-proteinaceous building blocks like nanoparticles.^{272,277}

Because the free energy of a salt bridge interaction is relatively small in well-solvated environments (3-4 kcal/mol),²⁸² a multivalent display of oppositely charged residues or a high overall net charge may be required to exploit electrostatically driven self-assembly. In a recent example, Simon et al. found that positively and negatively “supercharged” variants of the asymmetric superfolder green fluorescent protein (sfGFP) assembled via electrostatic

interactions to form discrete particles with eight-fold symmetry (Section 3.2.2).²⁷⁶ In this context, highly symmetric protein building blocks are particularly appealing, as they require fewer surface mutations to obtain highly charged states and can more readily arrange into desired assembly geometries due to their symmetry.²⁷² These advantages are well illustrated by the electrostatically driven, binary 3D lattices of cowpea chlorotic mottle virus (CCMV) capsid²⁷⁹ and ferritin.²⁷⁷ In addition to proteins, researchers have utilized RNA, nanoparticles, and dendrimers as building blocks for electrostatically stabilized lattices.^{272,274,277,278,280} Electrostatic interactions represent an intuitive and powerful tool for the design and construction of tunable, multicomponent assemblies.

2.2.7 Hybridization with non-proteinaceous components

The design strategies for artificial protein assemblies discussed thus far rely on optimizing the geometric arrangement of protein building blocks and/or engineering the protein-protein interfaces between subunits. Such protein-centered design approaches can be complemented by the incorporation of non-proteinaceous components to build hybrid assemblies (**Figure 4g**). As evident in many natural protein assemblies and biomaterials (e.g., nucleoprotein assemblies like the ribosome and nucleosome),^{283,284} and extensively demonstrated in nanotechnology, polymer chemistry, and supramolecular chemistry, hybrid systems are often synergistic.^{285,286} The combination of two different types of building blocks gives rise to emergent properties that would not be accessible using just one type of building block.²⁸⁷⁻²⁸⁹ Importantly, hybridization of proteins with non-proteinaceous components (e.g., nucleic acids, inorganic nanoparticles, synthetic polymers)²⁹⁰⁻²⁹⁴ expands the structural and functional scope of the proteins, creating new-to-nature functions^{295,296} and enabling access to self-assembly modes beyond what is possible based on a polypeptide-only composition.

Non-proteinaceous building blocks can be biological (i.e., nucleic acids, sugar) or abiological (i.e., synthetic polymers, metal nanoparticles, and carbon nanotubes),^{272,277,278,280,290,295,297-302} and impart a high degree of tunability in the control of protein self-assembly. In this regard, the high specificity of Watson-Crick base-pairing interactions is quite powerful, as highlighted by the remarkable structural diversity and programmability of artificial DNA-based nanostructures.³⁰³⁻³⁰⁶ In one of the earliest examples of DNA-dependent protein self-assembly, the Finn Group prepared CMPV-DNA conjugates, which could be tunably arranged into 2D hexagonal arrays and extended 3D aggregates by varying the assembly temperature (Section 3.6.4).³⁰⁷ More recently, the Mirkin Group demonstrated, through several examples, exquisite control over the 3D lattice arrangements of proteins covalently conjugated to single-stranded DNA sequences (Section 3.6.4).^{290,308} Like DNA, synthetic polymers and inorganic nanoparticles offer considerable structural tunability in terms of protein self-assembly, while also imparting novel mechanical and functional properties. As described in section 4.3.4, the Kostianen Group harnessed electrostatic interactions between spherical or cylindrical protein building blocks (ferritin, CCMV, tobacco mosaic virus (TMV)) and inorganic nanoparticles or polymeric dendrimers to mediate the assembly of hybrid 1D wires and 3D lattices with new optical, magnetic and mechanical features.^{296,309} As detailed in section 4.4, Zhang et al. integrated ferritin crystals with hydrogel networks to render 3D crystalline hybrid materials that were capable of isotropic expansion/contraction and self-healing.²⁹⁵

Potential size mismatches between the protein and non-proteinaceous building blocks and the inherent flexibility of protein-DNA or protein-polymer linkages can be an obstacle to the formation of ordered assemblies. Similarly, site-specific conjugation of DNA or polymers to proteins is often laborious and low-yielding. Nevertheless, as in the case of host-guest-interaction-driven self-assembly, the hybridization of proteinaceous and non-proteinaceous components brings the important advantage of modularity and the possibility of creating inherently multifunctional materials without extensive interface design and modification.

3. Finite and extended protein assemblies

3.1 Dimers

Dimeric proteins are the simplest and by far the most abundant form of protein assemblies in nature.³¹⁰ The design of homo- and heterodimeric proteins is a stringent test of our understanding of the principles underlying protein-protein interactions. Importantly, the approaches developed for designing dimeric protein assemblies are relevant for higher-order protein assemblies, which inherently consist of binary protein interfaces. One widely used approach to designing dimeric assemblies, chemically induced dimerization, is exclusive to dimers and has been extensively covered in previous reviews.^{89,311,312} In this section, we will discuss the distinct design strategies for protein dimerization including domain swapping, computational design and metal-mediated assembly, which have also been applied to form protein oligomers and higher-order protein architectures.

3.1.1 Dimerization by domain swapping

Domain swapping is a mode of oligomerization found in natural proteins in which one protein domain is exchanged with the identical domain from a second copy of the same protein.^{313,314} In essence, the pre-existing intra-molecular interface between two domains within the monomeric protein is repurposed into a new inter-molecular interface between two proteins.³¹⁵ As such, domain swapping typically requires minimal interface redesign and has been exploited to design new protein assemblies.

Early cases of domain swapping in engineered proteins involved stabilizing domain-swapped states that were discovered serendipitously. For example, a six-residue deletion in a surface loop of *staphylococcal* nuclease was found to result in a solution-stable dimer.³¹⁶ A crystal structure revealed that the deletion strained the loop and forced an α -helix into an unfavorable extended conformation. This conformation was stabilized by a second copy of the protein, which, through domain swapping of the α -helix, allowed the helix to make native-like contacts with the second protein. Meanwhile, recombinant expression of the protein CD2 fused to glutathione S-transferase (GST) yielded a domain-swapped dimer of CD2.³¹⁷ While mutations intended to stabilize the dimeric form of CD2 had little effect, mutations that destabilized the monomeric form significantly increased the fraction of protein that assembled into the domain-swapped dimer in solution. Similarly, a single-point mutation to the IgG-binding domain of protein

L resulted in low yields of a domain-swapped dimer.³¹⁸ To obtain an obligate dimer of protein L, the interface of the domain-swapped dimer was computationally re-designed to incorporate three mutations, which led to a dimer dissociation constant that was comparable to that of many naturally occurring protein dimers.

Geiger and coworkers have explored the domain swapping behavior of human Cellular Retinol Binding Protein II (hCRBPII) in detail and used it to design protein switches. Initially, they found that mutation of Tyr60 to hydrophobic residues produced domain-swapped dimers, with L60 and I60 variants producing 80-100% dimer.³¹⁹ Crystal structures revealed that the orientations of Asn59 and residue 60 are reversed in the dimers compared to the monomers and that breaking a hydrogen bond between Tyr60 and Glu72 was key to promoting the domain-swapped state (**Figure 5a-b**). Because Asn59 partially occupies the hCRBPII ligand-binding site in the domain-swapped state, the authors then hypothesized that ligand binding could result in conformational changes in the domain-swapped state. They identified an additional residue, Thr51, where mutations to bulkier amino acids also resulted in domain-swapped dimers which, significantly, could be crystallized in both the apo and ligand-bound states (**Figure 5c**).³²⁰ Incorporation of a disulfide bond along the interface of the two domains led to two new apo and ligand-bound conformations, thus expanding the total number of available conformations. Finally, installing a His₂Cys metal-binding site along the domain interface resulted in a Zn-binding site with micromolar Zn affinity in the apo state, but a five-fold lower affinity in the ligand-bound state.

The Loh Group developed an approach to induce domain swapping termed “mutually exclusive folding”, which involves inserting a small protein domain, called the “lever”, into a surface loop of a second protein, called the “assembler” (**Figure 6**).³²¹ If the distance between the N- and C-termini of the lever is greater than the distance between the ends of the selected loop, then the lever in its folded state exerts strain on the assembler, disfavoring its monomeric state. While this was initially predicted to lead to unfolding of the “assembler”, insertion of ubiquitin (“lever”) into barnase (“assembler”) instead resulted in a mixture of oligomers, of which the purified dimers were structurally confirmed to assemble via domain swapping.³²¹ The same approach was applied with several other proteins as either the “lever” or “assembler”. Insertion of ubiquitin in to any one of four loops in ribose binding protein (RBP) led to domain swapping and partial re-activation of a de-activated RBP for ribose binding.³²² Replacement of ubiquitin with the protein FKBP, which transfers from a partially unfolded apo state to a folded state upon the binding of the ligand FK506.9, enabled FK506-dependent activation of RBP, as well as *staphylococcal* nuclease, upon insertion of FKBP into a loop.³²³

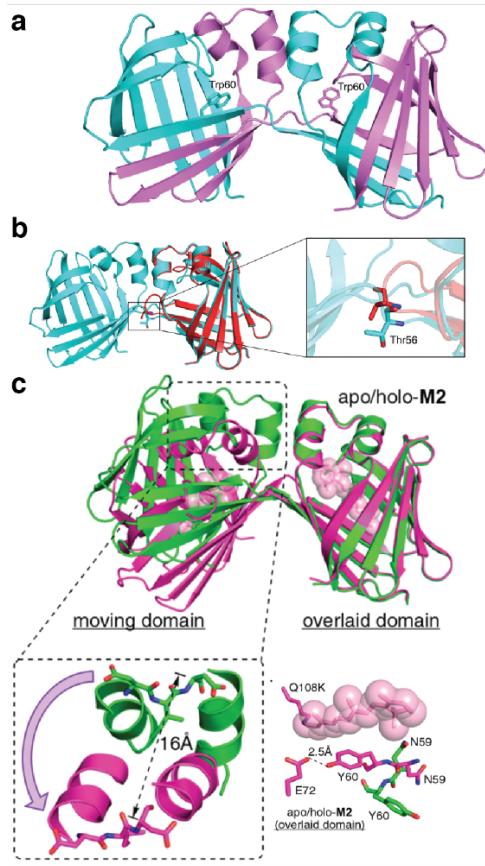


Figure 5. Conformational changes in domain-swapped hCRBPII. **a)** Structure of the W60 hCRBPII domain-swapped dimer. **b)** Overlay of the hCRBPII monomer (red) and the W60 hCRBPII domain-swapped dimer (cyan). **c)** Comparison of the apo (green) and holo (magenta) states of the D51 hCRBPII domain-swapped dimer. Binding of retinal results in significant conformational changes. **(a)** Adapted with permission from Ref. 319. Copyright 2016 Elsevier. **(b)** Adapted with permission from Ref. 319. Copyright 2016 Elsevier. **(c)** Adapted with permission from Ref. 320. Copyright 2019 ACS.

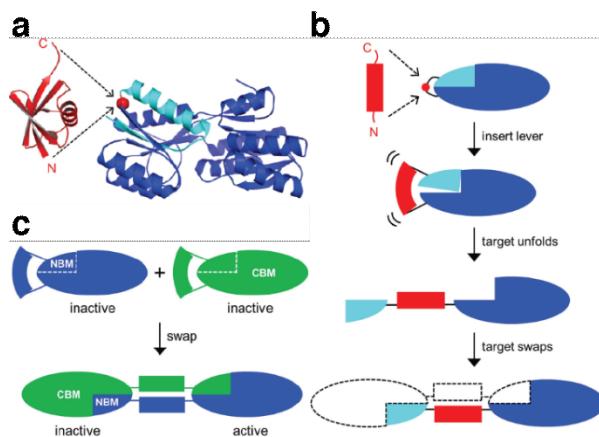


Figure 6. Domain swapping by mutually exclusive folding. **a)** Insertion of Ub into a loop in RBP. **b)** Schematic of domain swapping by mutually exclusive folding. **c)** Activation of protein function by domain swapping. Adapted with permission from Ref. 322. Copyright 2015 Elsevier.

In addition to the mutually exclusive folding strategy above,³²¹ a second domain-swap approach that has been applied to multiple proteins was reported by Gosavi and coworkers.³²⁴ This approach utilizes a hydrophobic QVVAG hinge loop that has been implicated in domain swapping of the natural protein stefin B (**Figure 7a**).³²⁴ The authors first created three variants of single-chain monellin, a protein which does not domain-swap, where each variant had a different loop replaced by the QVVAG motif. All three loops in question connected two β -strands. Of the three variants, two gave a mixture of monomers and dimers, while a third resulted in pure dimers. A crystal structure of the most successful variant revealed that it did indeed dimerize via domain swapping, as the inserted motif formed an extended β -strand instead of a loop (**Figure 7b**). The authors then created an additional variant of monellin in which two of the loops were replaced by the QVVAG motif and obtained a double domain-swapped dimer in high yield (**Figure 7c**). Finally, the motif was engineered into a loop in the proteins MK0293, Sso7d, and ubiquitin, which yielded solely dimers, dimers with a small fraction of monomers, and a mixture of oligomeric states, respectively. Thus, the QVVAG motif can be engineered into loops to induce domain swapping in a variety of proteins, though it does not always yield a single oligomeric state.

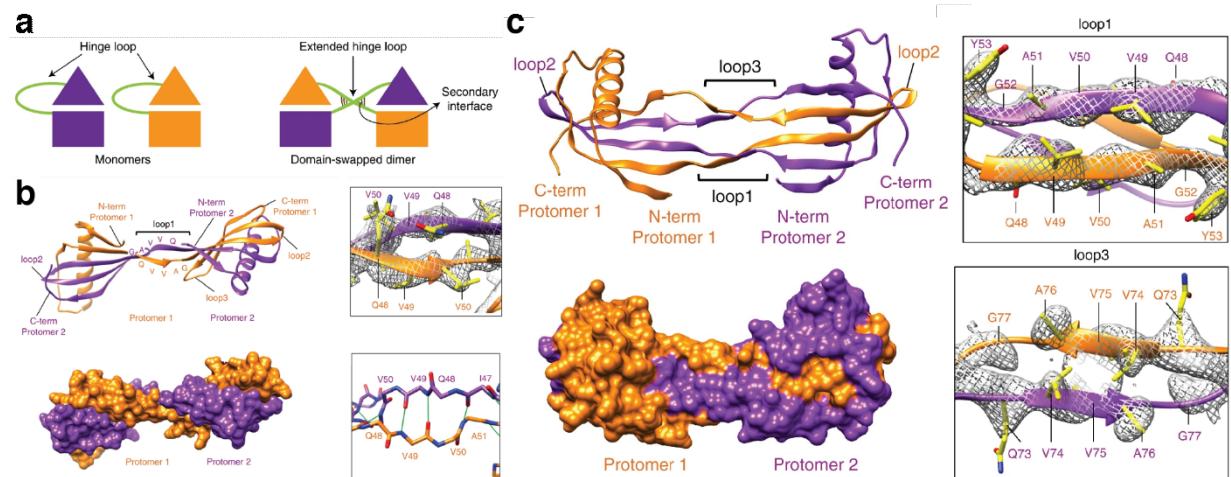


Figure 7. Domain swapping by insertion of a QVVAG motif. **a)** Cartoon depiction of domain swapping via the QVVAG hinge loop. **b)** Structure of the MNEI single domain-swapped dimer. **c)** Structure of the MNEI double domain-swapped dimer. Adapted with permission from Ref. ³²⁴. Copyright 2019 NPG.

3.1.2 Dimerization through computational interface design

One approach to designing a dimeric protein complex involves grafting a structural motif from one member of an existing protein-protein interface onto a new, evolutionarily unrelated protein. While the resulting protein-protein interface should mimic the pre-existing interface, the new protein scaffold can provide benefits over the natural protein scaffolds, such as increased stability. In general, motif grafting involves several steps (**Figure 8**). First, a starting protein scaffold is selected, which can be done by choosing a certain protein fold (e.g., a three-helix bundle)³²⁵ or by searching the Protein Data Bank (PDB) for proteins with similar backbone structures to the motif to be transplanted.³²⁶⁻³²⁹ Next, the desired motif is computationally inserted

into the selected scaffold. The insertion process can be carried out in a number of ways. Initial approaches involved grafting just the amino acid side chains of the structural motif onto regions of the scaffold protein with similar backbone structures to the initial motif.^{326,327} Later approaches enabled the entire structural motif, including both the side chains and the backbone conformations, to be transplanted into the scaffold, either by deleting segments of the target scaffold and replacing them with the motif^{328,329} or by folding simple protein topologies around the motif.³²⁵ Once the motif has been inserted into the new protein scaffold, residues around the motif are computationally designed to optimize the scaffold for the incorporation of the motif. Finally, the top-scoring designs are selected for experimental validation. Any variants that are able to bind the target can be further optimized by generating libraries and performing phage display or related high-throughput binding assays. Motif grafting has been used to design proteins capable of binding HIV³²⁶⁻³²⁹ and respiratory syncytial virus (RSV)³²⁵ monoclonal antibodies, as well as cancer-associated pro-survival proteins.^{330,331} The designed antibody-binding proteins recapitulate the native antibody-antigen interface and have the potential to lead to new vaccine candidates.

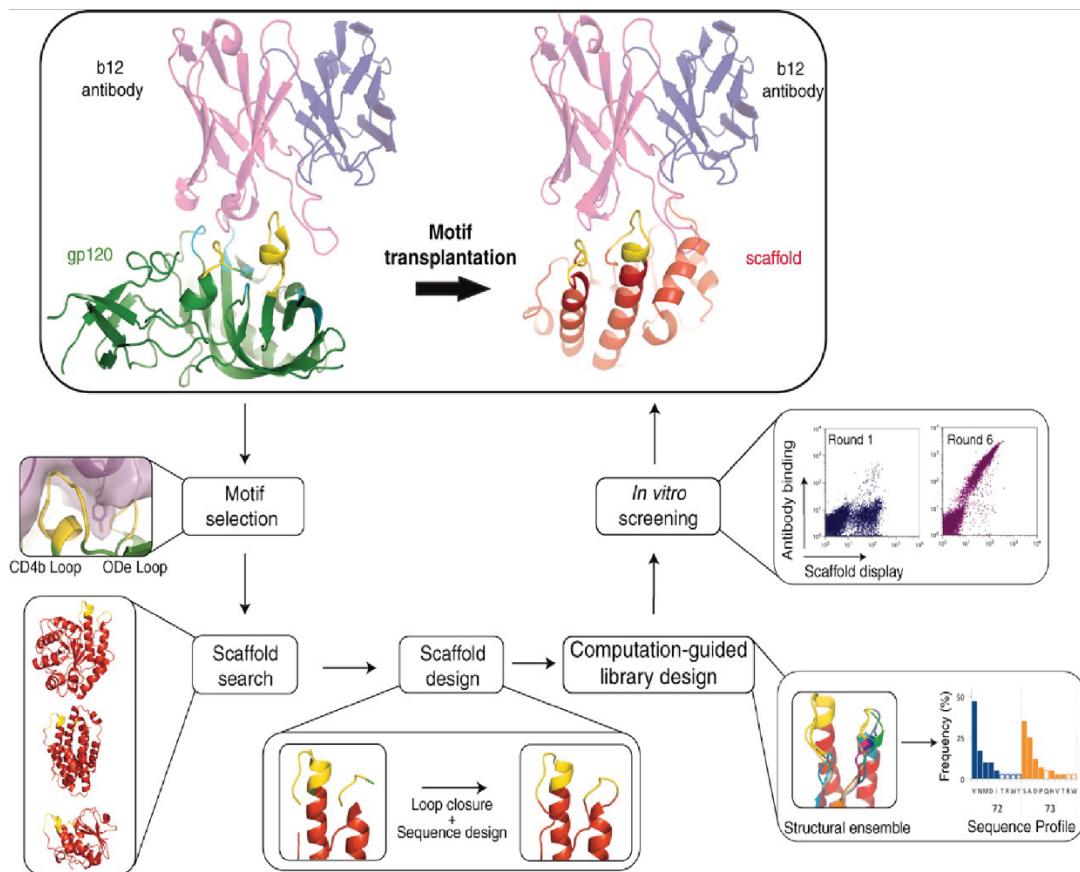


Figure 8. Computational and experimental steps for the transplantation of a structural motif to a new scaffold. Scaffolds can either be chosen by searching the PDB for existing structures or by designing a desired fold around the motif using *ab initio* folding calculations. Adapted with permission from Ref. ³²⁸. Copyright 2011 AAAS.

Motif grafting can be an effective approach in part because it relies on a pre-existing protein-protein interface. Designing protein-protein interactions from scratch presents a more difficult challenge. Several groups have demonstrated the redesign of monomeric proteins to generate solution-stable homodimers. Kuhlman and coworkers took advantage of β -strand pairing to redesign the γ -adaptin appendage domain into a symmetric homodimer via the formation of an intermolecular β -sheet.³³² Of the four computationally designed candidates that were experimentally tested, one expressed well and formed a homodimer in solution. The crystal structure of that dimeric variant closely matched the design model. Mayo and coworkers redesigned the engrailed homeodomain into a symmetric homodimer by designing an α -helical interface instead.¹⁷⁶ Following library generation and screening, they obtained a variant that formed a dimer in solution, and an NMR structure of the dimer aligned well with the design model.

In designing new protein-protein interfaces from scratch, researchers have frequently attempted to recapitulate general properties of natural interfaces. Most protein-protein interfaces feature some degree of shape complementarity between the constituent proteins' backbones, which contributes to the burial of large portions of solvent-exposed surface area along the interface. Therefore, protein docking calculations are often an important first step in designing new protein-protein interfaces. Once backbone-complementary orientations of the two proteins are found, residues along the nascent interface can be designed to stabilize the dimeric conformation. Several groups have used protein docking calculations followed by computational sequence design to generate heterodimeric complexes, though these have not been thoroughly structurally characterized.^{333,334}

More recently, Baker and coworkers have used computational docking and interface design calculations, followed by affinity maturation, to develop Ankyrin repeat proteins capable of binding Frizzled subtypes with high affinity and selectivity for certain subtypes over others.³³⁵ Key to the selectivity of the binding proteins was the computational docking (**Figure 9**), which allowed for the design of large interfaces that included both the highly conserved Frizzled lipid-binding site and regions with less conservation among Frizzled subtypes. Maly and coworkers also used docking, interface design, and affinity maturation to develop designed helical repeat proteins that selectively bind to different drug-bound states of the NS3a protease.³³⁶ These different dimeric complexes could subsequently be used to translate drug inputs into diverse outputs.

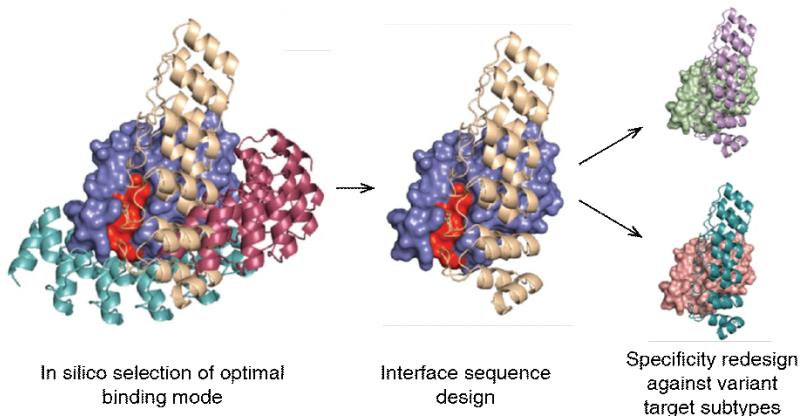


Figure 9. Computational docking calculations can be used to sample possible binding modes, from which the resulting protein-protein interface can be designed to obtain dimeric complexes. Adapted with permission from Ref. ³³⁵. Copyright 2019 NPG.

In addition to shape complementarity, natural protein-protein interfaces typically contain several “hot-spot” residues that contribute a large fraction of the total binding energy.³³⁷ To exploit this concept, a protocol for designing protein-protein interfaces was developed. It first places disembodied amino acids to form such key interactions with a target protein (**Figure 10**).³³⁸ The PDB is then searched for suitable scaffolds, which are docked against the target protein to identify binding modes that can accommodate the hot-spot residues. Once the hot-spot residues are incorporated into the scaffold protein, the remaining residues around the interface are redesigned and the best designs are selected for experimental characterization and affinity maturation, as necessary. Using this protocol, Baker and coworkers designed proteins with nanomolar affinity toward the conserved stem region of influenza hemagglutinin,³³⁸ a heterodimeric complex between an Ankyrin repeat protein and the protein PH1109,¹⁸⁴ as well as proteins capable of binding to and inhibiting hen egg lysozyme.³³⁹ In all three studies, affinity maturation was used to obtain the final, optimized proteins. A crystal structure of a complex of influenza hemagglutinin and one of the evolved binders revealed good agreement with the computational model,³³⁸ as did a crystal structure of the designed hen egg lysozyme inhibitor.³³⁹ Structural characterization of the Ankyrin repeat-PH1109 complex, however, revealed that, following affinity maturation, the relative orientation of the two proteins was flipped 180° compared to the computational model.¹⁸⁴ Nonetheless, many of the designed hot-spot interactions along the protein-protein interface were preserved. The diversity of targets for which new binding proteins have been designed via this approach demonstrates the importance of hot-spot residues in the design of new protein-protein interfaces.

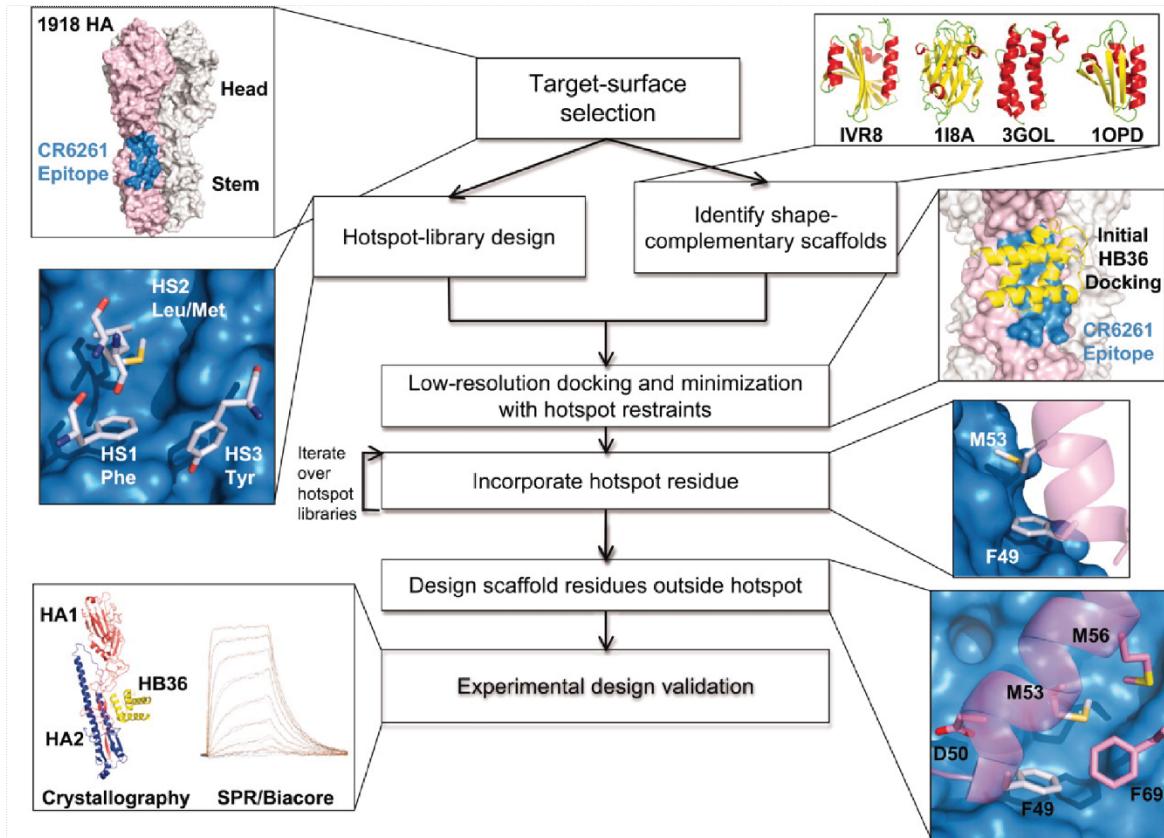


Figure 10. Overview of protein-protein interface design via “hot-spot” residues as applied to the design of proteins that bind to influenza hemagglutinin. Adapted with permission from Ref. ³³⁸. Copyright 2011 AAAS.

3.1.3 Dimerization through metal-coordination

The relatively low success rate of individual computational designs and frequent need for affinity maturation to obtain high-affinity protein complexes underscores the challenges of designing protein-protein interfaces strictly through non-covalent interactions. An alternative approach is to use metal-ligand interactions, which are stronger than non-covalent interactions and capable of driving protein self-assembly without requiring an extensively designed interface.^{198,340,341} The resulting interfaces can then be redesigned to stabilize the oligomeric assembly with or without metal ions.¹⁹³

Tezcan and coworkers have applied metal-mediated assembly to design homodimeric protein complexes, along with other homo-oligomeric states, using the four-helix bundle hemoprotein cyt *cb*₅₆₂ as a scaffold protein. In one study, two *i/i+4* bis-His motifs were designed on the surface of helix 3 of cyt *cb*₅₆₂ to yield a variant termed MBPC1.¹⁹⁴ Addition of Cu²⁺ resulted exclusively in the dimerization of MBPC1 in solution, and a crystal structure of the copper-bound protein revealed an antiparallel dimer held together by two Cu²⁺:His₄ sites with square planar geometries (Figure 11a).¹⁹⁴ In another study, metal-mediated oligomerization was expanded to include the use of non-natural ligands.¹⁹⁷ A cyt *cb*₅₆₂ variant was designed with a single His residue and a hydroxyquinoline chelate covalently attached to a Cys residue in an *i/i+7*

arrangement. The addition of half-molar equivalents of the divalent metal ions Co^{2+} , Ni^{2+} , Cu^{2+} , and Zn^{2+} resulted in the formation of dimeric species in solution, with dissociation constants on the order of 10^{-9} M or lower. A crystal structure of the nickel-bound dimeric complex revealed a V-shaped dimer with a single Ni^{2+} ion bound to the His and hydroxyquinoline moieties of the two monomers in a distorted octahedral geometry (**Figure 11b**).

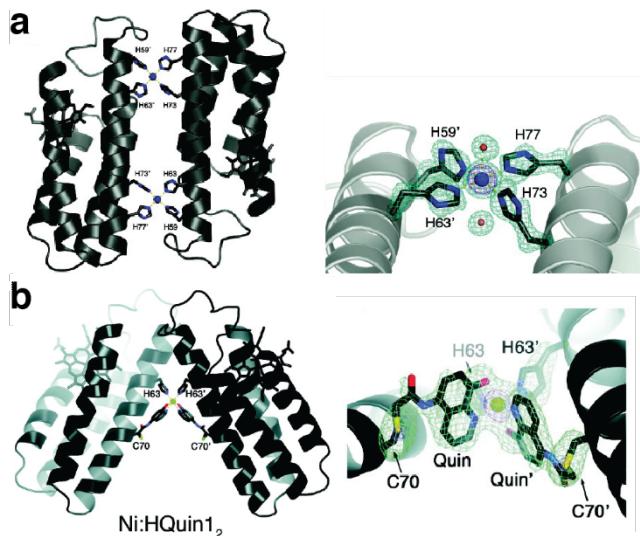


Figure 11. Metal-mediated dimerization of cyt *cb*₅₆₂ via **a**) Cu^{2+} binding to a pair of *i/i+4* bis-His motifs or **b**) Ni^{2+} binding to a His residue and a hydroxyquinoline chelate in an *i/i+7* arrangement. **(a)** Adapted with permission from Ref. ¹⁹⁴. Copyright 2009 ACS. **(b)** Adapted with permission from Ref. ¹⁹⁷. Copyright 2010 ACS.

Kuhlman and coworkers designed a homodimeric assembly by incorporating metal-binding functionalities into the computational interface design process.³⁴² Rosetta was used to place pairs of His residues on the surface of α -helical scaffold proteins. The second copy of each monomer was then placed to form C_2 -symmetric dimers with two tetrahedral metal-binding sites, and the resulting dimeric interfaces were re-designed. Of eight tested computational designs, one design, termed MID1, expressed well and formed a dimer with or without Zn^{2+} , though Zn^{2+} binding significantly increased the affinity of two monomers for each other. Crystal structures revealed that the overall conformation of the Zn^{2+} -bound dimer closely resembled the design model, although only three out of four His residues coordinated to Zn^{2+} (**Figure 12**). Though unanticipated, these coordinatively unsaturated Zn sites enabled the design and evolution of enzymatic activity toward hydrolysis and Diels-Alder reactions, as discussed in more detail in Section 4.^{207,343,344}

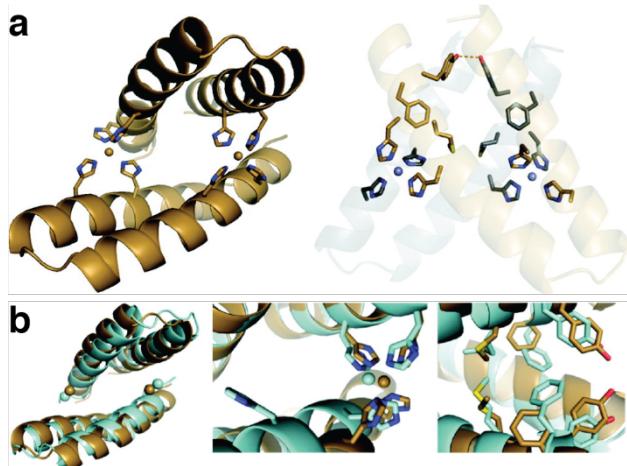


Figure 12. Design model and structure of MID1. **a)** Computational design model of MID1. **b)** Comparison of the MID1 design model (tan) and crystal structure (cyan). Adapted with permission from Ref. ³⁴². Copyright 2012 ACS.

The function of most natural metalloenzymes relies on a stable protein scaffold that exerts control over the metal coordination environment and tunes metal reactivity.^{345,346} Such a metal-independent, highly preorganized architecture is challenging to achieve via metal-mediated protein self-assembly, because in this approach the structure of the protein scaffold tends to be directed by the coordination preferences of the metal ions instead. To overcome this challenge, Rittle et al. developed a simple approach, termed Metal Active Sites by Covalent Tethering (MASCoT), in which two proteins are first covalently tethered by the formation of a disulfide bond between single Cys residues on their surfaces (Figure 13a).³⁴⁷ Incorporation of metal binding residues subsequently allows the conformation of the nascent protein-protein interface to be locked into place upon addition of metal. Using this approach, the researchers made a suite of metal-binding proteins with tunable primary and secondary coordination spheres (Figure 13b–e). These metal binding sites featured unusual asymmetric coordination environments due to constraints imposed by the disulfide bond and bound all first-row transition metals from Mn^{2+} to Zn^{2+} with unusually high affinities for designed proteins. Furthermore, a variant with a penta-His metal binding site was able to bind nitric oxide, suggesting that metal binding sites developed through the MASCoT approach have the potential to bind ligands and act as active sites for catalysis.

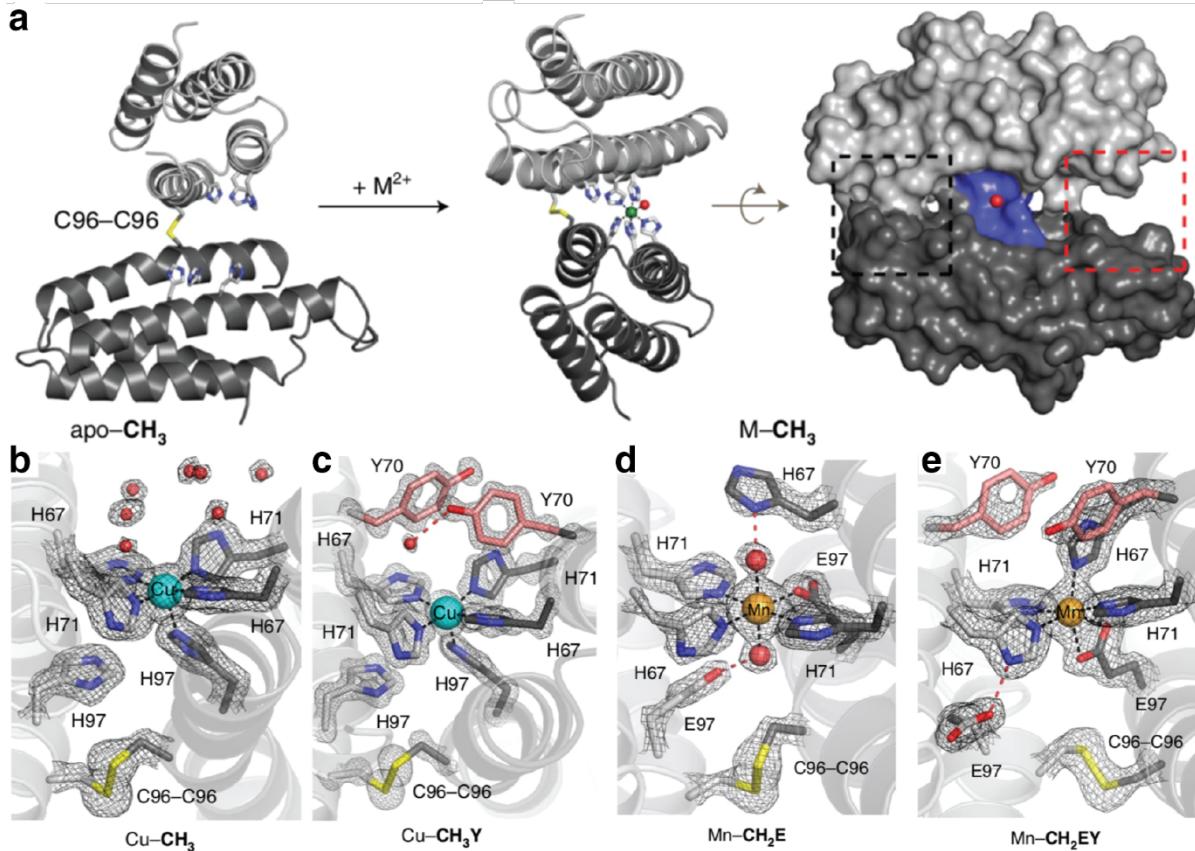


Figure 13. Metal Active Sites by Covalent Tethering (MASCoT). **a)** Implementation of MASCoT. The C96-C96 disulfide bond covalently tethers two proteins, forming a nascent protein-protein interface that is locked into place upon metal addition. **b-e)** Formation of metal binding sites with different primary and secondary coordination spheres. Adapted with permission from Ref. ³⁴⁷. Copyright 2019 NPG.

3.2 Small oligomers

Nature uses protein oligomerization to generate structural and functional complexity through the self-assembly of individual protein domains into larger protein oligomers.^{45,54} From an engineering perspective, designing new oligomeric protein assemblies is beneficial for (1) increasing the intrinsic stability of proteins, (2) controlling shape, composition and size of protein complexes for target function, and (3) creating novel inter-molecular interfaces for new enzymatic active sites or allosteric regulation of supramolecular structures.¹¹⁰ In this section, we will focus on protein oligomerization aided by computational design, based on the generation of genetic fusion and domain swapping constructs, and mediated by electrostatic interactions and protein-metal interactions.

3.2.1 Oligomerization through computational interface design

Symmetry enables the formation of large architectures from simple building blocks through the design of a small number of associative surface patches.³⁴⁸ Schulz and coworkers

took advantage of symmetry by rationally designing protein-protein interfaces between homo-oligomeric protein building blocks, which enabled the formation of larger homo-oligomers with only a few mutations.³⁴⁹ The researchers first attempted to build dimers out of the monomeric protein 6-phospho- β -galactosidase (Pga) by stabilizing two different crystallographic interfaces through the introduction of additional large, non-polar residues (i.e., Phe, Trp, Met). This resulted in four variants with dimer yields between 3% and 56%, although none of the dimers yielded crystals for structure determination. The researchers then set out to generate tetramers out of the homodimeric proteins O-acetylserine sulfhydrylase (Oas) and urocanase (Uro), and construct octamers out of the homotetrameric protein RhuA. In each case, the researchers aligned two copies of the protein building block along its molecular symmetry axis (two-fold rotational axis for the dimers, four-fold rotational axis for the tetramer) and performed a one-dimensional search along the relative rotation angle to find suitable orientations for the two copies of the protein to form a new protein-protein interface. Mutations were then rationally made to stabilize the interface. Three variants of the Oas dimer yielded only small amounts of tetramer in solution and failed to crystallize. In contrast, a variant of the Uro dimer yielded a tetramer in 80% yield, with the crystal structure revealing that the tetramer was similar to the designed assembly. However, a slight shift in the relative orientation of the monomers broke the designed D_2 symmetry and instead resulted in four local C_2 symmetries. Meanwhile, three variants of the tetramer RhuA, which had only one or two mutations, all resulted in quantitative yields of the octamer in solution, and crystal structures indicated that one formed the octamer as designed, whereas another formed an octamer with displaced fourfold axes. This example illustrated the possibility of designing higher-order homo-oligomers with only a small number of mutations through the use of symmetric building blocks.³⁴⁹

Many natural proteins contain modular, monomeric structures and are believed to have evolved by the duplication of structural elements to form symmetric homooligomers, followed by the fusion of these domains to form a monomeric protein.^{350,351} This process can be reverse engineered by designing symmetric proteins, then breaking them into smaller monomers that assemble to form oligomeric analogues of the original monomer. Meiler and coworkers used this approach to design a perfectly symmetrical eight-stranded $\beta\alpha$ -barrel protein.³⁵² The researchers took half of an asymmetric $(\beta\alpha)_8$ -barrel protein, HisF, and connected two copies of the half-protein to reform the full $(\beta\alpha)_8$ -barrel, but with perfect two-fold symmetry (**Figure 14a**). They then used the Rosetta suite to optimize the protein sequence and sidechain packing and expressed the variant with the best score according to Rosetta energy units. A monomer consisting of half of the designed protein assembled into a dimer whose crystal structure closely matched that of both the single-chain $(\beta\alpha)_8$ -barrel and the design model.³⁵² Similarly, Voet et al. designed a perfectly symmetrical β -propeller protein.³⁵³ Starting with a six-bladed β -propeller protein, the sensor domain of a protein kinase from *Mycobacterium tuberculosis*, the researchers replaced each blade with the third blade of the protein to form a perfectly symmetrical template. They optimized the sequence of the blade using Rosetta and fused the identical blades to form two, three, and six-bladed monomers, named Pizza2, Pizza 3, and Pizza 6. All monomers expressed well, and the two and three-bladed proteins assembled into six-bladed oligomers whose structures almost exactly matched the six-bladed monomer (**Figure 14b**).

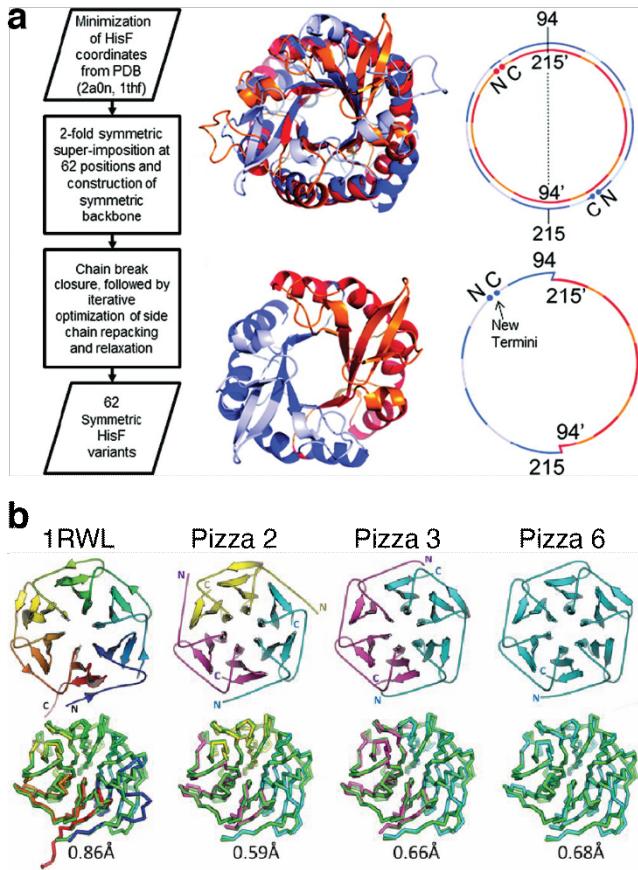


Figure 14. Design of symmetric homo-oligomers from proteins with repeating units. **a)** Design of a two-fold symmetric $(\beta\alpha)_8$ -barrel. The two halves are soluble as monomeric proteins that assemble into a dimeric barrel. **b)** Crystal structures of Pizza6 proteins with different oligomeric states. From left to right: wild-type protein used as a template for Pizza6 design, Pizza2 (trimer), Pizza3 (dimer), Pizza6 (monomer). **(a)** Adapted with permission from Ref. ³⁵². Copyright 2011 ACS. **(b)** Adapted with permission from Ref. ³⁵³. Copyright 2014 National Academy of Sciences.

While both of the previous examples used pre-existing proteins as building blocks, Bradley and coworkers developed a computational strategy (**Figure 15**) for the *de novo* design of symmetrical proteins consisting of repeating structural motifs that form closed, toroidal structures.¹⁸⁵ The protocol first defined the secondary structure of the repeat motif, as well as the number of repeats of that motif and the geometrical parameters of the final structure (inter-repeat rise and curvature). *Ab initio* folding calculations were then performed to generate backbone conformations that match those criteria and determine the amino acid sequence to yield each backbone conformation. The resulting protein candidates were filtered to remove poor designs and clustered to identify recurring packing arrangements. Finally, low-energy designs from those clusters were further assessed by re-predicting their structures based on their sequences. The researchers applied this protocol to design left-handed α -helical repeat proteins that formed closed toroids of various sizes.¹⁸⁵ Variants corresponding to four toroidal architectures were crystallized and revealed structures that closely matched the design models. A 9-repeat design was then split into a 3-repeat fragment that was expected to assemble into a 9-repeat trimer.

Instead, the fragment formed a 12-repeat tetramer, suggesting that the repeat sequence was compatible with both 9- and 12-repeat structures.

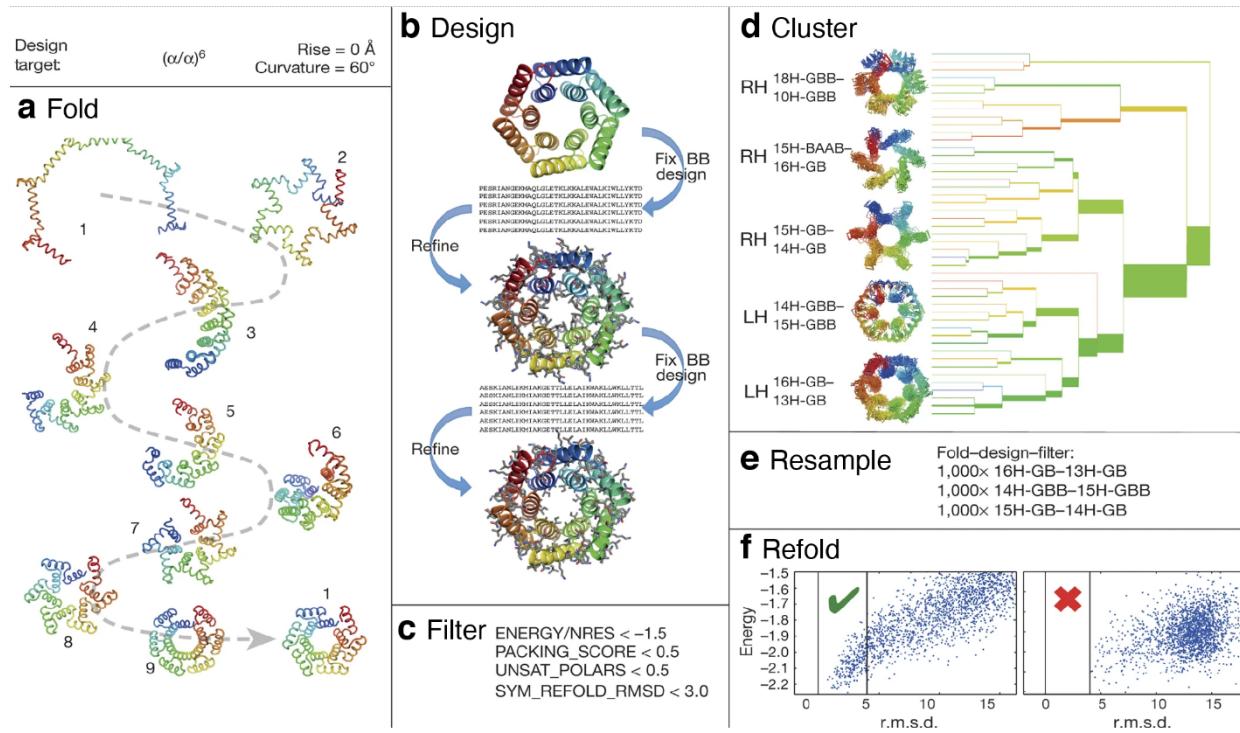


Figure 15. Computational design of α -helical toroids. The protocol consists of **a)** *ab initio* folding to generate backbone conformations, **b)** sequence design of conformations, **c)** filtering to remove poor designs, **d)** clustering of results to identify recurring packing arrangements, **e)** resampling of structures from the clusters, and **f)** a final assessment of results by re-predicting the designed structures from their sequences. Adapted with permission from Ref. ¹⁸⁵. Copyright 2015 NPG.

The aforementioned studies involved the design of oligomeric assemblies by breaking a symmetric protein into smaller monomers that assemble to re-form the initial protein. In an alternative approach, Baker and coworkers sought to design cyclic homo-oligomers by starting with *de novo*-designed $\alpha\beta$ -proteins³⁵⁴ or monomeric repeat proteins.³⁵⁵ Although oligomers based on $\alpha\beta$ -protein building blocks formed multiple oligomeric species in solution, several monodisperse oligomers were formed with the repeat protein scaffolds. The researchers began by docking the selected scaffold proteins in cyclic geometries using a low-resolution docking protocol. Because low-resolution docking ignores side chains, predicting which docked geometries yielded the most readily designable interfaces presented a challenge. This challenge was addressed by precompiling a database of favorable interactions between pairs of residues and binning them based on the rigid-body orientations of the pairs of residues. For a pair of residues along a nascent interface in a docked model, their side-chain-independent orientations were used to find the Rosetta energy of the most favorable potential interaction in the database, and the sum of these energies over all pairs of residues along the nascent interface was used to score the docking geometry (**Figure 16**). The best scoring geometries were then subjected to sequence design to optimize the protein-protein interfaces, and ~100 circularly symmetric, homo-

oligomeric designs (ranging from dimers to hexamers) were selected for experimental characterization. A large fraction of these variants was soluble, and ~20 % had oligomeric states that matched the design model as determined by SEC-MALS. Furthermore, crystal structures and SAXS patterns of several designs were consistent with computed models. Mohan and coworkers then used this approach to design dimers of an Ankyrin repeat protein that could bind to the cytokine receptor EpoR, which dimerizes in response to its protein-ligand.³⁵⁶ By modifying the size and orientation of the Ankyrin repeat dimers, the researchers were able to examine and tune the cellular response induced by EpoR dimerization.

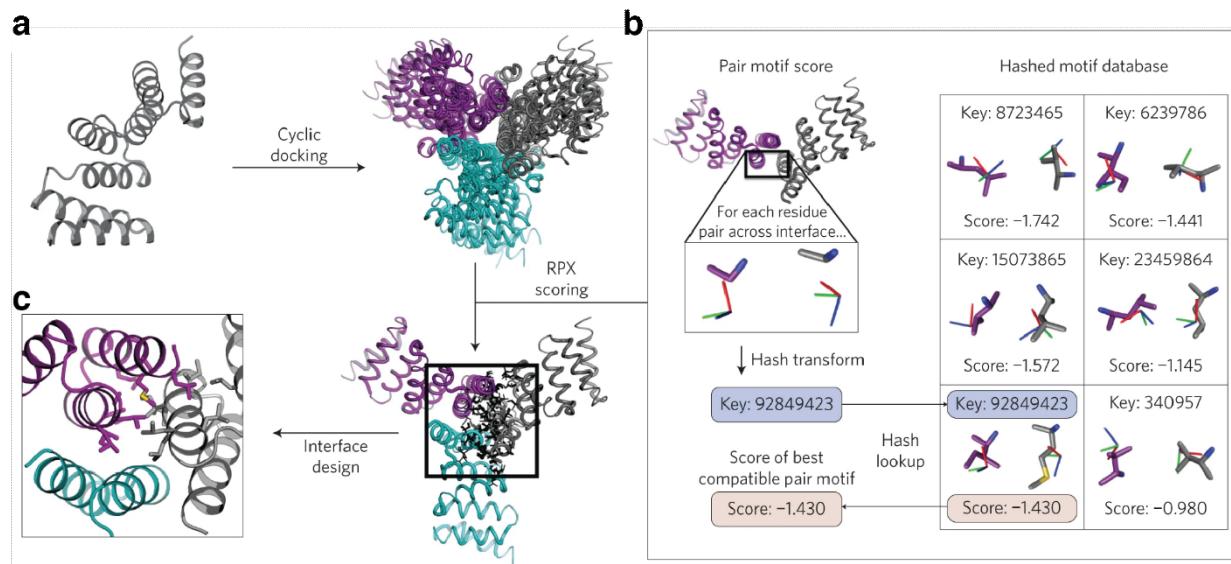


Figure 16. Computational design of cyclic homo-oligomers. **a)** Monomers are first docked in cyclic geometries using low-resolution, symmetric docking. **b)** Docked oligomers are then scored using a residue-pair transform (RPX) method which models side-chain interactions implicitly. **c)** The best scoring geometries are finally subjected to interface design. Adapted with permission from Ref.³⁵⁵. Copyright 2016 NPG.

There has also been progress in designing oligomeric assemblies by focusing on hydrogen bond networks, which can impart specificity in protein-protein interactions. However, hydrogen bond networks can be challenging to design due to the precision with which polar residues must be placed and the energetic penalties associated with buried, unsatisfied hydrogen bond donors or acceptors. Boyken et al. therefore developed a computational sequence design method, HBNet, which precomputes all possible hydrogen bonding interactions between all possible polar residues at a given set of positions and is thus more efficient at identifying possible hydrogen bonding networks than traditional sequence design methods.²¹¹ To design homo-oligomers with HBNet, the researchers first generated a large number of α -helical, coiled-coil backbone arrangements consisting of multiple copies of two-helix monomeric subunits. They then used HBNet to identify backbone arrangements capable of accommodating hydrogen bond networks spanning the protein-protein interface(s) and involving at least three residues. Finally, traditional Rosetta sequence design was employed to optimize the remaining residues outside of the hydrogen bond networks (**Figure 17**). Of the >100 top-ranked designs that were experimentally tested (including dimers, trimers, and tetramers), more than half formed

the designed oligomeric state. Furthermore, the crystal structures and solution SAXS scattering patterns of several variants confirmed that most of the hydrogen bond networks closely matched their corresponding design models. This study demonstrated that it is possible to computationally design a wide range of hydrogen bonding networks at protein-protein interfaces to achieve different assembly geometries, which is reminiscent of the sequence-dependent programmability of DNA. HBNet has since been applied to design a set of orthogonal heterodimers that can be used to make pH switchable assemblies and protein logic gates.^{189,190,357}

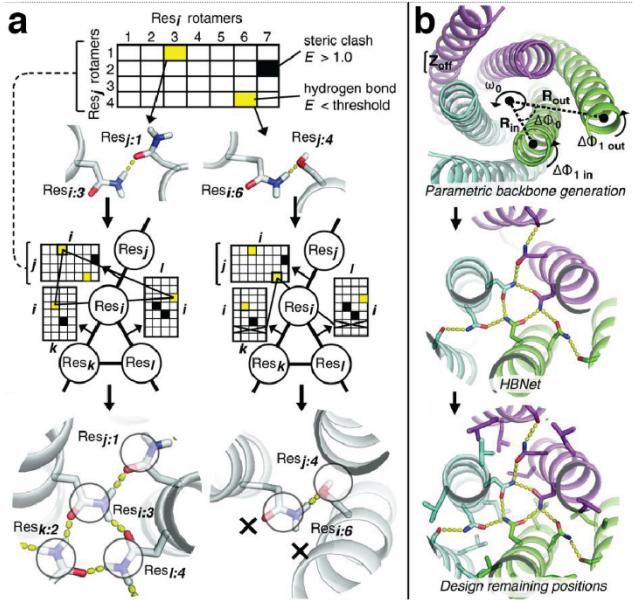


Figure 17. Design of coiled-coil oligomers via HBNet. **a)** HBNet precomputes the hydrogen bond and repulsive interaction energies between side chains for all possible rotamers at a set of selected positions. This information is stored in a graph structure; traversing the graph reveals the combinations of hydrogen-bonding rotamers that form a hydrogen bond network. **b)** Design of homo-oligomers by sequentially designing coiled-coil backbones, using HBNet to identify hydrogen bonding networks that connect two-helix monomers, and designing the remaining residues around the hydrogen bonding network to accommodate the network. Adapted with permission from Ref. ²¹¹. Copyright 2016 AAAS.

Having developed a method for computationally designing hydrogen bond networks, Lu et al. surmised that they could use the method to design transmembrane proteins.³⁵⁸ Because of the hydrophobicity of the lipid bilayer, specific protein-protein interactions in transmembrane proteins cannot be designed by burying hydrophobic residues along the interface. Instead, buried hydrogen bonds can play a key role in determining protein interaction specificity in the membrane environment. The researchers therefore used HBNet to create buried hydrogen bond networks in four-helix bundles and subsequently added a ring of amphipathic aromatic residues to define the extracellular side and a ring of positively charged residues to define the cytoplasmic side. Hydrophobic residues were then placed at surface-exposed positions between the amphipathic and positively charged rings to generate transmembrane regions. The researchers expressed the four-helix bundles as either monomers or as dimers of two-helix monomers. Dimers with longer transmembrane regions and a monomer expressed well, localized to the membrane of cells, had high thermal stability and exhibited the expected oligomeric state. The crystal structure

of one of the dimers closely resembled the design model (**Figure 18a-b**). Additionally, the researchers used the same approach to design a six-helix transmembrane trimer and a transmembrane tetramer with eight membrane-spanning helices (**Figure 18c**), both of which were expressed and formed the expected oligomeric state.

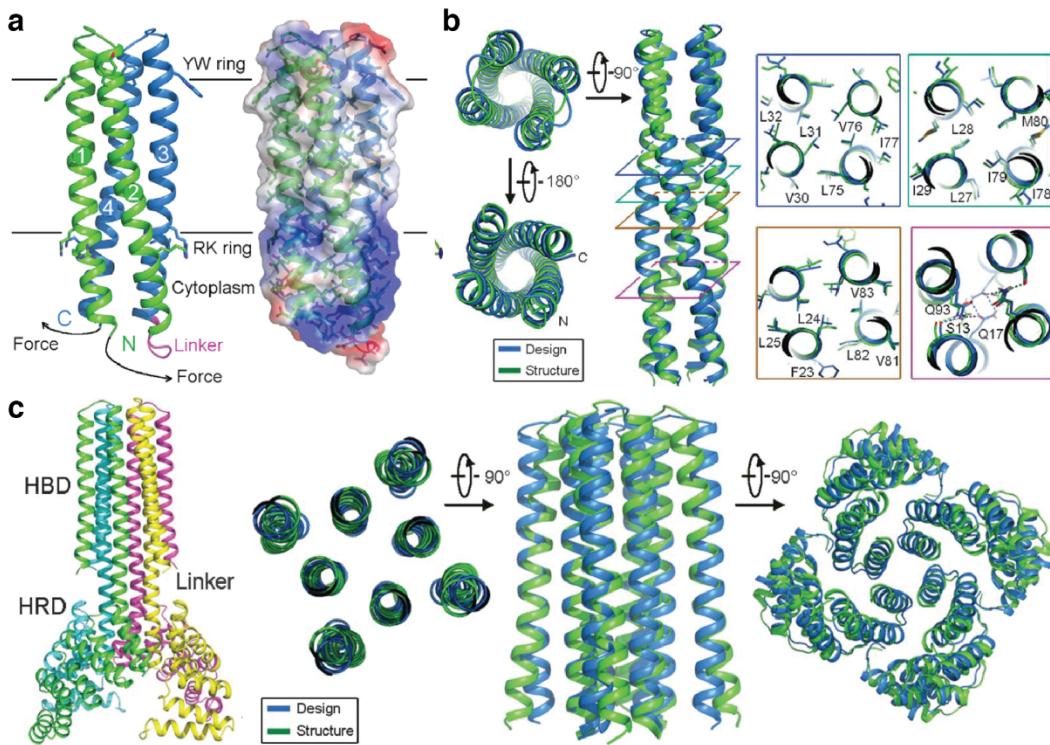


Figure 18. Crystal structures of *de novo* designed transmembrane oligomers. **a)** Model of the dimeric transmembrane protein, indicating the position of the amphipathic (YW) and positively charged (RK) rings. **b)** Comparison between the crystal structure and design model of a transmembrane dimer. **c)** Crystal structure and design model of the transmembrane tetramer. Adapted with permission from Ref. ³⁵⁸. Copyright 2018 AAAS.

Building on these studies, Xu et al. designed α -helical proteins that assemble into transmembrane pores.³⁵⁹ Using HBNet and protocols they had previously developed for the design of α -helical oligomers, the researchers started by designing water-soluble two-helix proteins that assembled into hexamers or octamers with central solvent channels. They experimentally tested three hexameric and fifteen octameric designs and found one hexameric and two octameric variants that formed homogeneous assemblies with the correct oligomeric state. Crystal structures of the hexameric variant (**Figure 19a**) and one of the two successful octameric variants (**Figure 19b**) confirmed that the structures of the assemblies closely matched the design models. The researchers then converted the structurally characterized, water-soluble oligomers into transmembrane pores by replacing solvent-exposed residues in the intended transmembrane region with hydrophobic residues. Both the hexameric variant and the octameric variant were expressed and purified from the membrane fraction of *E. coli* cells, and the hexameric variant formed the correct oligomeric state. Meanwhile, the oligomeric state of the

transmembrane version of the octameric variant was ambiguous, but a tetrameric version formed by linking two monomers more clearly formed the correct oligomeric state (**Figure 19c**).

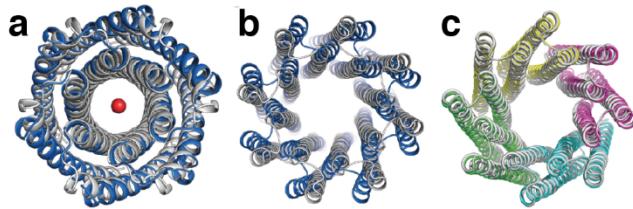


Figure 19. Computationally designed, oligomeric transmembrane pores. **a)** Overlay between the crystal structure (blue) and design model (grey) of the water-soluble hexamer. **b)** Overlay between the crystal structure (blue) and design model (grey) of the water-soluble octamer. A slight tilt in the monomers results in a greater deviation from the design model than in the hexameric case. **c)** Cryo-EM structure of the tetrameric transmembrane pore. Adapted with permission from Ref. ³⁵⁹. Copyright 2020 NPG.

3.2.2 Oligomerization through electrostatic interactions

Forming oligomeric protein assemblies strictly through non-covalent interactions typically requires the design of extensive protein-protein interfaces. By contrast, Ellington, Glotzer and coworkers demonstrated that discrete oligomers can also be formed simply by combining pairs of oppositely supercharged proteins.²⁷⁶ Using GFP as a model system, the researchers observed that various pairs of oppositely charged GFP variants, such as Ceru+32 and GFP-17 (with total charges of +32 and -17, respectively), formed large oligomeric species when combined in solution (**Figure 20a–b**). The sizes of these species depended on solution conditions. At high salt concentrations and/or low pH, the predominant species were either monomeric with a diameter of ~6 nm or formed oligomers with a diameter of 12 nm. In contrast, at low salt concentrations and pH 7.4, the predominant species were large, 1,300-nm diameter oligomers. These solution behaviors suggested that the oligomers formed by supercharged GFP variants assembled via specific intermolecular interactions, even though no such interactions were explicitly designed. A high-resolution cryogenic electron microscopy (cryo-EM) structure of the ~12 nm oligomer revealed a D_4 symmetric oligomer consisting of two stacked octameric rings (**Figure 20c**). The researchers hypothesized that the large number of charged residues on the monomers provide an abundance of potential interaction sites, enabling the monomers to find modes of interaction that mediated the formation of larger assemblies. The ability of multiple supercharged GFP variants to form oligomeric assemblies when paired with monomers of the opposite charge, along with the fact that supercharging has been demonstrated for various proteins in addition to GFP, suggests that supercharging could serve as a simple, broadly applicable method to obtain non-covalent oligomers.

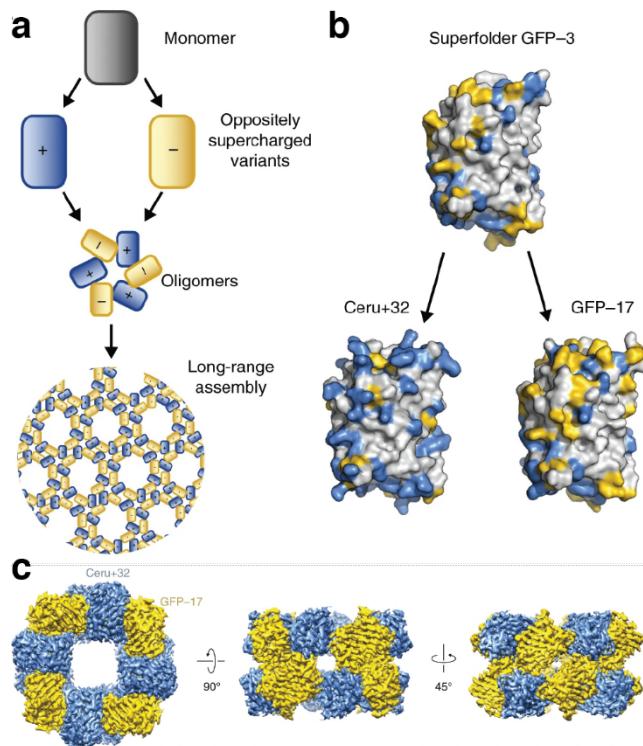


Figure 20. Oligomerization of oppositely supercharged Cerulean (Ceru) and GFP variants mediated by electrostatic interactions. **a)** Hypothesized model of symmetric oligomers and long-range assembly of oppositely supercharged monomeric proteins. **b)** Supercharged GFP variants engineered by mutating the surface residues of sfGFP to basic (blue) or acidic (yellow) residues. **c)** Cryo-EM reconstruction of the Ceru⁺³²/GFP⁻¹⁷ 16-mer at 3.47 Å resolution in three different orientations. Adapted with permission from Ref. ²⁷⁶. Copyright 2019 NPG.

3.2.3 Oligomerization by genetic fusion and domain swapping

Globular proteins have evolved both intramolecular and intermolecular interactions (with other biomolecules) to stabilize and maintain the rigidity of their 3D structures.¹⁵⁷ From the perspective of artificial protein self-assembly design, the intra/intermolecular interactions of pre-existing motifs/domains can be repurposed to form new assembly interfaces between protein building blocks. In this subsection, we will discuss how genetic fusion can be used to this effect.

The genetic fusion strategy for designing artificial protein assemblies is based on introducing connectivity between two independent domains. Arai and coworkers used WA20-Foldon fusion proteins as a building block: the WA-20 unit served as the main framework in the homodimeric state and Foldon was a connector to mediate homotrimer assembly (**Figure 21**).³⁶⁰ Trimerization of monomeric Foldon in the fusion protein induced the self-assembly of the building blocks into four unique nanostructures: S (6-mer), M (12-mer), L (18-mer), and H (24-mer). Solution SAXS results indicated that S-form oligomers had a barrel shape, while M-form oligomers formed a tetrahedron. The same research group also utilized the homodimer assembly of WA20 to construct both closed and open structures.³⁶¹ Two WA20 monomers were linked by two types of peptide linkers whose amino acid sequences prefer the formation of an α -helix or random coils. Such fusion dimers preferentially formed cyclized, chain-like homo-oligomers

ranging from dimer to pentamer regardless of linker types. Conversely, when the fusion WA20 dimer was mixed with the original WA20 dimer as a stopper and all intermolecular interactions were reconstituted, the fusion WA20 dimers preferentially formed extended, chain-like assemblies. Furthermore, shape analysis of the repeating units using SAXS indicated that the conformational details of the chains depend on the types of linkers used at the fusion point. The fusion dimer with the helix-favoring linker showed a linear alignment whereas the random coil-favoring linker showed a compact alignment.

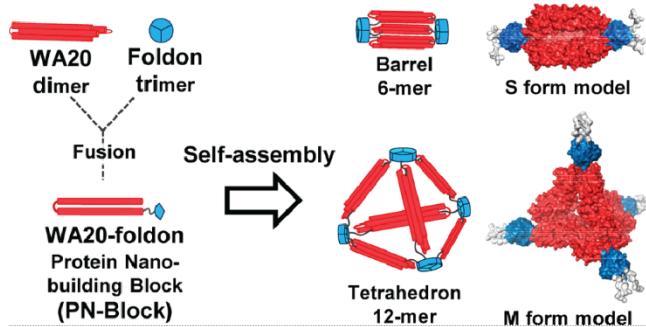


Figure 21. Genetic fusion strategy for self-assembly of WA20-Foldon fusion protein using native trimerization of Foldon. Adapted with permission from Ref. ³⁶⁰. Copyright 2015 ACS.

Taking advantage of the high stability and folding efficiency of green fluorescent protein (GFP), Jung and colleagues utilized the sfGFP protein as a building block for designed protein oligomers inside cells (**Figure 22**).³⁶² The 11 β -strand GFP was first split to two fragments: β -strand 11 of GFP (GFP11, residues 215-330) and the rest of the GFP structure (GFP1-10, residues 1-214). Then, GFP11 was transplanted to the N-terminus of GFP 1-10 via a short peptide linker, which did not allow the intramolecular association between GFP11 and GFP1-10 within the same monomer. This genetic fusion construct underwent intermolecular incorporation of the β -strand from GFP11 into the empty space in the β -barrel of GFP10, thereby forming polygonal GFP assemblies ranging from dimer to decamer (**Figure 22**). The linker length was the most critical factor to promote cellular self-assembly. This strategy was further developed to stabilize the polygonal GFP assemblies by incorporating negative charges on the GFP surface, and to control polygonal morphologies by adding a functional fusion pair to the N- and C-termini of GFP or co-expressing capping GFP, which suppressed circular oligomerization of the fusion constructs. This study demonstrated that genetic fusion could be used to reengineer a protein monomer to promote self-assembly into oligomeric states with defined structures or varied higher-order assemblies inside cells.

Muñoz and coworkers designed a variant of chymotrypsin inhibitor 2 (CI2) to assemble into a hexameric ring upon fold switching.³⁶³ They began by noticing that CI2 always crystallized in a hexagonal space group, and that some crystallization conditions resulted in domain-swapped dimers. The researchers therefore set out to engineer a variant of CI2 that destabilized the native fold in the hope that a change in the fold of the protein would expose new residues that could stabilize the hexameric conformation. The resulting protein, CI2_{eng}, was largely monomeric at low concentrations but formed mostly hexameric and some dodecameric species at high concentrations. Circular dichroism (CD) and solution NMR experiments strongly

implicated a fold switch in the Cl2_{eng} monomer as being responsible for its assembly into the larger oligomers, and a cryo-EM structure of the dodecamer revealed that the alternate fold of Cl2_{eng} exposed a previously buried hydrophobic patch that formed a new interface with an adjacent monomer in the assembly. Further mutagenesis increased the yield of hexameric and dodecameric species in solution, and the researchers could take advantage of the conformational changes involved in oligomerization to reversibly assemble and disassemble the oligomers (see also Section 4.2.1).³⁶³

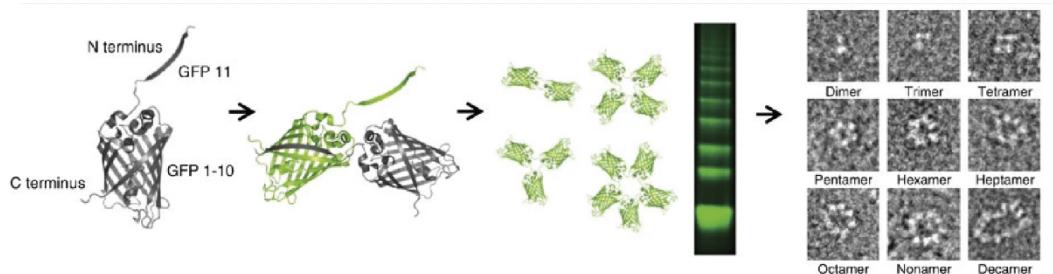


Figure 22. Discrete formation of GFP polygons by GFP1-10 with N-terminal-fused GFP11. The intermolecular incorporation of β -strand 11 into the β -barrel of GFP 1-10 guides stable oligomerizations from dimeric to decameric structures. Adapted with permission from Ref.³⁶². Copyright 2015 NPG.

3.2.4 Metal-mediated oligomerization

Metal-binding interactions are considerably stronger than common non-covalent interactions, and metal-mediated oligomerization can capture many salient features of non-covalent protein-protein interactions (e.g., stability, directionality, symmetry, reversibility) on a much smaller surface area than that required for stable non-covalent interfaces.^{88,198} Various research groups have utilized metal-mediated oligomerization to create novel protein assemblies.^{364,365} In particular, the Tezcan Group has extensively studied metal-mediated oligomerization design principles using *cyt cb*₅₆₂, a four-helix bundle hemoprotein with high stability and solubility, as a scaffold.^{84,193,194,196,199,205,208,210,232,366,367} *Cyt cb*₅₆₂ is natively monomeric even in the mM concentration range, implying that, with regards to protein self-assembly, its surface is evolutionarily naïve. Tezcan and coworkers have developed two complementary approaches for the design of metal-mediated oligomers. In the first approach, metal-directed protein self-assembly (MDPSA), metal-chelating motifs are installed on the surface of a target protein. Upon metal addition, the strength, directionality, and selectivity of the resulting metal-ligand interactions help guide protein self-assembly (**Figure 23**). Metal-directed *cyt cb*₅₆₂ assemblies can be further modified using a second approach, metal-templated interface redesign (MeTIR), in which metal-directed protein assemblies are redesigned using non-covalent interactions and covalent linkers along the protein-protein interface(s). Sequential interface design through MDPSA and MeTIR results in stable, metal-independent oligomers, mimicking the time course of a hypothetical evolutionary pathway for the formation of stable protein assemblies from an initial metal-templated structure.³⁶⁸⁻³⁷⁰

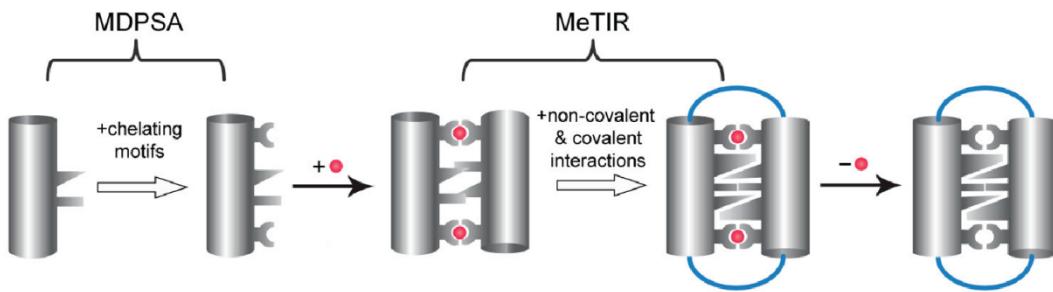


Figure 23. MDPSA and MeTIR strategies for metal-mediated oligomerization. In MDPSA, metal-chelating motifs are installed on the surface of target proteins, thereby guiding protein self-assembly through metal-coordination. In MeTIR strategy, metal-directed protein assemblies are redesigned using non-covalent interactions and a covalent linker at the protein-protein interface. The sequential design approach through MDPSA and MeTIR results in the formation of stable oligomers without the aid of metal ions. Adapted with permission from Ref. 83. Copyright 2016 Elsevier.

Using MDPSA, Salgado et al. demonstrated that the installation of two bis-His motifs (H59/H63, H73/H77) on helix 3 of cyt *cb*₅₆₂ guides protein self-assembly mediated by various first-row transition metal ions (**Figure 24a**). The resulting construct, MBPC1, self-assembled into various oligomeric states dictated by the coordination preferences of nucleating metal ions. For example, due to its preference for a tetrahedral coordination geometry, Zn²⁺ induced *D*₂ symmetric tetramerization of MBPC1 (i.e., Zn₄ : MBPC1₄) by forming four 3His/1Asp coordination sites, His73/His77 from one protomer, His62 from a second, and Asp74 from a third.^{199,366} Meanwhile, Ni²⁺ induced *C*₃ symmetric trimerization of MBPC1 (i.e., Ni₂ : MBPC1₃) via octahedral 6His coordination, while Cu²⁺ mediated *C*₂ symmetric dimerization (i.e., Cu₂ : MBPC1₂) via the formation of square planar 4His coordination sites.¹⁹⁴

Close examination of the tetrameric Zn₄:MBPC1₄ complex revealed an extensive buried surface area within the architecture (~5000 Å²). To examine how the surface interactions between the protomers collectively influence the assembly, Salgado et al. made a small modification to the metal coordination sphere.³⁶⁶ Within Zn₄:MBPC1₄, Zn²⁺ coordination to the Asp74 residue was instrumental for crosslinking the MBPC1 monomers at the helix 3 C-termini to give rise to V-shaped dimers. Assuming a negligible effect from non-covalent interactions within the interface, the authors developed the variant MBPC2, where the coordinating Asp residue was moved from the C-terminal His73/His77 clamp to the N-terminal His59/His63 clamp, predicting that the modification would invert the crosslinking position in the assembly. MBPC2, characterized by the mutations R62D/D74A indeed formed a Zn₄:MBPC2₄ complex characterized by ultracentrifugation and crystallography, wherein the V-shaped dimers were linked at the N-terminus rather than the C-terminus (**Figure 24b**). However, Asp62 was revealed not to participate in coordination. Instead, the Zn²⁺ coordination environment was composed of the 73/77 bis-His clamp from one protomer, His59 from a second and His63 from a third. These results revealed that the coordination environment alone did determine the outcome of metal-mediated protein assembly. Upon closer examination of MBPC1 and MBPC2 Zn-mediated tetramers, salt bridging interactions involving Arg34 were shown to be critical in guiding the precise geometry of the self-assembled architectures.³⁶⁶

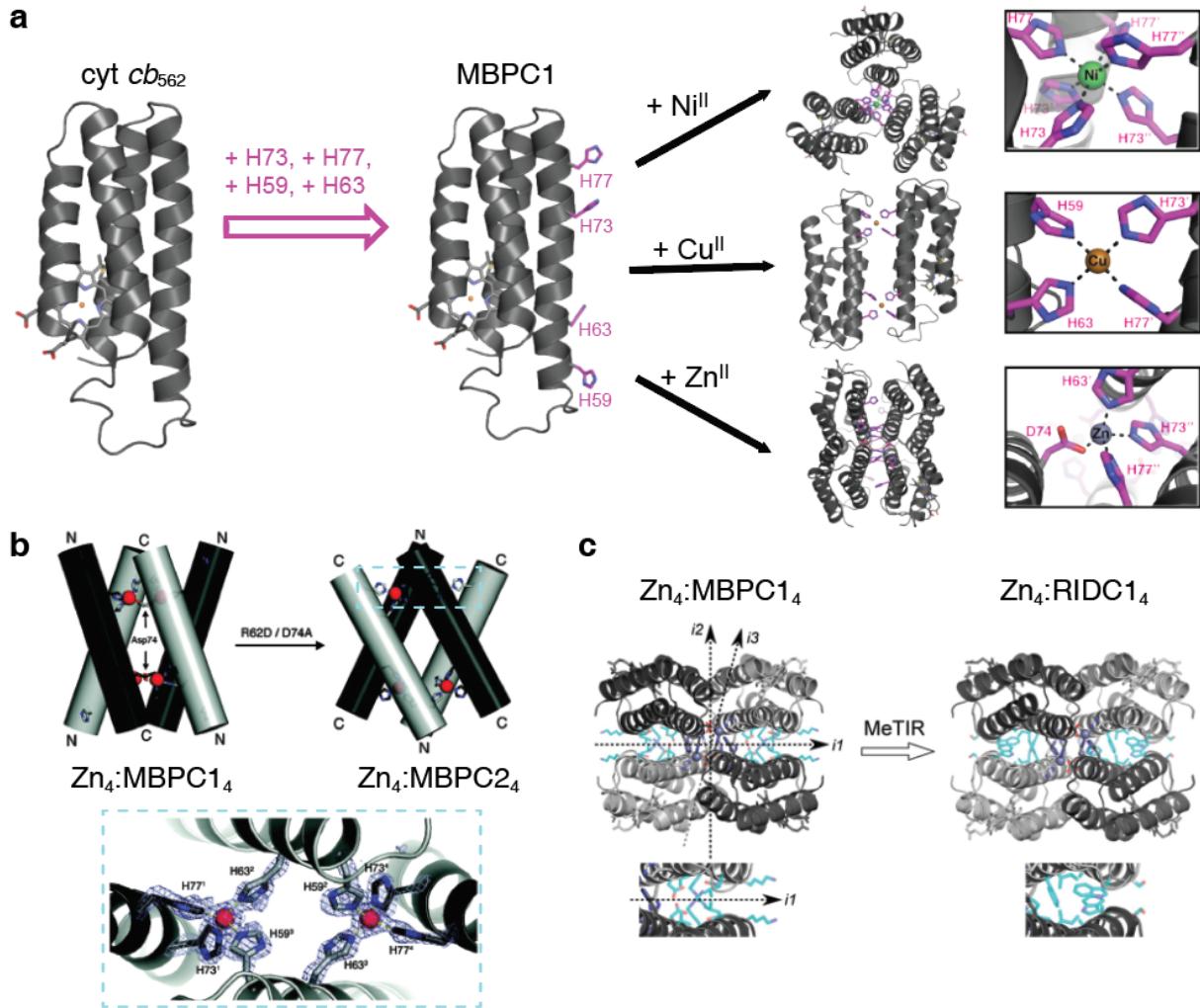


Figure 24. Metal-mediated protein oligomers based on cyt *cb*₅₆₂ as a building block. **a)** Design of cyt *cb*₅₆₂ oligomers using the MDPSA approach. In the MBPC2 variant, surface-exposed bis-His clamps are located at residues 59/63 and 73/77. Depending on the metal ion coordination preferences, different oligomerization states are achieved – Ni₂:MBPC1₃ trimer (top), Cu₂:MBPC1₂ dimer (middle), and Zn₄:MBPC1₄ tetramer (bottom). **b)** In the MBPC2 variants, the R62D and D74A mutations give rise to the Zn₄:MBPC2₄ tetramer where the orientation of the protein monomers is inverted. **c)** MeTIR strategy for redesigning the interface of Zn₄:MBPC1₄. *i*1 interface of Zn₄:MBPC1₄ tetramer is redesigned to install hydrophobic residues and create the stabilized Zn₄:RIDC1₄ tetramer. **(b)** Adapted with permission from Ref. ³⁶⁶. Copyright 2008 ACS. **(c)** Adapted with permission from Ref. ⁸⁸. Copyright 2019 ACS.

These studies demonstrated that it was feasible to mediate protein self-assembly simply through installation of a small number of metal coordination motifs on a protein's surface. Recognizing that such metal-directed protein assemblies possessed extensive buried interfaces between the protein monomers, Salgado et al. developed the MeTIR strategy to obtain stable protein assemblies wherein the metal-templated interfaces were further engineered for installation of favorable interactions. Based on the structure of Zn₄:MBPC1₄, which has three different sets of C₂ symmetric interfaces (*i*1, *i*2, and *i*3), the researchers redesigned the surface

of MBPC1 using rotamer optimization based on Rosetta (**Figure 24c**).¹⁹³ The computational analysis of *i1* identified six mutations (R34A/L38A/Q41W/K42S/D66W/V69I) that were incorporated to yield the construct Rosetta Interface Design Cytochrome1 (RIDC1) with a well-packed hydrophobic core. This was followed by the computational redesign of *i2*, which included six more mutations (I67L/Q71A/A89K/Q93L/T96A/T97I) and produced the second-generation variant RIDC2. Based on crystal structures, both RIDC1 and RIDC2 constructs formed tetrameric assemblies with the same supramolecular geometry as $Zn_4:MBPC1_4$, while sedimentation velocity/analytical ultracentrifugation (SV-AUC) measurements showed that the interfacial mutations in RIDC1 and RIDC2 led to the stabilization Zn-induced tetramer by nearly two orders of magnitude (i.e., tetramerization at ~100-fold lower protein/Zn concentrations) compared with the parent MBPC1 in solution. The extensive hydrophobic interactions in RIDC1 also sustained stable monomer-monomer interactions in the absence of metal ions, with a dissociation constant of ~25 μ M for RIDC1. As an indirect means to evaluate the increase in Zn-binding affinity/specifity, the authors investigated the oligomerization behavior of RIDC1 with the addition of Cu^{2+} , which prefers a square planar rather than tetrahedral coordination geometry. The crystal structure of $Cu_2:RIDC1_2$ indeed demonstrated that Cu^{2+} was bound in an unfavored, distorted coordination configuration, which suggests that the “memory” of tetrahedral Zn-coordination had been programmed onto the RIDC1 surface.

The researchers observed that the incorporation of hydrophobic interactions into *i2* did not considerably stabilize the Zn-induced tetramer, which they ascribed to the large intermonomer separation in this interface, preventing efficient hydrophobic packing. Thus, Brodin et al. added a covalent linkage in the form of a disulfide bond (C96-C96) to the *i2* interface of RIDC1 which, along with the redesigned non-covalent interactions in *i1*, further stabilized the tetrameric assembly.³⁶⁷ Such stabilization of the both the *i1* and *i2* interfaces enabled the formation of a stable tetramer ($C^{96}RIDC1_4$) even in the absence of metal ions. As intended with Zn templating, $C^{96}RIDC1_4$ exhibited nanomolar binding affinities for Zn^{2+} and selective Zn^{2+} coordination over other divalent metal ions (except Cu^{2+}). In a follow-up study, Salgado et al. demonstrated that RIDC1 variants crosslinked with flexible, bis-maleimide crosslinkers also formed stable tetramers with high selectivity for Zn^{2+} coordination.³⁷¹

Both MDPSA and MeTIR approaches have been expanded to develop metal-mediated protein oligomers with broad functionalities such as stable metalloprotein assembly *in vivo*,²¹⁰ artificial metalloenzyme activity,²⁰⁵ allosteric metalloprotein activity,^{88,232} and chemically switchable metalloprotein design.¹⁹⁶ Medina-Morales et al. demonstrated that metal-mediated oligomers designed by MDPSA and MeTIR approaches could self-assemble upon Zn^{2+} complexation in bacterial cells. When a C81 mutation was applied to $C^{96}RIDC1$ ($C^{81/C^{96}}RIDC1$), the resulting variant showed Zn^{2+} -induced self-assembly in the periplasmic environment of *E. coli*.²¹⁰ Inspired by *in vivo* metal-mediated oligomerization, Song et al. modified the surface of the Zn^{2+} -complexed $C^{96}RIDC1$ tetramer with a tripodal Zn^{2+} binding site to generate the tetramer AB3, which exhibited β -lactamase activity in prokaryotic cells (**Figure 25a-b**).²⁰⁵ This tetramer had four structural Zn^{2+} sites which stabilized the protein self-assembly (**Figure 25c**), and four catalytic Zn^{2+} sites which promoted the hydrolysis of para-nitrophenyl acetate (pNPA) and ampicillin, an antibiotic agent commonly used as a selection marker. Based on this observation, directed

evolution of the AB3 tetramer was carried out to improve the catalytic efficiency of ampicillin hydrolysis, which will be further discussed in Section 4.5.3.

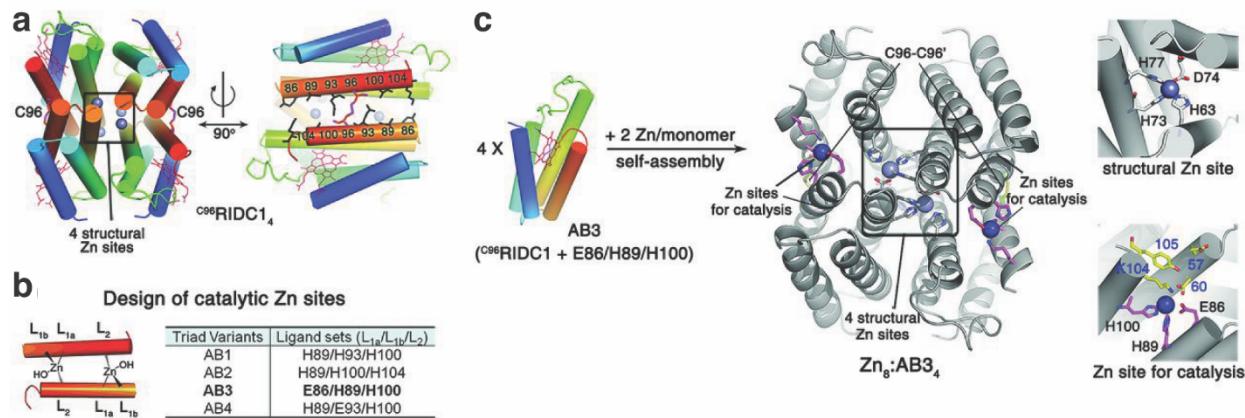


Figure 25. Overview of a designed supramolecular protein assembly with catalytic metal sites. **a)** Tetrameric structure of Zn^{2+} -bound $C^{96}\text{RIDC1}_4$. **b)** List of $C^{96}\text{RIDC1}_4$ variants with potential catalytic Zn^{2+} -binding sites on the outer surface of the tetramer. **c)** Crystal structure of AB3 which forms the stable tetramer with catalytic Zn^{2+} binding sites. Adapted with permission from Ref. ²⁰⁵. Copyright 2014 AAAS.

Kakkis et al. modified MBPC-1, one of the early variants created using the MDPSA approach, to yield a new metal-mediated trimer architecture through three generations of interface redesign.¹⁹⁶ In the first generation, TriCyt1, G70W and D66N mutations in the interface of MBPC1 induced metal-mediated trimerization in which hexa-His residues (H73/H77) ligate mid-to-late first row transition metal ions from Mn^{2+} to Zn^{2+} (**Figure 26a**). The yields of trimerization were highly metal-dependent: Co^{2+} , Ni^{2+} , Cu^{2+} , and Zn^{2+} showed high trimer populations whereas the trimer abundance was reduced in the presence of Mn^{2+} and Fe^{2+} . In the second generation, TriCyt2, Rosetta sequence design was used to add hydrophobic patches and hydrogen bonding interactions to the trimer interface, increasing the trimer abundance following association with all first-row metal ions (**Figure 26b-d**). In TriCyt3, the third-generation variant, additional hydrogen bonding and electrostatic interactions were incorporated into the trimer interface, generating a stable metal-independent trimer (**Figure 26e-g**). This stepwise “titration” of interfacial interactions into metal-templated interfaces (1) provided tunable coupling between protein quaternary structure and metal coordination, and (2) enabled the construction of high-affinity metal-binding sites.

Overall, metal-mediated oligomerization has led to significant advances in the design of artificial protein self-assemblies. This approach can be further expanded to build novel protein assemblies whose shapes and dimensions are predictable in a rational manner, and to construct diverse functional metal centers within these assemblies.

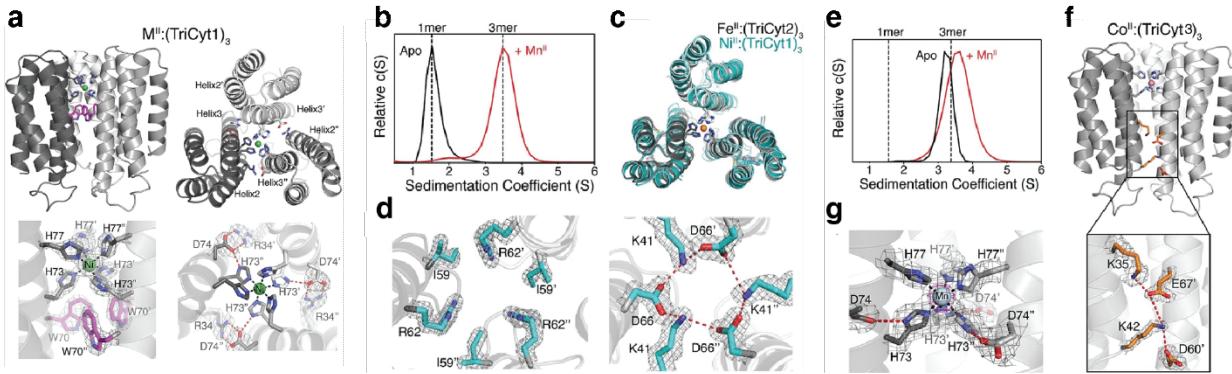


Figure 26. TriCyt series of metal-mediated protein trimers. **a)** Crystal structure of $\text{Ni}^{2+}:(\text{TriCyt1})_3$. **b)** Sedimentation coefficient distributions of apo-TriCyt2 (black) and $\text{Mn}^{2+}:(\text{TriCyt2})_3$. **c)** Crystal structure of $\text{Fe}^{2+}:(\text{TriCyt2})_3$ superimposed to $\text{Ni}^{2+}:(\text{TriCyt1})_3$. **d)** Hydrophobic packing and electrostatic interaction newly added to the trimer interface of $\text{Fe}^{2+}:(\text{TriCyt2})_3$. **e)** Sedimentation coefficient distributions of apo-TriCyt3 (black) and $\text{Mn}^{2+}:(\text{TriCyt3})_3$. **f)** Crystal structure of $\text{Co}^{2+}:(\text{TriCyt3})_3$ and hydrogen-bonding networks in the trimer interface. **g)** Hexa-his coordination environment of Mn^{2+} in $\text{Mn}^{2+}:(\text{TriCyt3})_3$. Adapted with permission from Ref. ¹⁹⁶. Copyright 2020 Wiley.

3.3 Protein cages

Cage-like proteins are hollow, well-defined nanoparticles that carry out functions that relate to encapsulation and delivery, sequestration, and containment.^{94,372-374} Inspired by the highly symmetric and aesthetically pleasing structures of these assemblies as well as their diverse functions, protein engineers have developed various strategies to build artificial protein cages, ranging from *de novo* designed structures to reengineered natural assemblies.

There are many comprehensive reviews focusing on the modification of natural protein assemblies such as virus-like particles (VLPs), chaperonin barrels, and ferritin without altering the overall architectures.^{85,94,375} In this section, we will focus on artificial or semisynthetic protein cages with new structures that have been obtained by *de novo* design or the reengineering of natural protein cages. We also note that many of these artificial protein cages display intriguing functional properties, which are discussed in Section 4.

3.3.1 Protein cages through genetic fusion

Natural protein cages are generally characterized by polyhedral symmetries (e.g., tetrahedral, octahedral, icosahedral). They are typically composed of one (or two) monomeric subunits which present several self-associative surfaces that simultaneously fulfill the multiple symmetry requirements of polyhedral assembly geometries. Guided by the principles of natural protein cage assemblies, researchers have used symmetric protein fusion to design protein cages of a wide structural variety.^{87,132,133} As described in Section 2.2.1, the fused protein constructs are obtained by linking two protein oligomers with α -helical peptides.¹²⁷ They are usually semirigid, contain naturally evolved surface patches with high selectivity for non-covalent interactions and satisfy the symmetry requirements for the target structure. Moreover, the length and sequence of the peptide linker between different domains need to be optimized so that the

fused proteins are flexible enough to avoid clashes during self-assembly and rigid enough to prevent structural heterogeneity.⁸³

To meet the symmetry requirement of target structures, protein fusion exploits the symmetry arising from the quaternary structures of the fused domains. One approach is to fuse two oligomeric proteins such that the resulting building block possesses appropriate symmetries and takes advantage of their native interfacial interactions to drive the formation of higher-order assemblies. α -helical peptides have been preferred linkers for forming rigid fusion structures and to meet the geometric design rules for target cage assemblies (**Figure 27a**). The Yeates Group was the first to apply this strategy to construct a tetrahedral protein cage.¹²⁷ Two natural oligomeric domains, dimeric M1 matrix protein from the influenza virus and trimeric bromoperoxidase, were joined through a semirigid helical linker in a predetermined orientation based on computational models. To fulfill the geometric design rules for a tetrahedral structure, symmetry axes of the trimeric and dimeric domains need to intersect at an angle of 54.7°. Achieving this specific angle is dependent on the structural rigidity of the linker. The fused protein building blocks self-assembled into a 12-subunit tetrahedral cage with a molecular weight of 550 kDa, although minor components larger or smaller than 12-subunits were also detected, reflecting possible flexibility or polymorphism in the assembled particles. To resolve the cage-like assembly in atomic detail, the design was later improved by mutating two critical residues on the trimeric component to increase the homogeneity and crystallizability of the cage. These efforts culminated in a crystal structure of a tetrahedral protein cage with a 16-nm diameter (**Figure 27b**).³⁷⁶ To further understand the origin of structural flexibility and heterogeneity, the authors prepared a double and a triple mutant of the initially designed sequence, which crystallized in five distinct forms as tetrahedral cages.³⁷⁷ After examining the oligomeric interfaces and the helical linkers in the crystal structures, the researchers found that the interfaces between the trimeric components of the fusion proteins were strictly conserved while the dimer interfaces were heterogeneous. The α -helical linkers also potentially contributed to the structural flexibility.³⁷⁷

In another study, the same group reported a porous cubic protein structure obtained using a different trimer and dimer. Trimeric KDPGal (2-keto-3-deoxy-6-phosphogalacto) aldolase and dimeric N-terminal domain of FkpA protein were genetically fused at an angle of 35.3°.³⁷⁸ A crystal structure revealed that the designed 24-subunit cubic cage (**Figure 27c**) was highly porous with openings of ~10 nm to a central cavity with a diameter of 13 nm. Such a large porous protein scaffold was proposed to serve as a container for macromolecules or a chaperone for crystallization. Notably, two additional protein structures, a 12-subunit tetrahedron and an 18-subunit triangular prism, were also observed in solution by native mass spectrometry (MS) and negative stain transmission electron microscopy (ns-TEM). These results demonstrated that while it was feasible to design protein cages of a specific shape/size, it was still challenging to achieve structural homogeneity due to the flexibility of fused protein building blocks. To comprehensively understand how solution environment influences the structural features of designed cages, Lai et al. carried out high-throughput SAXS measurements to characterize protein cages in solution. They subsequently developed force plots to measure based on their SAXS data how cage conformations varied in response to solution conditions, such as pH and

salt concentration.³⁷⁹ Along with crystallography and EM, these tools offer promise in analyzing and optimizing solution behavior of designed protein assemblies.

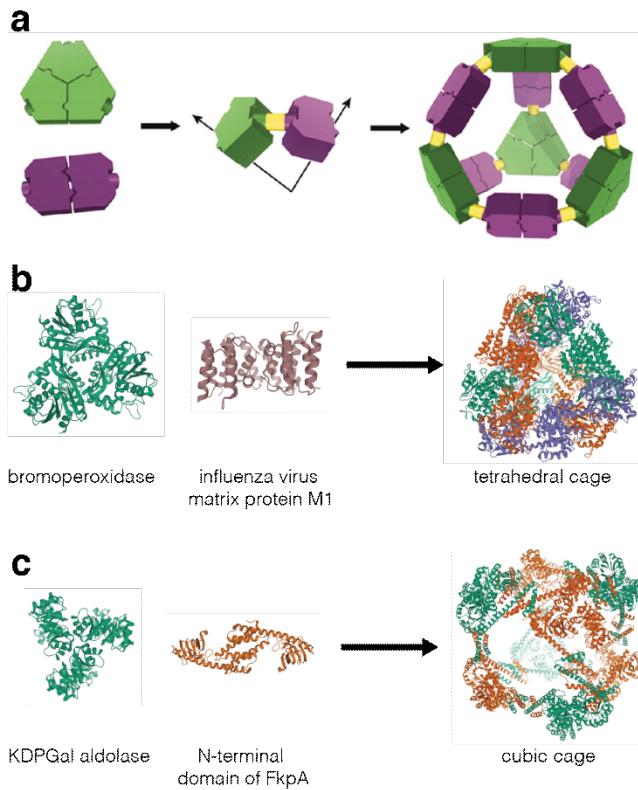


Figure 27. Design of protein cages by genetic fusion using α -helical peptides as linkers. **a)** Illustration of the geometric design principle of genetic fusion. Crystal structure of designed 12-mer tetrahedral cage **(b)** and 24-mer cubic cage **(c)**. Natural oligomers used as building blocks (left) and genetically fused components, assembled into symmetric cage structures (right). **(a)** Adapted with permission from Ref. ³⁷⁶. Copyright 2012 AAAS. **(b)** Adapted with permission from Ref. ³⁷⁶. Copyright 2012 AAAS. **(c)** Adapted with permission from Ref. ³⁷⁸. Copyright 2014 NPG.

Increasing the versatility of protein building blocks can enhance the geometric diversity of polyhedral protein assemblies. Cannon et al. designed a symmetric fusion comprised of dimer-, trimer- and pentamer-forming domains to create a self-assembling icosahedral protein cage built from sixty copies of the protein subunit (Figure 28).¹⁴³ The three-component protein building block was generated from a rigid dimer-pentamer α -helical fusion and a flexibly-linked dimer-trimer fusion (Figure 28). After the computational screening of suitable oligomeric components, a construct with favorable design properties was tested and successfully formed a 60-subunit icosahedral cage with a 2.5 MDa mass and a 30 nm diameter. Cryo-EM and dynamic light scattering (DLS) analyses suggested substantial degrees of flexibility and asymmetric deformation of the assembled cages in solution, which is not unexpected given the hollow nature of the cage and the flexibility within the double-fusion protein. This study was the first to report a designed protein cage built from components containing three different rotational symmetry elements.

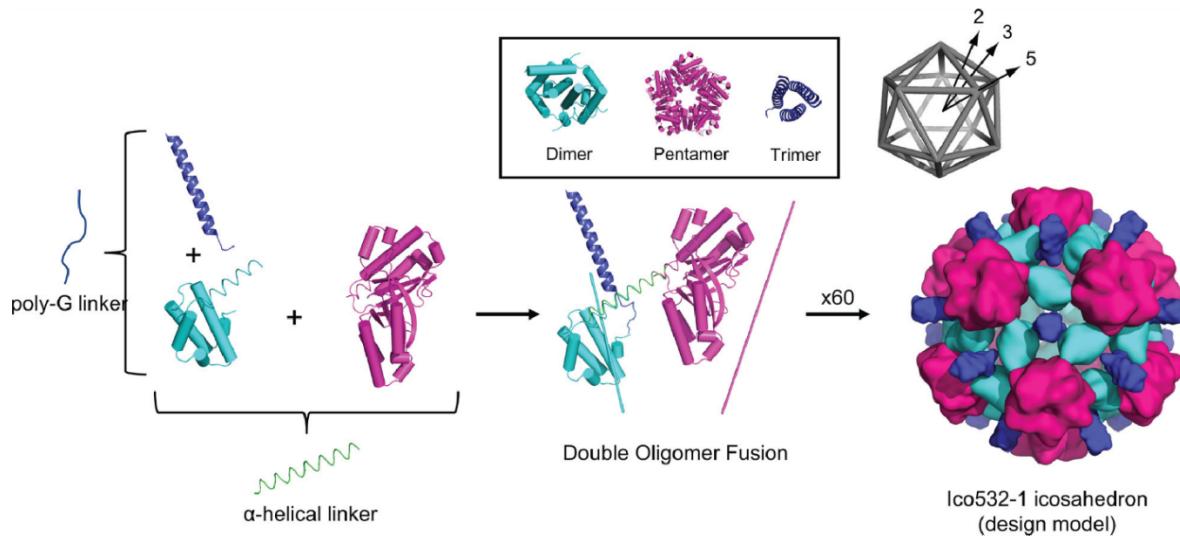


Figure 28. Design of protein icosahedral cage by a double-fusion protein with three symmetry components. Adapted from Ref. ¹⁴³. Copyright 2020 ACS.

An alternative to using helical linkers to connect two oligomeric proteins is fusing one protein domain with a *de novo* designed coiled-coil peptide domain, as introduced by the Marsh Group.³⁸⁰ The utilization of coiled-coils as building blocks for cage assembly has significant advantages.^{103,381} First, coiled-coil domains are stable, highly modular and offer access to a wide variety of modalities for oligomerization and symmetrization.^{103,381-386} Second, in the intended designs, coiled-coil domains are mainly distributed on the outer surface of the cage. With one end of the coiled-coil peptide fused to the oligomerizing protein and the other end remaining free, an extra handle is available to link a second protein on the exterior of the cage for target applications. Finally, the length of the coiled-coil domain can be adjusted, enabling the facile construction of cages with different dimensions. However, the inherent flexibility of the coiled-coil/oligomer linkages can also lead to high flexibility, precluding crystal formation and high-resolution structure determination by X-ray crystallography. In early studies, Marsh fused a trimeric KDGP-aldolase and a pair of designed complementary peptide sequences that adopt an antiparallel heterodimeric coiled-coil structure (**Figure 29a**). Upon mixing the trimeric building block fused to the positively charged peptide with its negatively charged fusion counterpart, the researchers obtained a heterodimer and a heterotetramer as determined by SEC and DLS analyses. Analysis by ns-TEM and analytical ultracentrifugation (AUC) further indicated the presence of several oligomeric states, including dimeric, tetrameric and octameric complexes.^{380,387}

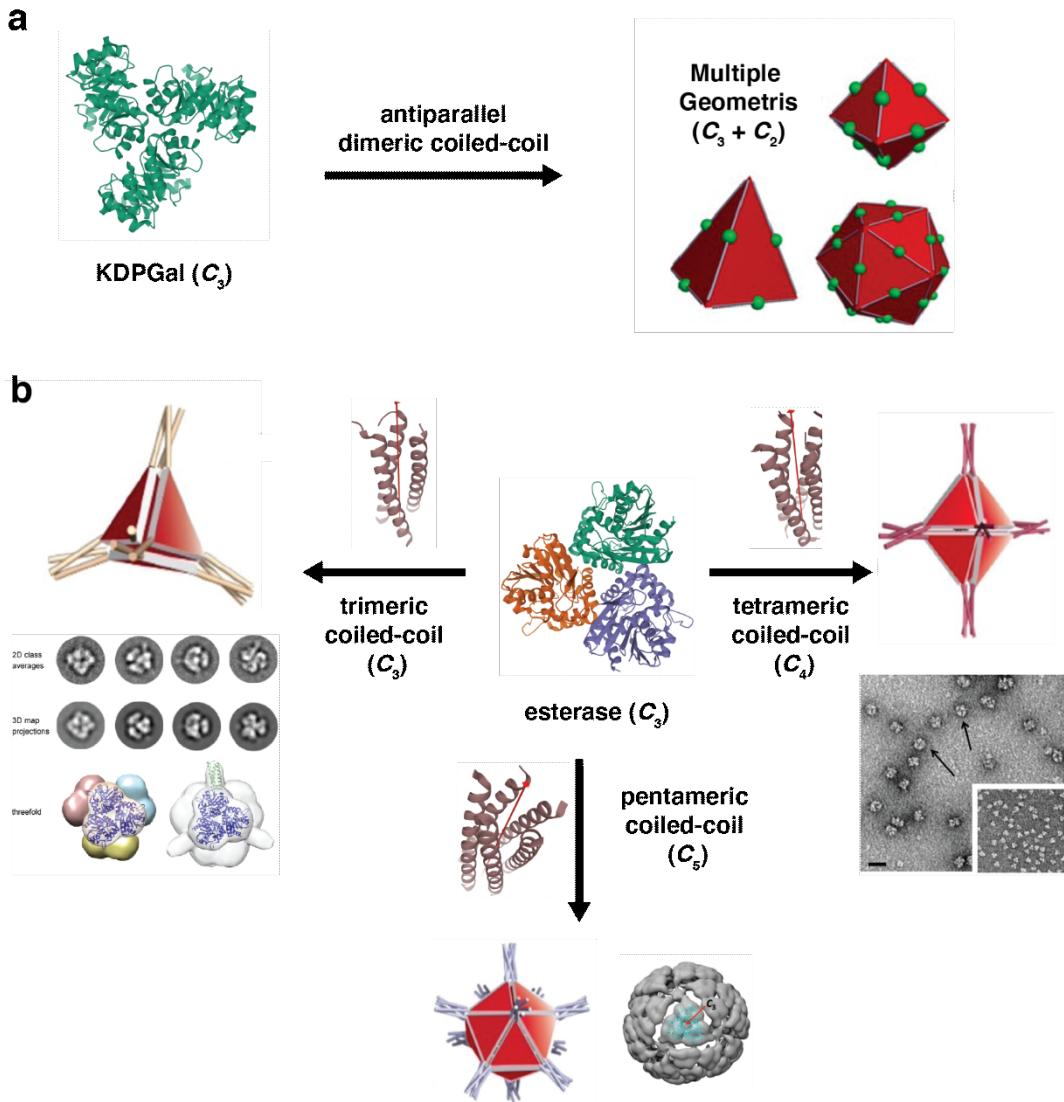


Figure 29. Design scheme of protein cages by genetic fusion using oligomers and coiled-coil domains. **a)** Trimeric KDPG-aldolase building blocks connected with antiparallel coiled-coil domains self-assemble into heterogeneous structures. **b)** Different combinations of symmetry elements are obtained by fusing the C-terminus of a trimeric protein to coiled-coil-forming peptides with different oligomerization states. Protein cages of different geometry are observed by ns-TEM or cryo-EM. **(a-b)** Adapted with permission from Ref.¹⁴¹. Copyright 2017 Wiley. **(b)** Adapted with permission from Ref.¹⁴⁰. Copyright 2016 National Academy of Science. Adapted with permission from Ref.¹⁴². Copyright 2019 ACS.

Following the proof-of-principle studies with KDPG-aldolase, the Marsh Group also employed a C_3 -symmetric esterase trimer as a building block for protein cage construction (**Figure 29b**) through fusion with a C_4 -symmetric coiled-coil domain.¹⁴⁰ A screening of different linker lengths led to the discovery of a construct with a four-Gly linker that yielded the desired octahedral cage (Oct-4). The architecture of the 24-subunit assembly was verified using AUC, native electrospray ionization mass spectrometry (ESI-MS) and ns-TEM, but its flexibility precluded the determination of a high-resolution structure by cryo-EM.¹⁴⁰ In later studies, the

researchers demonstrated the modularity of their design strategy by combining the esterase domain with trimeric and pentameric coiled-coil motifs, which yielded the desired 12-subunit tetrahedral (Tet8-5H) and a 60-subunit icosahedral cage (Ico8), respectively.^{142,388} As described in more detail in Section 4.1, Ico8 displayed very high thermal stability.¹⁴²

3.3.2 Computationally designed protein cages

As summarized in Sections 3.1.2. and 3.2.1, powerful methodologies have been developed for the computational design of binary protein-protein interfaces as well as symmetric oligomeric architectures. These methodologies have also led to several successful examples of *de novo* designed protein cages.

The Baker Group has developed a general computational approach for designing self-assembling protein cages that consists of two steps: 1) Symmetrical docking of protein building blocks in a target symmetric architecture; 2) Design of low-energy protein-protein interfaces between the protomers to drive self-assembly. Oligomeric proteins that share an element of symmetry with the target architecture have been used to reduce the number of new protein-protein interactions that need to be designed and the number of distinct new interactions that are required to overcome the entropic cost of self-assembly. They first employed natural trimeric protein building blocks to form a 24-subunit, 13 nm diameter octahedral cage (O3-33) and a 12-subunit, 11-nm diameter tetrahedral cage (T3-08/T3-10).¹⁸⁶ Based on the crystal structures of O3-33 and T3-10, the designed materials adopt the target models with high accuracy and the designed interfaces are responsible for driving self-assembly.

A subsequent study expanded the structural and functional range of designed protein architectures by using two distinct protein building blocks as subunits.³⁸⁹ A new simulation protocol related to symmetric docking was introduced and enabled the different protein components to be arranged and moved independently according to distinct sets of symmetry operators. Specifically, two distinct tetrahedral architectures (T33 and T32) were constructed from five pairs of designed proteins (**Figure 30**). In T33, the three-fold symmetry axes of two different trimeric building blocks (four copies of each) were aligned along the three-fold symmetry axes of a tetrahedron (**Figure 30a-c**), which placed one set of trimers at the vertices of the tetrahedron and the other at the face centers. Similarly, in the T32 architecture, four trimeric and six dimeric proteins were aligned along the three-fold and two-fold symmetry axes passing through the vertices and edges of a tetrahedron, respectively (**Figure 30e-h**). Subsequently, the amino acid sequences were designed by Rosetta at the new interface to stabilize the modeled tetrahedron and drive co-assembly of the two building blocks (**Figure 30d**). The structural characterization of designed structures (XRD and ns-TEM) revealed that the inter-building-block interfaces closely match with the computational models.

Although the computational design of novel interfaces between oligomeric units has led to many successful outcomes, this method generally requires screening many candidate variants, as computationally designed interfaces, which mainly rely of hydrophobic interactions, often produce poorly soluble proteins. Yeates and coworkers used the preliminary HBNet protocol (section 3.2.1) that favors the formation of hydrogen-bonding networks over exclusively hydrophobic interactions to stabilize the designed protein-protein interfaces.³⁹⁰ Two tetrahedral

cages (T33-51 and T33-53) were formed from a pair of trimeric building blocks in solution and characterized by ns-TEM. Uniform particles of ~13-nm diameter were observed for both tetrahedral cages. However, the crystal structures revealed that the interfaces of the tetrahedral cages are somewhat different from the designed models, illustrating the challenges of developing computational methods for polar interface design.

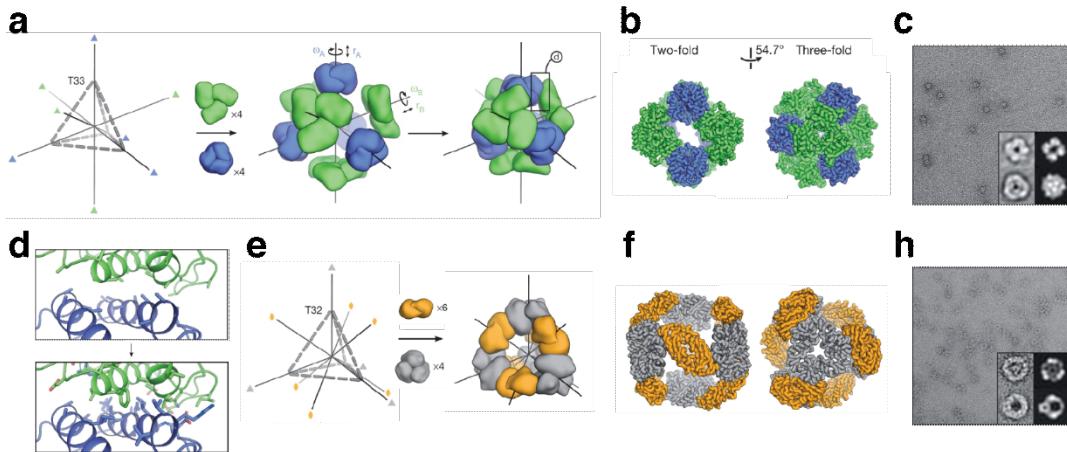


Figure 30. Overview of two-component tetrahedral protein cages developed by computational design. **a)** Symmetric docking of two distinct trimeric proteins to design a tetrahedral cage (T33). Atomic structure (**b**) and ns-TEM images (**c**) of T33-15. Two-fold and three-fold views are shown. **d)** Redesign of protein-protein interfaces between two trimers. **e)** The T32 cage is constructed with one trimeric and one dimeric building blocks. Crystal structure (**f**) and ns-TEM results (**h**) are shown. Insets in (**c**) and (**h**) show projections calculated from the computationally designed model (left) and class averages of the particles from microscopy (right). Adapted with permission from Ref. ³⁸⁹. Copyright 2014 NPG.

With the idea that protein cages with diameters above 25 nm are desirable as carriers for cargo packaging and delivery, computational methods have also been applied to construct large two-component cages (I53, I52 and I32) with icosahedral point group symmetries.^{391,392} As demonstrated in **Figure 31a**, the I53 model was formed from a combination of 12 pentameric building blocks and 20 trimeric building blocks aligned along five-fold and three-fold icosahedral symmetry axes, respectively.³⁹¹ Similarly, the I52 architecture (**Figure 31d**) consisted of 12 pentamers and 30 dimers, and the I32 architecture (**Figure 31g**) comprised 20 trimers and 30 dimers. After screening based on symmetrical docking, 71 designs of type I53, 44 of type I52, and 68 of type I32 were selected for experimental characterization. Ten designs formed stable assemblies based on SAXS and negative-stain EM and three complexes were characterized by single-crystal XRD (I53-40, I52-32 and I32-28) (**Figure 31**). It is worth noting that the I52 and I32 architectures were structurally different from any natural protein complexes characterized to date. A follow-up study investigating the assembly process of two icosahedral cages using a combination of biochemical, biophysical, and theoretical methods, revealed the dominance of cooperative assembly of two components at various stoichiometries.³⁹³

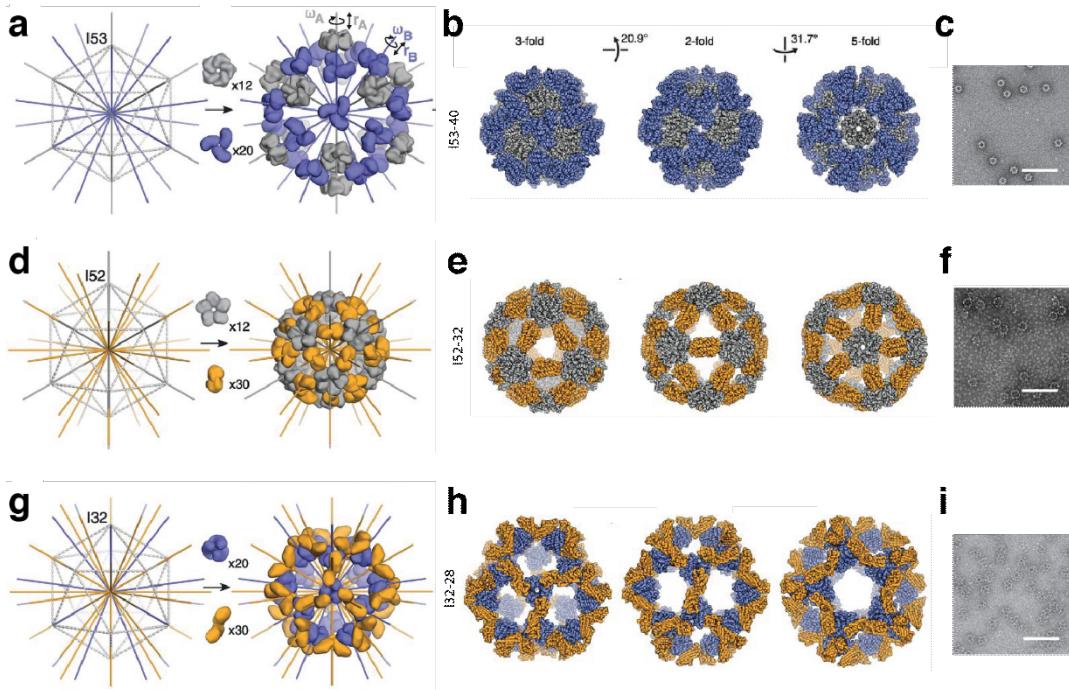


Figure 31. Overview of two-component icosahedral protein cages constructed through a computational design method. Icosahedral cages created by combining building blocks with different rotational symmetry were designed: I53-40 (**a-c**), I52-32 (**d-f**) and I32-28 (**g-i**). (**a, d, g**) show icosahedra outlined in gray dashed lines with three different combinations of symmetry axes (left). Models created by aligning pentameric, trimeric and dimeric proteins along target symmetry axes. Translational and rotational parameters are optimized by systematic screening. (**b, e, h**) Crystal structures of the designed cages. Views are shown along three-fold, two-fold and five-fold axes. (**c, f, i**) ns-TEM characterization of the designed cages (100 nm scale bar). Adapted with permission from Ref.³⁹¹. Copyright 2016 AAAS.

Along with the two-component icosahedral architectures, Hsia et al. also reported a 60-subunit protein icosahedron (I3-01) built from a single trimeric protein. The design strategy started with aligning the three-fold axes of the trimers with the three-fold axes of an icosahedron, followed by protein-protein interface optimization to drive the assembly (**Figure 32**). The resulting icosahedral cage demonstrated high stability, as further discussed in Section 4.1.³⁹²

The library of two-component protein nanocages has been further expanded with emergent functions by using antibodies as building blocks. Divine et al. constructed a series of antibody nanocages (AbCs) including dihedral, tetrahedral, octahedral and icosahedral architectures using two structural components, in which one building block was an antibody and the second a designed antibody-binding homo-oligomer (**Figure 33**).³⁹⁴ For all the target architectures, dimeric IgG antibodies served as the two-fold symmetric protein to fulfill the geometry requirement of the designed nanocages. The designed homo-oligomers were prepared by rigidly fusing antibody Fc-binding domains, helical repeat connectors, and cyclic oligomer-forming modules together. Depending on the point symmetry of the desired architectures, the symmetries of the homo-oligomer could be varied between C_2 , C_3 , C_4 and C_5 . The fused helices were tuned to optimize the intersection angles between the two-fold axis of the antibody and the principal symmetry axis of the homo-oligomers to drive the formation of

desired nanocages. The biological activity of AbCs on cell signaling and viral neutralization was investigated, as further discussed in Section 4.5.1.

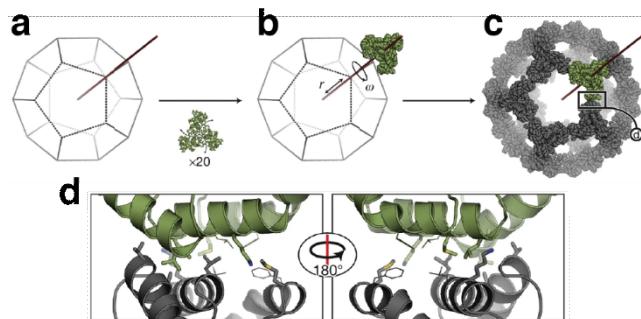


Figure 32. Design of an icosahedral cage based on single component. **a-c)** Symmetric docking of trimeric building block to an icosahedral structure. **d)** Sequence design yields low-energy interfaces after mutation on five residues. Adapted with permission from Ref. ³⁹². Copyright 2016 NPG.

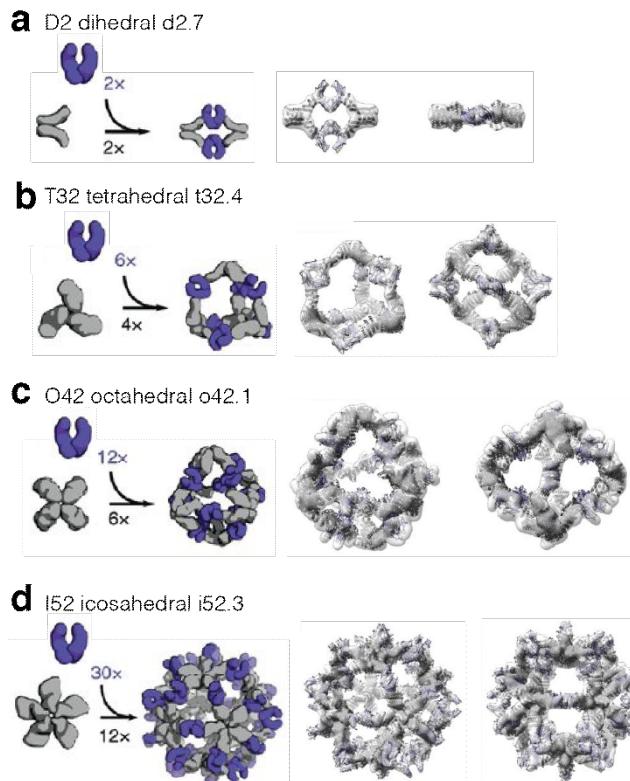


Figure 33. Model designs and structural characterizations of AbCs by cryo-EM. **a)** D2 Dihedral d2.7; **b)** T32 tetrahedral t32.4; **c)** O42 octahedral o42.1; **d)** I52 icosahedral i52.3. Combination of building blocks into a target cage geometry (left) and cryo-EM reconstructions (right). Adapted with permission from Ref. ³⁹⁴. Copyright 2021 AAAS.

3.3.3 Metal-directed protein cages

Although computational design provides a powerful means for the *de novo* design of heterogeneous, non-covalent protein-protein interactions, accurate replication of multiple, extensive, self-associative surface patches, as found in natural protein assemblies, remains challenging.³⁹⁰ Strong, reversible and directional metal-coordination interactions can be used to bypass the necessity of designing large, non-covalent protein interfaces while leveraging symmetry-directed assembly. Such strategies have enabled the construction of synthetic protein cages with unique structural, functional, and dynamic properties without depending on extensive computational and experimental work.

In early studies, Ni et al. reported tetrahedral protein cages that were assembled via Zn^{2+} coordination from monomeric cyt *cb*₅₆₂ building blocks in the crystal lattice.^{395,396} A variant of cyt *cb*₅₆₂ (CFMC1) was originally designed to form dimers through Zn^{2+} coordination and computationally designed hydrophobic interactions. While this targeted dimer structure was indeed confirmed through solution studies, the crystal structure of the Zn -adduct revealed tetrahedral, cage-like units with diameters of 8 nm, composed of twelve monomers and thirty metal ions ($Zn_{30}:\text{CFMC1}_{12}$). While serendipitous, this tetrahedral cage formed within an ordered lattice (**Figure 34**) and was used as a host to immobilize large and flexible targets such as microperoxidase for X-ray crystallographic structural interrogation (see Section 4.3.2 for more details).³⁹⁵ Notably, the $Zn_{30}:\text{CFMC1}_{12}$ cage was subsequently used as a structural template to design new cyt *cb*₅₆₂ variants that could stably assemble into dodecameric cage-like assemblies through the simultaneous coordination of Fe^{3+} and Zn^{2+} ions in solution (*vide infra*).¹⁹⁵ In a related example, Hirota and colleagues designed a domain-swapped cyt *cb*₅₆₂ dimer which formed a cage-like structure in crystal lattices through metal coordination. This pseudo- D_3 -symmetric, 6-nm diameter structure consisted of three domain-swapped dimers and was formed through coordination of several acidic residues to a $Zn\text{-SO}_4$ cluster encapsulated inside the cage cavity.³⁹⁶ The cage-like assemblies were further stabilized by the connection of the protein subunits via a hinge loop in the domain-swapped dimers.

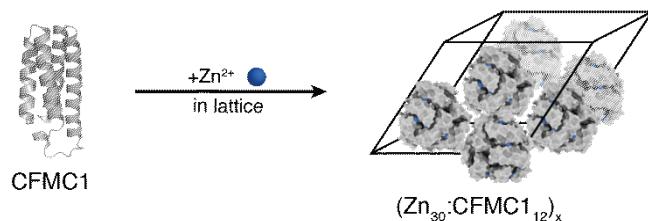


Figure 34. Zn^{2+} -mediated tetrahedral protein cage in ordered lattice. Adapted with permission from Ref. ¹⁹⁵. Copyright 2020 NPG.

Beyond the formation of protein cages within a crystal lattice, metal ion-dependent assembly of protein polyhedra in solution has been pursued by several research groups. Building on their protein oligomer-coiled-coil fusion strategy (Section 3.3.1), the Marsh Group reported a tetrahedral protein cage assembled from four copies of a trimeric esterase building block fused to a peptide sequence for metal-dependent coiled-coil formation (Tet8-M).³⁹⁷ The addition of divalent metal ions, in particular Ni^{2+} , led to the self-assembly of Tet8-M into dodecameric

assemblies in solution. Although the yield for the formation of the cages was not quantitative, the assembly process could be reversed through metal chelation or the acidification of the solution, providing a first example of stimuli-responsive protein cages obtained fully by design.

More recently, the Heddle Group reported a highly stable, Au-directed protein cage.²⁰⁰ In this work, a toroidal protein, tryptophan RNA-binding attenuation protein (TRAP), was used as the building block (**Figure 35a**). With the addition of a gold(I)-triphenylphosphine compound (**Figure 35b**), the Cys-substituted 11-mer TRAP ring assembled into monodisperse cage structures (**Figure 35c-d**) with masses greater than 2 MDa. Detailed cryo-EM analyses revealed that the TRAP-cages were held together by 120 S-Au⁺-S staples between 24 uniform rings. Notably, the TRAP cages existed in two different chiral forms (**Figure 35e-f**). The geometry of these structures is based on the Archimedean snub cube and belongs to a class of “paradoxical geometries”, considering that C_{11} -symmetric building blocks are precluded from assembling into regular polyhedra. Additionally, the Au-mediated TRAP-cage assembly displayed high efficiency, with a yield over 80%, as well as high chemical and thermal stability. Meanwhile, it could also be disassembled into the capsomer units by addition of reducing agents. Such a stable yet reversible structure highlighted the utility of metal-mediated protein-protein interactions in constructing assemblies with externally controllable assembly and disassembly behavior, making them suitable for packing and releasing cargo molecules as potential delivery platforms.

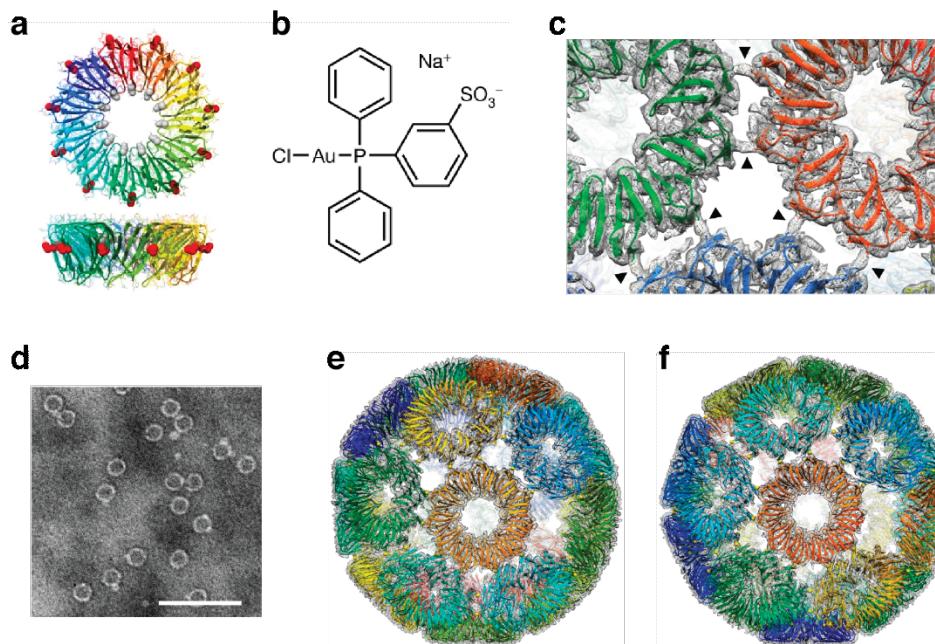


Figure 35. Metal-mediated TRAP-cage assembly. **a)** Cys-substituted 11-mer TRAP ring. **b)** Au-TPPMS structure. **c)** Left-handed TRAP-cage model and electron density map. The arrowheads indicate density bridges connecting neighboring TRAP rings. **d)** ns-TEM images of cages purified by SEC after incubating TRAP with Au-TPPMS for 3 days. Overall fits of the final TRAP-cage models onto their respective density maps: left-handed (**e**) and right-handed (**f**) structures. Adapted with permission from Ref.²⁰⁰. Copyright 2019 NPG.

As it can be gleaned from the examples of artificial protein cages described thus far, all design approaches have certain limitations. For example, the computational design or protein

fusion approaches exclusively depend on the use of natively oligomeric proteins, which eventually restricts the structural scope of the assemblies obtained and typically yields highly porous architectures that are not stimuli-responsive. Despite providing facile access to responsive behavior, metal-mediated protein self-assembly also often relies on inherently symmetric building blocks. Importantly, metal coordination does not provide sufficient selectivity for the simultaneous formation of different protein-protein interfaces which is necessary for the formation of high-symmetry architectures like protein cages. This is due to the fact that the typical metal-coordinating amino acid residues on protein surfaces (e.g., His, Cys, Glu and Asp) are considered as soft (or intermediate-soft) according to Pearson's Hard-Soft Acid-Base (HSAB) classification³⁹⁸ and have considerable overlap in terms of their coordination preferences for soft transition metal ions (e.g., Ni^{2+} , Cu^{2+} , Zn^{2+}). Owing to this lack of chemical discrimination, it has been challenging to design a heterometallic protein complex whose self-assembly is selectively guided by multiple metal ions that mediate different protein-protein interactions.

To address this challenge, Golub et al. turned to a "hard" bidentate chelating motif, hydroxamate (HA, the conjugate base of hydroxamic acid), a common functional group found in bacterial siderophores to enable exceptionally stable coordination of the "hard" metal ion (e.g., Fe^{3+}).^{399,400} HA groups preferentially form octahedral Fe^{3+} complexes with an inherent C_3 symmetry that the researchers sought to impose on protein oligomerization.¹⁹⁵ Importantly, the formation constants of $\text{Fe}^{3+}:(\text{HA})_3$ complexes ($>10^{28} \text{ M}^{-3}$) are vastly higher than those of HA-complexes of any other metal ion such that they can be considered as orthogonal to the aforementioned soft metal-ligand combinations.^{399,400} For protein derivatization, a small reagent, iodo-hydroxamic acid (IHA), was synthesized for site-selective attachment to Cys residues. The resulting Cys-HA sidechain is isosteric with that of Arg, furnishing a pseudo-natural amino acid functionality with the unique ability to chelate hard metal ions and induce C_3 symmetry on a single-residue footprint. With the IHA reagent in hand, the researchers engineered the monomeric building block cyt cb_{562} such that it could self-assemble into a polyhedron through simultaneous coordination of Zn^{2+} (soft) and Fe^{3+} (hard) ions to generate C_2 and C_3 symmetries, respectively. For this, they used the structure of the aforementioned $\text{Zn}_{30}:\text{CFMC1}_{12}$ crystalline cage assemblies as a template (Figure 34)³⁹⁵ and strategically placed Zn-coordinating His/Asp motifs and Fe-coordinating Cys-HA groups on the cyt cb_{562} surface to create the construct BMC3 (**Figure 36a-b**).¹⁹⁵

BMC3 was found by ns-TEM to self-assemble into 9-nm dodecameric cages with tetrahedral symmetry in the presence of Zn^{2+} and Fe^{3+} . Importantly, AUC experiments showed that the self-assembly of the dodecameric cages in solution required the simultaneous presence of both Zn^{2+} and Fe^{3+} . The crystal structure of resulting cage protomers ($\text{Fe}_8:\text{Zn}_{21}:\text{BMC3}_{12}$) confirmed the formation of the desired tetrahedral architecture through selective $\text{Fe}^{3+}:(\text{HA})_3$ coordination at the C_3 symmetric vertices and $\text{Zn}^{2+}-(\text{His})_3(\text{Asp})_1$ coordination in the C_2 symmetric edges (**Figure 36d-e**). A 2.6-Å cryo-EM structure revealed that the solution structure was nearly identical to that observed in the solid state. The researchers further showed that a modification of the Zn-mediated interfaces yielded a second variant (BMC4), which formed Fe/Zn-mediated hexameric cages, enabled by the flexibility of $\text{Fe}^{3+}:(\text{HA})_3$ mediated vertices.¹⁹⁵ In contrast to the typically porous structures of many artificial protein assemblies, the polyhedral BMC3 cages have tightly packed shells without large apertures and are responsive to diverse stimuli such as EDTA

and chemical reductants due to the reversible nature of the assembly process (see Section 4.2.1). These assemblies also represent the first examples of artificial protein cages constructed from simple, asymmetric building blocks like their natural counterparts.

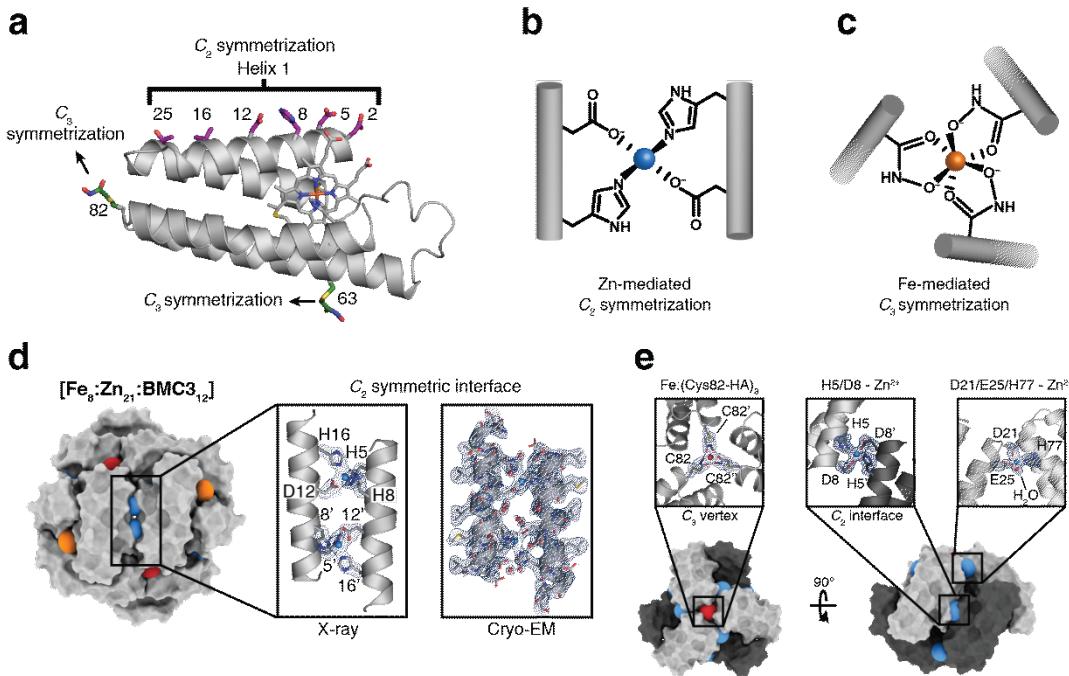


Figure 36. Metal-mediated protein cage via orthogonal chemical interactions. **a)** Structural overview of the cyt cb₅₆₂ scaffold. Salient structural elements are shown as sticks. **b)** C₂-symmetric protein dimerization induced by tetrahedral Zn²⁺ coordination of native amino acid side chains. **c)** C₃-symmetric protein trimerization induced by octahedral Fe³⁺-tris-hydroxamate coordination. Surface representations of the BMC3 cage (**d**) and BMC4 cage (**e**), with metal ions shown as colored spheres. Insets show atomic details of each metal coordination site, with the $mF_o - DF_c$ electron density omit map (blue mesh) contoured at 3 σ . Adapted with permission from Ref. ¹⁹⁵. Copyright 2020 NPG.

3.3.4 Reengineering of natural cages

Constructing cage-like protein assemblies is not limited to bottom-up design. In a complementary approach, natural cage-forming proteins can be reengineered to obtain protein architectures with alternative structures and/or new properties.

Huard et al. reported a protein-protein interface engineering strategy called “Reverse Metal-Templated Interface Redesign” (rMeTIR), which transforms a natural protein-protein interface into one that only engages in selective response to a metal ion (**Figure 37a**).⁴⁰¹ As shown in **Figure 37b**, an obligate protein complex (ferritin in this case) was used as a starting species and targeted for conversion into a form that requires metal binding for self-assembly. The first step of rMeTIR was to graft metal-coordination sites that possess the right geometry to fit into one of the three types of symmetrical (C₂, C₃ or C₄) ferritin interfaces in an unstrained fashion. The next step involved the elimination of key interfacial interactions to destabilize the interface so that the formation of the cage could only be triggered by the addition of the

appropriate metal ion. The authors applied rMeTIR to the C_2 interface of ferritin to engineer a Cu-dependent protein cage.⁴⁰¹ Specifically, a ferritin variant was first engineered with two pairs of His residues at the C_2 -symmetric interface to enable the formation of square-planar Cu coordination sites upon cage formation. The successful grafting of two interfacial Cu centers was confirmed by XRD. Subsequently, the researchers engineered the C_2 -interfacial surfaces of ferritin monomers with destabilizing mutations to create a variant that would exclusively form the cage upon Cu²⁺ binding (**Figure 37c**). The resulting variant MIC1 was indeed monomeric when isolated, but self-assembled efficiently into spherical cages upon selective binding of Cu²⁺. Other divalent metal ions, which do not favor square-planar geometries, could not induce cage formation, illustrating the requisite geometric specificity of Cu coordination (**Figure 37d**). Notably, Cu²⁺ ions acted as structural templates for cage formation, as they could be removed post-assembly without disrupting the cage architecture. The resulting apo-MIC1 cage was characterized by XRD, AUC, and ns-TEM (**Figure 37**). This example provided a convenient strategy to convert obligate protein assemblies into stimuli-responsive architectures and enabled the study of the monomeric assembly components in isolation.⁴⁰¹

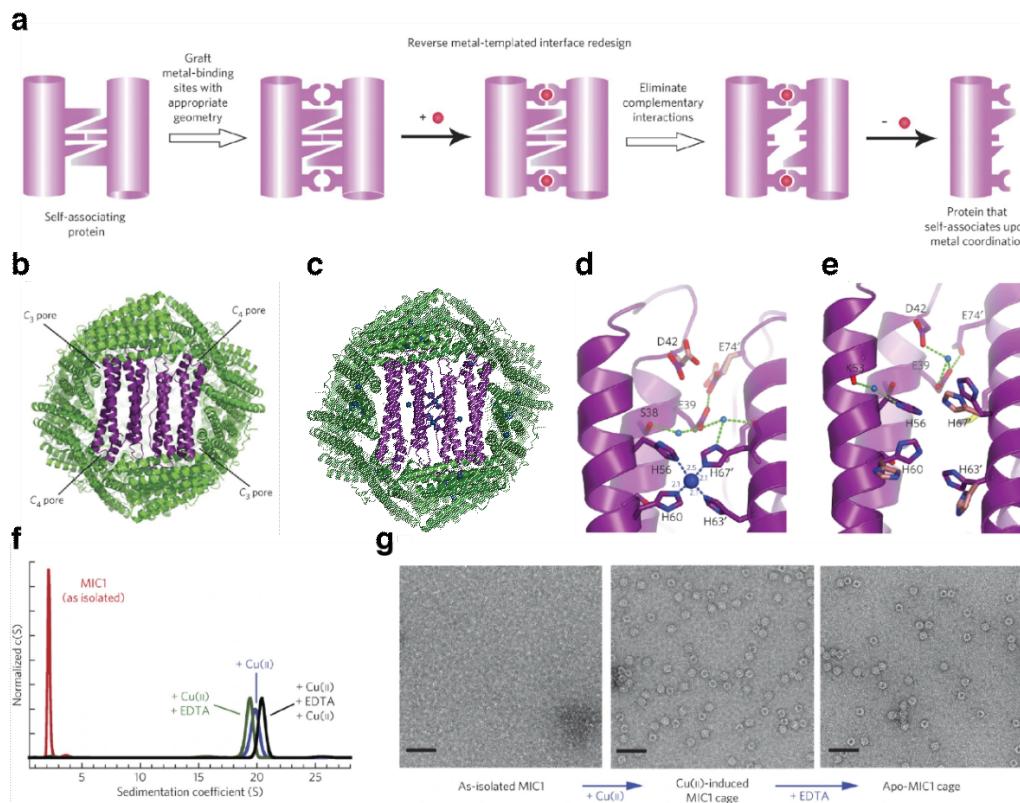


Figure 37. Copper-inducible ferritin cage assembly. **a)** Schematic illustration of the reverse metal-templated interface redesign (rMeTIR) process. **b)** Ferritin cage viewed down the C_2 symmetry axis. **c)** The crystal structure of Cu²⁺-bound MIC1 cage. **d)** Close-up view of the C_2 interface of the Cu²⁺-bound MIC1 cage. **e)** Close-up view of the C_2 interface of the MIC1 cage (apo-MIC1) obtained after the chelation of copper ions with EDTA. **f)** Sedimentation velocity profile of MIC1 in the different states: as-isolated, monomeric; Cu²⁺-induced cage; EDTA-treated cage; Cu²⁺-reconstituted cage. **g)** ns-TEM images of MIC1 in different states: monomeric, Cu²⁺-induced cage, and apo- cage. Scale bars, 50 nm. Adapted with permission from Ref.⁴⁰¹. Copyright 2013 NPG.

In a similar vein, Zhao and coworkers reported an engineering strategy named Key Subunit Interface Redesign (KSIR) to control the geometry of natural protein cages.⁴⁰² KSIR alters natural protein-protein interactions by selectively deleting a small number of “silent” amino acid residues which are not involved in interfacial interactions. Specifically, this reengineering process was carried out in three steps: 1) Determination of the key subunit interfaces in a target symmetrical protein architecture based on the crystal structure; 2) Identification of silent amino acid residues (SAAR) that were located at the key subunit interfaces; 3) Deletion of SAAR and redesign the key subunit interface, which drove the formation of new protein architecture.⁴⁰² The KSIR approach was applied to the C₃-C₄ interface of the octahedral ferritin cage. Through the selective deletion of six amino acid residues, the researchers converted the native 24-meric ferritin cage with a 12-nm diameter into an non-native 48-mer nanocage with a 17-nm diameter, whereas the insertion of six residues led to a 16-meric lenticular cage.⁴⁰³

The same group demonstrated that complete elimination of an inter-subunit interface (C₃-C₄ interface) in ferritin cages resulted in the formation of 8-meric nanorings in solution. In the crystal lattice, these nanorings were observed to stack in a face-to-tail pattern through hydrogen bonding to form nanotubes, which staggered with each other to form porous 3D protein materials as shown in **Figure 38a**.⁴⁰⁴ In a follow-up study, the 8-mer nanoring was further engineered to generate a set of discrete protein cages with different sizes and geometries via disulfide bond formation (**Figure 38b**).⁴⁰⁵ Specifically, the deletion of an intra-subunit disulfide bond in the octameric protein architecture yielded 24meric, ferritin-like nanocages in solution, while selective insertion of an inter-subunit disulfide bonds into the octamers triggered its conversion into 16-meric lenticular nanocages. Deleting the intra-subunit disulfide bond and inserting the inter-subunit disulfide bond at the same time promoted the formation of 48-mer protein cages in solution. Taken together, the examples using the KSIR strategy demonstrate the diverse protein architectures that could be obtained by altering the assembly modes of the highly associative components of natural cages.

Circular permutation is a widely used strategy to change the connectivity of secondary structure elements in a protein while maintaining the overall three-dimensional shape.⁴⁰⁶ Conceptually, this type of sequence rearrangement allows for relocation of the N and C termini of capsid subunits and can lead to morphology alternation of natural protein cages. Inspired by naturally permuted assemblies, the Yeates Group reengineered PduA, a major shell component of the propanediol-utilization (Pdu) bacterial microcompartment, by circular permutation.⁴⁰⁷ The modification caused a dramatic change in the quaternary structure of the capsomer subunits from symmetric cyclic hexamers to pentamers. Consequently, the higher-order assemblies of PduA varied from an outer shell of the microcompartment with thousands of hexamers to a dodecahedral cage with icosahedral symmetry made of pentamers.⁴⁰⁷

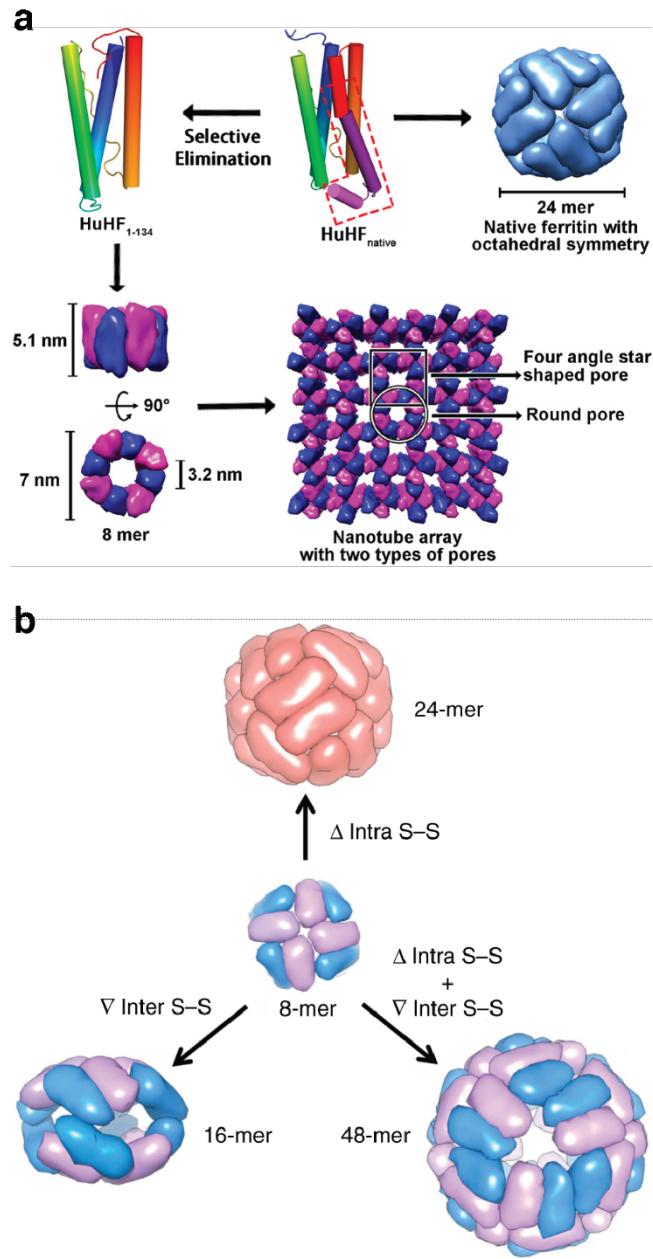


Figure 38. Ferritin cage reengineering into structures with different sizes and geometries. **a)** Conversion of the 24-mer ferritin cage into 8-mer nanorings with D_4 symmetry and a 3D porous protein lattice. **b)** Conversion of the 8-mer bowl-like nanoring into higher-order structures via disulfide bond formation. **(a)** Adapted with permission from Ref.⁴⁰⁴. Copyright 2018 ACS. **(b)** Adapted with permission from Ref.⁴⁰⁵. Copyright 2019 NPG.

The Hilvert Group reported a circularly permuted variant of *Aquifex aeolicus* lumazine synthase (cpAaLS) that self-assembled into spherical and tubular cage structures with morphologies that could be controlled by the length of the linker connecting the native termini.⁴⁰⁸ Specifically, for cpAaLS, new chain termini were inserted into a loop that faced the luminal cavity via a flexible peptide linker between residues 119 and 120 (**Figure 39a**). The authors found that

peptide length could be used to alter the cone angle of the capsomer. Shorter linkers favored the formation of larger spherical structures and 1D tubular assemblies, while longer linkers promoted the assembly of smaller spheres (**Figure 39b**). Moreover, combining the cpAaLS variant with wild type AaLS and other engineered variants enabled the co-assembly of patchwork cages in *E. coli* cells. This approach enabled encapsulation of guest proteins (GFP) in the lumen, modification of the interior and exterior surfaces of the cage via genetic fusion, and the tuning of the size and electrostatics of the particles. These studies highlight the utility of circular permutation as a potentially general strategy for tailoring the properties of cage-forming proteins and altering the self-assembly of natural protein cages.

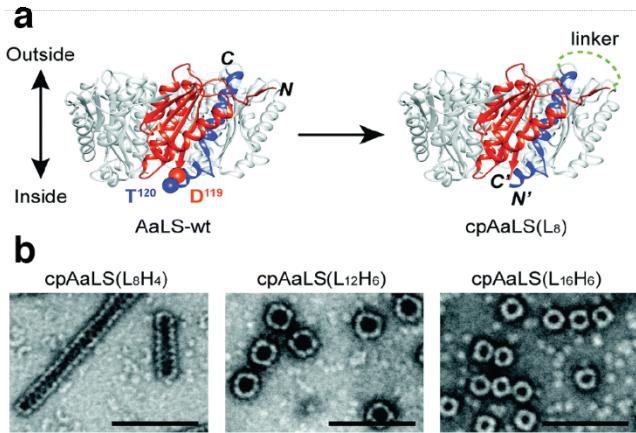


Figure 39. Circular permutation of *Aquifex aeolicus* lumazine synthase (AaLS). **a)** Design of circularly permuted AaLS (cpAaLS). **b)** ns-TEM images of cpAaLS with linkers of varying length, cpAaLS(L_xH_y), where x and y represent the number of total amino acids and His residues, respectively (100 nm scale bar). Adapted from Ref. ⁴⁰⁹. Copyright 2018 RSC.

3.4 Extended 1D assemblies

The self-assembly of proteins into extended filaments is essential for the formation of natural cytoskeletal structures (e.g., actin filaments and microtubules) in all domains of life. In nature, extended 1D assemblies have diverse functions, serving as structural supports, transport highways for molecular cargo and containers of genetic material.^{410,411} On the other hand, uncontrolled fibrillation can contribute to disease states.⁴¹² At a first glance, 1D structures may appear like readily accessible assembly design targets due to simpler design requirements than other higher-order structures with multiple symmetry elements. However, rational construction of 1D assemblies, which involves control over length, width and assembly dynamics, remains a significant challenge. Extended 1D assemblies encompass several common morphologies, from flexible filaments and nanowires^{260,291,413-416} to nanotube structures.^{202,224,270,417} There are also several subtypes of nanotubes, as they can be composed of stacked protein rings,^{417,418} assembled through the association of helical filaments,²⁷⁰ or formed when nascent 2D sheets roll up into a hollow structure due to inherent curvature or kinetic effects.^{202,419} Potential applications for extended 1D assemblies vary with morphology. They span broad functions across different length scales, from scaffolding and encapsulation of functional moieties at the nanoscale to bulk material such as gelation. It should be noted that the discussion below focuses on ordered

protein assembly generated *de novo* and does not touch on modification of naturally occurring 1D protein scaffolds such as tobacco mosaic virus (TMV) and M13 bacteriophage rods,^{374,420,421} or the extensive literature in the field of peptide filament assembly.^{104,422-430}

Runaway domain swapping and end-to-end domain stacking are the two main structural mechanisms leading to the emergence of protein filaments that have been categorized in nature.⁴³¹ In the context of protein assembly engineering, domain swapping, discussed in Section 3.1.1 in the context of protein dimer design, is also relevant to the development of protein materials formed by fibrilization.⁴³² Genetic protein fusion is another strategy that relies on existing protein interfaces for assembly design (Sections 2.2.1 and 3.2.3). Fused constructs capable of forming filaments were among early examples of designed 1D assemblies, developed as proof-of-principle illustrations toward more complex structures like designed cages and 2D protein arrays which are described in Sections 3.3.1 and 3.4.1.^{127,128} The discussion on extended 1D assemblies below focuses on protein filaments, nanowires, nanotubes, and microtubules assembled via host-guest or receptor-ligand interactions, covalent bonding, metal coordination, electrostatic interactions, and computational interface design.

3.4.1 1D assemblies mediated by host-guest and receptor-ligand interactions

Hayashi and coworkers have used hemoproteins reconstituted with artificial heme analogs as functional and structural units in protein assemblies. Taking an approach based essentially on host-guest association, the group exploited apoprotein-cofactor interactions to generate various types of protein fibers, networks and clusters by removing the native heme cofactor and replacing it with an artificial heme cofactor covalently linked to the protein surface.²⁵¹ The building blocks were produced from cyt *b*₅₆₂^{250,414,415,433} and myoglobin.^{415,434,435} The two heme *b* (protoporphyrin IX iron complex) containing hemoproteins were modified by site-selective mutagenesis to introduce a Cys residue for the covalent attachment of heme via iodoacetamide or maleimide reactive groups (**Figure 40a**). The reversible head-to-tail assembly of these building blocks gave rise to oligomeric clusters, nanorings, and fibrils depending on the length of the linker between the protein surface and appended heme. Alternatively, synthetically generated heme dimers⁴¹⁵ and trimers²⁵⁰ were used to change the connectivity between the protein subunits. Oohora et al. extended this strategy by combining heme-directed assembly with the orthogonal biotin-streptavidin ligand-receptor pair, obtaining heteromeric fibrillar co-assemblies where myoglobin and streptavidin take on alternating alignments (**Figure 40b**).⁴³⁵

The ability of macrocyclic host cucurbit[8]uril (CB[8]) to form a 1:2 complex with peptide tags such as Trp-Gly-Gly (WGG) or Phe-Gly-Gly (FGG) with binding constants on the order of 10⁹–10¹¹ M⁻² has also been exploited to generate protein nanowires (**Figure 41**). The Liu Group combined CB[8] with the dimeric glutathione-S-transferase (GST) protein fused to the FGG tripeptide at the N-termini to obtain uniform nanowires up to ~120 nm in length from an equimolar mixture of the two components.⁴¹³ The assembly strategy was extended to generate nanowires with spring-like response to Ca²⁺ addition.²⁶⁴ The protein building block was modified to incorporate a recoverin domain, which adopts a contracted conformation in the absence of Ca²⁺ and an extended conformation upon binding to Ca²⁺. The FGG-recoverin-GST construct

underwent assembly into nanowires with CB[8] as expected and the conformational change triggered by Ca^{2+} addition produced a reversible extension of fiber length by as much as 50%.

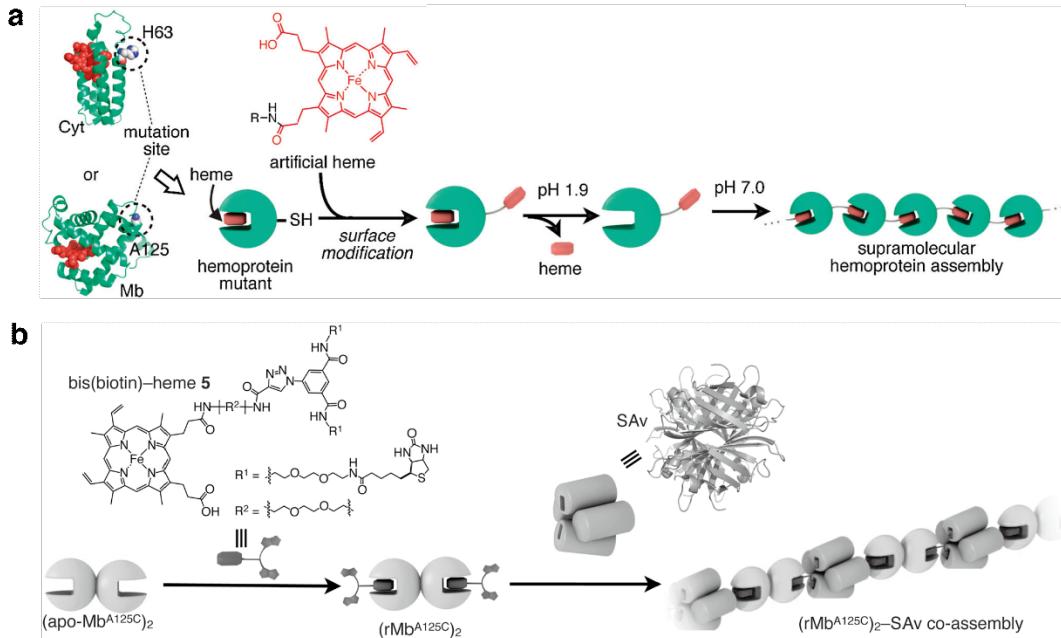


Figure 40. Fibrillar protein self-assembly through apo-hemoprotein/heme interactions. **a)** H63C and A125C mutations are incorporated into cytochrome b_{562} (Cyt) and myoglobin (Mb) respectively to position Cys residues on the protein surface for further conjugation to artificial heme derivatives. Under low pH conditions, the native heme cofactor is removed from the heme pocket. After reconstitution at physiological pH conditions, hemoprotein assemblies are obtained via artificial heme-heme pocket interactions. **b)** Heterotypic co-assembly of dimerized apo-myoglobin (apo-Mb^{A125C})₂ and streptavidin (Sav) is achieved using a bis(biotin)-heme bifunctional ligand. Adapted with permission from Ref. ²⁵¹. Copyright 2012 RSC.

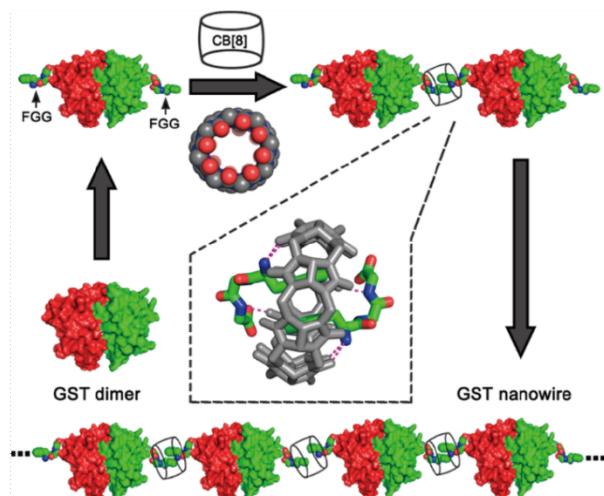


Figure 41. Formation of GST homodimer nanowires via host-guest interactions. The phenylalanine-glycine-glycine (FGG) motifs fused to the N-termini of GST serve as guest molecules binding to the cucurbit[8]uril (CB[8]) macrocyclic host. Adapted with permission from Ref. ⁴¹³. Copyright 2013 Wiley.

Yang et al. combined the specific affinity of a protein building block to a ligand with π - π stacking interactions to develop well-defined helical microtubules.²⁷⁰ Each monomer of their homotetrameric building block, soybean agglutinin (SBA), has a site that binds to N-acetyl- α -D-galactosamine (GalNAc) or α -D-galactopyranoside (Gal) in the presence of Ca^{2+} or Mn^{2+} . Within the tetramer, the sites are not oriented in the same plane, preventing the formation of 1D fibrillar structures. Instead, when combined with equimolar concentrations of designed ligands composed of (1) GalNAc or Gal, (2) a variable-length oligo(ethylene oxide) spacer and (3) a Rhodamine B (RhB) moiety that induces ligand dimerization by π - π stacking, SBA assembled into regular micron-scale tubular structures with a 26 nm diameter (**Figure 42a**). A 7.9- \AA resolution cryo-EM reconstruction revealed that the tetramers adopted a left-handed helical structure within the tubules, which were composed of 3 proto-filaments wound around a hollow, with each helical repeat being composed of nine SBA tetramers (**Figure 42b**). Ligand dimerization through RhB stacking was only observed along the length of the protofibers, while interactions between protofibers occurred between two pairs of ligands. The formation of the microtubules was reversible through the addition of β -cyclodextrin (β CD), a competing ligand that complexes with RhB, while subsequent addition of adamantane, a stronger β CD binder, led to nanotube recovery. In addition, the kinetic and thermodynamic characteristics of the assembly indicated that growth proceeded in a directional, pseudo-1D fashion.

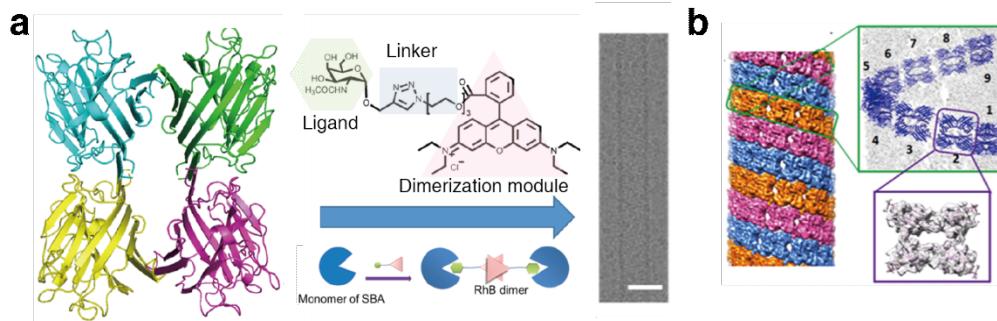


Figure 42. Formation of helical microtubules based on dual interactions. **a)** Soybean agglutinin (SBA) tetramers (left) associate with dual-function ligands to form microtubules via protein-sugar binding and π - π stacking. The ligand (middle) is composed of a protein-binding sugar moiety, a variable-length linker and a dimerizing RhB moiety. Microtubule cryo-EM micrograph (right), scale bar: 25 nm. **b)** Model of microtubule based on cryo-EM reconstruction. Three helical filaments compose the microtubule, where each helical turn consists of nine tetramer units. Adapted with permission from Ref. ²⁷⁰. Copyright 2016 ACS.

3.4.2 1D assemblies through covalent bonding

Several studies have used Cys modifications on the faces of toroidal or ring-shaped building blocks to build covalently linked protein nanotubes. Ballister et al. applied this approach to Hcp1 from *Pseudomonas aeruginosa*, a homohexameric ring structure with a height of 4.4 nm and outer and inner diameters of 9 and 4 nm, respectively (**Figure 43a**).²²⁴ While under physiological conditions the rings do not undergo self-association, they stack (non-helically) into extended tubular structures upon crystallization into a *P6* honeycomb lattice (**Figure 43b**). After examining the ring-ring interfaces within the crystal structure, the researchers targeted residues Gly90 and Arg157 for mutation to Cys. These residues were selected because they are located

on opposite sides of the ring in a suitable orientation to accommodate disulfide bond formation. Covalently linked nanotubes corresponding to the design (**Figure 43c**) were obtained both from purified protein and *in vivo*, although their lengths were limited to a few connected units and aggregates were also observed *in vivo*. Optimization of *in vitro* assembly conditions through pH and ionic strength screening, as well as addition of chaotropic polyethylene glycol (PEG) and reductant, led to the growth of longer tubes that consisted of more than 10 subunits and up to 100 nm in length. Additional controls over nanotube structure were enacted by using end-differentiation capping units, rings bearing reactive Cys residues only on one side, or pore-plugging polyamidoamine (PAMAM) dendrimers functionalized with bifunctional maleido-NHS-ester linkers, which were designed to react with inward-facing Q54C residues. Both the length-control and pore-plugging features were designed in view of potential drug encapsulation applications of the covalently linked Hcp1 nanotubes.

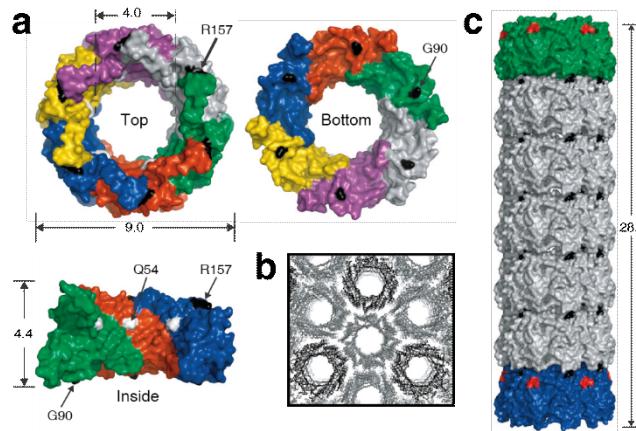


Figure 43. Design of covalently linked Hcp1 nanotube based on crystal packing. **a)** Structure of the Hcp1 hexameric ring viewed from top, bottom, and in cross-section exposing the interior surface of the pore. **b)** The honeycomb lattice packing of Hcp1 crystal viewed from the z-axis, showing stacking of Hcp1 rings into nanotubes. **c)** Side view of the nanotube formed by five $\text{G}90\text{C/R}157\text{C}$ Hcp1 rings (grey) and two capping units (green and blue) modified with Cys residues only on one side. Adapted with permission from Ref. ²²⁴. Copyright 2008 National Academy of Sciences.

Similar to the design process for Hcp1 nanotubes, Miranda et al. installed V69C and E50C mutations on the opposing faces of TRAP from *Bacillus stearothermophilus*, an 11-subunit homooligomeric ring, with the goal of assembling nanotubes through multivalent disulfide bond formation between the ring faces.⁴¹⁸ The construct, bearing additional mutations for improved solubility and future biomimetication of the cavity, was found to form nanotubes hundreds of nanometers in length in the presence of dithiothreitol (DTT) or dimercapto-1-propanol (DMP). DTT or DMP addition was essential for tube formation, as the use of alternative reducing agents β -mercaptoethanol (β -ME) and tris(2-carboxyethyl)phosphine (TCEP) did not promote nanotube formation. This finding suggested that DTT or DMP, which both bear dual thiol functionalities, might be integrated into the nanotubes as crosslinkers. The authors proposed a mechanism for nanotube formation wherein the rings first self-associated in a head-to-head fashion via hydrophobic surface burial, and the resulting dimers then underwent DTT-mediated covalent crosslinking to form nanotubes. A follow-up study by Nagano et al.⁴³⁶ confirmed the proposed

mechanism through molecular modeling, particle reconstruction from ns-TEM micrographs and control experiments performed with TRAP mutants displaying Cys residues only on one face of the ring. This study demonstrated that the C50 residues were not required for crosslinking due to stable association between ring surfaces through hydrophobic packing, while C69 residues did participate in DMP-mediated crosslinking.

Whereas the previous examples used crystal packing of ring-shaped proteins to determine the position of potential crosslinking residues but carried out nanotube assembly in solution, the Ueno Group has used protein crystals as non-equilibrium scaffolds for the formation of protein nanotubes which cannot be obtained from solution.⁴³⁷ The crystal provides a controlled microenvironment for site-specific crosslinking, ensuring uniform orientation and precise connectivity between building blocks. The covalently linked nanotubes are then recovered upon crystal dissolution. This approach was applied to generate nanotubes based on *P3₁21* crystals of ribulose-1,5-bisphosphate carboxylase/oxygenase type III (RubisCO) from the archaeon *Thermococcus kodakaraensis* KOD1. RubisCO is a homodecamer composed of double pentameric rings with *C₅* symmetry. It was modified with a single mutation, I419C, located face-to-face on the decamer surfaces such that the Ca-Ca distance for the Cys residues was expected to be 6.5 Å, or within range for bridging by a disulfide bond. Nanotube formation was carried out by co-oxidation of the ¹¹⁴⁹C RubisCO crystals with the crosslinkers 1,2-ethanedithiol (ED) triggered by the addition of hydrogen peroxide (H₂O₂). The addition of H₂O₂ alone was not effective for nanotube elongation, as mostly monomeric rings and stacks composed of up to five rings were recovered under these conditions. On the other hand, the treatment with DTT or ED yielded nanotubes up to 400 nm in length. The necessity of using flexible crosslinkers could be explained by the constrained conformation of Cys side chains within the crystal.

Abe et al. later used a similar approach to isolate covalently linked protein filaments from *Trypanosoma brucei* cysteine protease cathepsin B (TbCatB) crystals, formed by overexpression of ^{R92C/T223C}TbCatB in baculovirus-infected insect cells upon oxidation (**Figure 44a**).⁴¹⁶ Monomers form 1D arrays within the *P4₂2₁2* crystals of TbCatB assembled *in vivo*. The mutant R92C/T223C was selected because the 5.9 Å Ca-Ca distance for the residues in the wildtype crystal was deemed appropriate for disulfide bond formation by oxidation (**Figure 44b**). Due to the reducing conditions within the cell, filament formation did not occur *in vivo*. However, the oxidation reaction took place during crystal isolation under aerobic conditions, as confirmed by the structural characterization of purified crystals by X-ray diffraction (XRD). The covalently linked filaments were recovered following crystal dissolution under acidic conditions. Notably, the purified protein crystallized *in vitro* in the *P2₁* space group, wherein the monomer orientation did not allow for C92-C223 crosslinking. Interestingly, the extended 1D structures recovered upon dissolution were bundles of two designed filaments with a zigzag arrangement of proteins (**Figure 44c**). The two filaments stayed bound together due to the large contact surface area between subunits, with interface stabilization provided by non-polar contacts and hydrogen bonding.

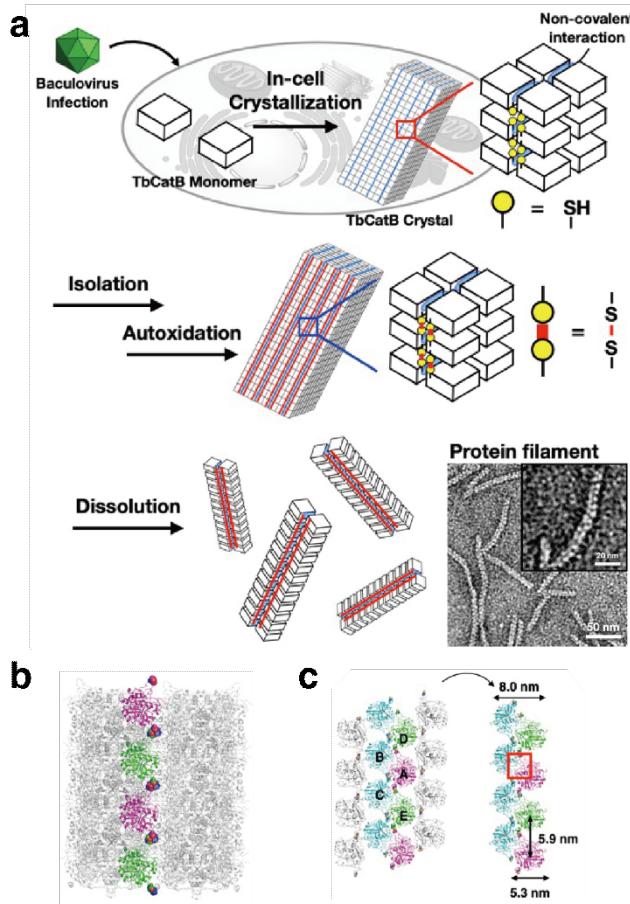


Figure 44. Covalently linked protein filaments based on the TbCatB protein. **a)** Illustration of the process to generate covalently linked $R^{91C/T^{223C}}$ TbCatB protein filament bundles. Overexpression of TbCatB gives rise to protein crystals *in vivo*. During isolation, the crystals undergo autoxidation which links monomers by disulfide bonding into filaments that can be recovered by dissolving the crystals. **b)** Design of $R^{91C/T^{223C}}$ TbCatB filament based on the $P4_2212$ crystal structure. **c)** Zig-zag protein arrangement in the two-filament bundle isolated upon crystal dissolution. Adapted with permission from Ref. ⁴¹⁶ Copyright 2021 Wiley.

3.4.3 Metal-directed 1D assemblies

Building 1D protein assemblies directed by metal-protein interactions is an appealing method due to, among other reasons, the small footprint of the required protein modification, the directionality of the metal-protein interactions, and the inherent modularity of the approach. In an early example, the Ward Group was inspired by metal-mediated assembly of coordination polymers and networks to design a protein fiber composed of streptavidin and a linear biotin-bearing metal-binding linker.²⁰⁹ The selected bis-biotinylated terpyridine (Biot₂-terpy) ligand was combined with Fe²⁺ to form the complex $[\text{Fe}(\text{Biot}_2\text{-terpy})_2]^{2+}$, a preorganized linear connector bearing two pairs of biotin moieties at either end. With the addition of Ca²⁺, the mixture of streptavidin and the connector yielded 1D polymerized assemblies that bundled into fibers with diameters and lengths on the micron and millimeter-scale, respectively. The bundles also acted as template for CaCO₃ biominerization in the presence of CO₂ vapor.

In another example of metal-mediated protein nanowires, Liu and coworkers took advantage of the homodimerization of glutathione S-transferase from *Schistosoma japonicum* (*SjGST*) and the metal-affinity of His-tags to generate self-assembled nanofibers.²⁶⁰ The building block *SjGST*-6His was generated by fusing a 6His tag at the N-terminus of the monomer. The homodimerization of the protein gave rise to C_2 symmetric building blocks presenting the His-tag arms at opposing ends, and thus serving as two-way linear units that can be connected by the addition of His-coordinating Ni^{2+} ions. Mixtures of *SjGST*-6His and $NiSO_4$ yielded flexible nanowires with a broad size distribution and a uniform height of 4.9 nm as measured by AFM. Notably, the self-assembled wires retained the enzymatic activity of the native *SjGST* and were able to scavenge the cytotoxic compound 1-chloro-2,4-dinitrobenzene (CDNB) in the presence of reduced glutathione (GSH). In a related study, Bai et al. introduced an alternative metal-chelating motif on the *SjGST* surface in order to produce metal-mediated protein nanorings with a highly controlled size distribution.⁴³⁸ By positioning the chelating bis-His clamp at residues 137 and 138, they created a Ni^{2+} coordination environment and a new protein-protein interface that was further stabilized by a network of electrostatic interactions and a hydrogen bond. This stable interface connected protein building blocks at a consistent geometry with respect to each other, such that the resulting nanoring assemblies formed within a narrow size distribution. The strength of the non-covalent interactions within the protein-protein interface was tunable with ionic strength in solution, which in turn could be used to control nanoring diameter.

The chaperonin GroEL has proved to be a versatile and functional building block for the construction of protein nanotubes.^{201,203,292,417,439,440} GroEL is a barrel-shaped complex with D_7 symmetry composed of two stacked heptameric rings. In nature, it assists the process of refolding denatured proteins. The Aida Group has developed a metal-mediated approach to produce GroEL nanotubes by modifying the barrel entrance with Cys residues conjugated to photochromic spiropyran (SP) via maleimide chemistry.⁴¹⁷ In solution, SP undergoes isomerization to merocyanine (MC), which is known to coordinate divalent metal ions in a 2:1 ratio (**Figure 45a**). In the presence of divalent metal ions (Mg^{2+} , Ca^{2+} , Mn^{2+} , Co^{2+} , Zn^{2+}), the modified GroEL barrels stacked into micron-scale nanotubes through multivalent interactions mediated by metal ions and the isomerized MC units exposed on the apical surfaces of the GroEL building blocks (**Figure 45b**). The 15 nm width of the nanotubes was consistent with the protein barrel's outer diameter. It is important to note that monovalent cations (Na^+ , K^+ , Cs^+) did not induce assembly, while trivalent cations (Fe^{3+} , In^{3+} , Ce^{3+} , Eu^{3+}) gave rise to ill-defined aggregates. The metal-mediated GroEL nanotubes were further shown to have switchable assembly properties (Section 4.2.1).^{203,439} Furthermore, a C_7 , single ring-layer GroEL mutant was used to create nanotube capping units. This mutant was originally designed by attenuating the salt bridges between the ring surfaces.⁴⁴¹ Using the same SP/MC conjugation and metal-mediated assembly strategy, Sim et al. produced nanotubes of controllable length between 40 and 320 nm by varying the ratio of the single-layer capping and double-layer polymerizing GroEL units.²⁰¹ The switchability and length control properties of the GroEL nanotubes make them an interesting candidate for use as drug delivery vehicles as detailed in section 4.3.1.

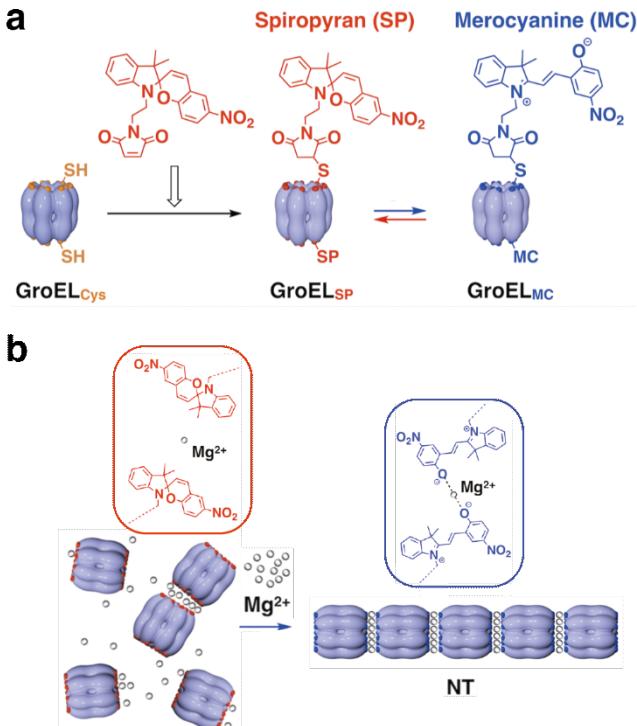


Figure 45. Mg^{2+} -mediated assembly of GroEL nanotubes. **a)** The apical domains of GroEL are labeled with spiropyran (SP) which undergoes spontaneous isomerization to merocyanine (MC) in solution. **b)** GroEL_{MC} nanotube formation induced by Mg^{2+} . Adapted with permission from Ref. ⁴³⁹. Copyright 2013 ACS.

In an alternative strategy for GroEL nanotube assembly, Kashiwagi et al. conjugated the Cys-modified GroEL barrels to DNA strands, which enabled nanotube formation through multivalent strand complementarity.⁴⁴⁰ The DNA-mediated assembly strategy was also applied to generate GroEL-DNA Janus nanoparticles bearing strands with different sequences at either end of the barrel.⁴⁴² Rather than forming discrete nanotubes, these constructs polymerized into extended structures with lamellar periodic features. The GroEL barrels were presumed to adopt a hexagonal packing when combined with additional DNA strands.

The Tezcan Group extended their metal-mediated interface design approach, in combination with computational interface design tools, to develop a protein building block capable of reconfigurable metal-mediated assembly into 1D nanotubes, 2D sheets, and 3D crystals.²⁰² This building block, termed Rosetta Interface Designed Cytochrome 3 (RIDC3), is based on an earlier metal-binding variant of cyt *cb*₅₆₂ and forms a *C*₂-symmetric dimer through Zn^{2+} coordination at a “high affinity” site composed of three His residues, H73/H77 on one monomer and H63 on the other. In this arrangement, a further “low affinity” site at the N-terminal Ala and Asp residue (A1 and D39) is available for metal coordination (**Figure 46a**). Brodin et al. screened assembly outcomes while varying pH (5.5 to 8.5) and $[Zn^{2+}]:[RIDC3]$ ratios, obtaining nanotubes with a diameter in the range of 80 nm and length up to 15 μ m and micron-scale 2D arrays under different conditions (**Figure 46b**). The nanotubes were observed at pH 8.5 and 100:1 $[Zn^{2+}]:[RIDC3]$, while 2D sheets were observed at lower $[Zn^{2+}]:[RIDC3]$ ratios and more acidic pH, closer to the RIDC3 pI (~5.3). Based on these findings and lattice packing parameters,

the authors proposed a Zn-dependent nucleation/growth mechanism for RIDC3 assemblies, where both 1- and 2D assemblies originate from small 2D nuclei. Under fast nucleation conditions (high $[Zn^{2+}]:[RIDC3]$ ratio or high pH, where high affinity sites are fully deprotonated), the large number of small nuclei preferentially form helical nanotubes. In contrast, when nucleation is slowed down, nuclei grow into large 2D sheets. The reconfigurable properties of the assembly are further discussed in section 4.2.2.

Characterization of the nanotubes by cryo-EM using helical, real-space reconstruction methods revealed that the hollow structures have C_9 helical symmetry. The surfaces of the nanotubes show a pattern of ridges and plateaus that is also seen in the 2D layers of the 3D RIDC3 crystals studied by XRD (PDB ID: 3TOM) (**Figure 46d**), supporting the mechanistic insights mentioned above. The conversion of nascent crystalline 2D sheets into helical nanotubes depends on curvature being induced along both lattice vectors within the sheet. While the coordination geometries at the Zn1 and Zn2 sites in the 2D plane (**Figure 46c**) appear to be maintained during the conversion, the coordination environment at the Zn3 site seems to deviate from the Zn3-Glu49₂ coordination observed in XRD due to the participation of adjacent acidic residues Asp2 and Asp50. Finally, the researchers also demonstrated that the stacking of the 2D layers into 3D crystals could be promoted by π - π interactions between site-specifically attached rhodamine functionalities. Collectively, the metal-directed RIDC3 arrays provided the first example for the externally tunable self-assembly of a designed protein building block into 1, 2 and 3D arrays with well-defined structures.²⁰²

Following the insight that RIDC3 nanotubes are derived from anisotropic rectangular 2D sheets, Brodin et al. redesigned the building block to generate a D_2 -symmetric tetramer that was used for metal-mediated formation of nanotubes with different diameters under kinetic control (**Figure 46e**).⁴¹⁹ They introduced residue C96 to generate covalently linked RIDC3 dimers, and residue H59 to stabilize the Zn^{2+} coordination environment that locks in the formation of the Zn-mediated tetramer (8 Zn^{2+} : 4 $H^{59/C96}$ RIDC3). The tetramers have additional Zn-binding sites with different affinities that were exploited to generate three different classes of nanotubes with different diameters (20, 48, or 68 nm) by changing pH and Zn^{2+} addition conditions in solution. The arrangement of tetramers within all nanotube types was the same, in addition to the identical structure of the tetramer monomers found in the nanotubes and 3D crystals. These results indicated that the same Zn-mediated contacts ensured the formation of all nanotube types. However, these contacts were also flexible enough to accommodate the range of curvatures that are associated with nanotubes of different diameters.⁴¹⁹

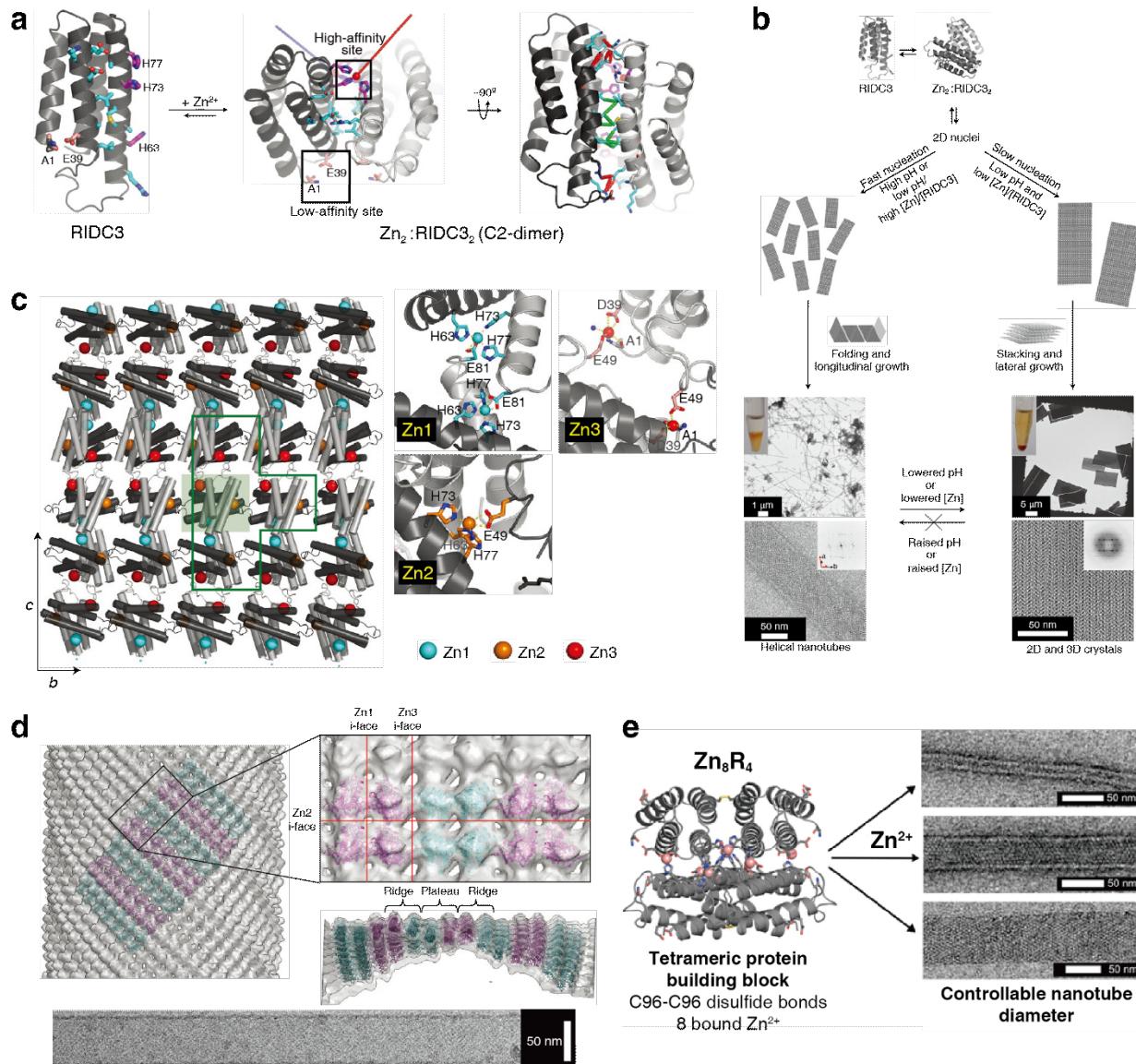


Figure 46. Zn^{2+} -mediated assembly of RIDC3, a metal-binding cyt *cb*₅₆₂ variant. **a)** The RIDC3 monomer undergoes dimerization upon Zn^{2+} coordination at the high affinity site. Amino acids that stabilize the interface in the C_2 -symmetric dimer are shown in cyan, the high- and low-affinity Zn binding sites are depicted in magenta and pink, respectively. **b)** Zn -mediated RIDC3 self-assembly into different morphologies under fast and slow nucleation conditions. High pH or high $[\text{Zn}]:[\text{RIDC3}]$ ratio enables fast nucleation, giving rise to more nuclei which form the helical nanotubes. Low pH and low $[\text{Zn}]:[\text{RIDC3}]$ give rise to fewer nuclei that grow into 2D and 3D crystals. **c)** Zn -mediated packing of RIDC3 monomers in the 2D crystal plane based on XRD (left). Close-up views of the three different Zn^{2+} binding environments found in the 2D RIDC3 crystals (right). **d)** RIDC3 nanotube reconstruction from cryo-EM micrographs (left). Zoomed-in representation of the packing of RIDC3 molecules highlighting interaction planes where different Zn^{2+} coordination environments are found (right). Cryo-EM micrograph of an RIDC3 nanotube (bottom). **e)** Structure of the tetrameric $\text{H}^{59}/\text{C}^{96}$ RIDC3 building block formed via disulfide bonding and Zn coordination (left). Three classes of nanotubes with controllable diameters form under different solution conditions. **(a-d)** Adapted with permission from Ref. ²⁰². Copyright 2012 NPG. **(e)** Adapted with permission from Ref. ⁴¹⁹. Copyright 2015 ACS.

3.4.4 Electrostatically directed 1D assemblies

Electrostatic interactions between protein building blocks and oppositely charged effector molecules have been widely employed in the design of extended protein assemblies, especially to generate ordered 3D lattices (Section 3.6.3), as well as for the assembly of 1D nanowires. Stable protein 1 (SP1) has been a choice protein building block for the formation of nanowires following early observations that it naturally undergoes stacking, a property that was exploited by Shoseyov and coworkers to form functionalized extended SP1 structures.⁴⁴³⁻⁴⁴⁵ SP1 is a homooligomeric dodecamer isolated from aspen (*Populus tremula*) plants. It is composed of two rings that stack by hydrophobic association into a double-layer structure with six-fold symmetry. The exposed ring surfaces are highly negatively charged due to the distribution of acidic residues (**Figure 47a**).

This structural property was exploited by the Liu Group in a series of reports to generate ordered SP1 nanowires by combining the negatively charged rings with non-proteinaceous building blocks bearing positively charged surfaces, including quantum dots (QDs),²⁹¹ soft nanoparticle PAMAM dendrimers (**Figure 47b**)^{281,446} and core-crosslinked micelles (CCMs).⁴⁴⁷ Such combinations of heterologous elements yielded nanowires formed by “sandwiching” the positively charged nanoparticles between SP1 rings through multivalent electrostatic interactions. The size match between protein ring and nanoparticle was a determinant factor in the structural outcomes,²⁹¹ an effect that has also been observed in the case of Cys-modified Hcp1 rings co-assembled with gold nanoparticles (AuNPs) of various sizes (section 4.3.4).⁴⁴⁸

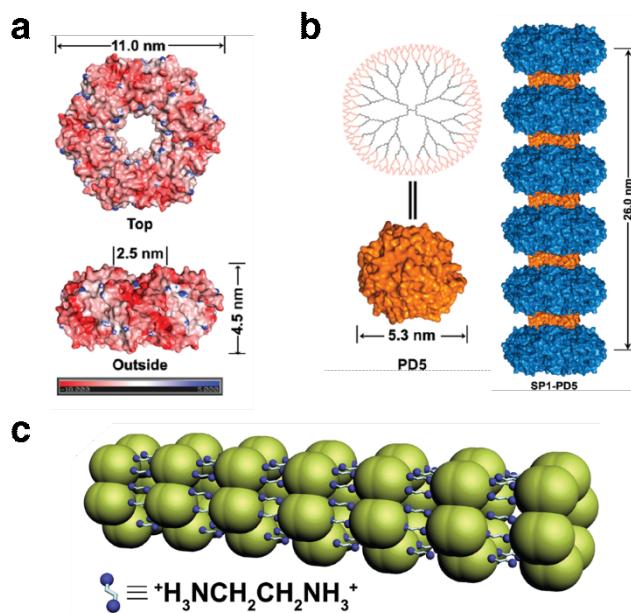


Figure 47. Assembly of SP1 nanowires via multivalent electrostatic interactions. **a)** Top and side views of the SP1 ring with structure and charge distribution on the surface. The color scale goes from negative (red) to positive (blue) charge. **b)** Illustration of a generation 5 PAMAM dendrimer (PD5) (left) and co-assembly of PD5 and SP1 into nanowires (right). **c)** Illustration of SP1 nanowire formation mediated by multivalent interactions with ethylenediamine. **(a-b)** Adapted with permission from Ref. 52. Copyright 2015 ACS. **(c)** Adapted with permission from Ref. ²³¹. Copyright 2016 RSC.

Miao et al. also showed that ethylenediamine (EDA) added to SP1 in 1:1 ratio with the exposed carboxyl groups on the top and bottom faces of the rings mediated multivalent electrostatic interactions that stabilize ordered nanotubes hundreds of nanometers in length (**Figure 47c**).²³¹ The authors further used a “zero-length” crosslinking strategy to covalently link the proteins to EDA by adding EDC and sulfo-NHS to the nanowires.

3.4.5 Computationally designed 1D assemblies

Nucleoprotein architectures, although ubiquitous and highly functional in nature,⁴⁴⁹⁻⁴⁵² remain somewhat rare among rationally designed protein assemblies. The Mayo Group applied computational protein interface design to generate co-assembling protein-DNA nanowires.²⁹⁹ The wires are composed of double-stranded DNA (dsDNA) and a protein building block that has the dual functions to bind to a dsDNA and homodimerize (**Figure 48a**). They selected engrailed homeodomain (ENH) from *Drosophila melanogaster* as the protein building block due to its high binding affinity to a dsDNA motif (TAATNN). Moreover, ENH, a three-helix protein, has a surface (the exterior faces of helices 1 and 2) amenable to computational homodimer interface design, as it is oriented opposite to the DNA-binding helix (helix 3). Fast Fourier transform-based docking was used to generate C_2 -symmetric ENH homodimer models that underwent computational redesign to minimize the energy of the interface and were iteratively screened and improved using a molecular dynamics protocol. The designed homodimer retained high, specific affinity to the target dsDNA sequence and oriented the binding domains opposite to each other. To generate linear wires from the co-assembly, the dsDNA component was designed to position two protein-binding sites 180° apart on the double helix. When imaged by AFM, the nanowires had lengths on the order of 300 nm, corresponding to ~60 repeat units. The co-crystal structure was also solved by X-ray diffraction, revealing that the wires indeed form following the proposed mechanism. However, two different protein-DNA binding configurations were observed, causing the infinitely repeated protein-DNA wires to be slightly kinked (**Figure 48b**). The strategy of combining the design of protein oligomers with DNA binding could be expanded to give rise to more complex and diverse ordered co-assemblies in 2D and 3D.

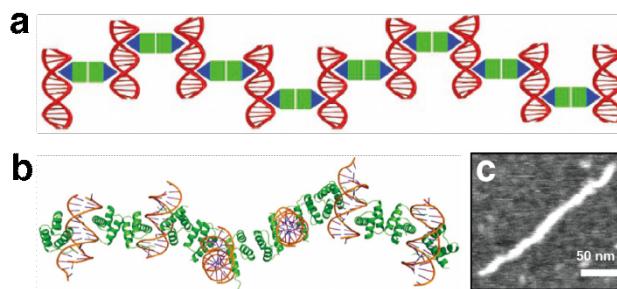


Figure 48. Protein-DNA nanowires. **a)** Illustration of the co-assembly of computationally designed ENH dimers and dsDNA into a linear nanowire. **b)** Co-crystal structure shows that kinked wires are formed. **c)** AFM image of the protein-DNA nanowire. Adapted with permission from Ref.²⁹⁹. Copyright 2015 NPG.

More recently, the Baker Group described a general computational approach for the design of self-assembled helical protein filaments with controllable geometries and diameters

based on *de novo* designed, idealized helical repeat protein building blocks.¹⁸⁷ Typically the protein interface design process is simplified by using building blocks with internal symmetry to reduce the number of interfaces that require redesign. However, reversibly forming protein filaments found in nature tend to be composed of asymmetric building blocks and thus have multiple interfaces. This problem was simplified by considering that helical symmetry originates from repeated application of a single rigid-body transform (six degrees of freedom) and accounting for cyclic symmetry. The design approach (**Figure 49a**) started with an asymmetric protein monomer structure and generated a second randomly oriented copy in physical contact with the first. This was accomplished by applying random rotation, choosing a direction, and sliding the second copy into contact with the first. The filaments considered were generated not only by the rigid-body transform relating the two contacting monomers, but also by the n^{th} root of this transform (where $n = 2 - 5$), and with cyclic symmetry generated by the application of cyclic symmetry operations (C_n) around the superhelical axis. In all cases, the next step generated several repeating turns of the full filament by re-applying the rigid-body transformation and cyclic symmetry operations, eliminating clashing geometries, and making sure that there is at least an additional interface beyond the one generated in the initial sliding-into-contact step. Filament architectures were selected based on low predicted energy at multiple interfaces, and Rosetta combinatorial sequence optimization was carried out on a central monomer, propagating the sequence to all other monomers.

The resulting designs were filtered for high shape complementarity, low monomer-monomer interaction energy and few or no buried unsatisfied hydrogen bonds. Following this design process, 124 designs were selected for experimental testing. Of these designs, 34 were found to form 1D assemblies and six were studied by cryo-EM. Among the six designs spanning a range of architectures, good correspondence was found between the designed model and experiment in terms of monomer orientations and packing in the filament. However, interfacial interactions varied considerably. While four of the designs had an excellent match to the model, monomers in another filament, designed with C_1 symmetry, shifted to produce a C_3 symmetric structure (**Figure 49b**).¹⁸⁷ Furthermore, a series of filaments with variable diameters were produced by changing the number of repeat units within the monomer (**Figure 49c**). The filament assembly dynamics were monitored both *in vitro* and in living cells. In addition, seeded fiber growth was achieved from a surface, and capping units were used to control growth and induce disassembly. The nucleating and terminating units were designed by selectively eliminating interaction interfaces.¹⁸⁷

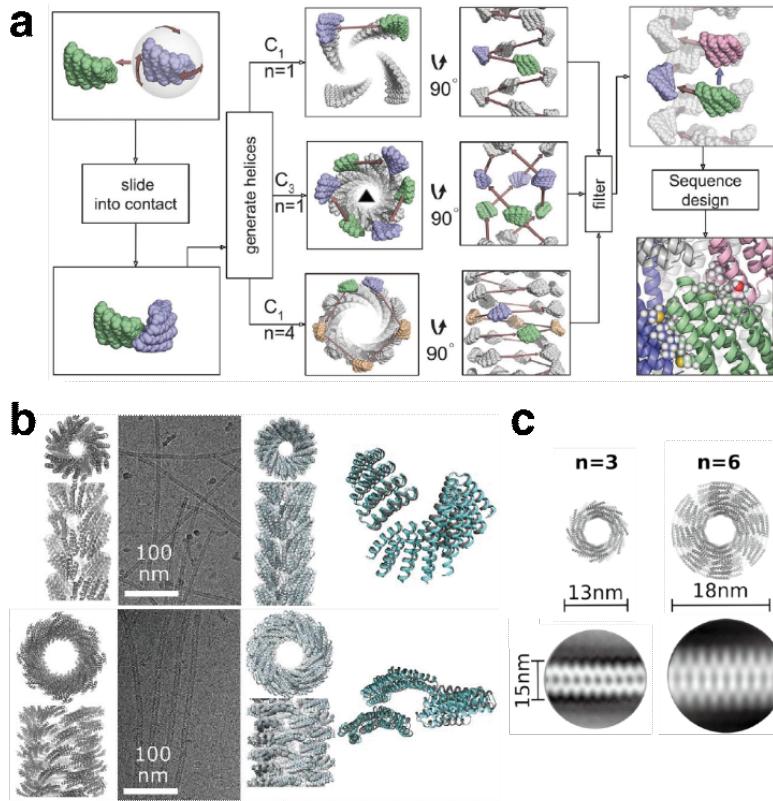


Figure 49. Computationally designed helical protein filaments. **a)** Computational protocol to design self-assembling protein filaments. A copy of an asymmetric protein monomer is randomly rotated and moved, then slid into contact with the first monomer. The operation is repeated to generate helices. Cyclic symmetry and the ordering of contacting units are screened, and sequence of the monomer is redesigned to optimize the interfaces. **b)** From left to right: Computationally designed models, cryo-EM micrographs, cryo-EM structures, and overlay of designed model and cryo-EM structure for the C3 symmetric DHF91 (top) and the C1 symmetric DHF79 (bottom) designs. **c)** Fibers with variable diameter can be generated by changing the number of repeat units within the monomer. Computationally designed models (top) and 2D class average structures (bottom) are shown for two variants of the DHF58 filament. Adapted with permission from Ref. ¹⁸⁷. Copyright 2018 AAAS.

3.5 Extended 2D assemblies

Two-dimensional protein lattices are common in nature, particularly in association with cellular membranes. Bacteria and archaea produce protective surface layers (S-layers) consisting of proteins or glycoproteins.^{117,453} Bacteriorhodopsin forms 2D patches, known as purple membranes, covering nearly half of the surface area of the archaeal cell to harvest light and pump protons.⁴⁵⁴ Other examples of naturally formed 2D protein lattices include gap-junction plaques⁴⁵⁵ and water channels.⁴⁵⁶ This section describes the efforts which, inspired by nature, have taken advantage of different design strategies to construct ordered 2D protein arrays. As in the case of other types of supramolecular and extended structures, symmetry takes on an important role in building block selection and array design, as 2D crystals are restricted to 17 possible plane group symmetry configurations.⁴⁵⁷ Certain lattice arrangements are more

amenable to design than others, as they require fewer unique interfaces or symmetry element combinations to be realized.⁴⁵⁸ In general, building blocks with two-, three- and four-fold symmetry have been highly represented in the area of 2D protein assembly so far. Given the technological relevance of self-assembled 2D materials^{459,460} and the functional diversity of proteins, ordered 2D protein assemblies built from the bottom up represent a highly sought-after platform that could find diverse applications, encompassing sensing, catalysis, light-harvesting, and nanomedicine.

For reviews of natural 2D assemblies functionalized for application development, or 2D materials based on peptides and proteins readers are encouraged to consult other recent publications.^{117,423,461-463} The section below reports on recent developments in 2D protein assembly design by way of diverse strategies such as genetic protein fusion, covalent bonding, metal coordination, receptor-ligand interactions, computational design, and interface-assisted assembly.

3.5.1 2D protein assemblies through genetic fusion

It is essential to consider the symmetries of the protein building blocks and the linkages between them to tile them into ordered 2D lattices. For example, square lattices are generated by combining a four-fold symmetric building block with a linking strategy that yields two-fold symmetry, as articulated by Yeates and colleagues.¹³¹ This simple combination can be achieved by several different approaches. The Noble Group proposed linking native protein assembly elements that have different rotational symmetries by genetic protein fusion. The elements are connected into a fusion construct at their termini along a shared symmetry axis, thus promoting a given assembly outcome by eliminating degrees of freedom from the fused building blocks.¹²⁸ The assembly elements can originate from homologous or heterologous systems. Using this strategy, the authors demonstrated the formation of ordered 1D and 2D protein arrays.¹²⁸ Specifically, two different sets of constructs were tested to produce 2D arrays. Aminolevulinic acid dehydrogenase (ALAD) was used as a homologous D_4 component in both cases. It was fused either to Streptag I peptide and assembled with streptavidin (heterologous assembly motif), or with the coiled-coil forming peptides Lac21E and Lac21K (heterologous assembly motif) for co-assembly into a 2D lattice with variable cell dimensions enabled by changing the length of the peptide (**Figure 50**). The authors noted that the length and flexibility of the linker, as well as the alignment of the fusion point to the shared symmetry axis along which the assembly is designed to grow, are all important parameters to promote formation of the desired 2D structures.

Poulos et al. implemented a genetic fusion strategy to design a crystalline 2D lattice⁴⁶⁴ by combining the concept of rotational symmetry matching described above¹²⁸ with the polymer-driven crystallization approach reported by Bowie and coworkers.⁴⁶⁵ They created a TTT-FUR fusion construct composed of three Tel-SAM domains (TTT) and the ferric uptake regulator (FUR) domain.⁴⁶⁴ Tel-SAM domains formed one dimensional fibers with a 2_1 screw axis under pH control, while the FUR domains underwent homodimerization. This combination allowed TTT-FUR fusion proteins to form 2D sheets with the FUR domains stabilizing contacts between polymerized TTT fibers. Screening of crystallization conditions yielded 3D crystals, where the 2D layers showed stacking one on top of the other. As described earlier (Section 2.2.1), the protein fusion strategy takes advantage of naturally evolved protein-protein interfaces, combining them

into building blocks designed for a target structure based on symmetry considerations. In extended assemblies, just as in finite assemblies built by genetic fusion, controlling the rigidity of the fusion construct and optimizing the orientation of different domains through linker design are important challenges as these parameters determine assembly outcomes.

Assembly types	Assemblies	Design	Visualization
a D ₄ Homologous D ₂ Heterologous	E. coli ALAD (1b4e) Streptavidin / Streptag I (1rlt)	14 nm	140 nm 500 nm
b D ₄ Homologous C ₂ Heterologous	E. coli ALAD (1b4e) Lac21E / Lac21K (model)	11 nm	110 nm

Figure 50. Genetic fusion approach to 2D protein assembly. Self-assembly of binary 2D arrays along the C₂ symmetry axis shared by the components. The homologous D₄-symmetric protein building block ALAD is combined with **a**) a heterologous D₂-symmetric assembly motif (Streptavidin/Streptag I) or **b**) a heterologous C₂-symmetric assembly motif (Lac21E/Lac21K). Adapted with permission from Ref. ¹²⁸. Copyright 2011 NPG.

3.5.2 2D assemblies through covalent bonding

In addition to imposing two-fold rotational symmetry, the formation of covalent bonds between building blocks provides a facile means to create stable protein arrays while only incorporating a minimal number of mutations into the building blocks. Tessellation of the C₄-symmetric RhuA building block, first employed by Ringler and Schulz (section 3.4.4),⁴⁶⁶ inspired Suzuki et al. to take advantage of strategically positioned Cys residues to create C₂-symmetric linkages for a straightforward path toward generating 2D protein lattices. The four corners of the RhuA tetramer were identified as appropriate positions to install a single Cys residue (C98) for lattice assembly via disulfide bond formation under slow oxidation conditions (**Figure 51a**).²²⁵ An octameric D₄-symmetric RhuA variant (F88/C98) was also investigated. Both building blocks formed extensive, highly regular 2D arrays with p42₁2 plane group symmetry. Notably, the self-assembly process was reversible under reducing conditions. An additional RhuA variant bearing a bis-His motif (H63/H98) at the four corners was shown to form p4 lattices through metal coordination, an alternative C₂-symmetric linkage (**Figure 51b**). The formation of this RhuA assembly could also be reversed by EDTA addition. Among the three variants, the ^{C98}RhuA lattices connected through a single disulfide linkage at the corners exhibited the largest array size, the lowest defect frequency, and dynamic lattice behavior, as discussed further in Section 4.2.2.

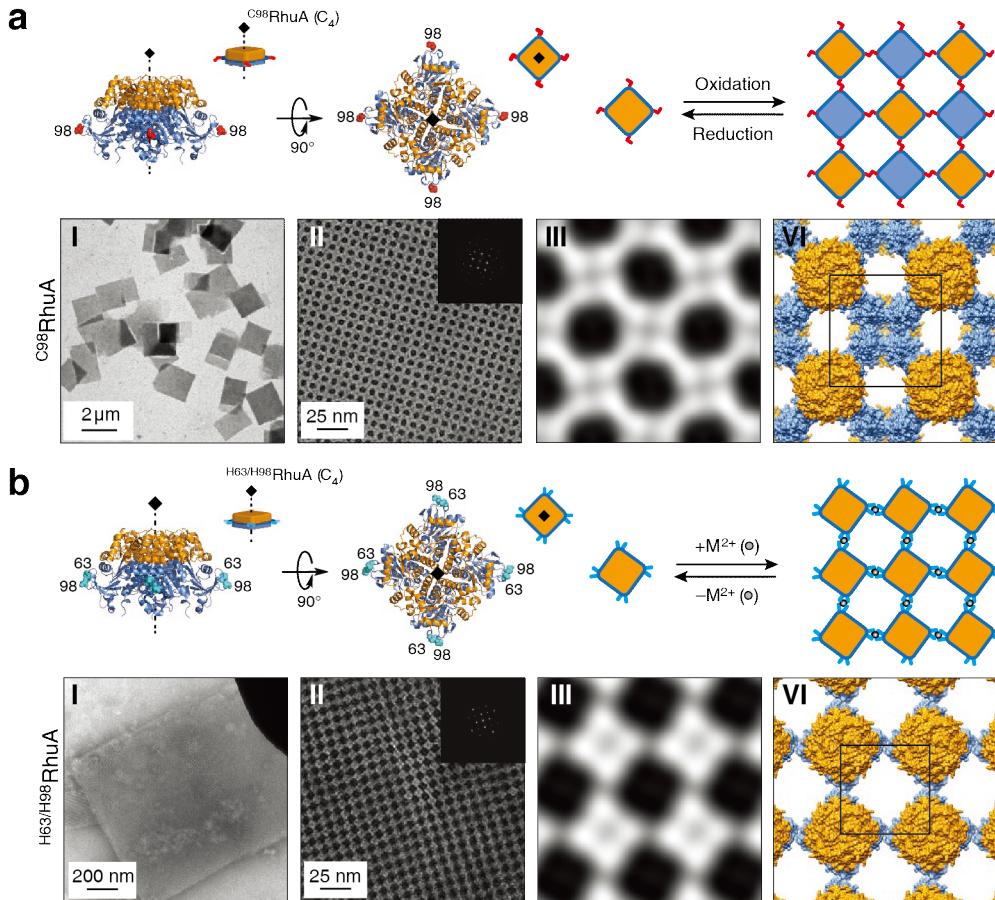


Figure 51. Assembly of 2D crystals based on C₄-symmetric RhuA modified at the corners to enable C₂ connectivity between building blocks. **a)** C⁹⁸RhuA forms a p42;2 lattice following oxidation. **b)** H⁶³/H⁹⁸RhuA forms a p4 lattice following divalent metal coordination. TEM characterization of 2D crystals of RhuA variants: (I) Low-magnification views of RhuA crystals, (II) high-magnification views of RhuA crystals with the Fast Fourier transforms (inset), (III) reconstructed 2D images, and (IV) structural models based on (III). Adapted with permission from Ref. ²²⁵. Copyright 2016 NPG.

Zhao et al. exploited oxidative coupling of tyrosine residues as an alternative strategy to produce covalently crosslinked 2D protein arrays.²³⁰ They selected the ring-shaped building block SP1, discussed above (Section 3.4.4) in the context of extended 1D assemblies, as a building block. Previously, Shoseyov and coworkers had observed the formation of long-range 2D arrays from wild type⁴⁴⁵ and polyHis-tagged⁴⁶⁷ SP1 rings following assembly at an air-phospholipid and an air-water interface, respectively. In contrast, Zhao et al. carried out the assembly in bulk solution after installing an S98Y mutation at the SP1 ring periphery (Figure 52). They reported an enzymatic crosslinking strategy using either a single enzyme (horseradish peroxidase (HRP) in the presence of H₂O₂), or a cooperative dual enzyme system (HRP and glucose oxidase (GOx) in the presence of glucose).²³⁰ In both cases, hexagonally packed sheets, hundreds of nanometers in size, were formed following the crosslinking reaction. Sheet stacking into multilayered structures could be controlled by adjusting pH, with lower pH conditions closer to SP1 pI (~4.3) favoring stacking due to reduced repulsion between the charged sheet surfaces.

In a follow-up study, the authors demonstrated that ^{S98Y}SP1 could be rapidly converted to covalently linked nanosheets with light-controlled oxidative crosslinking of tyrosine residues by using $[\text{Ru}(\text{bipy})_3]^{2+}$ as a photosensitizer in the presence of ammonium persulfate (APS).⁴⁶⁸

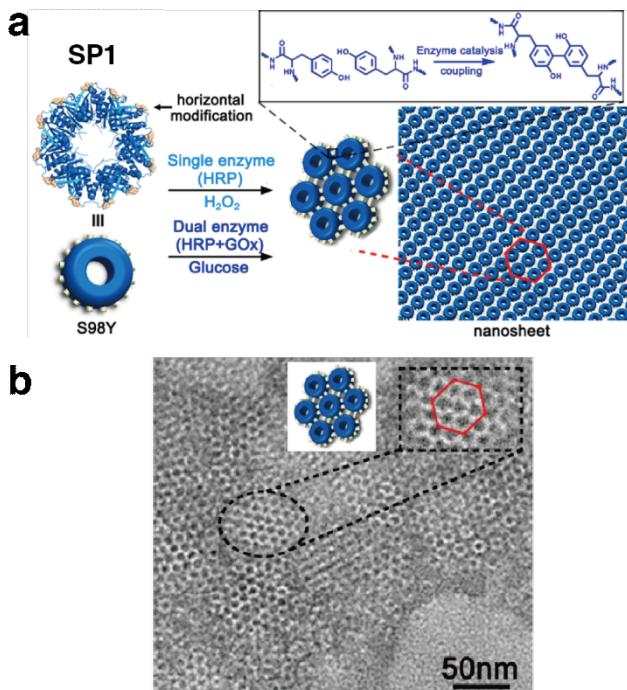


Figure 52. Covalently linked protein nanosheets based on the S98Y variant of cricoid stable protein 1 (SP1). **a)** Tyrosine crosslinking at the periphery of the SP1 disk can be carried either via a single enzyme pathway with horseradish peroxidase (HRP) and H_2O_2 , or a dual enzyme pathway with HRP, glucose oxidase (GOx) and glucose. **b)** TEM characterization of the protein nanosheets with inset showing an enlarged area of the micrograph and a model of hexagonal packing of the SP1 disks. Adapted with permission from Ref. ²³⁰. Copyright 2017 ACS.

Later, Li et al. used small molecule tethers for the preparation of covalently linked 2D nanosheets. The authors employed bis-maleimide-terminated PEG linkers of variable length to generate nanosheets composed of EBFP2 and EGFP, fluorescent proteins that constitute a Förster resonance energy transfer (FRET) pair.⁴⁶⁹ In this template-free system, the β -barrel-shaped proteins, modified with Cys residues at equatorial positions lying close to the C_4 rotational axes, were mixed with the thiol-reactive linkers to yield a covalently linked 2D monolayer. The length of the PEG linker significantly impacted nanosheet size, with longer linkers giving rise to larger sheets due to reduced electrostatic repulsion between adjacent proteins. Notably, the proteins adopted a fixed orientation and even distribution within the nanosheets, which exhibited light-harvesting properties.

Ferritin, a protein cage composed of 24 subunits, has been a popular building block for the formation of extended protein-based lattices due to its octahedral structure that exhibits 432 point group symmetry. Ferritin's robustness and tolerance to genetic and chemical modification have made it an ideal scaffold to test a broad range of assembly design strategies. It has arguably proven to be the most versatile building block available in the design and construction of protein assemblies, as demonstrated by the broad range of ferritin-based ordered 2D and 3D lattices

reported just in the last decade. The Zhao Group employed several types of modifications at the C_4 axis of the ferritin cage to enable the formation of square 2D lattices.^{228,273,470,471} Zhou et al. applied a disulfide bonding strategy to make covalently linked 2D lattices from homo-oligomeric human heavy chain ferritin (HuHF) by installing Cys residues near the C_4 axes of the cage (C162).²²⁸ This single mutation positioned four Cys residues in close proximity to each other, created a “hot spot” for interactions, to construct a superlattice upon slow oxidation in the presence of β -ME (Figure 53).

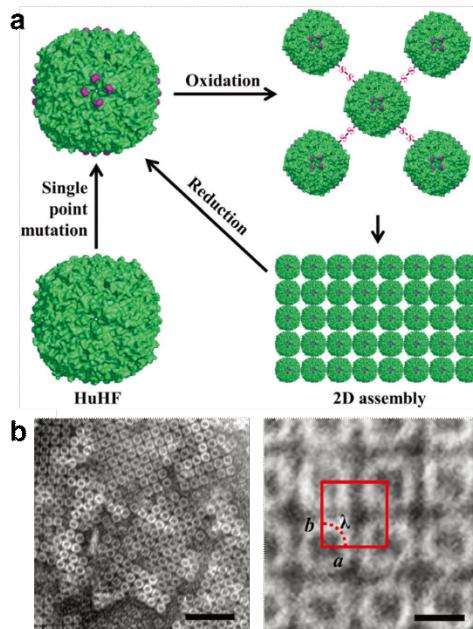


Figure 53. Covalently linked 2D HuHF arrays. **a)** A single mutation near the C_4 axis positions four Cys residues in close proximity to each other. Slow oxidation yields ordered 2D HuHF arrays through multivalent disulfide bond formation at the “hot spots”. **b)** ns-TEM characterization of the square lattice. Adapted with permission from Ref.²²⁸. Copyright 2019 RSC.

The tobacco mosaic virus coat protein (TMVCP) has a rich phase behavior that is intricately controlled by pH, salt concentration, and temperature.⁴⁷²⁻⁴⁷⁴ The double-layer TMVCP “20S” disk structure, composed of 34 subunits arranged into two stacked 17-mer rings, has an 18 nm outer diameter, a 4.7 nm height and 4 nm-wide central pore. Blum and coworkers observed that TMVCP disks bearing a C-terminal His-tag form dense hexagonally packed 2D arrays at pH 5.5-6.0, a behavior that markedly differs from the observations made with wildtype TMVCP.⁴⁷⁵

In the last few years, the Wang Group has used strategies such as disulfide bond- and metal-mediated assembly to generate ordered 2D and 3D lattices based on TMVCP. Notably, Zhang et al. reported that a variant bearing three Cys mutations (C1 and C3 at the periphery of the disk designed to promote in-plane assembly, and C103 on the inner surface of the central pore) exhibited four distinctive assembly behaviors depending on pH and ionic strength.⁴⁷⁶ The observed structures included rods, rod-bundles, and 2D arrays. The 2D arrays were observed in conditions ranging roughly from pH 6.0 to 8.0 with phosphate buffer concentration maintained below 200 mM. The same C1/C3/C103 mutant was later used to generate large monolayered

nanosheets tens of microns in size upon Cys oxidation with Cu^{2+} (**Figure 54**).⁴⁷⁷ Triclinic and hexagonal close-packed TMVCP 3D crystals were also assembled based on a C1/C3/C103 variant assembled through disulfide bonding and a C-terminal 4-His tagged variant (also bearing the C103 mutation on the inner pore surface) via Zn-binding respectively.⁴⁷⁸ Most recently, the C103/C-terminal 4-His variant was used to generate ordered 2D co-assemblies of TMVCP and various functional nanoparticles such as AuNPs and QDs as discussed further in section 4.3.4.⁴⁷⁹

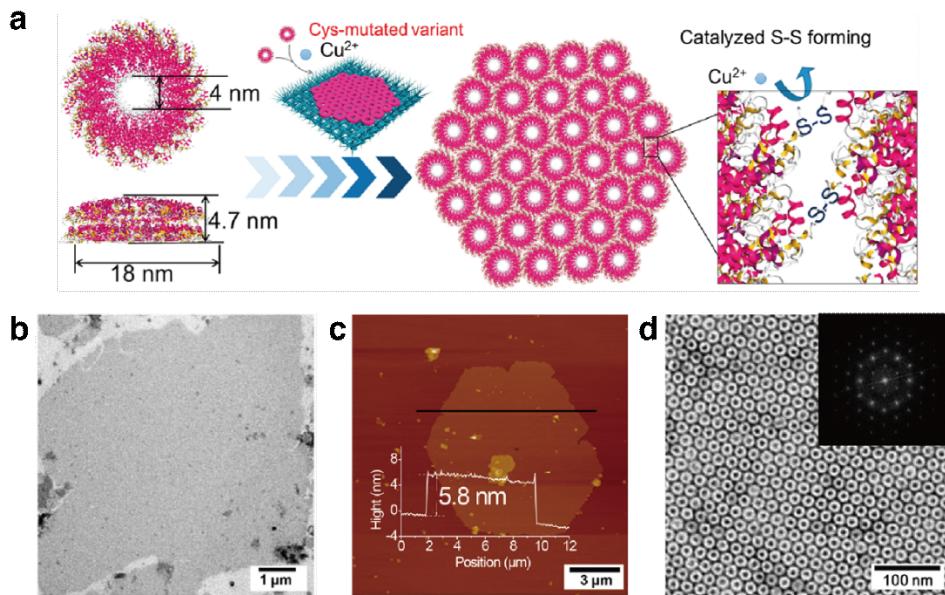


Figure 54. Covalently crosslinked TMVCP nanosheets. **a)** A schematic diagram of the formation of TMVCP 2D nanosheets upon oxidation of the peripheral Cys residues by Cu^{2+} . Characterization of the nanosheets by **b)** ns-TEM, **c)** AFM, and **d)** high resolution ns-TEM. Inset shows the corresponding FFT. Adapted with permission from Ref. ⁴⁷⁷. Copyright 2018 ACS.

3.5.3 Metal-mediated 2D assemblies

Metal coordination is a powerful tool to build interfaces between protein building blocks in large part due to the strong and directional nature of the interaction. As noted earlier, metal-mediated assemblies can be designed by introducing metal-binding motifs composed of natural amino acids on the protein surface, by conjugating metal-chelating functionalities, or by introducing non-canonical metal-binding amino acids. The Tezcan Group extended their metal-mediated interface design approach to generate periodic protein arrays based on a metal-binding variant of cyt *cb*₅₆₂.²⁰² As outlined in Section 3.4.3, the variant RIDC3 assembles into extended 2D crystals under the slow nucleation conditions of low $[\text{Zn}^{2+}]:[\text{RIDC3}]$ ratios and acidic pH, close to the *pI* of RIDC3 (~5.3). The crystal structure of the 3D crystal, determined at 2.3 Å, revealed the central role of Zn^{2+} coordination in the assembly and confirmed that the 3D crystals form through stacking of the 2D layers at pH conditions where inter-layer repulsion is minimized. As shown in **Figure 46c** in Section 3.4.3, the assembly of the 2D crystals occurred through the tiling of antiparallel RIDC3 dimers mediated by the coordination of two Zn^{2+} ions (Zn1) at the high affinity sites described previously. The key determinants of crystal growth were orthogonal

coordination vectors at Zn1 and Zn2, where inter-dimer connections within the plane occur through Zn1-E81 and Zn2-E49, producing an infinite chain along one axis. Finally, a second set of interactions at the lower affinity site with Zn3 and another E49 residue connected the chains along the orthogonal axis.

Subramanian et al. later developed an artificial DNA-protein hybrid architecture based on the RIDC3 building block, which displayed intricate assembly behavior that depended on the synergistic effects of Watson-Crick base pairing, protein-metal coordination, and nucleic acid-protein interactions.²⁹⁸ A Cys residue (C21) was first installed on RIDC3 for site-specific bioconjugation to short DNA strands. Protein-DNA chimeras bearing complementary 10 base-long DNA strands were prepared (**Figure 55a**) and their co-assembly was screened under a range of conditions where temperature, pH, protein-DNA conjugate concentration, and $[Zn^{2+}]:[RIDC3-DNA]$ ratio were varied. In contrast to the metal-mediated assemblies based on unmodified RIDC3 discussed above, protein-DNA chimeras formed ordered arrays only in a very narrow parameter window owing to the delicate balance of contributing interactions. Structural characterization of the arrays was carried out through a combination of AFM, SAXS, cryo-EM and molecular dynamics modeling to reveal that the layered architecture consisted of V-shaped dimeric modules linked through Watson-Crick base-pairing and a four-coordinate Zn-binding motif (two E27/E31 pairs). The dimers were further linked in an antiparallel fashion to four neighbouring modules through tridentate Zn^{2+} coordination (E8/D12/H63). The resulting arrangement created a corrugated 2D sheet where the dsDNA domains served as staples between the proteins above and below the plane (**Figure 55b**). The sheets were capable of stacking in register, such that the DNA domains fit into the open protein-protein interfaces of the sheets above and below where they form non-covalent contacts with the protein surfaces (**Figure 55c**). This unique, sophisticated architecture illustrated the complexity in the interplay of diverse interactions that contribute to the intricate association of biopolymers in natural nucleoprotein structures such as the ribosome.²⁹⁸

In a related approach, Qiao et al. reported the Zn^{2+} -mediated assembly of SMAC, a homodimeric V-shaped building block, into nanowires and wavy 2D layers.⁴⁸⁰ They installed two bis-His motifs on each monomer (H75/H79 and H137/H141), which adopted a quadrilateral in-plane orientation upon SMAC dimerization. As observed in the case of RIDC3, the two metal coordination sites on SMAC had different binding affinities to Zn^{2+} ions. Upon metal addition, the H75/H79 sites became saturated first, leading to the formation of zig-zag nanowires, while higher Zn^{2+} concentrations were required to occupy the orthogonally oriented H137/H141 sites and induce the formation of wavy 2D sheets.

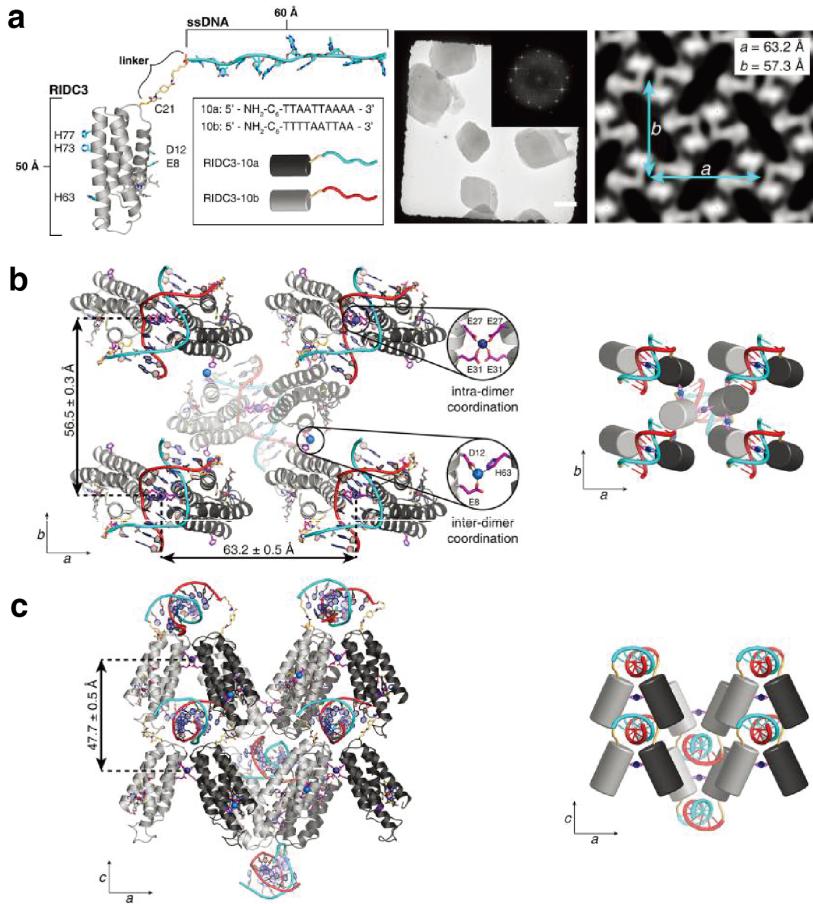


Figure 55. Artificial metal-dependent nucleoprotein assemblies based on a chimeric RIDC3-DNA building block. **a**) RIDC3-DNA hybrids (left). TEM characterization (middle) and reconstructed 2D cryo-EM map (right) of the RIDC3-DNA 2D crystals. MD-minimized models and cartoon illustrations of **b**) a single 2D RIDC3-DNA layer and **c**) the 3D stacking of two RIDC3-DNA layers. Adapted with permission from Ref. ²⁹⁸. Copyright 2018 ACS.

Recently, Yang and Song reported on the genetic incorporation of the unnatural chelating amino acid bipyridine-alanine (bipy-Ala) to drive assembly of various extended structures in 1D and 2D through the selective formation of $[\text{Ni}(\text{bipy-Ala})_2]$ complexes.⁴⁸¹ Acetyltransferase from *Bacillus antharsis*, a D_3 symmetric homohexamer, was selected as a building block for this study based on a number of criteria, including its symmetry being amenable to multidirectional assembly. Amber codon and orthogonal aminoacyl tRNA synthases/tRNA pairs were used to incorporate bipy-Ala at several positions on the top/bottom faces and the lateral faces of the hexamer to screen for 1D and 2D assemblies respectively. Ni^{2+} was chosen as the target metal ion to drive assembly for its high affinity to bipy chelating groups and its preference for selective formation of $[\text{Ni}(\text{bipy})_2]^{2+}$ complexes. Ordered 1D and crystalline 2D assemblies were achieved with optimization of assembly conditions, in addition to combinatorial and hierarchical structures. The incorporation of unnatural chelating amino acids promises to be a powerful approach to expand the repertoire of metal-dependent protein assembly motifs and, by the same token, to diversify the repertoire of orthogonal interaction types available to tune and control protein assembly.

3.5.4 2D assemblies mediated by protein-ligand interactions

The first example of a designed 2D protein assembly harnessed the specific recognition between streptavidin and biotin, a natural receptor-ligand pair with exceptionally high affinity. Ringler and Schulz combined the C_4 -symmetric, tetrameric RhuA, and a C_2 -symmetric linker (streptavidin), to design a 2D array with square geometry.⁴⁶⁶ RhuA was mutated at three positions (N133C, K261C, and C126S), and the newly installed Cys residues were further labeled with biotin (bR) to enable directional binding of one RhuA molecule to four streptavidin units, producing a cross-shaped construct (bRS_4) (Figure 56). Further association of RS_4 with bR yielded small 2D arrays limited to roughly 50 x 50 nm in size. Tethering the bR building block to a lipid monolayer interface enabled the growth of arrays up to 200 nm in width. Furthermore, addition of bis-biotinylated streptavidin spacers (^{bb}S) was used to increase spacing between bRS_4 hubs, thus controlling mesh size. This early example of 2D protein assembly achieved modest domain sizes, which can be attributed to the low potential for reconfigurability of the associations between proteins due to high biotin-streptavidin affinity and the rigidity of the building blocks. These limitations have since been circumvented through various protein assembly design approaches, as described throughout Section 3.5.

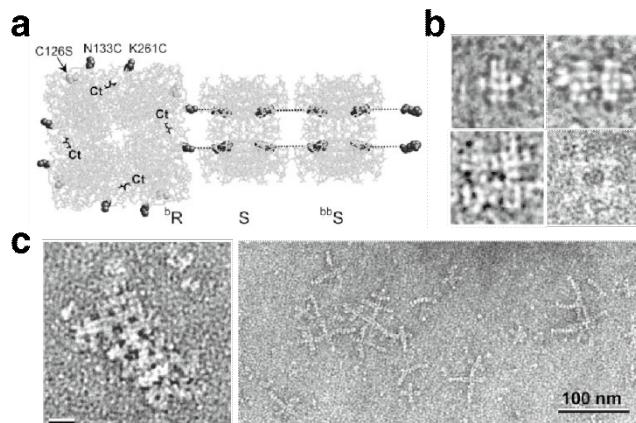


Figure 56. Assembly of biotin-labeled RhuA with streptavidin. **a)** Schematic representation of the C_4 -symmetric enzyme RhuA with point mutation for biotin labeling (bR). Streptavidin (S) binds to two biotin labels on each side of bR , to form the building block $^bR\bullet S$. ^{bb}S , S bound to bis-biotin linkers, is further bound to $^bR\bullet S$. **b)** Association of bR and S building blocks imaged by ns-TEM. Clockwise from top left: $^bR\bullet S_4$, $^bR_2\bullet S_7$, $^bR_4\bullet S_{16}\bullet^{bb}S_{12}$, and $^bR_4\bullet S_{12}$. **c)** Self-assembled networks produced from bR with $^bR\bullet S_4$ (left) and $^bR\bullet S_4$ with $(^{bb}S\bullet S)_x$ (right) imaged by ns-TEM. Adapted with permission from Ref.⁴⁶⁶. Copyright 2003 AAAS.

The high affinity of many natural protein-ligand interactions and precise positioning of receptors within the protein architecture provide an attractive means to design protein assemblies by combining native proteins with multivalent ligand mimics. Over the past few years, Chen and coworkers have exploited the specific protein-ligand interactions of sugar-binding lectin proteins in combination with $\pi\text{-}\pi$ stacking-driven dimerization of RhB to build a range of extended protein assemblies. Following earlier success using dimerizing sugar-RhB conjugates as linkers in the design of 3D crystals (section 3.6.2.2)⁴⁸² and 1D microtubules (section 3.4.1),²⁷⁰ they have targeted more diverse structures spanning multiple dimensions in recent years. In the first study seeking to diversify the types of assemblies achieved using this linking strategy from

a single building block, Yang et al. selected lectin A (LecA) from *Pseudomonas aeruginosa* as the protein component, a cuboid homotetramer that specifically binds to galactose, and varied the length of the oligo-(ethylene oxide) tether within the sugar-RhB assembly-inducing ligand (**Figure 57a**).⁴⁸³ They were able to obtain three different LecA packing patterns (**Figure 57b**) depending on the length of the tether, which corresponded to extended assemblies such as 1D nanoribbons, 2D sheets, and 3D layered structures on the extended assembly scale. 2D sheets were observed with linkers containing 2, 4 or 5 ethylene oxide repeat units within the inducing ligand and arose from a diagonal-diagonal protein packing scheme (**Figure 57c**).

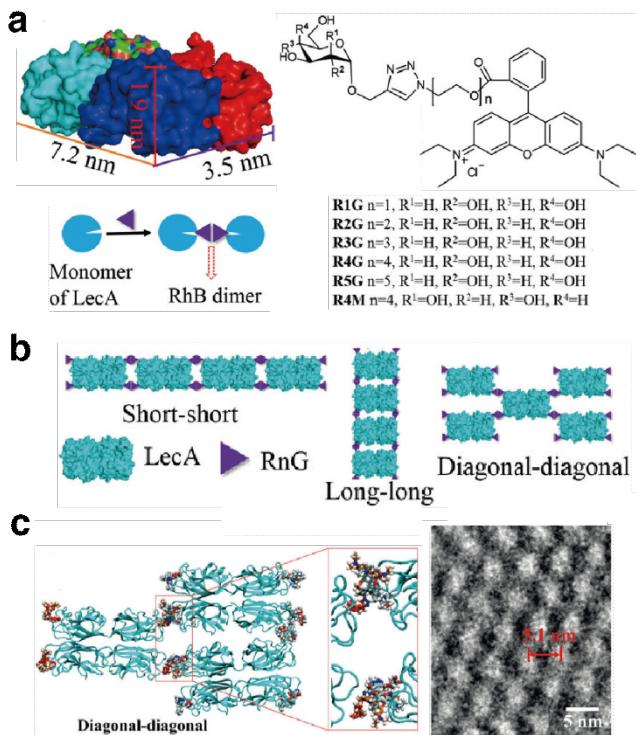


Figure 57. Assembly of LecA via a combination of ligand binding and RhB dimerization. **a**) Structure of the tetrameric protein, LecA and inducing ligand RnG (n = 1 to 5). Cartoon representation of LecA/RnG dimerization. **b**) The three packing patterns of LecA/RnG based on the dimerization of RnG. **c**) Schematic representation of diagonal-diagonal packing of LecA/R2G (left) and enlarged cryo-EM images of a LecA/R5G 2D lattice. Adapted with permission from Ref. ⁴⁸³. Copyright 2017 Wiley.

In a further study, the researchers targeted the formation of a Pascal triangle lattice using wheat germ agglutinin (WGA) as the building block and sialyllactoside-RhB conjugate as the assembly-inducing ligand (**Figure 58a-b**).⁴⁸⁴ WGA is an anisotropic, horse-shoe-shaped homodimer that bears eight independent carbohydrate-binding sites. The four sites on the top face of the dimer have at least two-fold stronger binding affinity to the carbohydrate ligand than the sites on the bottom face. The anisotropic shape of the WGA dimer, the disparity in ligand affinity, and RhB dimerization all contributed to the success of the design. Triangular units formed first through dimerization of ligands bound to high affinity sites. This was followed by formation of clusters and in-plane polymerization into 2D lattices via the ligands bound to weaker sites in a subsequent step, yielding the targeted periodic Pascal triangle lattice structure at equimolar

concentrations of WGA and inducing ligand (**Figure 58c–e**). Stoichiometric excess of the inducing ligand eventually led to the formation of 3D crystals through saturation of the available binding sites, where both intralayer and interlayer contacts were mediated by RhB dimerization.

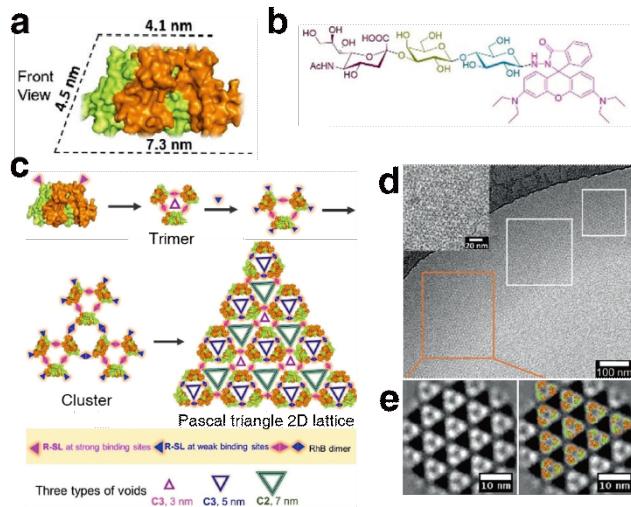


Figure 58. Pascal triangle lattice formation from wheat germ agglutinin (WGA) and sialyllactoside-RhB (R-SL). **a)** Structure of the WGA dimer. **b)** Structure of assembly-inducing ligand (R-SL). **c)** Proposed mechanism of WGA assembly with R-SL. **d)** Cryo-EM characterization of the WGA 2D lattices. Inset: enlarged image. **e)** Cryo-EM images of the 2D lattice (left), and the overlay of the structural model (right). Adapted with permission from Ref. ⁴⁸⁴. Copyright 2020 Wiley.

In a departure from the typical structurally stable and immutable protein building blocks used in extended assembly, Xu et al. recently exploited the allosteric ligand binding properties of adenylate kinase (AKe) to design a ligand-dependent assembly system that switches from 1D to 2D configuration.⁴⁸⁵ AKe contains dual recognition domains for ATP and AMP, bisubstrates that induce a conformational change in the enzyme from an open to closed state upon association and undergo a phosphotransfer reaction ($ATP + AMP \rightarrow 2 ADP$) (**Figure 76c** in section 4.2.1). The dynamic protein assembly was designed by converting AKe to an amphiphile through the conjugation of a maleimide-terminated poly-L-valine tail to C77. In the open conformation, the AKe amphiphile assembled into extended 1D fibers of uniform diameter above its critical aggregation concentration under the influence of the hydrophobic effect. Upon addition of equimolar diadenosine-5-pentaphosphate (Ap5A), a tethered unreactive substrate mimicking the dual binding of ATP and AMP, the 1D assemblies were converted to ordered 2D sheets following the conformational shift of AKe from the relaxed open state to the rigidified closed conformation. This study highlights the knowledge gaps in the protein assembly design field when it comes to inherently dynamic building blocks. This area is ripe for investigation toward the development of active protein-based materials and artificial devices based on active building blocks.

3.5.5 Computationally designed 2D assemblies

As outlined previously (Section 2.1), natural protein assemblies are characterized by extensive non-covalent interfacial contacts. Along these lines, the Baker Group focused on a general computational approach to design ordered 2D protein arrays mediated by non-covalent interactions.¹⁸⁸ To simplify the design process, Gonen et al. targeted layer group symmetries requiring only two unique interfaces between identical, homo-oligomeric building blocks (6 of the 17 possible layer group symmetries). Furthermore, the scope of potential building blocks was limited to proteins with internal point symmetry, as they already contain one of the two required interfaces, and cyclic rather than dihedral symmetry, because of a greater abundance of candidate proteins possessing this characteristic in the PDB. A symmetric docking procedure was applied in Rosetta to first position the cyclic oligomers into the target layer groups. Monte Carlo trajectories for protein orientations were generated and the most shape-complementary solutions with the largest number of contacting residues and the fewest steric clashes were selected. After successive minimization and optimization sequence design steps, 62 candidates were selected for experimental investigation. Three of the designs were reported to form extended, regular, planar arrays, which could be produced both *in vitro* and *in vivo*. High resolution structural characterization of these arrays with *p*321, *p*42₁2, and *p*6 layer group symmetries was carried out by cryo-EM (Figure 59).

Matthaei et al. combined symmetry docking calculations with a fusion strategy to generate a tiled 2D surface based on naturally-occurring oligomeric proteins with a central pore.⁴⁸⁶ Following screening of oligomeric proteins from the PDB according to symmetry, shape, termini orientation, and additional criteria, a homohexameric protein (STM4215) was selected for fusion via a flexible hexaglycine linker and interface redesign for a tighter arrangement within the lattice. Addition of Ca²⁺ ions, which naturally binds to the monomers, rendered the designed fused dimers competent for 2D assembly. Large sheets, typically in the 1 to 10 μm range, but as large as 100 μm, were observed by fluorescence microscopy and ns-TEM allowed to characterize the designed *p*3 lattice. In a follow-up study, the building block was modified with different functional tags (e.g., hexa-His tag, gold-binding peptide, biotinylation tag) and, upon Ca²⁺-triggered assembly, was used as a 2D scaffold for the fabrication of hybrid materials by binding AuNPs and proteins to the functionalized array surface.⁴⁸⁷

Chen et al. later extended the 2015 study¹⁸⁸ to a general method for developing 2D protein assemblies based on *de novo* designed pseudosymmetric protein building blocks.⁴⁸⁸ The homodimeric *de novo* designed helical bundle protein was first connected into a single chain. Possible 2D lattices with pseudo-C12 symmetry were screened based on lattice dimensions and rotation of the building block around its central axis. Surface redesign of the *de novo* protein was then carried out in Rosetta using a standard fixed backbone approach. After identifying a promising design capable of forming extended ordered 2D assemblies that contained exclusively hydrophobic residues at the interface, an effort to improve binding specificity and reduce non-specific aggregation was made by applying the Rosetta HBNet algorithm (see Section 3.2.1) to introduce buried hydrogen bonds at the interface. A design selected after this procedure indeed formed more extended regular assemblies than the fully hydrophobic design.

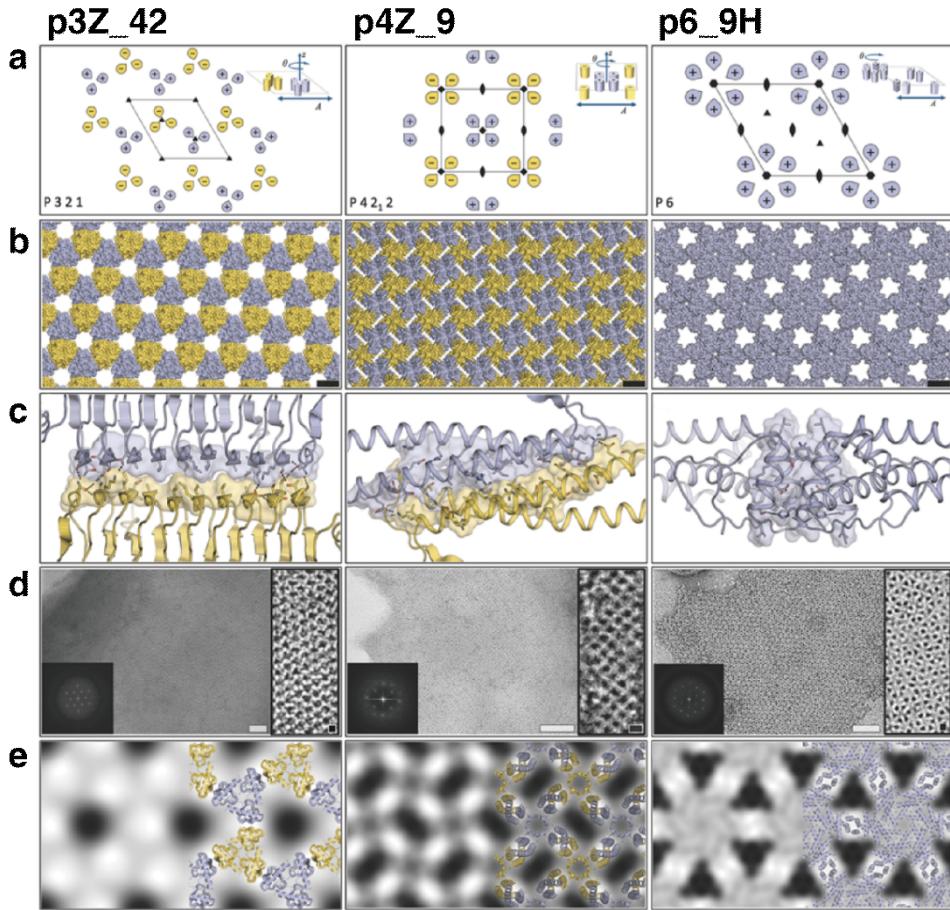


Figure 59. Computationally designed 2D lattices with $p321$, $p42_2$, and $p6$ layer group symmetry (designs $p3Z_42$, $p4Z_9$, $p6_9H$). **a)** Packing of the designed lattices with the representation of the unit cells. **b)** Schematic representations of the designed 2D arrays. **c)** Designed interfaces between protein building blocks. **d)** ns-TEM characterization of the designed 2D lattices. Inset: enlarged image and FFT of the ns-TEM image. **e)** Projection map calculated from **(d)**, and the overlay of the designed model on the projection map. Adapted with permission from Ref. ¹⁸⁸. Copyright 2015 AAAS.

Most recently, the co-assembly of two protein building blocks into binary two-dimensional layers was achieved by Ben-Sasson et al. by using computational redesign of rigid interfaces between pairs of dihedral protein building blocks.⁴⁸⁹ Although it required sampling of more degrees of freedom, proteins with dihedral rather than cyclic symmetry were chosen to take advantage of the additional in-plane two-fold rotation axes present in the dihedral structures, which intrinsically corrected for any deviation from the design model that might cause out-of-plane curvature. After arranging the two protein building blocks with their symmetry axes aligned to the 2D layer group, the spacing between them and their orientations were sampled to identify arrangements with contact regions meeting a 400 \AA^2 surface area threshold and composed mainly of aligned helices. Subsequently, the amino acid sequences at the interfaces were optimized with Rosetta combinatorial sequence design to give low-energy interfaces with a hydrophobic center surrounded by polar residues. Screening experiments led to the discovery of a successful design, a system composed of D_3 and D_2 homo-oligomers that formed $p6m$

symmetric two-dimensional lattices (**Figure 60**). In isolation, the building blocks were soluble in the millimolar concentration range. However, once combined at concentrations as low as 10 nM, they quickly assembled into ordered arrays. Assemblies could be formed both *in vitro* and *in vivo*. Notably, when interfaced with cell surfaces, the arrays exhibited properties with therapeutic relevance (Section 4.5.2).

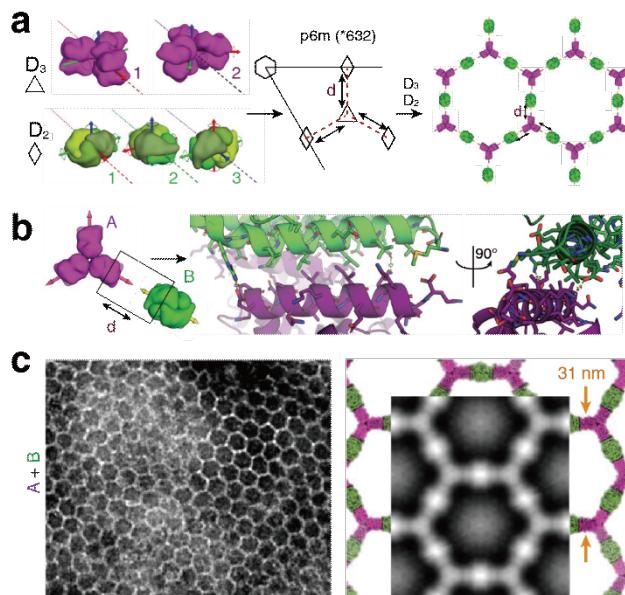


Figure 60. Computationally designed binary 2D protein arrays. **a)** Orientation of the *D*₃ and *D*₂ symmetric protein building blocks to form a heterogeneous *p6m* protein assembly (left). Top view of the *p6m* symmetry operators and the lattice spacing degree of freedom, *d* (middle). A plausible arrangement of *D*₃ and *D*₂ symmetric building blocks to form *p6m* lattices (right). **b)** Interface between the two protein building blocks. **c)** ns-TEM characterization of *p6m* arrays formed in *E. coli* upon co-expression of the two protein building blocks. Adapted with permission from Ref. ⁴⁸⁹. Copyright 2021 NPG.

3.5.6 2D protein assembly at interfaces

In the 2D protein assemblies described above, the formation of protein arrays was mostly carried out in bulk solution. However, the potential for extensive 2D protein organization through their assembly at interfaces has long been recognized. For example, 2D protein crystallization at lipid-water interfaces has been investigated as early as the 1970s.^{490,491} More recently, the assembly of SP1 at air-water interfaces, with or without the presence of phospholipids, has yielded large, ordered, hexagonally packed arrays of SP1, a ring-shaped protein that has proved to be a versatile building block for the design of extended protein assemblies (Sections 3.3.4 and 3.5.2).^{445,467} The 2D crystallization of membrane proteins is also an important goal for investigating this class of protein that can be recalcitrant to forming 3D crystals.^{492,493} General conditions favoring 2D crystallization at an interface include: 1) the ability of molecules to stay in plane, which reduces translational degrees of freedom and increases local concentration, 2) the mobility of molecules within the plane to enable sampling of possible bonding configurations and correct lattice defects, and 3) the identical orientation of all molecules to favor regular contacts.⁴⁹⁴

A pair of recent studies have examined the potential of using a solid-liquid interface for programmable control over protein assembly and access to structures that cannot be obtained from bulk solution. Specifically, these studies were carried out on muscovite mica (*m*-mica) (001), an atomically flat mineral substrate which exhibits pseudo-hexagonal tessellation of charged cavities. These cavities bind to K^+ and other cations, an association which affords control over the charge state of the substrate depending on cation concentration.⁴⁹⁵ Baker, De Yoreo and coworkers were inspired by the carboxylate-rich nature of proteins that interact with minerals and the lattice matching property of ice-binding proteins to investigate the design of protein-inorganic mineral interfaces where the locations of carboxylate residues match the electrostatic and structural features of the *m*-mica lattice for optimal binding and control of assembly.⁴⁹⁶ The *de novo* designed helical repeat protein DHR10, which possesses a flat surface and a regular repeating backbone with spacing equal to a multiple of 5.2 Å, or the nearest-neighbor distance between K^+ sites on *m*-mica, was the chosen protein scaffold. The surface of DHR10 was redesigned with either Glu or Ala residues to match the pattern of K^+ ions occupying cavities on *m*-mica, thus enhancing protein binding to the mineral template (**Figure 61a**). By changing the length of the protein building block and the mobility of the proteins on the substrate by adjusting salt concentration, the orientational order and patterning of proteins on the surface could be tightly controlled (**Figure 61b-c**). At low K^+ concentrations, individual proteins were observed to bind independently to the surface in designed orientation, whereas at high K^+ concentrations, liquid-ordered protein lattices that line up with the *m*-mica sublattice were obtained, exhibiting ordering over a length scale of many millimeters. Protein interface redesign was used to stabilize different monomer arrangements, allowing formation of 1D wires and 2D honeycomb lattices (**Figure 61d-f**). Overall, matching protein to substrate structure and tuning the charge state of the substrate allowed the authors to manipulate 2D assembly at the interface with respect to orientation, morphology, and size.

In a collaboration with the De Yoreo Group, Tezcan and coworkers revisited the disulfide-linked ^{C98}RhuA lattices described above (section 3.5.2)²²⁵ to explore their assembly behavior at the *m*-mica interface.²²⁷ The surface of RhuA has a highly anisotropic charge distribution: the flat C-terminal surface is enriched in negatively charged residues, while the N-terminal side (which has a four-legged profile) is more positive (**Figure 62a**). This distribution endows the protein with a sizeable dipole moment (~1200 D) and simulations show that long-range interactions between the protein dipoles control the symmetry of the *p42_12* lattice that is observed in solution-grown crystals. By tuning the charge state of the *m*-mica surface with K^+ concentration, this charge anisotropy was harnessed to control the outcome of surface-templated ^{C98}RhuA assembly (**Figure 62b**). At low K^+ concentrations (5 mM) both faces of the protein were seen binding to the surface and formed *p4*, or all parallel, lattices with small domain sizes. In contrast, at high K^+ concentrations (3 M), large ordered monolayered *p4* arrays were obtained with the C-terminal side interacting with the surface. At increased protein monomer concentrations and 3 M K^+ , protein bilayer lattices formed under the influence of a salting-out effect. This rich phase behavior (**Figure 62c**) emerged under the influence of interactions that operate at different length scales. Access to the new *p4*-symmetric lattice prompted an examination of the properties emerging from an all-parallel arrangement of protein dipoles in the crystal. These properties are further discussed in section 4.4.

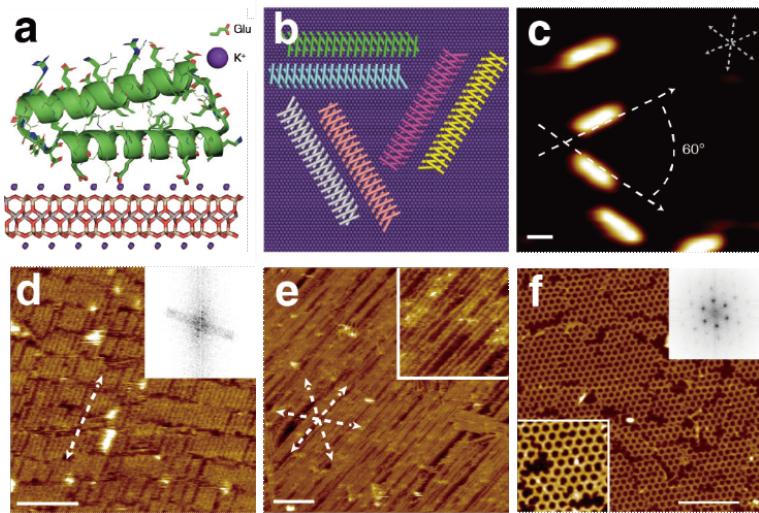


Figure 61. Assembly of *de novo* designed helical repeat proteins at the muscovite mica interface. **a)** Schematic representation of DHR10-mica18 protein binding to the mica substrate through the modulation of a layer of K^+ sublattice. **b)** The protein-mica interface design model predicted the orientation of DHR10-mica18 on K^+ -saturated mica surface. **c)** The three predominant orientations of DHR10-mica18 bound to the mica surface with the layer of K^+ sublattice characterized by AFM. Assemblies of different variants of *de novo* designed helical proteins observed by AFM on mica treated with 3 M KCl: **d)** DHR10-mica18, **e)** DHR-mica6-NC and **f)** DHR-mica6-H. Adapted with permission from Ref. ⁴⁹⁶. Copyright 2019 NPG.

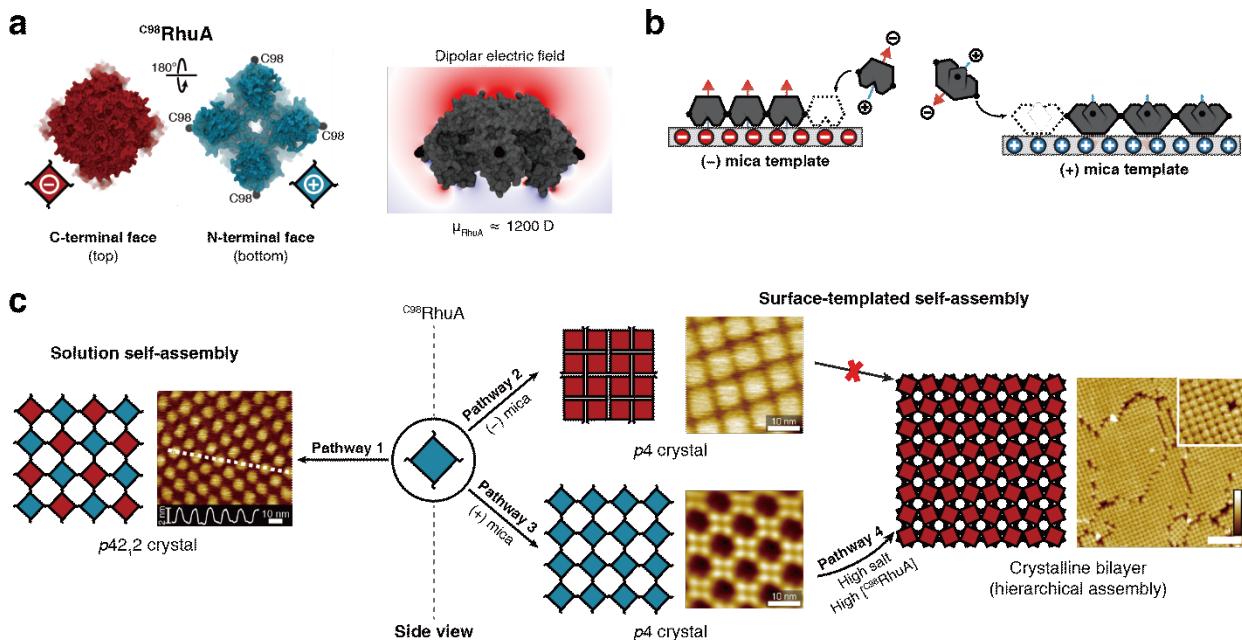


Figure 62. C^{98}RhuA assembly at the mica interface. **a)** Schematic diagram of the negatively charged C-terminal face and partially positively charged N-terminal face of the C^{98}RhuA tetramer (left), and the large dipole moment of the tetrameric protein (right). **b)** Preferential adsorption of either the N-terminal or the C-terminal face of RhuA depends on the charge state of the mica template. **c)** C^{98}RhuA tetramer assembly pathways in solution and templated by the mica surface. Adapted with permission from Ref. ²²⁷. Copyright 2020 NPG.

3.6 Extended 3D protein assemblies

Current approaches for the design of ordered 3D protein assemblies are largely informed by decades of effort to understand and improve protein crystallization. For a long time, obtaining large, high-quality single crystals to elucidate protein structure by X-ray diffraction was the primary motivation for optimizing protein crystal growth. Conventionally, researchers have varied the physical, chemical, and biological parameters involved in crystallization mixtures (e.g., temperature, pH, ionic strength, addition of ligands) to obtain well-diffracting protein crystals. However, the advent of recombinant DNA technology and the accumulation of structural information from homologous proteins has allowed researchers to better design rational modifications to the protein structure (e.g., improve crystal contacts or induce symmetry that promotes crystallization).⁴⁹⁷ Protein modification approaches that have improved protein crystallization success rates include: systematic truncation of the primary sequence, removal of post-translational modifications, screening of homologous proteins, or co-crystallization of the target protein with binding proteins or antibodies.⁴⁹⁸

In recent years, the design of 3D protein assemblies has proceeded mainly through two paths: rational mutagenesis of protein surface residues and addition of effector molecules that promote ordered assembly. The section below briefly surveys the approaches to rationally design protein surfaces and assembly strategies that incorporate different types of intermolecular interactions mediated by small molecules, nanoparticles, or polymeric effectors to drive ordered 3D protein assembly. With the wealth of available protein crystallization approaches, the field of 3D protein assembly has evolved to build ordered materials with potential applications that include storage, nanoreactor development, and control of soft matter properties.

3.6.1 3D protein assemblies through rational design of protein surfaces

Rational mutagenesis of protein surface residues has been a fruitful approach to improve protein crystallization efforts and enable designed 3D protein assembly. The methods employed broadly target the design of intermolecular interfaces or the symmetrization of the building blocks.

One of the first examples of protein engineering studies to generate 3D crystals was the work of Lawson et al., who mutated HuHF to install metal ion-mediated contacts that facilitate crystal formation.⁴⁹⁹ Inspired by the propensity of horse spleen ferritin to readily crystallize upon the addition of divalent metals, the authors sought to reproduce the metal-mediated interface in HuHF.⁵⁰⁰ Specifically, the K86Q mutation, in combination with native D84, led to the formation of large single crystals of HuHF through Ca^{2+} -mediated interactions (**Figure 63**).

Chemical modification approaches have similarly been employed to facilitate protein crystallization. For example, reductive methylation of Lys residues on lysozyme enabled Rypniewski and coworkers to obtain a 1.8 Å resolution crystal structure in 1993.⁵⁰¹ Soon after, Rayment et al. reported the first crystal structure of the head portion of myosin by implementing a methylation approach.⁵⁰² Surface entropy reduction is a related approach which replaces flexible surface residues such as Lys, Glu, or Gln with less flexible residues like Ala to promote crystallization.⁵⁰³

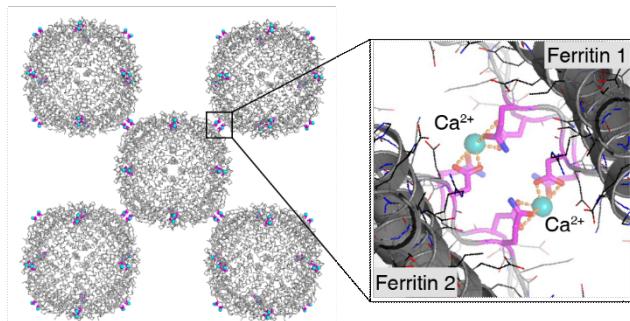


Figure 63. HuHF crystallization facilitated by Ca^{2+} contacts at K86Q residues.

Whether by introducing protein point mutations or bulk modifications, these early mutagenesis approaches are still active areas of research today and have informed subsequent innovations in the development of 3D protein assemblies.^{497,504} Inspired by a leucine-zipper sequence known to promote formation of a stable protein-protein interaction, Yamada et al. reported the first crystal structure of a mutant human pancreatic ribonuclease 1 (RNase 1) protein.⁵⁰⁵ Using a strategy termed ‘crystal lattice engineering’, they installed four leucine residues on helix 2 of RNase 1, which promoted interfacial interactions via the designed hydrophobic leucine patches. Based on the 1.7 Å resolution crystal structure, they showed that human RNase 1 was structurally similar to bovine RNase A.

In a recent study, Du et al. took advantage of the symmetry of RhuA to build 3D arrays of different morphologies with minimal sequence modification.⁵⁰⁶ RhuA has a C_4 symmetry axis, with Pro residues located at the C_4 -symmetric nodes. Notably, its N- and C-terminal faces (or top and bottom faces) have different morphologies and surface charge distributions, conferring anisotropic character to the protein. Upon installing His residues at the N- and C-termini, the RhuA mutants formed ribbons and sheets that assembled into directional 3D arrays. Specifically, the accessible His residues on the top and bottom of RhuA were able to form π - π interactions along the *c*-dimension to facilitate the stacking of RhuA molecules. Perpendicularly, the Pro residues at the corners of RhuA formed the edge-to-edge contacts along the *a*- and *b*-dimensions. By implementing selective point mutations, the researchers were able to obtain a range of 3D constructs that assembled through synergistic interactions, with minimal perturbation to the initial amino acid sequence.

Mutations at specific residues and patches along protein surfaces have led to the successful design of 3D assemblies. In contrast, the Beck Group has relied on changes to the global surface charge of ferritin cages to direct assembly.²⁷⁷ By enriching the surface of HuHF with Arg/Lys or Glu/Asp residues, they created positively and negatively charged ferritin cages, respectively. Mixing the oppositely charged cages led to the self-assembly of binary protein crystals. One unit cell of the resulting tetragonal crystal lattice was composed of four positively and one negatively charged ferritin cages (**Figure 64**). In a follow-up study, the team explored the impact of divalent Mg^{2+} on 3D assembly.⁵⁰⁷ While the binary lattices were maintained at low Mg^{2+} concentrations, high concentrations of Mg^{2+} led to the preferential formation of unitary cubic lattices composed solely of the negatively charged ferritin cages with Mg^{2+} -mediated interfaces.

These unitary lattices were used as a protein scaffold for the organization of metal oxide nanoparticles synthesized within the cages prior to crystal formation (Section 4.3.4).

Design of 3D protein assemblies can be simplified by using symmetric building blocks, which are more likely to form long-range, ordered protein assemblies than asymmetric building blocks. Yeates and coworkers have proposed a “synthetic symmetrization” strategy (**Figure 65a**) motivated by the observation that protein oligomers tend to crystallize into space groups that support the point group symmetry of the oligomer.²²³ Synthetic symmetrization involves installing residues that promote self-association on the surface of asymmetric proteins to promote their crystallization. Accordingly, the researchers installed Cys residues at various positions on lysozyme to enable the formation of covalently-linked dimers.²²³ The lysozyme dimers were then screened for crystallization, leading to the isolation of six crystal morphologies that had not been previously observed in lysozyme crystals. This work demonstrated that protein symmetrization did not only improve protein crystallization but also expanded the list of accessible crystal morphologies. In a follow-up study, the monomeric protein CelA, known to be a difficult crystallization target, was subjected to the same disulfide bond-mediated dimerization approach (**Figure 65b-d**).⁵⁰⁸ This method of creating synthetic homodimers enabled the crystallization of CelA and structure determination at 2.4 Å resolution.

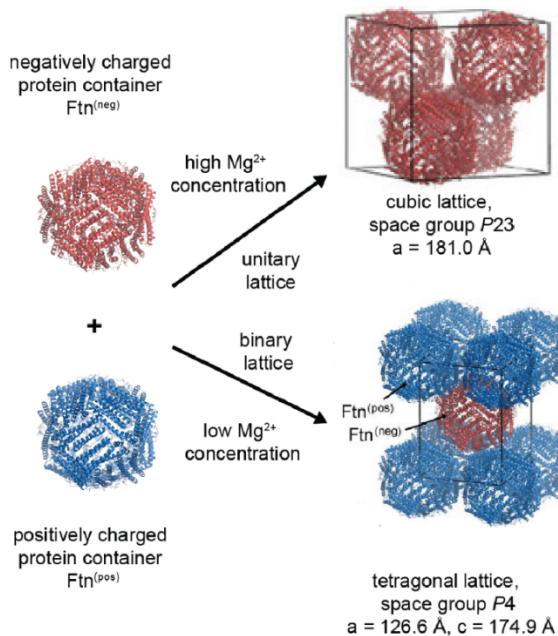


Figure 64. Assembly of charged HuHF lattices. Under low Mg^{2+} concentration conditions, the mixture of positively and negatively charged HuHF forms a tetragonal binary lattice, whereas when Mg^{2+} concentration is raised, cubic lattices composed only of the negatively charged variants and mediated by Mg^{2+} contacts are observed. Adapted with permission from Ref.⁵⁰⁷. Copyright 2018 ACS.

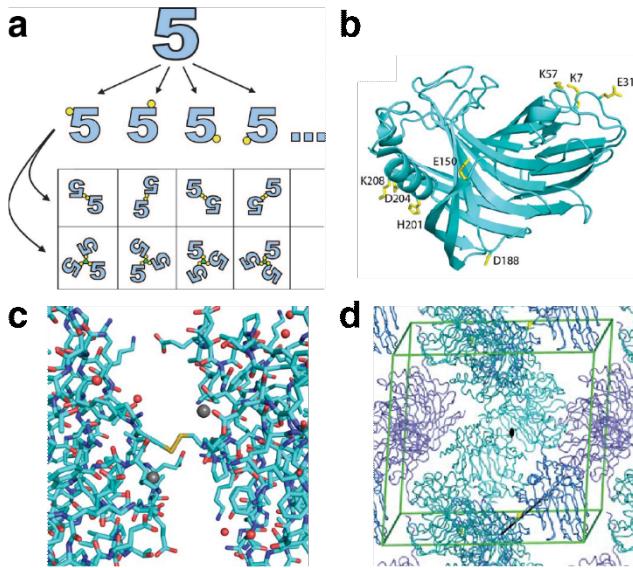


Figure 65. Synthetic symmetrization via Cys residues introduced by mutagenesis. **a)** Illustration of synthetic symmetrization. An asymmetric protein monomer is represented by the figure “5.” Cys residues introduced at different positions on the surface are shown in yellow. Dimerization is achieved via disulfide bond formation. Trimerization requires a trivalent mediator. **b)** Illustration of the approach using CelA. Eight positions were selected for mutation to Cys. **c)** Interface for disulfide-bonded monomers of the D188C variant obtained by XRD. Very few contacts, other than the disulfide bond, are observed. **d)** Crystal packing of the D188C dimer. Adapted with permission from Ref. ⁵⁰⁸. Copyright 2011 Wiley.

The synthetic symmetrization approach was expanded to use discrete metal-binding to promote self-association of asymmetric protein building blocks into ordered assemblies. Using T4 lysozyme (T4L) and maltose-binding protein (MBP) as model systems, Laganowsky et al. installed His and Cys residues and incubated the mutants with metal ions (Cu^{2+} , Ni^{2+} , Zn^{2+}) to form 3D protein crystals.⁵⁰⁹ The metal-binding sites were placed near the ends of the helices to form accessible binding sites and reduce steric hindrance. Upon metal addition, the proteins crystallized into 3D architectures featuring metal-ligand interactions between the exogenous metal and metal-binding amino acid residues at the protein surfaces (His, Glu, Asp). Notably, the metal-ligand complexes had octahedral, tetrahedral, and square planar coordination geometries.

A related example of synthetic symmetrization via metal binding was reported by Brodin et al., who developed ordered metal-mediated assemblies in 1D, 2D and 3D based on RIDC3, a variant of the asymmetric building block cyt *cb*₅₆₂.²⁰² As detailed in Sections 3.4.3 and 3.5.3, in this assembly system, the strong Zn^{2+} binding site first promoted association of two RIDC3 molecules into an antiparallel dimer (symmetrization of an asymmetric building block), and the dimers further underwent long-range assembly through additional Zn^{2+} -mediated interactions.

The Zhao Group has used HuHF surface modifications to generate 2D and 3D square arrays. By installing aromatic residues – phenylalanine (Phe), tyrosine (Tyr), or tryptophan (Trp) – at the 162 position of HuHF, Zhou et al. converted the C₄ axis “hot spots” to hydrophobic patches prone to self-association via π - π interactions (**Figure 66**).⁴⁷⁰ The Phe mutant gave rise to 2D assemblies, while both the Tyr and Trp variants underwent further assembly into cubic 3D

superlattices. The authors attributed the difference in assembly outcome on altered configuration of the π - π contacts depending on the aromatic ring substituents. The assembly process was reversible and sensitive to pH and salt concentration, parameters which influenced π - π stacking interactions and contributed to the screening of ferritin surface charges.

In a further study, Zheng et al. explored the pH-dependent formation of 2D and 3D arrays based on HuHF modification with the amyloidogenic “GLMVG” peptide motif at the C_4 symmetry axis.⁴⁷¹ The hydrophobic motif was installed by replacing the native residues 159-163 in a solvent-exposed region around the C_4 axes with the amyloidogenic sequence. At pH 9.0, the modified cages remained dispersed in solution. However, under more acidic conditions (pH 5.7), just above the theoretical pI for ferritin (~5.3), they quickly formed 2D sheets. With further incubation, the 2D assemblies first increased in size and later transformed into cubic 3D arrays. The same structures were observed at pH 6.5, albeit the assembly proceeded at a much slower rate.

With the diversification of approaches to protein surface engineering, the pace of growth of the protein assembly field has accelerated, producing an expanding range of ordered 3D assemblies. In addition to augmenting the protein units themselves, symmetrization of protein building blocks has been a useful means to drive long-range, ordered assembly.

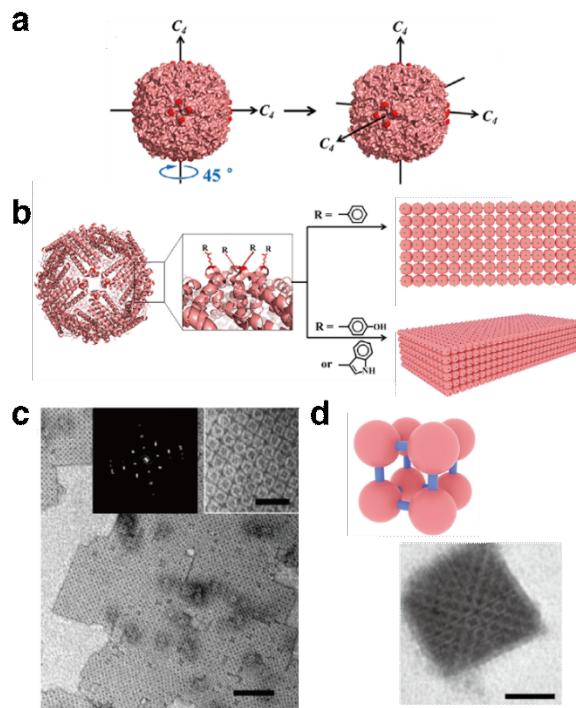


Figure 66. 2D and 3D arrays based on HuHF bearing aromatic residues at the C_4 axes. **a)** The three 4-fold symmetry axes of the HuHF nanocage, Glu162 near the axis is shown in red. **b)** HuHF cages bearing aromatic residues at the 162 position undergo assembly into 2D nanosheets ($R = \text{Phe}$) or 3D superlattices ($R = \text{Tyr}$ or Trp). **c)** ns-TEM characterization of 2D arrays formed by Phe162 HuHF variant. **d)** Unit cell obtained by SAXS and ns-TEM micrograph of 3D superlattices formed from Tyr162 HuHF. Adapted with permission from Ref. ⁴⁷⁰. Copyright 2018 ACS.

3.6.2 Effector-mediated 3D protein assembly

The addition of external functionalities as effectors to facilitate 3D protein assembly has been a widely studied method. Small molecules, nanoparticles, or polymers, whether added as part of a physical mixture or covalently conjugated to the protein of interest, widen the scope of potential interactions and symmetry elements available to mediate the contacts between proteins in ordered, extended assemblies.

3.6.2.1 *Metal-directed 3D assemblies*

The introduction of metal-binding groups on protein surfaces is a common approach to promote ordered protein assembly. In addition to providing small-footprint binding motifs for developing protein-protein interfaces, metal-coordination enables the introduction of directionality and symmetry that promotes organization.^{202,509} Although selective mutagenesis of the protein surface has proven to be successful for assembly design, conventional methods employing only the canonical amino acids are limited in terms of accessible metal-binding functionality and geometric diversity.

To combat this limitation, researchers have harnessed post-synthetic protein modification through conjugation of small-molecule ligands onto the protein surface. In an early study, Radford and Tezcan conjugated a Cys-specific iodoacetamide derivative of phenanthroline (IA-Phen) ligand onto Cys59 of a cyt *cb*₅₆₂ variant to form MBP-Phen1.²⁰⁸ Titration studies with first-row transition metals indicated that Ni(II) tightly associated with MBP-Phen1 to form quaternary structures. The MBP-Phen1 mixture with Ni(II), yielded crystals with two different space groups, *P*2₁ and *P*6₃22. Specifically, MBP-Phen1 assembled into unique triangular structures with each vertex formed through the coordination of a Ni atom to the phenanthroline ligand from one protein monomer and His77 from another (**Figure 67a**), whereby three Ni atoms were positioned in the same plane to form a Ni₃:MBP-Phen1₃ triangle. These trimers then stacked into a hexagonal geometry to form the *P*2₁ and *P*6₃22 crystals (**Figure 67b**). A closer examination of the crystal structures revealed that the Phen groups fit into a small hydrophobic crevice covered by a flexible loop and were further stabilized by a H-bond between the Cys59 amide nitrogen and the Pro53 backbone carbonyl. These interactions stabilized MBP-Phen1 and favored the open configuration of the Ni₃:MBP-Phen1₃ complex, as the hydrophobic crevice sterically hindered a second Phen group from coordinating to Ni and allowed only one other monomer to coordinate through His77. In a follow up study, the researchers investigated the influence of the hydrophobic environment on the Phen group and whether alternative protein oligomers/frameworks can be achieved.⁵¹⁰ In a similar study, Douglas and coworkers conjugated phenanthroline ligands onto DNA binding protein from starved cells (DPS) cages and induced iron-mediated assembly.⁵¹¹ Although the obtained assemblies were not crystalline, this study represented another early example of integrating exogenous metal-binding ligands on the protein surface.

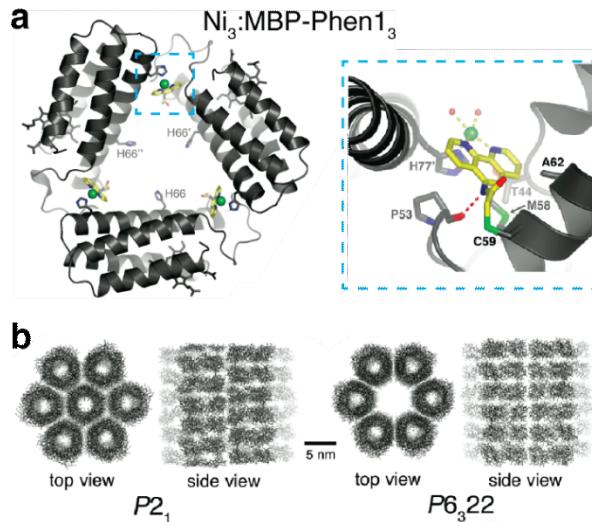


Figure 67. Ni-mediated assembly of phenanthroline-modified cyt cb₅₆₂ variant MBP-Phen1. **a)** Ni₃:MBP-Phen1₃ assembly (left) and coordination environment of Ni-PhenC59 (right). **b)** Ni₃:MBP-Phen1₃ lattice packing arrangement in the P2₁ and P6₃22 crystal forms. Adapted with permission from Ref. ²⁰⁸. Copyright 2009 ACS.

Sontz et al. applied MDPSA to create 3D crystalline protein frameworks akin to metal organic frameworks (MOF) (Figure 68).¹²⁹ Specifically, by exploiting the inherent symmetries of octahedral (432) HuHF cages and C₂-symmetric ditopic organic linkers, the researchers successfully formed large, modular crystalline frameworks that were termed protein-MOFs. They prepared the ^{T122H}HuHF variant, positioning His residues at the C₃ rotation axes of the cage, to generate tripodal coordination motifs for Zn²⁺ ions. Incubation of ^{T122H}HuHF with Zn²⁺ yielded crystals with face-centered cubic (fcc) packing, where Zn²⁺ ions located at the C₃ axis pores adopted near-ideal tetrahedral geometries between the three His122 residue ε-N's and a water molecule. Characterization of the mixture in solution by DLS suggested that the Zn ions did not drive solution-phase assembly. In contrast, addition of the rigid ditopic linker benzene-1,4-dihydroxamic acid gave rise to rhombic dodecahedral crystals in solution. Structure determination at 3.8 Å resolution revealed a body-centered cubic (bcc) lattice (I432 space group) with electron density at the expected organic linker position. This work introduced a new class of hybrid protein framework materials composed of a ternary mixture of proteins, metals and organic linkers.¹²⁹ In follow-up studies, Bailey et al. expanded the library of protein-MOFs. By screening different dihydroxamate linkers with various metal ions (Zn²⁺, Ni²⁺, and Co²⁺), they obtained an array of lattice arrangements (15 ferritin-MOFs), demonstrating the modularity of the system.^{512,513} One of the permutations led them to fdh-Ni-HuHF, a protein-MOF with unusual thermomechanical properties detailed in Section 4.4.

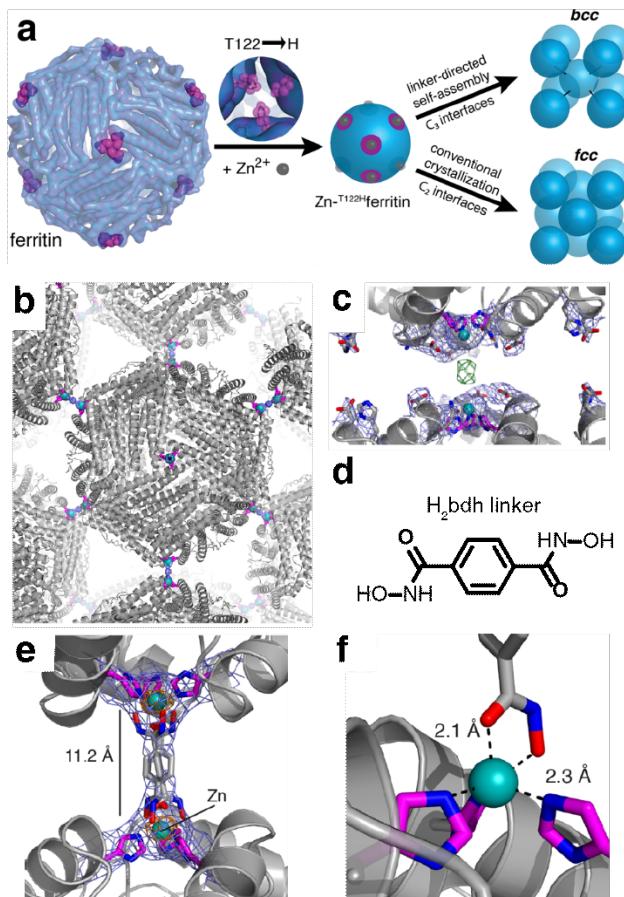


Figure 68. Protein-MOF assembly. **a)** Metal/linker-directed self-assembly of ferritin into 3D crystals. **b)** The *bcc* packing of the bdh-Zn-T¹²²H ferritin lattice, mediated by bdh²⁻ bridges across the *C*₃ symmetric ferritin interfaces. **c)** View of *C*₃ symmetric interfaces between neighboring ferritin molecules, showing the lack of direct protein–protein contacts and the presence of electron density between Zn²⁺ ions (2Fo–Fc map: blue-1 σ ; Fo–Fc map: green-3 σ). **d)** Structure of H₂bdh linker. **e)** Closeup view of the three crystallographically related bdh²⁻ rotamers that bridge neighboring ferritin cages. **f)** Modeled bdh-Zn coordination. Adapted with permission from Ref. ¹²⁹. Copyright 2015 ACS.

3.6.2.2 3D assemblies through receptor-ligand interactions

In addition to metal-binding organic linkers, receptor-ligand interactions have also been employed to mediate 3D protein assemblies. Early works by Sacchettini and Brewer investigated the sugar-mediated assembly of the lectin protein soybean agglutinin (SBA).^{514,515} SBA is a tetrameric protein that specifically binds to GalNAc/Gal sugars. In the presence of bivalent saccharides, SBA is crosslinked and forms a hexagonal lattice with a *P*6₄22 space group, which differs from the crystals of the free protein.⁵¹⁴ In a related study, different ligands were used to crosslink SBA, which gave rise to crystals with the same *P*6₄22 space group, but different unit cell dimensions.⁵¹⁵ A subsequent study by Dotan et al. described the formation of 3D protein assemblies exploiting another tetrameric lectin protein, concanavalin A (ConA).⁵¹⁶ Similarly to the work by Sacchettini and Brewer, Dotan et al. used a bivalent saccharide with *C*₂ symmetry to rationally connect ConA units. Interestingly, the quaternary structure of ConA is almost

tetrahedral, and the researchers surmised that the addition of a symmetric ligand would yield a diamond-like lattice. Bismannopyranoside, a bismannoside (bis-Man) unit with a C_2 spacer, was synthesized and incubated with ConA to form a protein crystal. Analysis of the 6-Å resolution structure revealed a pseudo-cubic orthorhombic unit cell (F_{222} space group). Results from single crystal XRD and auxiliary electron microscopy observations confirmed that ConA formed a diamond-like lattice as intended by rational design. These investigations of sugar-binding proteins provide some of the earliest examples of 3D protein crystals whose lattice arrangements were dictated by synthetic effector molecules.

In a more recent study, Sakai et al. displayed added control over ConA Man-mediated assembly.⁴⁸² Rather than using symmetric bis-Man linkers, the researchers conjugated RhB onto Man, with an oligo(ethylene glycol) spacer, and relied on the π - π stacking-mediated dimerization of RhB to drive protein assembly. Control of the spacing between the sugar groups and RhB, enabled the assembly of ConA into interpenetrating and non-interpenetrating lattices. When three repeating units of ethylene glycol were used as the spacer, ConA formed lattices with $P2_1$ packing, a non-interpenetrating protein crystalline framework. However, extending the linker to four repeating units led to the formation of an $P2_12_12$ interpenetrated lattice (**Figure 69**). As discussed in Sections 3.4.1 and 3.5.4, the coupling of protein-ligand binding and dimerization of RhB moieties by π - π interactions has been a versatile approach to generate ordered protein assemblies in 1D, 2D and 3D.^{270,483,484}

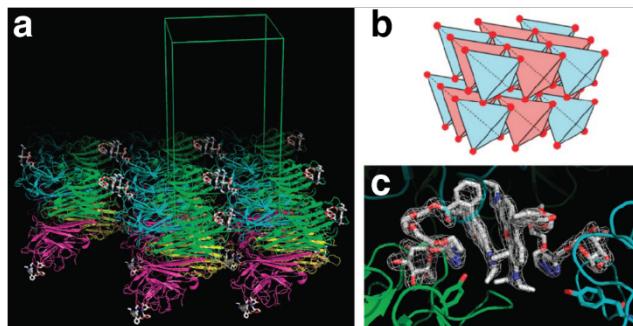


Figure 69. Molecular packing of the Concanavalin A (ConA) and Mannopyranoside-RhB ligands in the 3D crystal. **a)** Packing in one layer of the crystal. **b)** Simplified packing model. ConA units shown in red and blue. Molecules of the same color are crosslinked via dimerizing ligands. **c)** The conformation of the ligand at the dimerization interface. Adapted with permission from Ref. ⁴⁸². Copyright 2014 NPG.

3.6.3 Electrostatically directed 3D protein assembly

3.6.3.1 3D assemblies mediated by small molecules

Electrostatic charge complementarity between the protein surfaces and effectors is one of the most accessible approaches for forming ordered 3D assemblies. Kostiainen and coworkers have used incubation of ferritin with various cationic molecules toward this end.^{309,517} In one study, Mikkila et al. used zinc phthalocyanine (ZnPc) and pyrene derivatives to induce co-crystallization of a ferritin variant.³⁰⁹ Specifically, the pyrene and ZnPc small molecules formed a supramolecular complex carrying a positive charge, which acted as a molecular glue to facilitate

protein-protein contacts between the negatively charged patches on the ferritin surface to mediate the formation of co-crystals. The resulting ternary lattices displayed *fcc* packing and retained the optical properties of the dye molecules. In a conceptually related study, upon screening the co-assembly of apo ferritin from *Pyrococcus furiosus* with a library of cationic macrocyclic cyclophane hosts, Beyeh et al. found that by optimizing the number and orientation of electrostatic interactions between the protein cage and host molecule it was possible to generate porous cyclophane-cage networks.⁵¹⁷ A pillar[5]arene with ten protonatable amines was shown to be an optimal host, giving rise to porous framework structures containing large interconnected voids between the protein cages. The macrocycles were located at the C₃ axes of ferritin, electrostatically bridging the proteins to form a *fcc* lattice. The authors envisioned the use of the frameworks in heterogeneous catalysis by using the cyclophanes to trap guest molecules in close proximity to catalytic sites on the biomolecules, or for the development of water remediation materials.

Similar studies of protein-small molecule assemblies were carried out by Crowley and coworkers with the protein cytochrome *c* (cyt *c*) and the sulfonato-calix[8]arene (sclx₈) macrocycle.⁵¹⁸ Upon screening different crystallization conditions, the researchers found that sclx₈ acted like a molecular glue by facilitating the formation of cyt *c*-sclx₈ complexes in three different crystal forms (corresponding to the space groups *P*3₁, *H*3, and *P*4₃2₁2). Under various conditions, sclx₈ adopted similar extended conformations and was sandwiched between two cyt *c* units. Specifically, two double cone calixarenes were closely packed around the K4 residues of the protein (**Figure 70**). In a follow-up study, Engilberge et al. demonstrated that crystal porosity can be modulated with another positively charged small molecule, spermine.⁵¹⁹ Taking advantage of the host-guest interactions between spermine and sclx₈, the researchers could modulate sclx₈-mediated protein self-assembly by spermine addition and obtain four additional crystal structures. Applying a similar approach, in combination with metal-dependent interactions, Guagnini et al. were able to control the crystal lattice formation of *Penicillium chrysogenum* antifungal protein B (PAFB), an anionic protein rich in His residues.⁵²⁰ On its own, PAFB has been shown to crystallize into a C2-symmetric lattice with 1.1 nm wide pores. Upon the co-crystallization of PAFB with Zn²⁺ and sclx₈, crystals of *H*32 symmetry with 2.2 nm wide pores were formed. The Zn²⁺ ions formed trinuclear metal clusters and mediated protein assembly. Both supramolecular interactions – sclx₈ complexation and zinc coordination – worked in synergy to alter the resulting crystal lattice. Similarly relying on electrostatic contacts between protein and effector, Voet and colleagues used polyoxometalate clusters that carry a negative charge with *de novo* designed oligomeric proteins Pizza-6s and Tako8 to generate cocrystals with various packing geometries.^{521,522}

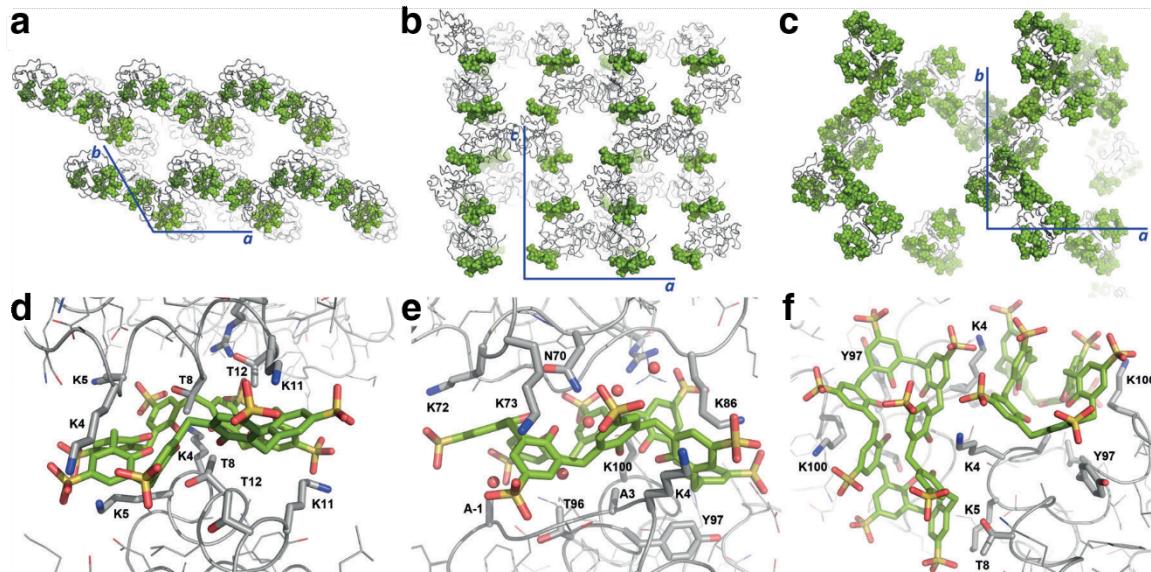


Figure 70. Co-crystallization of cyt c and sulfonato-calix[8]arene (sclx₈). Crystal packing in space groups **a)** P3₁ **b)** H₃, and **c)** P4₃212. Cyt c, sclx₈ and unit cell axes are depicted in gray, green and blue, respectively. **d-f)** The sclx₈ mediated protein-protein interaction for each lattice. Adapted with permission from Ref.⁵¹⁸. Copyright 2018 Wiley.

3.6.3.2 3D assemblies through interactions with nanoparticles

Further taking advantage of the tunable effective net charge on protein surfaces, researchers have exploited the complementary electrostatic interactions between proteins and nanoparticles of different composition (metal, polymer, other proteins) to control protein self-assembly. For example, Kostiainen et al. incubated positively charged, 7–9 nm AuNPs with CCMV or ferritin, protein cages that carry patches of negative charge, to form ordered binary 3D lattices.²⁷² The mixture of CCMV and AuNPs produced AB₈^{fcc} lattices, a lattice type that had only previously been observed for large colloidal polymer particles and did not have a known atomic or molecular crystal counterpart. On the other hand, the mixture of ferritin and AuNPs gave rise to interpenetrating simple cubic AB lattices akin to CsCl crystals. This study illustrated the potential of using small inorganic particles to organize patchy proteins into ordered 3D assemblies.²⁷² Further examples of functional 3D protein assemblies incorporating metal nanoparticles are described in Section 4.3.4.

Liljestrom et al. furthered this design approach by using a protein as the oppositely charged counterpart to the larger protein cages instead of AuNPs. Upon incubating CCMV with avidin, they observed self-assembly of binary 3D lattices integrating both protein types.²⁷⁹ Avidin is a 7.5-nm wide protein with an isoelectric point of ~10.5. Both the size and charge of avidin work synergistically to facilitate the co-assembly of a binary lattice with CCMV cages. The resulting lattice could be modified pre- and post-assembly through the binding of biotinylated fluorescent dyes to the avidin molecules incorporated in the lattice (**Figure 71**). The same biotinylation strategy was used to immobilize HRP or AuNPs within the lattice. This work represents another example of exploiting global electrostatic interactions between protein particles to form ordered 3D arrays.²⁷⁹

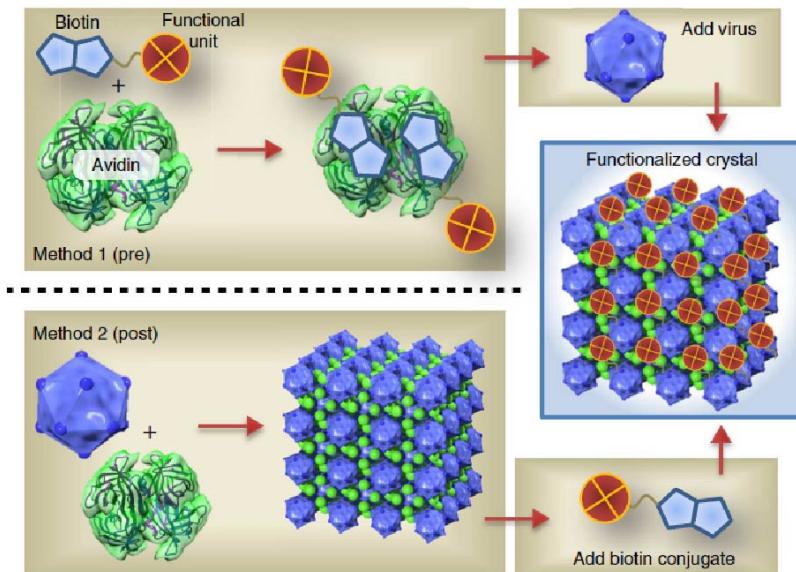


Figure 71. Pre- and post-functionalization of CCMV-avidin co-crystals through biotin-avidin interactions. Method 1 and 2 describe the pre- and post-functionalization approaches, respectively. Both methods yield a functionalized 3D crystalline array. Adapted with permission from Ref. ²⁷⁹. Copyright 2014 NPG.

Charged dendritic polymers provide a modular synthetic system for mediating protein assembly through electrostatic interactions. Upon the incubation of CCMV with a cationic PAMAM dendrimer, Kostiainen et al. formed a hybrid 3D protein-polymer array.²⁷¹ By screening differently charged cationic polymers and ionic environments, they determined that effectors carrying a charge greater than +4 (corresponding to four cationic amines) were necessary to achieve complexation. In a follow-up study, Liljestrom et al. incubated ferritin cages with PAMAM to yield ordered 3D lattices.⁵²³ Screening different salt conditions allowed them to control the lattice packing between cubic and hexagonal arrangements.

Following up on this work, Douglas and coworkers used cationic dendrimers to pre-organize the VLP P22.²⁹³ VLP P22 exhibits a negative charge on its surface and is able to bind to the capsid decoration protein (Dec). Upon creating C_2 symmetric Dec proteins and incubating them with P22, the researchers were able to lock the P22 cages into place after initially templating 3D lattice formation via interaction with the dendritic polymer (**Figure 72**). The strong binding of Dec to P22 produced a highly crosslinked protein material which contributed to capsid stabilization. The dendrimer-mediated templating step was important in the preparation of the material, as simple mixing of Dec with P22 cages led to the formation of amorphous protein aggregates.

Interactions between proteins and small molecules, proteins, or polymers bearing complementary charge states exploit electrostatic interactions to facilitate ordered assemblies. These assemblies highlight the generalizability of the approach for future development.

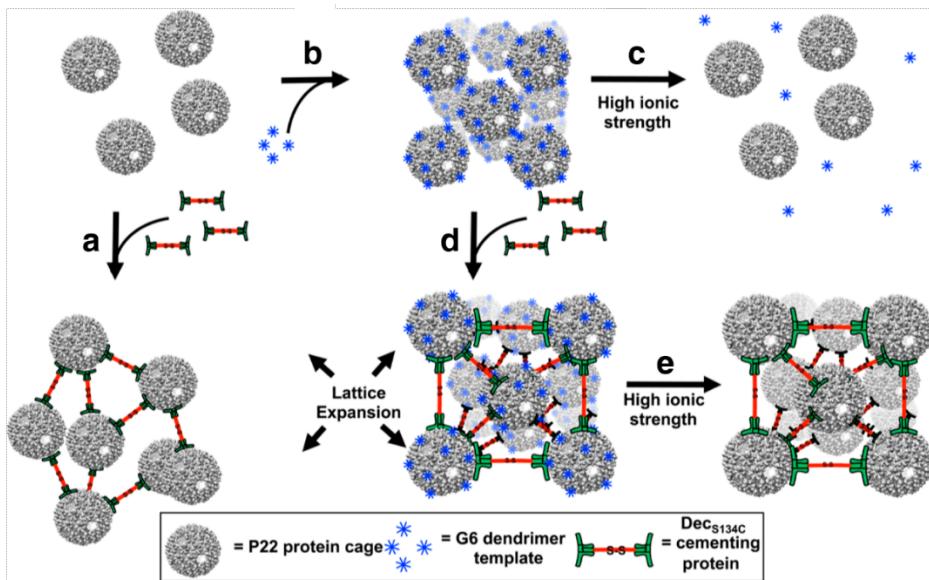


Figure 72. Construction of a VLP-based macromolecular framework. Negatively charged P22 VLPs are mixed with either **a)** Dec cementing proteins or **b)** positively charged PAMAM generation 6 dendrimers (G6) to form an amorphous array or an ordered array, respectively. **c)** High ionic strength conditions disrupt the P22-G6 interaction. **d)** Dec cementing proteins added to the P22-G6 lattice lock the P22 particles in place, preserving the ordered structure. **e)** High ionic strength conditions are used to wash away the G6 dendrimers, leaving behind a protein-only lattice. Adapted with permission from Ref. ²⁹³. Copyright 2018 ACS.

3.6.4 3D protein-DNA hybrid lattices

The base pairing of complementary nucleic acid sequences is another type of interaction that has been widely used to organize biomolecules into ordered arrays.^{524,525} Researchers have used nucleic acids to create an expansive library of nanoscale architectures, primarily through the highly specific Watson-Crick base-pairing interactions between DNA strands with designed sequences.³⁰³⁻³⁰⁶ Important advantages of incorporating nucleic acids into designed assemblies include predictable interparticle distances, controllable annealing/melting temperatures, and precise spatial addressability controlled by sequence design.^{306,526,527}

The predictability of nucleic acid sequence-dependent pairing interactions allows researchers to rationally design assemblies and has been used synergistically with proteins to create hybrid materials. Exploiting the specificity of DNA, Strable et al. conjugated single-stranded DNA onto the surface of cowpea mosaic virus (CPMV) cages and incubated them with another batch of CPMV labeled with complementary DNA strands.³⁰⁷ At 4 °C, the CPMV-DNA conjugates assembled into small hexagonal arrays and grew into 3D aggregates after heating to 35 - 40 °C. In later reports, researchers have been able to create more ordered lattices using DNA. For example, Cigler et al. successfully utilized DNA and AuNPs to organize VLP Q β cages into ordered 3D lattices.⁵²⁸ By decorating the surface of the Q β cages and AuNPs with DNA, the researchers were able to form particles with a corona of DNA strands having similar effective radii to the cages, and enabled the formation of 3D lattices detectable by SAXS.

Building on earlier studies in which complementary DNA strands conjugated to nanoparticle surfaces were used to build lattices,^{529,530} the Mirkin Group has explored the extended assembly of proteins into ordered assemblies mediated by DNA base-pairing.²⁹⁷ The chemically rich protein surfaces were selectively conjugated to DNA to facilitate interparticle connectivity and control spacing. Brodin et al. developed protein-DNA hybrid superlattices using two different variants of the enzyme catalase as building blocks. The surface amine residues were functionalized with azide moieties and further labeled with DNA strands via azide-alkene cycloaddition (Figure 73). The resulting protein-DNA conjugates and DNA-bearing AuNPs were mixed and thermally annealed in different combinations to yield single- and multi-component enzyme crystals that retain enzymatic activity upon assembly. Follow-up studies demonstrated that lattice arrangements of protein and nanoparticle components, as well as pathways of crystallization, could be tuned by controlling the DNA labeling location on the protein and modulating repulsive interactions between the DNA-bearing nanoparticles.^{290,531}

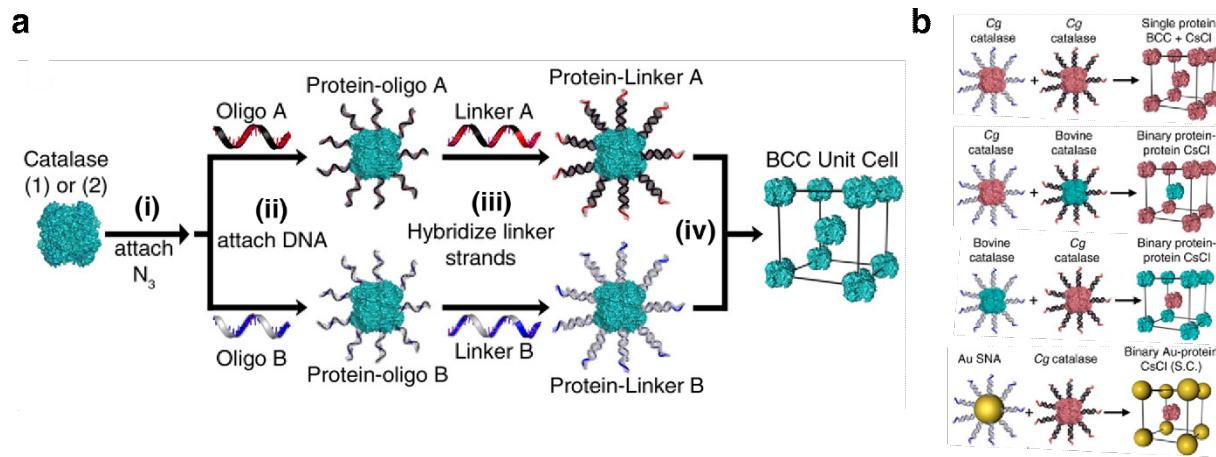


Figure 73. Multicomponent protein-DNA lattices. **a)** Synthesis and assembly of a protein-DNA lattice. (i) The surfaces of both catalases are first functionalized with N_3 moieties. (ii) They are then conjugated with two distinct DNA strands (Oligo A and B) via strain-promoted cycloaddition “click chemistry.” (iii) Hybridization of linker strands to the DNA-protein conjugates and subsequent mixing of the two (iv) results in the assembly of the proteins into a BCC unit cell. **b)** Illustrations of single protein, binary protein, and binary protein-AuNP lattices mediated by DNA hybridization. Adapted with permission from Ref. ²⁹⁷. Copyright 2015 National Academy of Sciences.

Hayes et al. increased the complexity of the protein-DNA conjugated building block by creating self-assembled Janus nanoparticles (i.e., protein-DNA constructs modified with two sets of DNA strands of distinct sequences capable of orthogonal assembly with additional DNA-bearing AuNPs).⁵³² First, they created a GFP variant with evenly distributed Lys residues and a single Cys on the surface. Disulfide chemistry was used to label the Cys residue with a single DNA strand while N-hydroxysuccinimide (NHS) ester chemistry was used to conjugate the Lys residues to DNA strands with an orthogonal sequence (Figure 74a). This procedure created a building block with a set of orthogonally addressable DNA-binding domains. Another batch of GFP was labeled with a complementary strand at the Cys and a different DNA sequence at the Lys residues. Annealing the two types of GFP-DNA building blocks yielded dimers linked by a

DNA duplex and exposing two sets of DNA strands of different sequence on either side. The protein-DNA Janus dimers were integrated into multi-component superlattices upon mixing with DNA-conjugated AuNPs and silver nanoparticles (AgNPs), or AuNPs of different sizes, bearing sequences complementary to both sides of the Janus particle (**Figure 74b-c**). The low symmetry of the Janus particle design enabled the formation of complex multicomponent lattices whose lattice dimensions could be selectively and anisotropically controlled by changing the length of one set of interparticle DNA connectors.

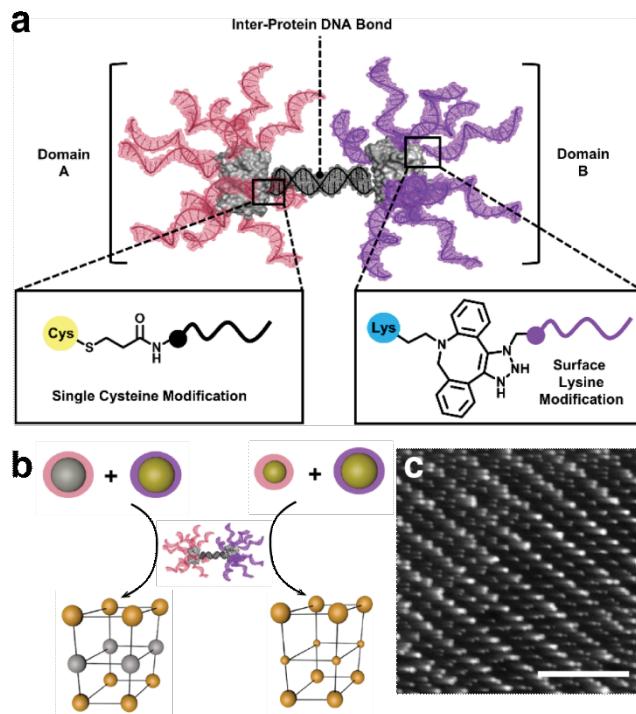


Figure 74. Protein-DNA Janus nanoparticles as building blocks for complex, multicomponent superlattices. **a)** Design of protein-DNA Janus nanoparticle. **b)** Multicomponent superlattices composed of Janus nanoparticles and (left) 10 nm AgNPs and AuNPs or (right) 5 and 10 nm AuNPs. **c)** TEM micrograph of the superlattice from **(b)** (right) embedded in silica reveals layers of 5 and 10 nm AuNPs. Adapted with permission from Ref. ⁵³². Copyright 2018 ACS.

Winegar et al. addressed DNA-directed protein crystallization in the context of single crystal formation. They systematically studied how DNA sequence, length, and protein-attachment position affect the formation and packing of GFP crystals.⁵³³ Upon crystallization of GFP and GFP-DNA conjugates, the researchers observed different packing arrangements, indicating that DNA affected the protein-protein interactions. Interestingly, the crystal contacts were observed between two proteins while electron density corresponding to DNA was not detected. Furthermore, the length and labeling position of DNA was found to affect the crystallization of GFP. Longer DNA strands inhibited crystallization, while different positions did not hinder the protein-protein interactions. Through this study, the effects of protein labeling on self-assembly were elucidated to better design 3D assemblies. Most recently, Partridge et al. showed that replacing a highly conserved protein-protein interaction with DNA-DNA interactions

provided a means to modulate the crystallographic packing of proteins through programmable DNA sequences, which in some cases could be crystallographically resolved.⁵³⁴

As described in section 3.5.3, Subramanian et al. have contributed a different approach to the design of nucleoprotein architectures by designing and studying the assembly properties of metal-binding RIDC3-DNA chimeras.²⁹⁸ The formation of the resulting 3D crystals depended on the intricate balance of Watson-Crick base-pairing, metal-protein binding, and nucleic acid-protein interactions.

Given the large body of work in the areas of protein and DNA assembly in isolation, the synergistic effects of combining the two types of biological building blocks represents an exciting avenue of research for the development of functional biomaterials.⁵³⁵

4. Properties, Functions and Applications of Designed Protein Assemblies

As summarized in Section 2.1, natural protein assemblies present an immense structural diversity across many length scales. Their structures and dimensionality are intimately connected to their properties and biological functions. However, while the structures of protein assemblies generally dictate their function, shape and structure alone are not sufficient to carry out biological roles. Indeed, the activities of protein assemblies often require stimuli-responsive behavior and dynamic interactions with the environment, which cannot be simply derived from a static structure.

Given the inherent challenge of designing protein-protein interactions and self-assembly, early efforts in designed protein assembly or protein nanotechnology have justifiably focused on developing tools and strategies to construct protein assemblies with desired shapes and structures (Section 3). Within the last decade, the field has begun to transition from purely structure-building to the generation of new properties and functions. This rapid progress has not only been enabled by advances in the field of biomolecular design, but also by the outstanding chemical and structural versatility of proteins as building blocks (compared to other biomolecular building blocks like DNA/RNA or peptides), their facile manipulation and production in large quantities by cells, and their capacity to be readily interfaced with biological and abiological components.

In this Section, we aim to provide an overview of designed protein assemblies with emergent properties, functions, and applications that reach beyond mere “structure”, ranging from highly stable and switchable/reconfigurable systems (Sections 4.1 and 4.2) to protein assemblies that serve as active scaffolds (Section 4.3), dynamic assemblies with novel material properties (Section 4.4), and architectures with sophisticated biochemical functions (Section 4.5). It is important to note that there are many protein- or peptide-based systems/materials that fit under these classifications.^{105,536-539} As in the previous Sections, for the sake of focus, we will only cover designed and structurally well-defined protein architectures whose properties, functions, and applications derive from their self-assembly. For natural protein assemblies which

have been engineered or tailored for added functions, the reader is referred to other excellent reviews.^{81,82,85,86}

4.1 Stability of protein assemblies

Although protein stabilization by self-assembly is not an added function *per se*, it is an emergent property that can have important implications for potential functions and applications of a protein. Most natural proteins are only stable in a narrow window of temperature, pH and other solution conditions, in addition to being prone to chemical and physical denaturation. Thus, their use in industrial processes often requires extensive engineering, chemical modification or immobilization.⁵⁴⁰⁻⁵⁴⁸ For example, covalent crosslinking of enzyme assemblies, aggregates or crystals is commonly applied to increase their stability and durability under processing conditions.^{129,549-552} Similarly, from a biological perspective, stabilization of a protein can impart resistance to degradation by proteases^{553,554} and raise its tolerance to mutations, thereby increasing the extent of its evolvability.

It is safe to predict that the assembly of any protein building block into well-ordered oligomers or extended structures is likely to improve its stability by increasing the number of favorable intermolecular interactions, decreasing the total solvent-exposed surface area, and restricting the conformational space of each building block.¹¹⁰ Yet, the stability of designed protein assemblies, or the stabilization of protein components due to self-assembly, has only been explicitly reported in a few instances. As mentioned in section 3.3.4, Huard et al. redesigned the intermonomer interfaces in the 24-meric ferritin cage to make its self-assembly dependent on Cu²⁺ coordination (**Figure 37**).⁴⁰¹ The engineered ferritin monomer MIC1 enabled the isolation of individual ferritin monomers and study of their properties. The authors found that MIC1 became significantly more resistant to unfolding upon Cu-directed self-assembly, as evidenced by an increase of its thermal denaturation midpoint from 39 °C (monomer) to ≥ 87 °C (cage) and the guanidine hydrochloride unfolding midpoint from 1 M (monomer) to 5.2 M (cage) (**Figure 75a**), based on CD spectroscopy results. Cage formation alone was estimated to stabilize each MIC1 monomer by 30 kJ/mol, a value comparable to the folding free energy of a stable, globular protein like cyt c.⁵⁵⁵ Cu²⁺ coordination made an additional stabilization contribution of 3-6 kJ/mol per monomer.

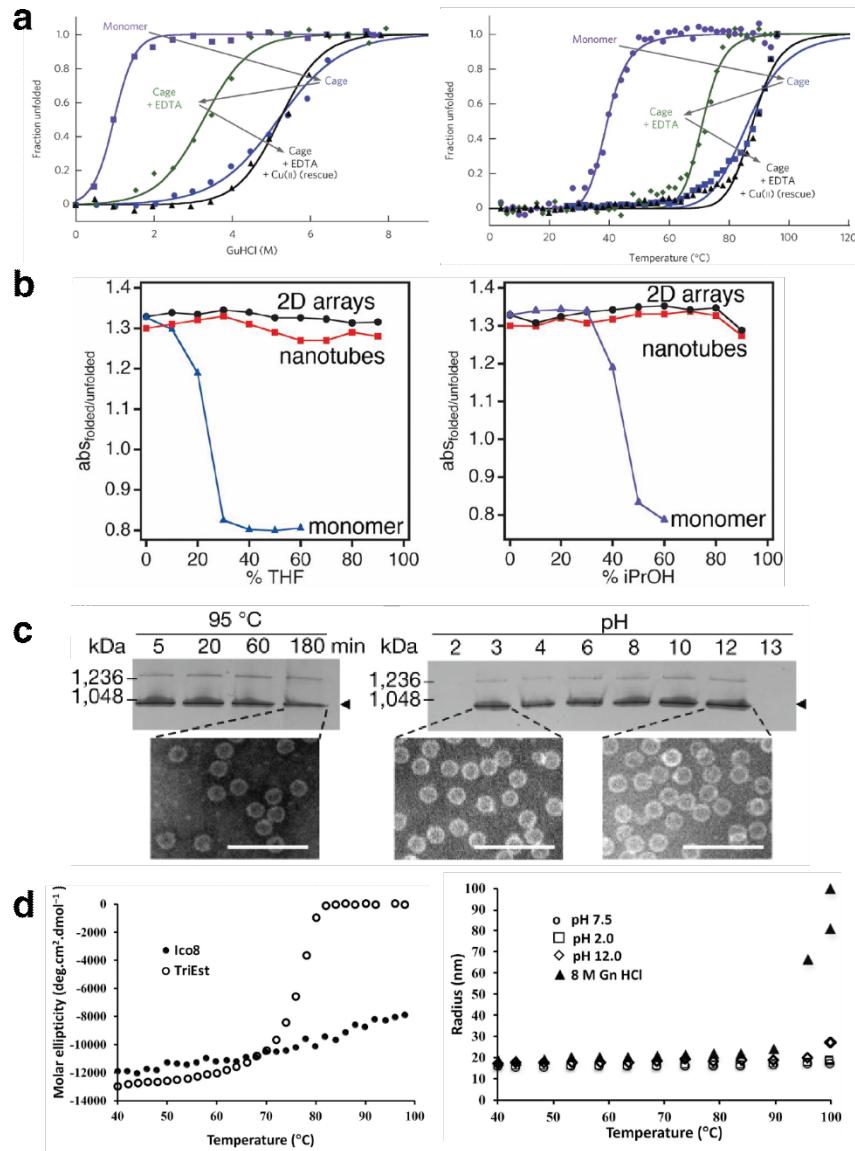


Figure 75. Thermal and chemical stability of designed protein assemblies and their components. **a)** Denaturation of MIC1 in various assembly states (monomer, Cu²⁺-mediated cage, cage + EDTA, cage + EDTA + Cu²⁺ (rescued assembly)) induced by GuHCl titration (left) or thermal denaturation (right) monitored by CD spectroscopy at 222 nm. **b)** Stability of monomeric and self-assembled RIDC3 in THF and iPrOH. **c)** Thermal and pH-dependent stability of TRAP cages tracked by native PAGE and ns-TEM. **d)** Thermally induced unfolding of the Ico8 protein cage and the TriEst subunit monitored by changes in molar ellipticity at 222 nm (left) and by light scattering (right). **(a)** Adapted with permission from Ref. ⁴⁰¹. Copyright 2013 NPG. **(b)** Adapted with permission from Ref. ²⁰⁴. Copyright 2014 National Academy of Sciences. **(c)** Adapted with permission from Ref. ²⁰⁰. Copyright 2019 NPG. **(d)** Adapted with permission from Ref. ¹⁴². Copyright 2019 ACS.

Using UV-vis spectroscopy and ns-TEM, Brodin et al. determined that their Zn-mediated 1D and 2D RIDC3 assemblies (Sections 3.4.3 and 3.5.3, **Figure 46**) were not only thermostable up to ~ 70 °C and ~ 90 °C respectively, but also maintained their structural integrity and dispersity

in polar organic solvents like tetrahydrofuran (THF) and isopropyl alcohol (*i*PrOH) at $\geq 90\%$ (v/v) concentrations. In contrast, individual RIDC3 monomers unfolded at 30% (v/v) THF and 50% (v/v) *i*PrOH, demonstrating the drastic stabilization effect of self-assembly (**Figure 75b**).²⁰⁴ Because the metal-directed assembly process is minimally invasive and the protein monomers can be recovered in their native form upon metal chelation, this strategy is promising for generating heterogeneous enzyme catalyst systems that can operate under industrial processing conditions.

More recently, Malay et al. reported that their Au^+ -directed, TRAP cages with an unusual snub-cube geometry (Section 3.3.3, **Figure 35**) were also highly stable.²⁰⁰ Through gel-electrophoresis and TEM experiments, these ~ 2 MDa cages were found to resist disassembly at high temperatures ($\geq 95\text{ }^\circ\text{C}$), in a wide range of solution pH conditions (pH 3 to 13), and in the presence of high concentrations of chaotropes ($\geq 7\text{ M}$ urea, $\geq 5\%$ SDS) (**Figure 75c**). Despite their high stability, the TRAP cages could be readily disassembled by adding chemical reducing agents that induce dissociation of the Au-Cys coordination bonds.

The stabilization effects are not limited to metal-directed protein assemblies. Hsia et al. observed that their computationally designed icosahedral cage (I3-01, Section 3.3.2, **Figure 32**) could maintain its structure at up to $80\text{ }^\circ\text{C}$ and 6.7 M guanidine hydrochloride, owing both to the stability of the individual trimeric building blocks and the hydrophobic inter-trimer interfaces.³⁹² The high stability of I3-01 also made it robust to genetic fusion, enabling its modification with multiple copies of GFP or designed protein pentamers without perturbation of the underlying icosahedral architecture.

Similarly, Cristie-David et al. reported an extremely stable icosahedral assembly (Ico8) designed through the fusion of a trimeric esterase (TriEst) with pentameric coiled coils (Section 3.3.1, **Figure 29**). The researchers determined by CD and DLS measurements that Ico8 cages remained intact at $120\text{ }^\circ\text{C}$ or in 8 M guanidine hydrochloride, while the TriEst building blocks were denatured at $75\text{ }^\circ\text{C}$ or in 1.5 M guanidine hydrochloride (GuHCl).¹⁴² (**Figure 75d**) Interestingly, Ico8 displays considerable conformational flexibility due to the flexible peptide linker between the TriEst and coiled-coil components, indicating that structural rigidity is not necessary for high stability as has been suggested for viral capsids.⁵⁵⁶⁻⁵⁵⁸ Importantly, like MIC1 and RIDC3 assemblies, the TriEst system provides another compelling example of how cooperative effects arising from the formation of multiple inter-subunit interactions during self-assembly can lead to a dramatic stabilization of the individual protein building blocks.

4.2 Stimuli-responsive and reconfigurable protein assemblies

A critical feature of natural proteins and protein assemblies is their ability to respond to environmental stimuli by changing their structure or assembly state.^{45,85,559} Such structural responsiveness and reconfigurability are imperative for the cellular control of protein functions in time and space. In chemical terms, responsiveness and reconfigurability are associated with reversible interactions that can be readily formed and broken without excessive energy input and can be controlled by mild chemical and physical stimuli. Such bonding fluidity within protein-protein interfaces enables, for example, microtubule extension (polymerization) and shrinkage (depolymerization) in a nucleotide-dependent fashion,⁵⁶⁰ viral capsid assembly/disassembly and

morphology changes,⁵⁶¹ and highly cooperative hemoglobin-O₂ binding that is at the same time modulated allosterically by pH, temperature, CO₂ and biphosphoglycerate.⁵⁶²

Traditionally, protein assembly design has focused on the optimization of energetically favorable interactions between the protein components (or within a protein sequence) to obtain the thermodynamically preferred products.^{145,563-565} This focus on energy minimization has favored singular architectures that assemble predominantly via the formation of quasi-irreversible, hydrophobic interactions and thus reside in deep energy minima. This mode of self-assembly is prone to produce kinetically trapped, disordered aggregates and yields structures that are rigid and static. In contrast, the free-energy landscape of a responsive/reconfigurable protein assembly should contain multiple minima that the system can occupy in a condition-dependent manner. Such a traversable energy landscape can, in essence, be realized with any mode of protein-protein association that is mediated by metal coordination (so long as the metal ions are substitution labile),^{195,200,202,204,397} electrostatic interactions,²⁷² ligands^{270,466,485} or host-guest interactions,^{264,413} reversible covalent bonds,²²⁵ nucleic acid base pairing,⁴⁴⁰ depletion forces,^{566,567} and surface interactions.^{227,496} These types of interactions are reversible and readily modulated by mild stimuli, enabling error-correction and greatly facilitating the formation of well-ordered protein architectures with proper tuning of assembly conditions.³⁵⁷

Another important advantage of externally controllable assembly strategies is their modularity. They open the door to multiple different assembly states from a single protein building block, which is difficult to achieve solely through the design of direct protein-protein interactions. Many such protein assemblies were described in Section 3 as well as in other reviews; some notable examples include the construction of alternatively structured lattices and arrays of ferritin,^{471,512} CCMV,²⁷² M13 phage,^{566,568} LecA,⁴⁸³ TMV,^{475,567} lectin²⁶⁸ and RhuA²²⁷. As we will discuss further in Section 4.3, such lattices and arrays have been used as versatile scaffolds for constructing functional materials.

In this section, we will highlight some studies that explicitly touch on the responsiveness and reconfigurability of protein assemblies and the functions/properties that stem from tunable protein-protein interactions or assembly strategies. We also note that there have been exciting advances in the rational design of protein-based switches and responsive signaling systems based on protein dimerization events;^{190,320,322,323,336,355,356,359} we refer the reader to the extensive literature on these topics.^{310,569}

4.2.1 Stimuli-responsive protein assemblies

It is important to note that responsiveness (a landscape with many energetic minima) does not preclude structural order and stability (a landscape with deep energetic minima). Indeed, the aforementioned Zn-directed 1D, 2D and 3D RIDC3 crystalline arrays^{202,204} and the Au-mediated TRAP cages²⁰⁰ are exceptionally stable with respect to chemical and thermal denaturation, yet they can be readily disassembled upon chelation of the metal ions or chemical reduction. In fact, RIDC3 arrays provide an apt example for reconfigurability, since the kinetics of the Zn-mediated self-assembly process can be controlled by adjusting solution pH or the metal concentration to direct the assembly toward kinetically (1D nanotubes) or thermodynamically (2D and 3D crystalline arrays) preferred products.^{202,204} Furthermore, the 1D nanotubular RIDC3

assemblies can be directly converted into 2D assemblies by raising the temperature to or above 80 °C, lowering the solution pH or lowering Zn concentration.²⁰²

In artificial cage-like protein assemblies, stimulus-responsive assembly/disassembly behavior is critically relevant to their proposed controllable molecular encapsulation and release functions. For example, such behavior is displayed by the Au-TRAP architecture as outlined above.²⁰⁰ Cristie-David et al. designed a trimeric TriEst-peptide chimera, based on their protein-fusion-mediated assembly strategy described earlier (Section 3.3.3), wherein the peptide domain was engineered with an *i/i+4* bis-His motif to form a trimeric coiled-coil domain upon metal coordination.³⁹⁷ The researchers observed a moderately efficient formation of the desired tetrahedral cage complex in the presence of Ni²⁺ (65-75% yield). As designed, treatment with metal chelators or lowering of solution pH induced complex dissociation.

Similarly, the disassembly of Fe³⁺- and Zn²⁺-dependent BMC3 cages (Section 3.3.3, Figure 36) designed by Golub et al. could be triggered not only by elevated temperatures (≥ 70 °C) or metal chelation, but also by treatment with the strong chemical reductant dithionite ($E_{\text{red}} < -500$ mV).¹⁹⁵ By contrast, ascorbate, a weaker reductant ($E_{\text{red}} > -100$ mV), was considerably less effective, as it has a higher reduction potential than the Fe³⁺-tris-hydroxamate centers responsible for the formation of the C₃ symmetric vertices. The redox-dependent switchability of BMC3 cages mimics a key feature of bacterial siderophores: in the cytosol, reduction of the tightly bound Fe³⁺ centers to Fe²⁺ stabilizes the ions, thus promoting their release.⁴⁰⁰ Importantly, owing to their tightly packed shells, BMC3 cages could encapsulate small fluorophores and release them upon chelation-driven disassembly.

Aida and coworkers have taken advantage both of their reversible assembly strategy and the inherent properties of their GroEL building block to design switchable 1D assemblies.^{203,417,439,440} As described in Section 3.4.3 (Figure 45), the barrel-shaped GroEL complex, modified with SP moieties that undergo isomerization to MC on its top and bottom faces, polymerized into micrometer-long nanotubes in the presence of divalent ions, which bind to MC in a 2:1 ratio.⁴¹⁷ As reversible SP-MC isomerization can be effectively triggered by UV (SP \rightarrow MC) and visible (MC \rightarrow SP) light, the researchers were able to control nanotube assembly and dissociation by irradiating the solution, thus providing the first example of light-controllable, artificial protein assemblies.⁴³⁹

Further expanding the versatility of the GroEL system, Kashiwagi et al. engineered an alternative strategy for nanotube assembly. The Cys-modified GroEL barrels were conjugated to complementary DNA strands (GroEL_{10a} and GroEL_{10b}) and assembled through multivalent strand complementarity into stable 1D nanotubes.⁴⁴⁰ In this case, nanotube disassembly could be triggered by a strand displacement reaction, wherein an invading DNA strand was added and formed a more thermodynamically stable duplex with the GroEL-DNA conjugate than the duplex connecting the building blocks, thus disrupting the binding between hybridized GroEL-DNA units and disassembling the nanotube.

In addition, Biswas et al. leveraged the natural ability of GroEL to undergo ATP-dependent conformational changes, which is essential for its biological function as a protein-folding chaperone, to establish an alternative trigger for nanotube disassembly.²⁰³ The researchers demonstrated that Mg-mediated GroEL_{MC} nanotubes, containing on average 20

GroEL_{MC} units, could be dissociated into smaller units upon incubation with ATP (Figure 76a). This observation was attributed to a chemomechanical effect, in which the ATP-mediated conformational changes in GroEL_{MC} led to a decrease in the multivalency of MC-Mg-MC interactions between the building blocks. They further demonstrated that this effect was specific to ATP, CTP, and UTP, which trigger conformational changes in GroEL , but was not observed with ITP or GTP, thus providing strong support for the proposed basis of the chemomechanical effect.

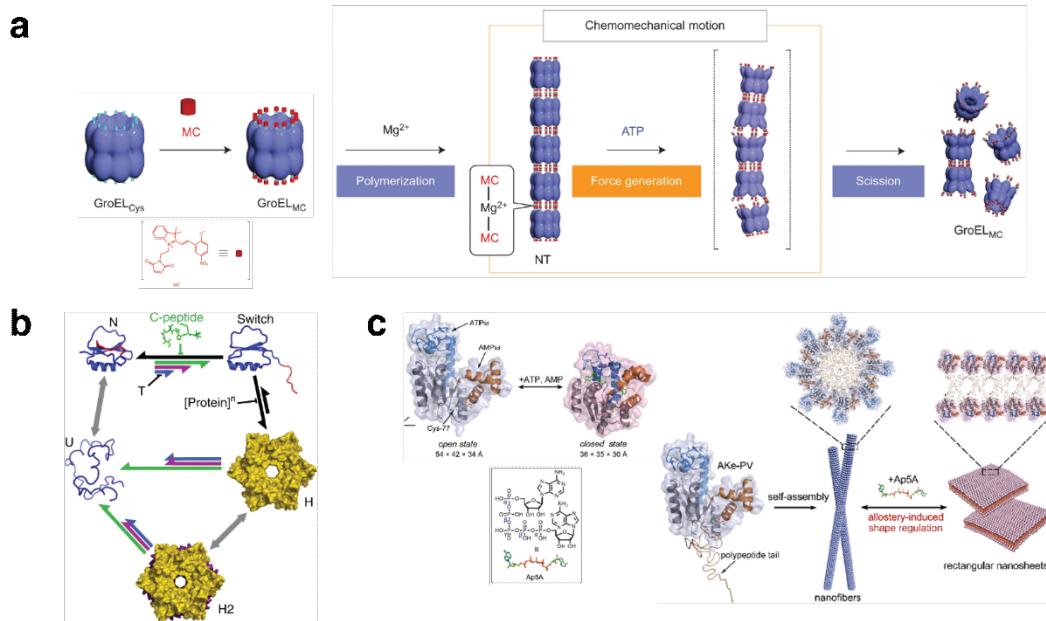


Figure 76. Stimuli-responsive, switchable protein assemblies. **a)** Mg^{2+} -mediated GroEL_{MC} nanotube formation and ATP-triggered chemomechanical nanotube scission. **b)** Thermodynamic model for allosteric assembly of the engineered CI2 variants controlled by temperature and C-peptide concentration. All the engineered variants populate the same five species: native monomer (N), fold-switched monomer (switch), hexamer (H), dodecamer (H2) and unfolded monomer (U). **c)** Adenylate kinase (AKe) in open and closed state (left) and structural transition of AKe-based protein amphiphile assembly from 1D nanofilament to 2D rectangular nanosheet upon Ap5A binding. **(a)** Adapted with permission from Ref. ²⁰³. Copyright 2013 NPG. **(b)** Adapted with permission from Ref. ³⁶³. Copyright 2019 NPG. **(c)** Adapted with permission from Ref. ⁴⁸⁵. Copyright 2019 ACS.

In a similar vein, Campos et al. designed a toroidal protein assembly whose formation could be induced by a fold-switching mechanism involving a small monomeric protein, chymotrypsin inhibitor 2 (CI2).³⁶³ The design was based on the observation that in the P622 symmetric crystal lattices of CI2 obtained under certain solution conditions, the C-terminal β -strand of each monomer was displaced from the hydrophobic protein core to associate with a crystallographically related neighbor. Based on the assumption that this domain-swapping event could stabilize the dodecameric, D_6 -symmetric substructures seen in the lattices, the researchers engineered CI2 to destabilize the interactions between the hydrophobic protein core and the C-terminal β -strand. The resulting variant CI2_{eng} was indeed observed to form hexameric and dodecameric assemblies in solution. Thermodynamic and structural analyses of CI2_{eng} and its

assemblies indicated that the assembly was promoted by the C-terminal displacement from the monomers to expose their hydrophobic cores, which in turn stabilized the key inter-monomer interactions. The researchers could assert control over Cl2_{eng} assembly both by temperature, which modulates Cl_{eng} conformational equilibrium, and the addition of a peptide sequence corresponding to the C-terminal strand, thus providing an exciting demonstration of metamorphic protein engineering⁵⁷⁰ for stimuli-responsive architectures (**Figure 76b**).

In another set of examples for conformationally gated protein assembly, Yan and coworkers took advantage of the switchability of the signaling protein calmodulin (CaM), which undergoes a large Ca^{2+} -dependent structural transformation to bind target proteins and peptides. In one study, the researchers developed a bifunctional linker composed of a phenothiazine headgroup that specifically associates with Ca^{2+} -bound CaM and a rhodamine group that dimerizes through π - π stacking interactions.⁵⁷¹ This linker was designed to induce Ca^{2+} -mediated CaM dimerization. In a second study, the rhodamine group was replaced by various collagen-like proline-hydroxyproline-glycine repeat domains ((POG)_n; n = 6, 10 or 14) intended to promote CaM homotrimerization.⁵⁷² In both cases, the researchers observed reversible, Ca^{2+} -dependent formation of CaM assemblies. With the rhodamine-based linkers, Ca^{2+} addition generated helical filaments, whereas with the POG-based linkers, spherical, cage-like assemblies were observed. The helical pitch of the filaments and the sizes of the cages could be tuned by adjusting Ca^{2+} concentration and the length of the POG domains, respectively. The researchers were also able to confirm the intended linker-mediated CaM dimerization and trimerization processes. They further suggested that CaM molecules must form an interlocked dimer to enable the formation of the observed supramolecular assemblies, although the structural basis for how these dimers form and interconnect remains to be experimentally established.

In another variation of their approach, the same research group employed adenylate kinase (AKe) as a conformationally-gated protein building block, which undergoes considerable compaction upon binding to the bifunctional ligand diadenosine-5-pentaphosphate (Ap5A).⁴⁸⁵ As described in Section 4.5.4, AKe modified with a polyvaline tail (AKe-MV) formed filamentous assemblies with a 20 nm diameter, which were reversibly converted into well-ordered 2D sheets upon Ap5A addition (**Figure 76c**). In analogy to other synthetic amphiphilic systems, this structural transformation was attributed to a change in the shape of the AKe-MV amphiphile from a conical to a cylindrical configuration, which favor tubular and planar packing, respectively.

As previously mentioned, because computational protein design approaches largely target structures that lie in deep free-energy minima, the *de novo* design of conformationally switchable proteins and protein assemblies represents a substantial challenge. Toward this goal, Boyken et al. recently reported the construction of pH-switchable protein assemblies by computational design.³⁵⁷ The researchers took advantage of parametric helical bundles as building blocks and of His side chains as pH-sensitive elements. The repeating geometric cross-sections of the helical bundles allowed extensive H-bonding networks to be built in a modular fashion by tuning the number of His residues, whose protonation at low pH would disrupt the H-bond networks and result in repulsion across the inter-monomer interfaces. Using the HBNet platform in Rosetta (Section 3.2.1), the researchers generated several stable dimeric and trimeric helical bundle architectures (termed pRO's) whose structures closely matched the computed

models. Native mass spectrometry experiments indicated that these structures displayed pH-dependent disassembly as designed. The position and cooperativity of the pH transition could be finely tuned over a pH range from 4 to 6 by modulating the number and positioning of H-bonding and hydrophobic interaction layers. The researchers further demonstrated that pRO's disassembled under low pH conditions could interact with lipid membranes through their exposed hydrophobic cores, efficiently disrupt liposomes *in vitro* and incorporate into lysosomal membranes in cells. In a related approach, Chen et al. used a set of orthogonal heterodimers they had previously designed with HBNet¹⁸⁹ to develop protein logic gates.¹⁹⁰ Fusion of two different monomeric domains with each other enabled the design of dimerization domains that could bring together two halves of a split reporter fused to the cognate monomeric domains. This enabled the development of two- and three-input logic gates, including AND, OR, and NOR gates, in which the activation of the split reporter was controlled by the presence or absence of the relevant dimerization domains according to the relevant logic gate.

A computational interface design strategy was also used to tune assembly and disassembly properties of metal-switchable protein oligomers. As described in Section 3.2.2 (**Figure 26**), Kakkis et al. engineered a cyt *cb*₅₆₂ variant termed TriCyt1, which assembled into homotrimers in low to near-quantitative yields (12% to 95%) upon coordinating various mid-to-late first-row transition metals (Mn²⁺ to Cu²⁺) within a hexa-His motif.¹⁹⁶ Using the crystal structures of the resulting trimers (M:TriCyt1₃) as a template, computationally prescribed mutations were incorporated to stabilize the inter-monomer interfaces with additional hydrophobic and H-bonding interactions. The resulting variants, termed TriCyt2 and TriCyt3, were shown by AUC to quantitatively switch between monomeric and trimeric states upon addition/removal of metal ions (TriCyt2) or tuning of solution pH (TriCyt3), thus furnishing a multi-responsive assembly by rational design.

4.2.2 Reconfigurable protein assemblies

The examples described in this section thus far have highlighted the diverse strategies for designing switchable protein architectures whose assembly and disassembly could be externally turned on and off. Arguably more challenging is the design of protein architectures which can interconvert between two or more well-defined structural states in a reversible and stimuli-responsive manner, while remaining in the assembled state. Such behavior is not only crucial for constructing protein assemblies with adaptive and regulable functions (such as allostery), but also for fabricating protein-based materials with emergent mechanical properties. In contrast to the rapidly growing number of switchable protein assemblies, relatively few such structurally adaptive systems obtained by design have been reported so far.

The nanoporous, 2D lattices of the engineered protein ^{C98}RhuA (Section 3.5.2, **Figure 51**) offer a notable example of an adaptive system.²²⁵ Owing to the short, but highly flexible disulfide bonds that link the ^{C98}RhuA monomers, these lattices are coherently dynamic, i.e., they can coherently transition between "open" and "closed" states upon application of minimal mechanical force (**Figure 77a**).

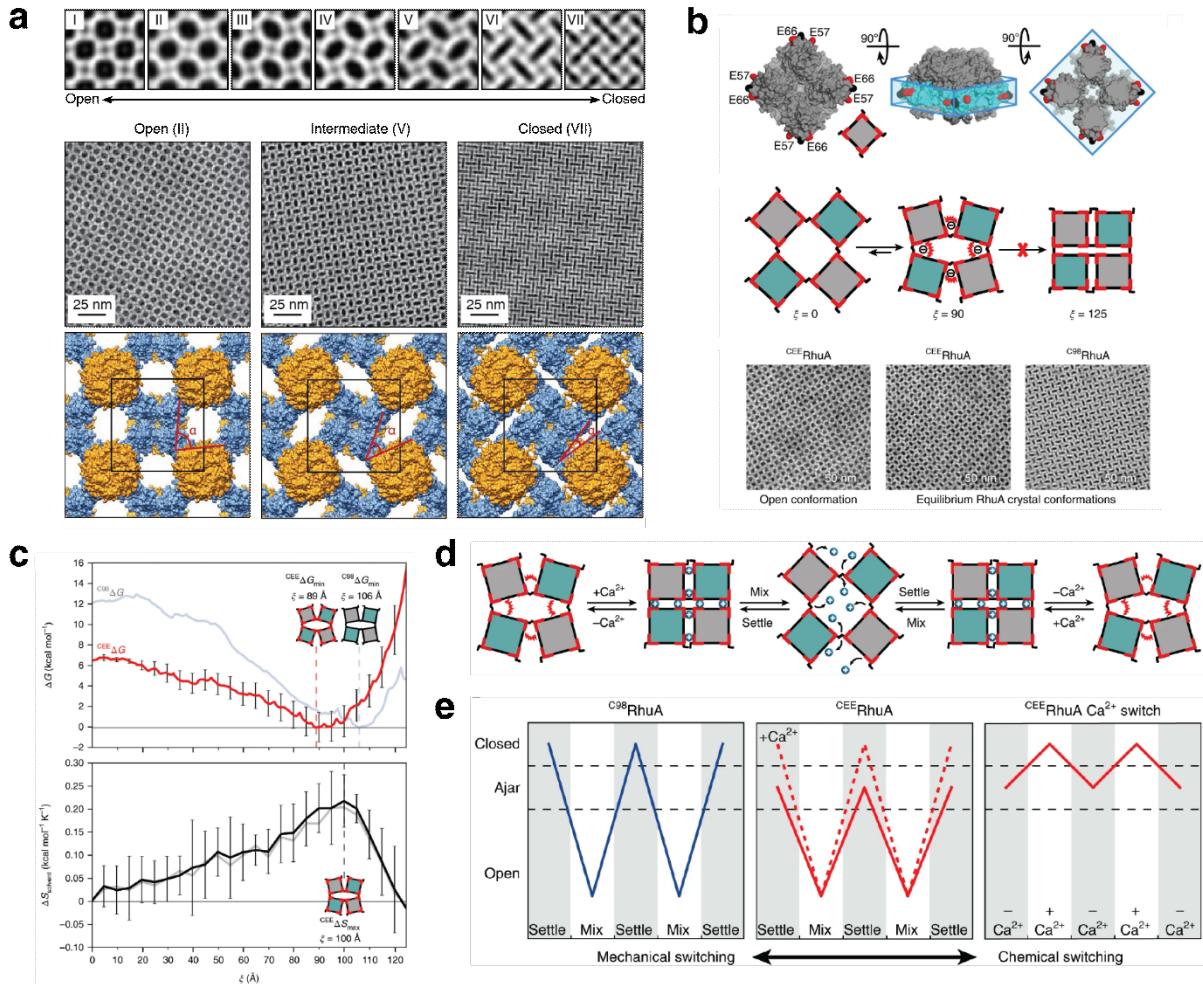


Figure 77. Lattice reconfiguration behavior of disulfide-linked RhuA crystals. **a)** Dynamic, auxetic nature of $C_{98}\text{RhuA}$ crystals. From top: reconstructed 2D images of seven distinct conformational states of the 2D crystals (I-VII); Magnified views of states II, V and VII; structural models of conformations II, V and VII with unit cells and hinge angles (α) between RhuA molecules highlighted in black and red. **b)** Surface representation (top) of a $^{CEE}\text{RhuA}$ tetramer with residues 98 and 57/66 colored in black and red, respectively. Illustration of $^{CEE}\text{RhuA}$ lattice dynamics (middle) and ns-TEM images of lattices (bottom) with open and equilibrium states for $^{CEE}\text{RhuA}$ and closed state for $C_{98}\text{RhuA}$. **c)** Free-energy and solvent entropy profiles for $^{CEE}\text{RhuA}$ (red and black lines) and $C_{98}\text{RhuA}$ (faint blue and grey lines). **d)** Chemical and mechanical reconfiguration behavior of $^{CEE}\text{RhuA}$ lattices. **e)** While $C_{98}\text{RhuA}$ crystals exhibit only one mechanical reconfiguration mode, $^{CEE}\text{RhuA}$ crystals have two modes of mechanical switching and an additional purely Ca^{2+} -induced, chemical switching mode. **(a)** Adapted with permission from Ref. ²²⁵. Copyright 2016 NPG. **(b-e)** Adapted with permission from Ref. ²²⁶. Copyright 2018 NPG.

The $C_{98}\text{RhuA}$ lattices are also noteworthy in that their conformational free-energy landscape has been quantitatively established. Alberstein et al. defined the free-energy landscape by a unidimensional reaction coordinate (ξ) that corresponded to the ellipticity of the lattice pores by taking advantage of the symmetry and coherent structural dynamics of $C_{98}\text{RhuA}$ lattices.²²⁶ All-atom MD simulations revealed a relatively shallow landscape with a distinct energy minimum corresponding to an almost fully closed state. The free energy landscape of $C_{98}\text{RhuA}$

lattices was determined through grid inhomogeneous solvation theory (GIST) calculations to be dictated almost entirely by solvent entropy. Upon confirming that the predicted equilibrium populations of lattice conformational states were quantitatively matched by experimental characterization, the researchers rationally engineered the ^{C98}RhuA surfaces with negatively charged Glu residues (^{CEE}RhuA) to favor a more open lattice conformation at equilibrium (**Figure 77b–c**). The Glu sidechains also created another handle for lattice dynamics actuation, as divalent metal coordination to ^{CEE}RhuA crystals triggered lattice closing. The ^{CEE}RhuA assembled into a lattice whose structure could be both mechanically and chemically actuated (**Figure 77d–e**). The simultaneous presence of structural order/integrity and flexibility, which are often mutually exclusive in designed systems, is an important design principle for adaptive protein assemblies. This requirement is also met in 3D lattices of cage-like proteins that are mediated by flexible linkages^{512,513} and adaptive polymer matrices.^{295,573}

Allostery is a key biological concept in the dynamic control of protein functions,^{110,574,575} which refers to the control of a binding or catalytic process by a chemical event at a distinct site in the protein. To the best of our knowledge, there are only a few examples of designed, allosteric protein assemblies that fit this definition, excluding the many examples of natural proteins that have been reengineered to be switchable.^{363,576,577} In such an example, Churchfield et al. sought to construct a flexible oligomeric protein assembly in which structural changes induced by metal binding could be mechanically transmitted onto a distant site to elicit a chemical event.^{232,233} For this goal, they employed a previously designed, Zn-templated, tetrameric cyt *cb*₅₆₂ assembly (Zn₄⁺^{C81/C96}RIDC1₄; Section 3.2.3),²¹⁰ whose interfaces were redesigned with fluid hydrophobic interactions and two pairs of disulfide bonds. Based on the observation that Zn₄⁺^{C81/C96}RIDC1₄ underwent a considerable separation of one of the intermonomer interfaces, the researchers incorporated an additional disulfide bond into this interface to create a mechanically strained quaternary architecture bearing three pairs of disulfide bonds (^{C38/C81/C96}RIDC1₄). Structural and biophysical analyses indicated that ^{C38/C81/C96}RIDC1₄ was indeed an allosteric system in which Zn²⁺ binding and dissociation were remotely coupled to the reversible formation and breakage of a disulfide bond over a distance >14 Å (**Figure 78**).

More recently, Ghanbarpour et al. exploited domain swapping as a means to design allosteric control of protein function in human cellular retinol binding protein II (hCRBPII).³²⁰ It was previously observed that certain mutations in hCRBPII yield domain-swapped dimers which undergo large conformational changes upon retinal binding.³¹⁹ After identifying key residues responsible for the ligand-induced structural transformation (i.e., mechanical coupling), the researchers installed a His₂Cys metal binding site at the interface of the mobile domain of the dimer distant from the retinal binding site. The resulting system displayed a five-fold difference in the Zn²⁺ binding affinity between the retinal-bound and -unbound forms, thus establishing an allosteric system with remote coupling between ligand- and metal-binding events.

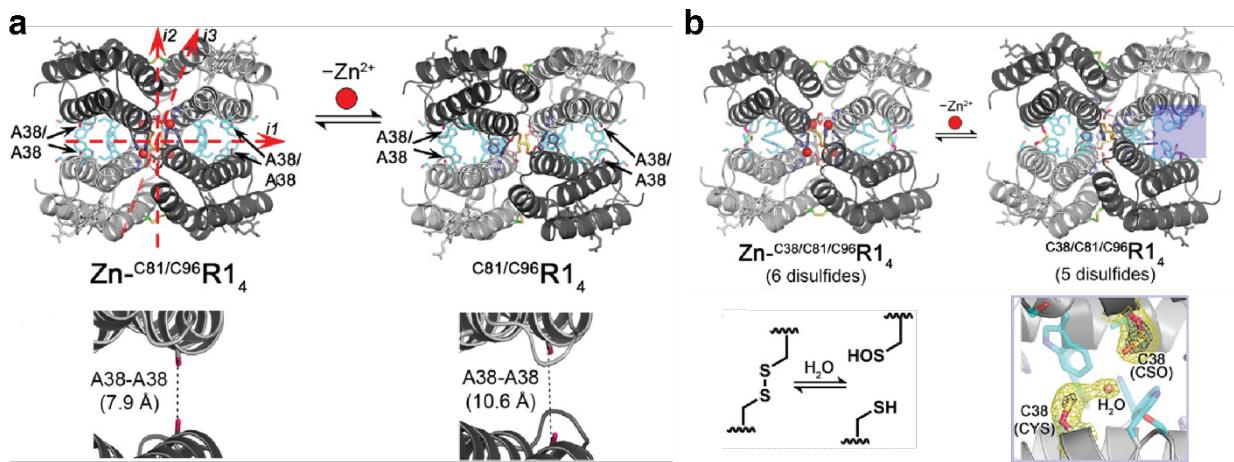


Figure 78. Allosteric metalloprotein assemblies with strained disulfide bonds. **a)** Scheme showing structural rearrangements in the $\text{Zn-C}_{81/\text{C}96}\text{R1}_4$ (labeled as R1_4) tetramer upon Zn^{2+} removal (top) and close-up view of A38-A38 residue pairs (bottom). **b)** Crystal structures of $\text{Zn-C}_{28/\text{C}81/\text{C}96}\text{R1}_4$ and the metal-free $\text{C}_{28/\text{C}81/\text{C}96}\text{R1}_4$ (top); hydrolytic dissociation of a disulfide bond and close-up view of the broken C38-C38 disulfide bond, forming a Cys and Cys sulfenic acid (bottom). Adapted with permission from Ref. ²³². Copyright 2016 ACS.

4.3 Encapsulation, scaffolding and structural organization by designed protein assemblies

Natural proteins and protein assemblies with well-defined shapes and higher-order geometries usually serve as active scaffolds for different functions, such as storage, protection from degradation, transport, and structural templating. The ordered structural features enable accurate control of the position, orientation, and geometry of all the components. Moreover, the specific shapes of protein assemblies often correlate with their functions. For example, hollow cages such as ferritin and viral capsids serve to store large amounts of iron minerals and genomic material and isolate them from the environment.

Inspired by such natural examples, many designed protein assemblies have been utilized as versatile scaffolds for constructing functional materials. Particularly in comparison to synthetic systems (e.g., polymers, MOFs, or covalent-organic frameworks (COFs)), protein architectures are attractive scaffolds because of their biocompatibility, immense structural diversity, ease of functionalization and versatility in function. For example, artificial protein cages can serve as biological containers for encapsulation, release or transport of different cargo such as nucleic acids, proteins, and drugs; they can also act as fluorescent candles for detection or nanoparticle vaccines when fused to specific functional domains.³⁹² 2D protein arrays can serve as structural templates for long-range nanoparticle organization; they can also act as mimics of natural light-harvesting membranes when photoactive proteins are used as building blocks.⁴⁶⁹

In this section, we highlight the use of designed protein assemblies as containers, scaffolds for structure determination and structural templates for the immobilization and organization of biological and synthetic components. We also note that there is a substantial body of research focusing on natural protein assemblies/arrays (e.g., viral or porous protein

lattices) as scaffolds that have been modified to serve as templates and containers.⁵⁷⁸⁻⁵⁸⁰ We refer the reader to the extensive literature on these topics.^{85,94,578-581}

4.3.1 Encapsulation

As summarized in Section 3.3.2, icosahedral protein cages with large diameters are desirable carriers for macromolecule packaging and delivery. In one study, Bale et al. reported two-component icosahedral protein cages with large diameters (24 to 40 nm) that assembled rapidly when the individual components were mixed *in vitro*.³⁹¹ Such structural and assembly features allowed the cage to encapsulate macromolecular cargo in a controllable manner. In further pursuit of this application, the authors engineered the trimeric and pentameric building blocks of an I53-50 cage with positively charged residues on its interior surface. By mixing negatively charged GFP (-30) with the two components of the I53-50 cage, seven to eleven GFP molecules were trapped per icosahedral cage due to electrostatic interactions, occupying roughly 11 to 17% of the interior volume in low salt solution based on fluorescence intensity and UV-vis absorbance measurements (**Figure 79a**). To precisely control the number of GFP molecules encapsulated per cage, the same group built GFP-tagged assemblies using genetic fusion instead of electrostatic interactions (**Figure 79a**).³⁹² They first fused sfGFP to one or both termini of the monomeric subunit from a trimer and the resulting construct formed icosahedral cages (I3-01 ctGFP) with 60 or 120 copies of sfGFP. Then, they applied the same design strategy to generate assemblies with 12 or 24 copies of sfGFP using a previously reported two-component tetrahedral cage (T33-21).³⁸⁹ All the designed sfGFP-nanocage assemblies with GFP copy numbers from 12 to 120 could be used as “standard candles” to correlate fluorescence intensity and GFP copy number.

Besides encapsulation of functional proteins such as sfGFP, icosahedral protein cages can provide a blank slate to evolve desired properties such as encapsulation of nucleic acids. Butterfield et al. started with a two-component icosahedral protein cage (I53-50-v1), containing a large internal volume and positively charged residues on its interior surface, to package its own full-length genome, but with low efficiency (**Figure 79b**).⁵⁸² After several rounds of selection, the evolved nucleocapsids (I53-50-v4) showed higher genome packaging efficiency, improved stability in blood, and an extended *in vivo* circulation time in mice. The resulting synthetic nucleocapsids packaged one full-length RNA genome for every 11 icosahedral cages. By combining computational design and optimization in the evolution process, this study demonstrated that artificial protein cages could acquire virus-like genome packaging and protection properties based on a bottom-up approach.

Edwardson et al. also took advantage of electrostatic interactions to convert a designed porous octahedral protein cage¹⁸⁶ (O3-33, as discussed in section 3.3.2) into a nucleic acid delivery vehicle. Six solvent-exposed residues on the luminal surface of O3-33 were mutated to Arg to create a positively supercharged capsule. The new variant of O3-33, the OP cage, demonstrated a high binding affinity to oligonucleotides. The OP cage was first tested as an siRNA delivery vehicle and shown to modulate gene expression in mammalian cells.²⁷⁵ The researchers then repurposed the OP cage to encapsulate poorly water-soluble small molecules through a two-tier host-guest approach.⁵⁸³ (**Figure 79c**) Anionic surfactants were introduced to the positively charged interior of the designed protein cage, forming protein-scaffolded micelles

within the cavity. The hydrophobic core of the resulting construct was suitable for encapsulating small nonpolar molecules such as the fluorescent dye Nile Red or the chemotherapeutic agent lapatinib. Moreover, altering the lipid composition enabled fine-tuning of the binding affinity and release kinetics for different cargo molecules. The resulting hybrid particles composed of protein cages and amphiphilic molecules are promising scaffolds for delivering nonpolar therapeutic cargo molecules to cells as they can establish a practical balance of serum stability, efficient cellular uptake, and intracellular release.

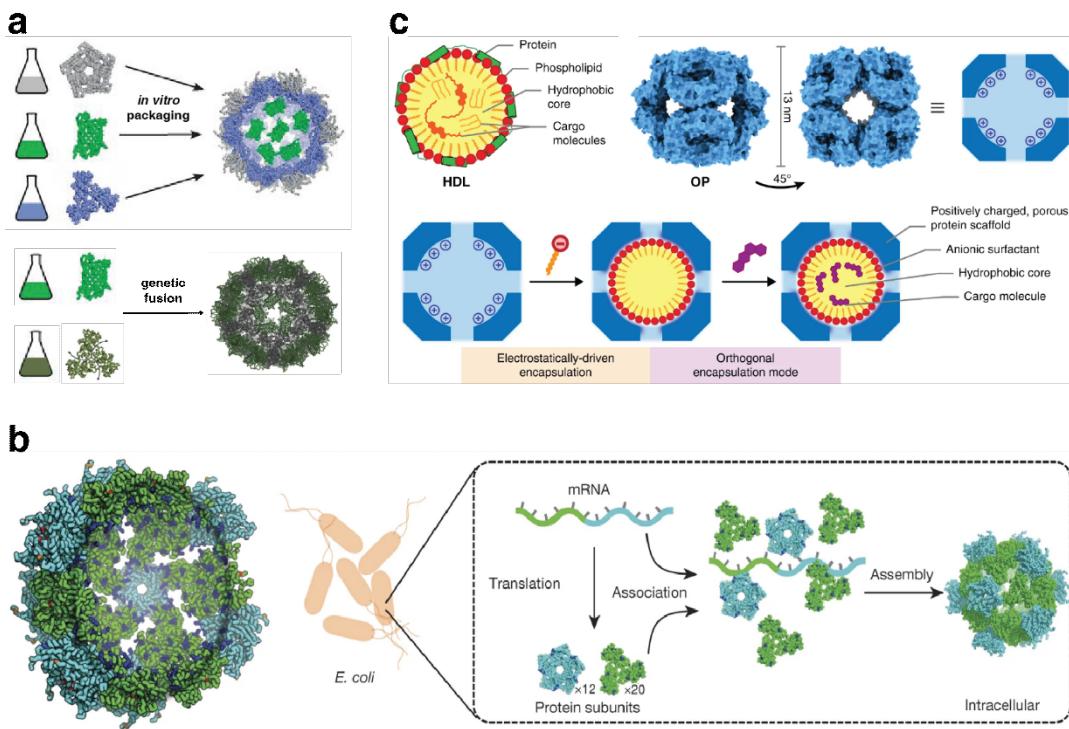


Figure 79. Encapsulation within artificial protein cages. **a)** Encapsulation of supercharged GFP in a positively charged I53-50 cage variant and formation of I3-01 ctGFP cage by genetic fusion. **b)** Design model of I-53-50-v1 (left). Synthetic nucleocapsids encapsulate their own mRNA genomes while assembling into icosahedral capsids inside *E. coli* cells (right). **c)** Self-assembly of lipoprotein-mimetic capsids. **(a)** Adapted with permission from Ref. ³⁹¹. Copyright 2016 AAAS; Adapted with permission from Ref. ³⁹². Copyright 2016 NPG. **(b)** Adapted with permission from Ref. ⁵⁸². Copyright 2017 NPG. **(c)** Adapted with permission from Ref. ⁵⁸³. Copyright 2020 NPG.

Alongside cage-like protein assemblies, GroEL nanotubes have also been utilized for encapsulation towards drug delivery applications. The nanotubes retain GroEL's inherent binding affinity for denatured proteins if the guest is added before Mg²⁺-mediated assembly.⁴¹⁷ This property was exploited for drug encapsulation. Irreversibly denatured α -lactalbumin, which binds to the barrel interior, was conjugated to a drug molecule through a cleavable linker and loaded into the nanotube. The drug release profile was studied *in vitro* under various ATP concentrations, as the nanotube underwent chemomechanical scission in response to ATP. The exterior surface of the nanotube was also functionalized with a boronic acid derivative known to activate the cytosolic uptake of proteins. The modified nanotubes were shown to enter HeLa cells and responded to intracellular ATP to disassemble and release loaded cargo. Preliminary

biodistribution studies in a tumor-carrying mouse model showed preferential accumulation of the nanotubes in tumor tissues.²⁰³ Furthermore, cell uptake studies carried out with length-controlled GroEL nanotubes showed tubes shorter than 100 nm to have efficient uptake into HEP3B cells, with a general trend indicating length-dependent uptake efficiency that favors shorter tubes.²⁰¹

4.3.2 Scaffolding for applications in structural biology

As demonstrated by an increasing number of examples, the structural order and uniformity of designed protein assemblies can be invaluable in scaffolding of molecular/biomolecular guests for structure determination by XRD and cryoEM.^{581,584,585} Compared to small molecule crystals, 3D protein lattices exhibit larger channel and pore diameters that permit the sequestration of large exogenous molecules and even proteins.⁵⁸⁶ Another advantage of protein building blocks lies in their chemically rich surfaces, which can be augmented through protein engineering and specifically functionalized by bioconjugation. Exploiting both characteristics, protein crystals have been used as crystallographic hosts to immobilize normally “uncrystallizable” molecules for structure determination. In an early example, Ni et al. immobilized a flexible heme peptide fragment, microperoxidase (MP9_{cb562}), within a designed, metal-mediated protein cage.³⁹⁵ The tetrahedral cage, which featured a 35 Å-wide cavity, was used to capture MP9_{cb562} in its interior via the coordination of the heme cofactor to a His residue located in the lumen. This anchoring strategy enabled the determination of the cage-peptide cocrystal at atomic resolution and the crystallographic characterization of a microperoxidase for the first time (**Figure 80a**).

To further develop scaffold-assisted crystallography techniques in an engineered protein crystal with large pores, Huber et al. used the CJ crystal system.⁵⁸⁷ CJ crystals are composed of a single protein, a variant of CJ0, a putative periplasmic polyisoprenoid-binding protein from *Campylobacter jejuni*, and feature 13 nm-wide solvent channels. After designing Cys residues pointing into the large pores, the researchers were able to immobilize an array of small molecules via disulfide bond formation at the designed Cys residues and visualize the electron densities of the anchored molecules by XRD analysis.

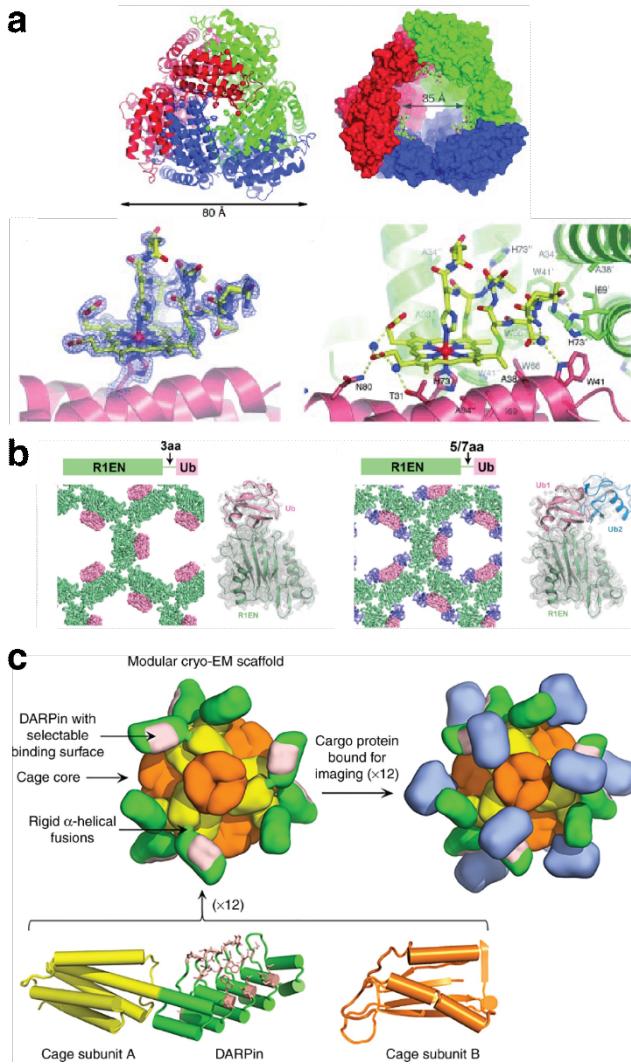


Figure 80. Protein assemblies as scaffolds for structure determination. **a)** Crystal structure of the tetrahedral protein cage architecture of $\text{Zn}_{30}:\text{CFMC-1}_{12}$ (top) and of the microperoxidase (MP9cb562) immobilized within the cage cavity (bottom). **b)** Diagrams and crystal structures of the fusion proteins R1EN-Ub. **c)** Design of a modular protein scaffold for cryo-EM imaging. **(a)** Adapted with permission from Ref. ³⁹⁵. Copyright 2010 Wiley. **(b)** Adapted with permission from Ref. ⁵⁸⁴. Copyright 2018 ACS. **(c)** Adapted with permission from Ref. ⁵⁸⁸. Copyright 2019 NPG.

In addition to peptides and small molecules, 3D protein lattices can also be used to host other proteins for crystallographic structure determination. Maita demonstrated the immobilization of ubiquitin through C-terminal genetic fusion to honeycomb lattice-forming protein R1EN.⁵⁸⁴ The large 11 nm pores within the R1EN lattice were able to accommodate several copies of the target in a symmetric fashion. Three R1EN-ubiquitin constructs were crystallized with various linker lengths under the same conditions as the original R1EN (**Figure 80b**). The ubiquitin structure was solved at 1.7-2.4 Å resolution and was almost identical to the previously published structure.

Following recent technical advances, cryo-EM has become a very powerful and now-common tool for macromolecular structure determination. In contrast to XRD, cryo-EM does not require crystalline samples and provides the distinct advantage of capturing snapshots of protein assemblies in a solution-like state, thus enabling the study of dynamic systems. However, cryoEM methods have their own intrinsic technical limitations and challenges.⁵⁸⁹ For example, image processing and 3D structure reconstruction can be particularly challenging for small proteins (<50 – 100 kDa),⁵⁸⁵ due to the low signal-to-noise ratio in single-particle imaging.⁵⁹⁰ To break through this size barrier in cryo-EM, one strategy is to attach the imaging target to a larger host or scaffold structure. Along these lines, the Yeates Group explored the potential of artificial tetrahedral protein cages (T33-21³⁸⁹ and T33-31,⁵⁹¹ Section 3.3.2) as modular, symmetrical scaffolding systems for cryo-EM.⁵⁸⁵ Using a rigid, continuous, α -helical linker, the researchers genetically fused 12 copies of a 17-kDa protein (Ankyrin repeat protein, DARPin) to the exterior of tetrahedral cages. The resulting construct was amenable to structural analysis by single-particle cryo-EM, revealing structural details of the DARPin adaptor component at resolutions ranging from 3.5 to 5.5 Å. Taking advantage of the ability of DARPin adaptors to recognize and tightly bind protein targets, the same scaffold was used to bind and symmetrically display 12 copies of GFP, a 26-kDa protein (Figure 80c).⁵⁸⁸ The GFP molecules were associated rigidly enough with the host cage to be resolved at 3.8 Å by cryo-EM/single-particle reconstruction. These results demonstrate that proteins considerably smaller than the proposed limit of 50 kDa for cryo-EM reconstruction can be visualized clearly when arrayed in an ordered fashion on a designed symmetric protein scaffold, thereby expanding the accessible target size range that can be studied by cryo-EM.

4.3.3 Scaffolding of biological molecules

The examples described above highlight the use of protein assemblies as structural scaffolds for encapsulation and structural determination. In many cases, the scaffolding effect of protein assemblies can also lead to new, synergistic functions that are unattainable with the individual components. For example, co-encapsulation of two enzymes in a protein crystal lattice can enable a cascade of catalytic reactions. Towards this end, Nguyen et al. redesigned a protein crystal to entrap lipase B and alcohol dehydrogenase to carry out a two-step reaction within a protein crystal *in vivo*.⁵⁹² The cytoplasmic polyhedrosis virus (CPV) programs infected cells to produce the polyhedrin monomer (PhM) protein, which forms the polyhedral crystal (PhC) during cellular replication. The PhC protects CPV particles from damage. The researchers redesigned PhM, by deleting a 38-residue portion of the monomer, to form hollow PhC cages with 5-nm wide cavities that are suitable for enzyme encapsulation. Overexpression of this PhM mutant along with the two enzymes targeted for encapsulation produced crystals with both enzymes entrapped in the crystal cavities inside the cell. Interestingly, the composite crystal demonstrated 1.9-fold and 3.8-fold higher reactivity of the cascade reaction compared to the wild-type crystal and the mixture of the free enzymes, respectively. The enhanced reactivity was attributed to the efficient diffusion of the substrate and intermediate through the expanded channels within the nanoporous crystal. This study demonstrates that *in vivo* protein crystallization is a promising approach toward generating artificial biocatalyst platforms by immobilizing different enzymes in nanoporous protein scaffolds.

McConnell et al. covalently attached two enzymes known to act synergistically in cellulose degradation to the surface of previously reported tetrahedral cages (Section 3.3.2, **Figure 30**) using a sortase enzyme to catalyze their tethering to the designed scaffold via a polyglycine tag.⁵⁹³ The enzyme-modified cages demonstrated enhanced activity in a cellulose degradation assay compared to free enzymes in solution or unmodified cages.

Crystalline 2D protein arrays are promising biotechnological scaffolds due to their ability to display polypeptides with high density and nanoscale tunability/reconfigurability. Engineering the surfaces of 2D protein lattices enables selective organization of the target biomolecules in a site-dependent manner. To generate functionalized 2D protein materials from the bottom up, Subramanian et al. developed an enzyme-directed surface modification approach to site-selectively tailor the surface of 2D protein crystals by using Sfp phosphopantetheinyl transferase (PPTase). Two different designed, self-assembled protein arrays (2D Zn-mediated RIDC3 crystals and disulfide-linked ^{C98}RhuA lattices), whose surfaces can be tagged with functional sites (short peptide ybbR or molecular tag CoA), were prepared for enzymatic modification by Sfp PPTase (**Figure 81a**).⁵⁹⁴ Notably, the site-specific modification of both 2D arrays could be carried out genetically or chemically without disrupting the underlying crystal lattice packing, characterized by TEM images of the protein crystals and confocal microscopy images of the labeled tags. This study highlights the potential for chemoenzymatic modification of 2D protein arrays towards the hierarchical construction of multicomponent protein systems.

Bradley and coworkers have designed a circular tandem repeat architecture with α -helical repeats to generate circular protein nanoparticles that could display multiple copies of functional protein domains.⁵⁹⁵ Their approach centered on deconstructing a large toroid of 24 left-handed, α -helical repeats (cTRP24) into monomers with 3, 4, 6, 8, or 12 repeats. Assembly of the monomers back into the 24-repeat structure allowed for multiple copies of a functional protein to be displayed on the larger oligomer via attachment to each monomer. Initially, only the 12-repeat monomer (cTRP24₁₂) assembled into the 24-repeat oligomer. Subsequent design of disulfide bonds to the putative interfaces in the oligomer yielded stable, disulfide-stapled dimers from the 12-repeat monomer (cTRP24₁₂SS) and tetramers from the 6-repeat monomer(cTRP24₆SS). Finally, various cargo proteins, including single-chain class I major histocompatibility complex molecule (scMHC) and single-chain tumor necrosis factor receptor superfamily ligand trimers, were displayed on the disulfide-stapled oligomers either by genetic fusion or by specific labeling with the SpyCatcher protein ligation domain or the SH2 peptide-binding domain (**Figure 81b**). The attached proteins remained active, and the formation of oligomers with multiple copies of each protein possessed properties such as avidity that were unattainable with monomeric proteins.

Covalent assembly of proteins into a 2D array can lead to emerging functions due to the highly ordered patterning and spacing between the components. Li et al. designed a 2D protein nanosheet as a light-harvesting system by covalent assembly of EBFP2 (donor) and EGFP (acceptor) proteins, in which the fluorescent chromophores were evenly distributed and adopted a fixed orientation (**Figure 81c**).⁴⁶⁹ By varying the length of the inducing linker, the distance between adjacent chromophores and the overall size of the assembly could be optimized to enhance energy transfer efficiency.

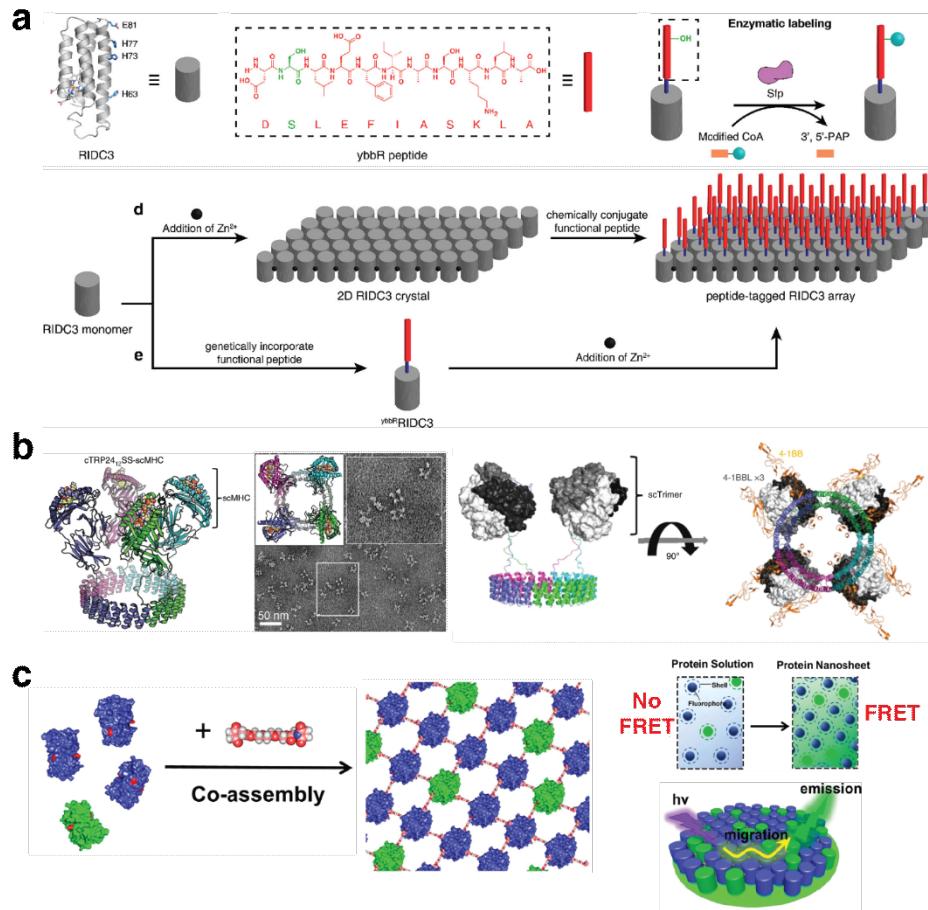


Figure 81. Scaffolding of biological molecules. **a)** Design of peptide-tagged RIDC3 arrays for enzymatic labeling. **b)** Structural model and ns-TEM image of cTRP246SS-scMHC (left); Structural diagram of the cTRP246SS-scTrimer^{4-1BB}, with the single-chain trimer rendered in space filling representation (right). **c)** Covalently linked nanosheets composed of EBFP2 (donor) and EGFP (acceptor) proteins. Within the nanosheets, energy absorbed by donors can be transferred to acceptors by direct FRET or successive donor-to-donor transfers, conferring light harvesting properties to the nanosheets. **(a)** Adapted with permission from Ref. ⁵⁹⁴. Copyright 2020 ACS. **(b)** Adapted with permission from Ref. ⁵⁹⁵. Copyright 2020 NPG. **(c)** Adapted with permission from Ref. ⁴⁶⁹. Copyright 2019 ACS.

Zhang et al. reported the design of single-layer porous protein nanosheets for the precise separation of nanoparticles.⁴⁷⁷ The researchers cross-linked a TMVCP variant with Cys residues on the outer surface of the ring via Cu²⁺-catalyzed disulfide-bond formation to yield ordered 2D nanosheets (Section 3.5.2). The resulting single-layer 2D nanosheets with regular 4 nm-wide pores extended over tens of micrometers in width. Based on the single-layer nanosheets, the authors prepared ultrafiltration membranes with a thickness of 40 nm that could precisely separate particles around 4 nm in diameter with high selectivity and exhibited water permeance up to ~7000 L m⁻²h⁻¹bar⁻¹.

4.3.4 Scaffolding of inorganic and synthetic components

The ability to control the spatial arrangement, orientation and geometry of assembled nano-objects such as metal nanoparticles with high precision is crucial to attain collective optical, magnetic and electronic properties.^{101,596} Given their ability to encapsulate various cargos and monodisperse structures as discussed above, protein cages such as ferritin and virus capsids have proven to be attractive scaffolds for the integration of synthetic components. Kostiainen et al. leveraged electrostatic interactions between iron oxide-loaded ferritin cages (recombinant magnetoferitin particles) with photodegradable Newkome-type dendrons that lead to self-assembly of micrometer-sized crystals with *fcc* lattices.⁵⁹⁷ The ordered superstructures could be disassembled by an optical stimulus that degraded the dendrons. Magnetometry studies revealed that the crystallographic order imparted by self-assembly induced dipole-dipole magnetostatic interactions between the magnetic nanoparticles, which manifest in the hysteresis loops of field-dependent magnetization and field-cooled temperature-dependent magnetization studies. The dipolar-coupled assemblies displayed magnetic properties that deviate from traditional superparamagnetism seen in isolated magnetoferitin particles. Upon UV-triggered disassembly, the released magnetoferitin particles regained their superparamagnetic properties. Okuda et al. synthesized cerium oxide nanoparticles with a narrow size distribution using apoferritin.⁵⁹⁸ Moreover, the authors used Ce³⁺ ions to bridge ferritin cages and form 2D (domain size over 500 nm) and 3D arrays (octahedral or prism-like) of CeO₂ nanoparticles encapsulated within the cages.

Maity et al. employed ferritin crystals as scaffolds to study Au sub-nanocluster nucleation under reducing conditions.⁵⁹⁹ They first replaced two residues at the metal accumulation center of ferritin with Cys to pre-organize Au ions within the cage lumen, then obtained Au-treated ferritin single crystals and cross-linked them with glutaraldehyde to facilitate structural investigation by XRD. Upon treatment of these crystals with a strong reduction agent (NaBH₄), the researchers were able to track the movement of Au ions toward the three-fold symmetric ferritin cage channels where the ions formed sub-nanoclusters. The ion migration and cluster formation were accompanied by changes in the side chain conformations of Au-bound amino acids, providing insights on dynamic metal-protein interactions involved in metal cluster formation.

Protein cage superlattice scaffolds are not limited to single-component systems. As described in Section 3.6.3.2 (**Figure 71**), Lijestrom et al. electrostatically assembled avidin and CCMV cages into binary crystals with *bcc* lattices responsive to external stimuli, such as pH and ionic strength.²⁷⁹ Furthermore, avidin-biotin interactions enabled selective lattice functionalization pre- and post-assembly with various synthetic molecules, including fluorescent dyes, enzymes, and plasmonic nanoparticles. Mikkila et al. combined ferritin, phthalocyanines (Pc) and 1,3,6,8-pyrenetetrasulfonic acid (PTSA) to form ternary *fcc* crystals with photoactive properties.³⁰⁹ The design strategy was based on the formation of a Pc/PTSA complex through electrostatic and π-π interactions, which in turn could bind to the ferritin cages via charge interactions to induce crystallization. The resulting crystals retained photoactive properties such as fluorescence at 695 nm and efficient light-induced singlet O₂ production, which hold great promise in various applications such as photodynamic therapy (PDT), water treatment and diagnostic array development. Using *Thermotoga maritima* ferritin (TmFtn) as a host protein for enzyme encapsulation, Chakraborti et al. generated binary crystals consisting of cargo-

encapsulated TmFtn and Au nanoparticles (AuNPs).⁶⁰⁰ The TmFtn cage was capable of salt-mediated reversible assembly/disassembly and its negatively charged lumen enabled encapsulation the positively charged guest proteins such as +36GFP and lysozyme. TmFtn-encapsulated lysozyme retained its enzymatic activity both in solution and within the superlattice of TmFtn-AuNP co-crystals.

As discussed in section 3.6.1 (**Figure 64**), the Beck Group has relied on electrostatic interactions to assemble binary protein lattices based on oppositely charged ferritin cages.²⁷⁷ This protein assembly strategy enabled the formation of mixed nanoparticle architectures, which was demonstrated by loading cerium oxide and cobalt oxide nanoparticles into positively and negatively charged ferritin cages, respectively (**Figure 82a**). The controlled assembly of these protein-nanoparticle composites via electrostatic interactions led to the formation of highly ordered binary nanoparticle superlattices as free-standing crystals hundreds of micrometers in size. The binary protein-nanoparticle crystals could function as versatile, highly modular materials for various applications due to the functional diversity of nanoparticles that could be loaded into the ferritin pores.⁶⁰¹

In a follow-up study, the same group demonstrated that the free-standing nanoparticle superlattices are catalytically active and can be reused for multiple reaction cycles.⁶⁰² Specifically, the CeO₂ nanoparticles were shown to display oxidase and peroxidase-like activity. Compared to free nanoparticles in solution, which are prone to aggregation and are hard to reuse, well-ordered CeO₂ nanoparticles in superlattices were stabilized by the protein scaffold without losing access to reactants due to channels within the crystal and protein cage pores, and therefore retained high activity after several reaction cycles. The researchers also found that the binary protein lattice can be transformed to a unary cubic lattice composed only of the negatively charged ferritin variant under high Mg²⁺ concentration conditions, which were also used to organize metal oxide nanoparticles.⁵⁰⁷ As illustrated in these studies, protein cage superlattices are versatile and tunable platforms to construct functional hybrid materials with modular biological/inorganic components.

Ring-shaped proteins tend to stack and form 1D nanotubes with a central hollow channel, which could serve as a template for 1D nanoparticle assembly (Section 3.3). Ardini et al. prepared 1D nanotubes by metal-induced self-assembly of ring-shaped peroxiredoxin with an engineered N-terminal His-tag and divalent metal ions such as Ni²⁺, Zn²⁺ and Co²⁺.⁶⁰³ The well-stacked protein rings have been applied to successfully capture and arrange colloidal Ni²⁺-modified AuNPs into 1D arrays. The formation of such nano-peapod complexes strictly depended on nanoparticle dimensions as the peroxiredoxin template could only capture ultrasmall AuNPs (~1.6 nm) in a size-selective manner. Similarly, Manuguri et al. installed a His-tag as a metal-binding site in the 7-nm diameter pore of the peroxiredoxin and used the engineered protein to organize iron oxyhydroxide nanoparticles.⁶⁰⁴ Changing the pH caused the peroxiredoxin bound to nanoparticles to stack, thus confining the nanoparticles to extended 1D assemblies. Both studies demonstrate that the peroxiredoxin-based 1D nanotube is a versatile template to induce one-dimensional nanoparticle assembly.

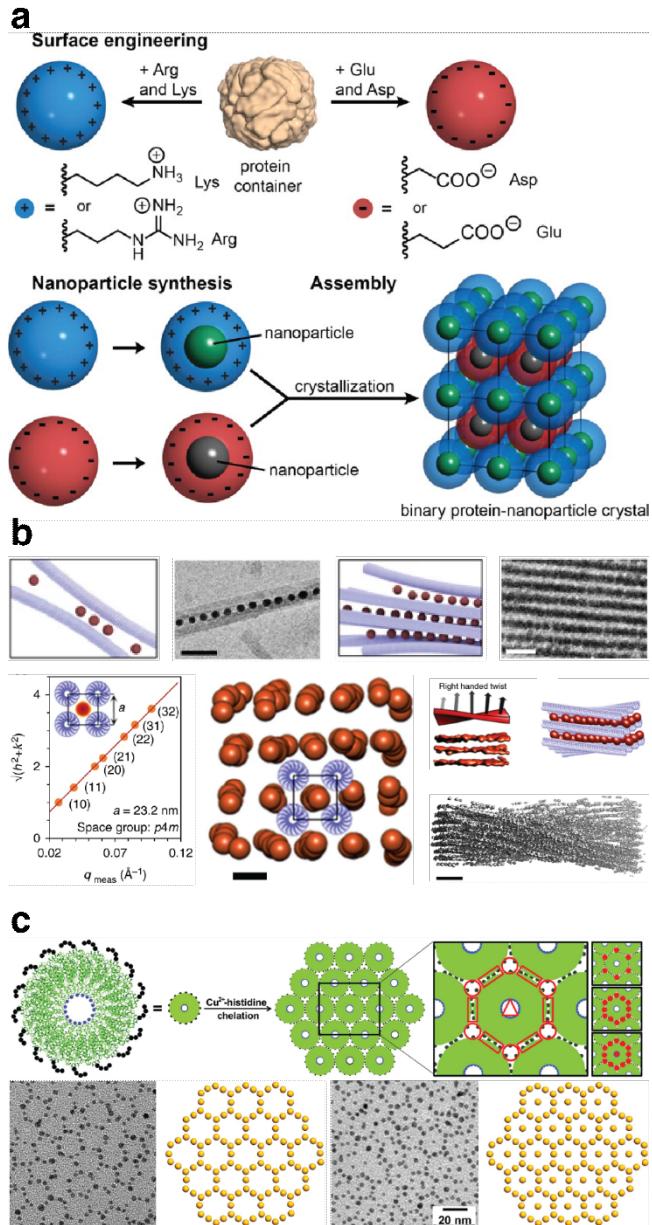


Figure 82. Protein assemblies as scaffolding of inorganic components. **a)** Scheme for the assembly of binary nanoparticle superlattices based on charged protein containers. **b)** Hierarchical structure of the superlattice wires composed of TMVs and AuNPs. **c)** Self-assembly of highly ordered 2D AuNP lattices directed by TMV monolayer sheets. **(a)** Adapted with permission from Ref. ²⁷⁷. Copyright 2016 ACS. **(b)** Adapted with permission from Ref. ²⁹⁶. Copyright 2017 NPG. **(c)** Adapted with permission from Ref. ⁴⁷⁹. Copyright 2019 Wiley.

The SP1 dodecamer, which naturally forms nanotubes, has been an appealing scaffold to organize QDs.²⁹¹ Miao et al. have taken advantage of electrostatic interactions between SP1 protein nanorings and CdTe QDs to make sandwich nanowires, bundles and irregular networks. Detailed characterization by AFM, TEM and DLS indicated that the size of QDs, as well as the structural topology of the SP1 nanoring, played critical roles in the formation of the

superstructures. Moreover, combining different sizes of QDs in the protein nanowires enabled efficient FRET, suggesting that the ordered QD arrays could be promising scaffolds for designing artificial light-harvesting systems. The design strategy was further applied to functionalize covalently linked 2D SP1 nanosheets with QDs.²³⁰ The nanosheets were decorated with CdTe QDs that bound electrostatically to the negatively charged surface of SP1-based assemblies. The ordered arrangements of QDs of different sizes on protein nanosheets, which served as donor and acceptor chromophores, yielded a pronounced FRET phenomenon. The light-harvesting behavior of these protein-QD nanosheets emulates the thylakoid membranes of natural chloroplasts.

Highly rigid GroEL nanotubes (Sections 3.4.3 and 4.2.1) have also served as scaffolds for nanoparticle organization.²⁹² Prior to metal-mediated assembly, MC-modified GroEL barrels were loaded with superparamagnetic iron oxide nanoparticles (SNPs) bearing a dopamine sulfonate zwitterionic ligand coating and hydrophobic, catechol-modified fluorescent dyes. The hydrophobic modification of the surface was required for the association of SNPs with the barrel cavity, which naturally binds to denatured proteins. Under a magnetic field, SNP-loaded nanotubes assembled into thick bundles, while removing the field returned them to the dispersed nanotube state. This process could be repeated over multiple cycles without causing denaturation of the nanotubes. Importantly, SNPs that were not housed in a protein nanotube sheath remained in an aggregated state when the magnetic field was removed. This work marked the first experimental observation of lateral aggregation of 1D SNP arrays, which had been predicted by theory. The rigid GroEL nanotubes proved to be an ideal SNP array scaffold for experimental observation of this phenomenon.

Superlattice wires represent another class of 1D protein assemblies. Such wires have been prepared through multivalent electrostatic association between anionic TMV rods and functional, cationic, glue-like components. The Kostianen Group first reported superlattice wire formation in a system composed of TMV and spherical AuNPs.²⁹⁶ The ordered association of the components was carried out by progressively lowering the ionic strength of the solution using sequential dialysis steps and was reversible with increased ionic strength. The superlattices assembled following a cooperative self-assembly pathway that proceeded in a zipper-like fashion, where nanoparticles crosslinked TMV rods into bundles. In cross-sections, the wires showed square lattice (*p*4m) packing of the TMV rods, rather than the hexagonal geometry commonly observed for the packing of rod viruses (**Figure 82b**). This observation was explained by the relatively large size of the AuNPs (*d* ~ 12 nm) in this system, which could bridge four TMV rods to form the close-packed square lattice. In contrast, smaller nanoparticles bound at most to three rods due to steric constraints, which induced hexagonal packing. The consistent interparticle distance within the lattice was controlled by competing repulsive (between AuNPs) and attractive (between AuNPs and virus rods) interactions. The right-handed helicity of the virus rods was imparted on the helical twist of the superstructure, leading to structure-dependent chiral plasmonic optical properties within the material. Proof-of-concept experiments were carried out to demonstrate the potential application of the superlattice wires as plasmonic polarizers by decorating the structures with cationic magnetic iron oxide nanoparticles.

In a related study, the same group generated photoactive TMV bundles with hexagonal packing using a peripherally crowded ZnPc as a cationic glue.⁶⁰⁵ Within the scaffolds, the ZnPc

acted as a photosensitizer to generate reactive oxygen species (ROS) under illumination with visible light. The immobilization of ZnPc within the fibers thus gave rise to a heterogeneous catalyst that could be easily purified by physical methods and had a high solvent-accessible area owing to the high aspect ratio of the structures. The bundles could be immobilized and irradiated over multiple cycles within a microfluidic device, showing consistent production of ROS. This self-assembled material was proposed as a promising, biodegradable catalyst system for applications in green organic synthesis and wastewater treatment.

Most of the examples described in this section thus far have highlighted the template-based design strategy to prepare nanoparticle assemblies. The Schiller Group developed a bottom-up approach, termed protein adaptor-based nano-object assembly (PABNOA), to prepare template-free assemblies of different plasmonically-active AuNP nanoarchitectures.⁴⁴⁸ Specifically, they designed protein adaptors with genetically encoded interaction sites to guide the assembly of AuNPs. The interactions between geometrically defined protein adaptors (modified Hcp1 toroidal protein building block) and AuNPs induced the self-assembly of different hybrid architectures including 1D chains, networks and stars. Interestingly, the interparticle distance between AuNPs could be controlled by different assembly conditions. The functionality of the different NP architectures could be extended by co-transcriptional encoding of unnatural amino acids as additional site-specific modification sites on the protein adaptors for covalent attachment of dye molecules.

The Trent Group introduced the strategy of using crystalline 2D protein assemblies as templates to generate ordered nanoparticle arrays.⁶⁰⁶ They used Cys-modified heat shock protein 60 (HSP60), a chaperonin subunit that assembles into hollow octadecameric double-ring structures, with nine subunits per ring and 3- or 9-nm apical pores depending on the presence of an apical loop in the monomer. The Cys residue was located in a solvent-exposed position on the apical side of the monomer. The engineered HSP60 monomers were crystallized into disk-shaped, hexagonally packed 2D templates with periodically arranged thiol groups serving as binding sites to organize AuNPs or CdSe-ZnS QDs into arrays in a size-selective manner. The size selectivity was attributed to the accessibility and positioning of Cys residues, along with the pore size of the templates. The same group further explored the templating and patterning of nanoparticle arrays using chaperonin assembled by heat shock protein TF55 β .⁶⁰⁷ The designed variant formed cage structures with a 20-nm diameter and a core containing 180 imidazole groups from the 18 N-terminal His₁₀ peptides. The chaperonin cages adopted hexagonal 2D packing and the His-rich cores within the lattices served as templates for the synthesis of bimetallic nanoparticle arrays (Ni-Pd or Co-Pd). The average size of the NPs corresponded to the interior diameter of the chaperonin cage and their patterning in arrays reflected the structure of the underlying 2D protein lattice.

Protein assemblies can also be used as chemically active scaffolds to induce nanoparticle assembly. Brodin et al. demonstrated that 2D Zn-RIDC3 arrays could function as redox-active scaffolds for templated growth of Pt⁰ nanocrystals.²⁰⁴ Specifically, the Zn-porphyrin (ZnP) cofactor in each Zn-RIDC3 building block is redox-inactive in its ground state but becomes a strong reductant/oxidant upon irradiation with visible light. Upon incubation with Pt²⁺ and light excitation, Zn-RIDC3 arrays displayed uniform coverage with Pt NPs with a narrower size distribution than a) Pt NPs produced in solution in the absence of the protein arrays or b) Pt NPs

on protein arrays formed in the absence of light, as characterized by SEM and ns-TEM images. This study illustrated that the supramolecular protein assemblies not only played an important role as structural templates but could also actively control the growth of redox-active NPs due to the specific functional activity of the component protein build blocks.

Thomas et al. used a 2D protein array as a versatile platform for the assembly of multicomponent nanostructures by functionalizing the building blocks with specific tags to recruit various components.⁴⁸⁷ Starting with a computationally redesigned protein called TTM that forms hexagonal 2D arrays upon Ca^{2+} triggered assembly, the researchers further modified the protein building block with different functional tags (e.g., hexa-His tag, gold-binding peptide, biotinylation tag). The resulting 2D arrays were used as 2D scaffolds for the fabrication of hybrid materials by binding AuNPs and proteins to the functionalized array surface. These proof-of-concept studies demonstrated that protein-based 2D arrays hold great potential for constructing versatile hybrid materials with promising applications in sensing, medicine, and energy harvesting.

Zhang et al. modified TMVCP disks to direct self-assembly of AuNPs and QDs into ordered nanostructures.⁴⁷⁹ TMVCP was modified with two distinct mutations: a Cys residue in the center of the TMVCP disk and a His residue on the outer surface of the disk. The combination of these mutations enabled both self-assembly of 2D TMVCP monolayer sheets via Cu^{2+} -His interactions and formation of highly ordered 2D AuNP lattices in three different geometries based on three different binding modes between AuNPs and TMVCP disks (**Figure 82c**). Similarly, assemblies incorporating two types of QDs arranged in honeycomb and hexagonal lattices were obtained by using the 2D protein nanosheets as templates. Moreover, because the TMVCP nanosheets contained two different functional groups, they could direct the co-assembly of AuNPs and QDs simultaneously, thus giving rise to binary functional NP lattices. Recently, Du et al. have also reported ordered binding of AuNPs to *p42₁₂C⁹⁸RhuA* lattices modified with an additional Cys residue.⁶⁰⁸

4.4 Mechanical properties of designed protein assemblies

From the mechanosensitive Piezo proteins⁶⁰⁹ and cytoskeletal fibers at the nm-to- μm scale,¹¹⁴ to extracellular protein composites (e.g., silk, skin, hair) with macroscopic dimensions,⁶¹⁰⁻⁶¹² protein-based materials possess outstanding mechanical properties (e.g., toughness, strength, elasticity, self-healing, etc.) that are central to their biological functions.^{114,613,614} These mechanical properties derive chiefly from how structural components within these materials are interconnected, hierarchically organized, and interfaced with other materials.^{114,613,614} Therefore, there is much interest in understanding and manipulating how intermolecular interactions are translated into mechanical properties upon assembly and organization at extended length scales. Much of the work in protein-materials engineering has centered on structural proteins including collagens, elastins, resilins and silks, which combine many attractive features such as biocompatibility, biodegradability and processability.^{615,616} These materials are characterized by peptide-repeat sequences that give rise to interconnected secondary structure motifs with local order, often dispersed within a non-ordered matrix.

In contrast, the protein assemblies discussed in this review are composed of building blocks with well-defined tertiary structures and have been intentionally designed to possess

structural order at all length scales. While these assemblies often result in crystalline materials, the highly tunable interprotein interactions that dictate assembly formation can manifest a wide range of mechanical properties. For example, artificial protein cages held by an extensive network of hydrophobic interfaces possess Young's moduli in the range of 0.1 – 1 GPa, typical of virus capsids with high mechanical strengths. In contrast, metal-directed, 1D protein nanotubes display moduli that are orders of magnitude smaller (0.1 – 30 MPa),⁴¹⁹ more similar to flexible natural materials such as fibrin and elastin. Notably, as alluded to in Section 4.2.2, there are several emerging examples of reconfigurable protein assemblies/arrays that display outstanding mechanical properties owing to their unique compositions and connectivities.

2D ^{C98}RhuA lattices were previously introduced as crystalline arrays with coherent in-plane dynamics, enabled by the short yet flexible disulfide linkages between the C₄ symmetric protein units. Due to the rotary motion of the proteins with respect to their neighbors, coupled with the overall P42₁2 plane group symmetry, ^{C98}RhuA lattices expand in the x dimension to the same extent as in the y dimension and *vice versa*. Thus, ^{C98}RhuA lattices are auxetic and possess the thermodynamically smallest possible Poisson's ratio (ν) for an isotropic crystalline material of $\nu = -1$, the first such material designed and constructed at the molecular scale. This behavior stands in contrast to common materials, which expand longitudinally when compressed transversely, corresponding to positive ν values.⁶¹⁷ Due to their unusual mechanical behavior, auxetic materials have been proposed for use in smart textiles, actuated filtration, sensing, and biomedical devices, as well as piezoelectric materials.^{618,619}

Indeed, Zhang et al. recently demonstrated a potential route to generate a piezoelectric system through surface-templated ^{C98}RhuA self-assembly.²²⁷ A structural analysis of the ^{C98}RhuA building blocks revealed a very large dipole moment (1200 D) aligned with the protein's principal C₄ symmetry axis (Section 3.5.6; **Figure 62**). As the ^{C98}RhuA lattices assembled in solution have p42₁2 symmetry, which is associated with an alternating up-down arrangement of the monomers, their overall dipole moment is zero. In contrast, lattices assembled at the mica interface are p4 symmetric, resulting in co-aligned ^{C98}RhuA dipole moments within the lattice. Formally, lattices of coaligned dipoles are "electrets", or polarized dielectric materials with a permanent dipole. Importantly, because ^{C98}RhuA lattices isotropically expand and contract in plane, they are calculated to carry a lattice conformation-dependent polarization density (0.008–0.016 C·m⁻²) and predicted to be piezoelectric.

The combination of crystalline protein lattices with flexible linkages can also be used to construct 3D protein materials with bulk-scale mechanical properties. To obtain thermoresponsive 3D lattices of *P. furiosus* ferritin ($\rho_l \sim 4.5\text{--}5.5$), Välimäki et al. synthesized linear-branched diblock copolymers consisting of a cationic multivalent dendron with a linear thermoresponsive polymer (pDEGMA) tail.⁶²⁰ Electrostatic interactions between the negatively charged ferritin and the dendron resulted in the formation of an *fcc* lattice ($a = 18.55$ nm). Notably, heating the lattice to 50 °C, above the cloud point temperature of pDEGMA ($T_{cp} \sim 31$ °C), induced a ~2% contraction to $a = 18.18$ nm owing to the compaction of the pDEGMA chains, thus yielding a thermoresponsive material displaying negative thermal expansion.

In a similar vein, Bailey et al. exploited the synthetic modularity of protein-metal-organic frameworks (protein-MOFs) to synthesize six different ferritin-MOF lattices from a combination

of two metal ions and three ditopic linkers.⁵¹³ The thermostabilities of all six ferritin-MOFs were tunable with the addition of molecular crowding agents. Furthermore, one of the *bcc* frameworks (fdh-Ni-ferritin) underwent a reversible and isotropic first-order phase transition near-room temperature, with the unit cell dimensions decreasing by 1.3% at 33-34 °C within a transition window of ≤ 1 °C, corresponding to a volumetric change of 4%. The lattice compaction was fully reversed at 25 °C, yielding a hysteresis window of nearly 10 °C (**Figure 83a**), which is a property that may be exploited in sensing and memory devices. Differential scanning calorimetry (DSC) and XRD measurements showed that these transitions were first-order and could be attributed to a diffusionless rearrangement of the organic linker connecting the metal nodes on the ferritin building blocks. Such extremely sharp phase transitions are rare in biological systems and typically associated with order-to-disorder transitions.⁶²¹⁻⁶²³ In solid-state materials, these phase transitions typically involve electronic/spin-state changes.^{624,625} The thermomechanical behavior of the fdh-Ni-ferritin framework demonstrates the benefits of modular protein design strategies in discovering novel material properties that cannot be predicted from first principles.⁶²¹⁻⁶²³

Finally, it is important to note that most biological materials such as skin, muscle or bone are not phase-pure or uniform. They are composites of different types of structural components hierarchically organized at different length scales, which enables the combination of properties such as strength, toughness, flexibility, damage tolerance, and self-healing.^{626,627} In contrast, the highly crystalline, compositionally uniform protein materials described above are highly brittle, have limited flexibility that depends on the bonding interactions which maintain the protein lattice, and are incapable of self-healing.

To address these limitations, Zhang et al. sought to create a composite material where 3D protein crystals are fully integrated with synthetic hydrogel polymers.²⁹⁵ The researchers took advantage of the mesoporosity of Ca^{2+} -directed, HuHF *fcc* crystals, to thoroughly infuse the lattice channels with acrylate/acrylamide monomers and form a polymer matrix that non-covalently bonded to the underlying protein lattice. The resulting polymer-integrated crystals (PIX) had a ~ 1 GPa modulus, typical of protein and small-molecule crystals. Despite this high stiffness, the ferritin-PIX displayed notable mechanical properties stemming from their flexibility: 1) They isotropically expanded to nearly 600% of their original volume while retaining crystalline periodicity and faceted polyhedral morphologies. 2) After substantial expansion (separation of the ferritin molecules in the lattice by more than 50 Å), the ferritin-PIX contracted back to their original state when solution ionic strength was increased and fully regained atomic-level periodicity (**Figure 83b**). Polymer integration and the expansion/contraction process were frequently observed to improve XRD quality, yielding a very high resolution (~ 1.1 Å) crystal structure of ferritin. 3) Due to the dynamic bonding interactions between the hydrogel network and ferritin molecules, the ferritin-PIX displayed efficient self-healing behavior (**Figure 83b**).

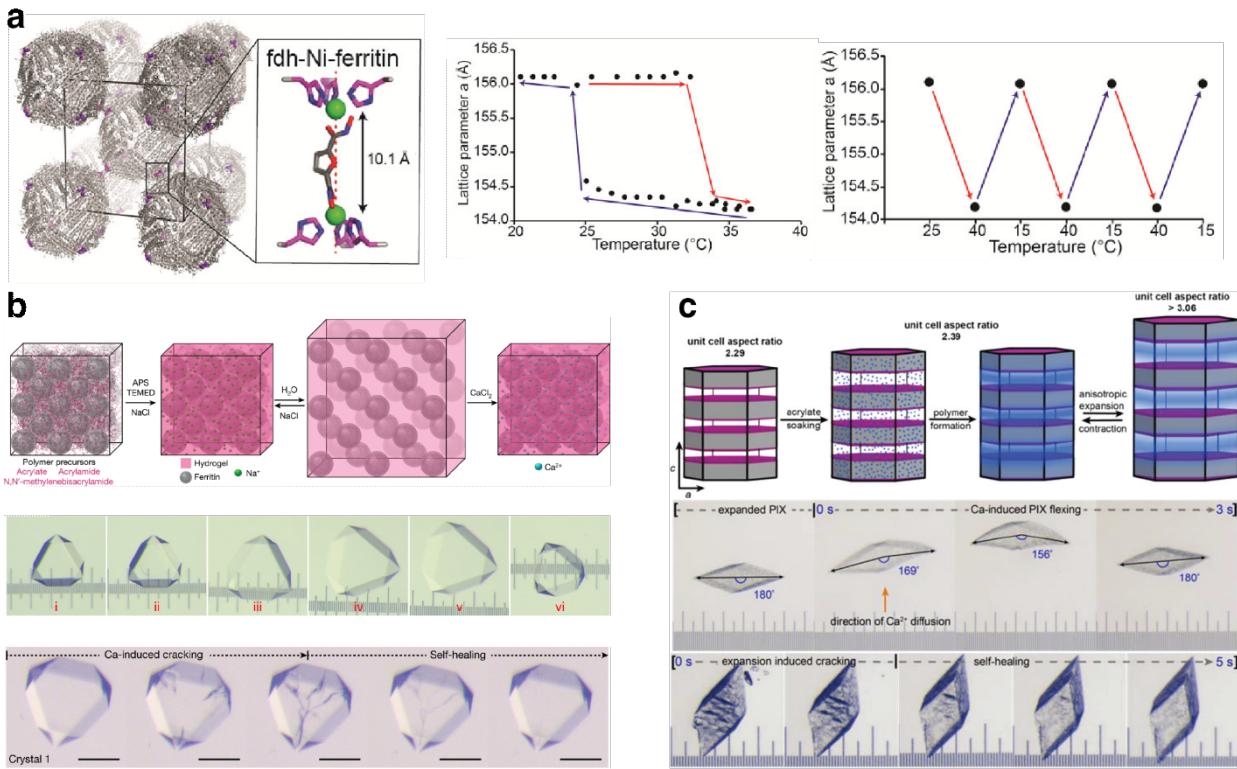


Figure 83. Mechanical properties of protein-MOFs and ferritin-PIX. **a)** fdh-Ni-ferritin lattice unit cell with a close-up view of the interfacial connectivity (left). Thermal hysteresis loop (middle) and reversible cycling (right) of the fdh-Ni-ferritin lattice expansion/contraction. **b)** Schematic representation (top) and light micrographs (middle) showing the formation, expansion and contraction of ferritin-PIX. Light micrographs of ferritin-PIX (bottom) showing the self-healing behavior of cracks that appear during Ca-induced contraction. **c)** Schematic representation of the reversible anisotropic expansion/contraction of rhombohedral *raft*-ferritin-PIX (top). Light micrographs of the rhombohedral *raft*-ferritin-PIX crystal showing cation-induced bending motion (middle) and self-healing behavior (bottom). **(a)** Adapted with permission from Ref. ⁵¹³. Copyright 2020 ACS. **(b)** Adapted with permission from Ref. ²⁹⁵. Copyright 2018 NPG. **(c)** Adapted with permission from Ref. ⁵⁷³. Copyright 2021 ACS.

Due to the cubic symmetry of the Ca^{2+} -mediated HuHF crystals and the isotropic expansion/contraction of the polymer network within the protein lattice, the structural dynamics of ferritin-PIX were also isotropic. In a subsequent study, Han et al. sought to achieve directional/anisotropic dynamic behavior by controlling the spatial distribution of hydrogel networks within ferritin-PIX.⁵⁷³ They employed a ferritin variant that self-assembled upon Ca^{2+} coordination into rhombohedral lattices composed of stacked hexagonal protein layers, which enabled an anisotropic patterning of the hydrogel matrix that formed *in crystallo*. The resulting ferritin-PIX indeed displayed directional expansion/contraction in response to changes in solution ionic strength, whereby the aspect ratio of the crystals increased/decreased by more than 30% (**Figure 83c**). These PIX also displayed rapid bending motions in response to directional influx of cation gradients (bending rates of $>10^\circ$ per second) and were capable of rapid self-healing like the isotropic ferritin-PIX. These studies collectively indicate that it is possible to attain some of the salient mechanical properties of sophisticated biological devices like skeletal muscles

through the proper design of self-assembled protein architectures and their simple integration with synthetic materials.

4.5 Biochemical functions of designed protein assemblies

Alongside the construction of biomaterials with emergent physical and mechanical properties, a major goal in designing protein assemblies is to engineer new biochemical functions that can ultimately be integrated into living systems. As we discuss in this section, designed protein assemblies not only enable the generation of new-to-nature biochemical activities that would be difficult to achieve with small (non-assembled) proteins, but also provide a unique platform for understanding and manipulating protein structure-function relationships without evolutionary constraints. The studies described below also highlight how rapidly the field of protein assembly design has transitioned from structure-building to function-building, thereby producing systems that demonstrate *in vivo* activities and successful interfacing with cellular systems.

4.5.1 Binding and recognition

Perhaps the most mechanistically straightforward biochemical function for a protein or protein assembly is the recognition and binding to targets for downstream effects. As summarized in Section 4.3.3, an important advantage of designed protein assemblies in this regard is the ability to display biological recognition elements with potentially controllable valency and periodicity. Particularly exciting is the design of protein arrays for scaffolding antibodies that may mimic and even go beyond the valences of IgG (bivalent), IgA (tetravalent), and IgM (decavalent) in the context of stable and readily modifiable platforms. In an early study, Wagner and colleagues reported a series of ring-shaped assemblies of genetically fused dihydrofolate reductase dimers (DHFR₂), which were self-assembled via chemical dimerizers (bis-MTX) and varied in size and composition from two to eight monomers.⁶²⁸ Based on these nanoring architectures, Li et al. constructed self-assembled antibody nanorings (CSANs) from fusions of DHFR₂ with single-chain anti-CD3 antibodies, affording valences of eight to ten (**Figure 84a**).⁶²⁹ Through flow cytometry measurements, it was determined that the octavalent anti-CD3 CSANs in particular demonstrated improved affinity to CD3 expressing cells compared to the parental monoclonal antibody (mAb) UCHT-1. Moreover, the researchers found that anti-CD3 CSANs and the parental mAb UCHT-1 demonstrated a similar cellular internalization mechanism. In contrast to mAb, no significant T-cell proliferation was observed after the treatment of peripheral blood mononuclear cells (PBMCs) with the bivalent and octavalent anti-CD3 CSANs. mAb and anti-CD3 CSANs also exhibited distinct effects on T-cell receptor internalization and IL-2 receptor expression.

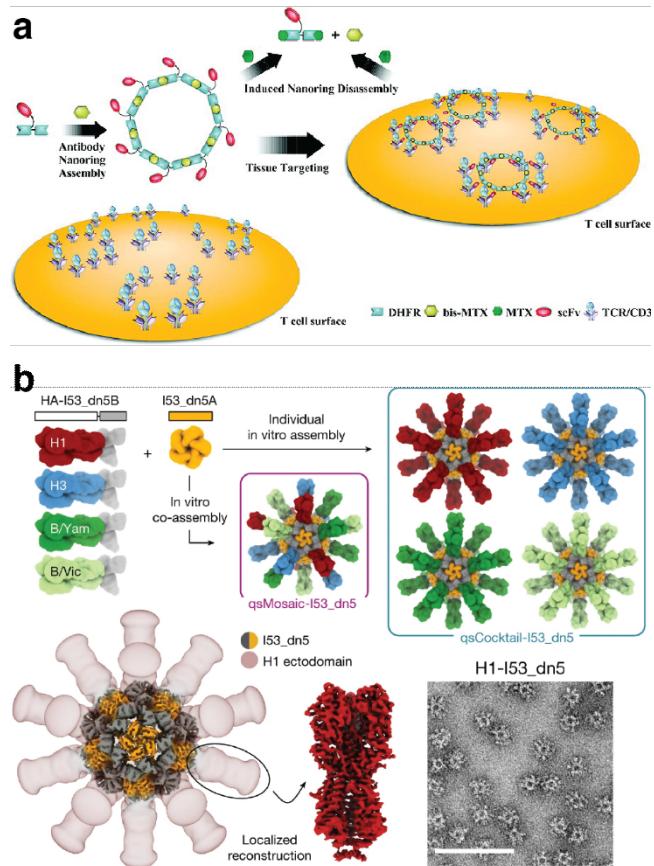


Figure 84. Design of protein assemblies with specific target recognition and binding properties. **a)** Scheme of the assembly and disassembly of octavalent anti-CD3 scFv antibody targeting T cell receptors. **b)** Design and characterization of HA nanoparticle immunogens (qsMosaic-I53_dn5 and qsCocktail-I53_dn5). **(a)** Adapted with permission from Ref. ⁶²⁹. Copyright 2010 ACS. **(b)** Adapted with permission from Ref. ⁶³⁰. Copyright 2021 NPG.

To further evaluate the efficacy of CSANs as prosthetic antigen receptors (PARs) *in vivo*, the same group developed bispecific PARs that selectively target the human CD3 receptor and human epithelial cell adhesion molecule (EpCAM), which is overexpressed on multiple carcinomas and cancer stem cells.⁶³¹ The designed bispecific CSANs stably bound to T cell surfaces for > 4 days *in vitro*, while being easily disassembled on the cell membrane by treatment with the nontoxic FDA-approved drug, trimethoprim. Furthermore, the study also demonstrated that CSANs could nongenetically generate reversibly modified T cells that were capable of eradicating target solid tumors.

Based on the design principles developed to build symmetric two-component nanocages,^{186,389,391} Divine et al. recently described a general approach for building precisely oriented antibody assemblies without the need for covalent modification.³⁹⁴ As described in Section 3.3.2 (Figure 33), such “antibody nanocages” (AbCs) were produced by leveraging the inherent two-fold symmetry of IgG antibodies, which drove nanocage self-assembly in combination with properly designed Fc-binding homo-oligomers. The authors prepared a series of AbCs with dihedral, tetrahedral, octahedral, and icosahedral architectures bearing 2, 6, 12,

and 30 antibodies per nanocage, respectively. Compared to free antibodies or Fc-fusions, the binding of designed antibody nanocages to cell-surface receptors showed enhanced signaling in DR5-mediated apoptosis, Tie2-mediated angiogenesis, CD40 activation, and T cell proliferation. Moreover, antibody nanocages composed of α -SARS-CoV-2 monoclonal antibodies and FcACE2 fusion proteins also increased SARS-CoV-2 pseudovirus neutralization.

In a slightly different structural context, Boyoglu-Barnum et al. transformed previously reported, artificial two-component icosahedral protein cages into nanoparticle immunogens that induce potently neutralizing and broadly protective antibody responses against influenza viruses.⁶³⁰ The researchers genetically fused hemagglutinin (HA) ectodomains from the four strains in licensed 2017-2018 seasonal influenza vaccines to the N terminus of a trimer (I53_dn5B). The resulting trimeric construct (HA-I53_dn5B) was mixed with I53_dn5A pentamer to generate a mosaic nanoparticle immunogen that co-displayed the four HAs (qsMosaic-I53_dn5). Similarly, nanoparticle immunogens with individual HAs were purified and mixed in equimolar amounts to prepare a “cocktail” immunogen that contained four individual HA-displaying nanoparticles (qsCocktail-153_dn5) (**Figure 84b**). In several animal models, both qsMosaic-I53_dn5 and qsCocktail-153_dn5 elicited antibody responses against vaccine-matched strains that were equivalent to or better than commercial quadrivalent influenza vaccines. Simultaneously, these immunogens induced broadly protective antibody responses to heterologous viruses by targeting the subdominant yet conserved HA stem. Following a similar design approach, protein nanoparticles displaying virus antigens such as glycoprotein^{632,633}, HIV envelope trimers⁶³⁴ and SARS-CoV-2 spike receptor-binding domains⁶³⁵ have also shown promise in inducing potent immune responses as potential vaccine development candidates.

In addition to the multivalent display of proteins that bind to biological targets, protein assemblies can also be designed to directly bind small molecules. Park et al. redesigned a C_3 -symmetric homotrimer, which they had previously designed with HBNet, to bind amantadine, an FDA-approved drug that is also C_3 -symmetric.⁶³⁶ The resulting protein bound amantadine with an affinity of 24 μ M and a crystal structure revealed decent similarity with the design model (RMSD = 0.63 \AA), though there were differences in the hydrogen bonding interactions between the protein and the ligand. Rittle et al. reported that a penta-His Fe(II) binding site designed through their MASCoT approach was capable of binding nitric oxide (NO).³⁴⁷ Addition of the NO donor diethylammonium NONOate to the iron-loaded protein in anaerobic conditions resulted in spectroscopic features consistent with an $\{\text{FeNO}\}^7$ species. A crystal structure of the complex featured electron density above the iron center that was modeled as a bound NO ligand.

4.5.2 Membrane-related functions

Aside from displaying antibodies for protein binding, protein assemblies have also been designed for interactions with cellular membranes and membrane components. Sundquist, King and colleagues reported artificial protein assemblies (enveloped protein nanocages or EPNs) that directed their own release from human cells and could deliver cargo to other cells.⁶³⁷ For this purpose, the building blocks of a previously designed icosahedral cage (I3-01, Section 3.3.2, **Figure 32**) were genetically fused to a signal peptide sequence for *N*-myristoylation in its N-terminus to enable membrane binding. In addition, a sequence (p6Gag from HIV-1) for the recruitment of endosomal sorting complexes required for transport (ESCRT) machinery was also

attached to its C-terminus to generate the construct EPN-1 (**Figure 85a**). When expressed in human embryonic kidney 293T cells, a small but significant amount (13.3%) of EPN-1 was found to be released into the culture medium in the form of multiple, 25-nm cages contained inside 107 ± 44 nm membrane envelopes (EPNs) (**Figure 85a**). Mutational analyses indicated that all three protein components were necessary for EPN release from the kidney cells. Vesicular stomatitis viral glycoprotein (VSV-G) could be recruited into EPNs, allowing their fusion with HeLa cells and delivery of cargo proteins into these cells. The researchers demonstrated the generality and modularity of EPN design with several more successful constructs that carried a variety of membrane-binding, self-assembly and ESCRT-recruiting elements, including designed, self-assembling protein cages with octahedral symmetry (O3-33).

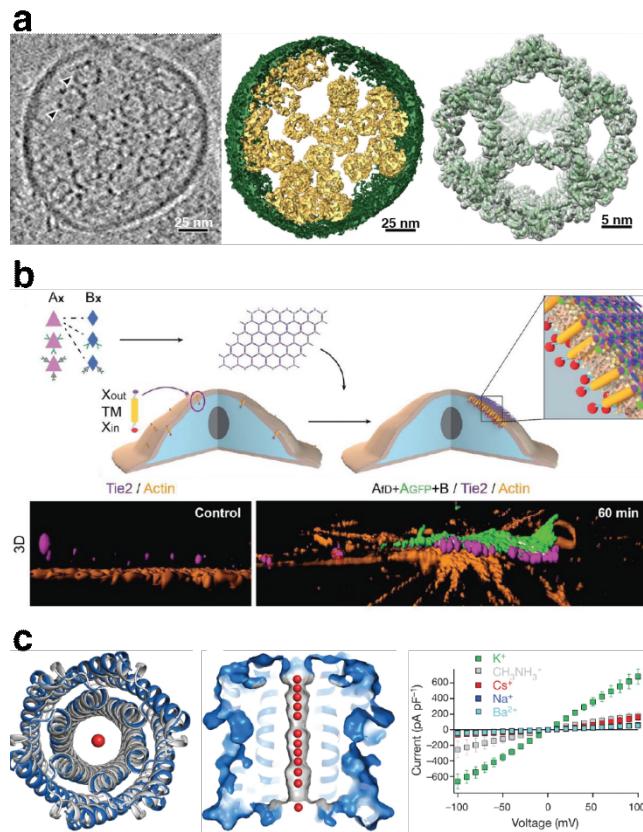


Figure 85. Protein assemblies designed for interactions with cellular membranes and membrane components. **a)** Central slice from a cryo-EM tomographic reconstruction of a released EPN (left), structural models of the 3D cryo-EM reconstruction from EPN-1 (middle) and I3-01 nanocage (right). **b)** 2D array functionalization by genetic or post-translational fusions (top) and 3D reconstruction of clustered TIE2 with or without the presence of 2D arrays (bottom). **c)** Structure of water-soluble hexameric WSHC6 determined by XRD and the ion conductivity of the 12-helix TMHC6 transmembrane channel (left) and ion conductance of TMHC6 with different cations (right). **(a)** Adapted with permission from Ref. ⁶³⁷. Copyright 2016 NPG. **(b)** Adapted with permission from Ref. ⁴⁸⁹. Copyright 2021 NPG. **(c)** Adapted with permission from Ref. ³⁵⁹. Copyright 2020 NPG.

More recently, Ben-Sasson et al. reported that computationally designed, binary 2D protein arrays could interact with cell membranes and control membrane receptor clustering.⁴⁸⁹

They took advantage of the two-component, hexagonal protein arrays composed of D_2 and D_3 symmetric building blocks, which were shown to self-assemble *in vitro* and *in vivo* with high fidelity (Section 3.5.5, **Figure 60**). An advantage of the two-component system is that the co-assembly process can be controlled by mixing. Due to their robustness, the 2D arrays could be genetically and post-translationally modified with other proteins and ligands for cell-surface receptors, such as GFP and TIE2, respectively. This allowed the researchers to demonstrate that these arrays can induce the clustering of membrane proteins and activate downstream signaling processes (**Figure 85b**). Specifically, super-resolution microscopy experiments revealed extensive remodeling of the actin cytoskeleton underneath the 2D-array-induced TIE2 clusters, whereas high-resolution AFM imaging confirmed that the *in vivo*-assembled arrays maintained the same hexagonal architecture observed under *in vitro* conditions. It was also observed that the extensive 2D arrays tailored with ligands for epidermal growth factor receptors (EGFR) suppressed endocytosis in a tunable fashion, whereas smaller particles such as designed protein cages were readily taken up by cells and degraded in lysosomes. The researchers proposed that the ability of designed 2D arrays to inhibit endocytosis without inducing signaling could help improve the efficacy of signaling-pathway antagonists by extending receptor engagement and immune evasion.

In another recent development, Xu et al. engineered oligomeric transmembrane pores that assemble in membranes *in vitro* and *in vivo*.³⁵⁹ Using a two-step strategy, the researchers first designed a water-soluble hexamer of coiled-coil motifs to generate a 12-helix barrel (WSHC6) whose crystal structure closely matched the computational design. The outward-facing residues of WSHC6 were then redesigned to promote membrane insertion, yielding the construct TMHC6 (**Figure 85c**). This construct was expressed as a hexamer in the membranes of *E. coli* cells and was highly thermostable. Whole-cell patch-clamp experiments with *Trichoplusia ni* insect cells expressing the TMHC6 pore indicated ion conductance, with higher selectivity observed for K^+ compared to other monovalent cations such as Na^+ . (**Figure 85c**) This selectivity was likely due to the size of the transmembrane pore, which was large enough to allow partially dehydrated K^+ ions, but not fully hydrated Na^+ ions, to pass through.

4.5.3 Catalysis

As illustrated by the aforementioned cage-like and 2D scaffolding systems, the enhancement of binding between protein assemblies and their targets can have many potential biological applications. However, binding and recognition alone cannot trigger more complex biochemical functions such as catalysis. Such functions require proteins to not only bind and organize substrate and solvent molecules, but also act upon them in a choreographed fashion to enable chemical transformations. In a way, the difference in complexity between “a binder” and “a catalyst” is similar to the difference between a stable protein structure sitting in a deep energy well and a dynamic one that can traverse a complex free-energy landscape with many minima (see Section 4.2). Given the inherent reactivities of metal ions, the majority of successes in enzyme design has involved the construction of protein structures with active sites containing metal ions or metallocofactors.⁶³⁸⁻⁶⁴⁵ Most of these engineered metalloenzymes are constructed by repurposing the active sites or cavities of pre-existing protein scaffolds in order to circumvent the need for designing a protein structure from scratch.^{638,640} An important advantage of designed

protein assemblies for enzymatic functions is that they provide extensive protein-protein interfaces where catalytic sites can be built without evolutionary constraints. This advantage has been successfully used in recent years.

The first example of a designed protein assembly with enzymatic activity was reported by Der et al.³⁴³ In their efforts to generate a stable C_2 symmetric dimer of a model helix-turn-helix domain (Rab4-binding domain of rabenosyn), the researchers incorporated two $\text{His}_4\text{-Zn}^{2+}$ coordination sites and computationally modeled the dimeric interfaces (Section 2.2.2, **Figure 12**, **Figure 86a**).³⁴² One of the resulting constructs, termed MID1, formed a stable dimer with high Zn^{2+} affinity at the interface ($K_d < 30 \text{ nM}$). However, the crystal structure of MID1 revealed that the two Zn centers were coordinated by a His_3 rather than the intended His_4 coordination motif. Noticing the similarity of the resulting $\text{His}_3\text{:Zn}^{2+}$ centers positioned in hydrophobic niches to the catalytic sites of natural Zn-dependent enzymes, the researchers examined the catalytic activity of the MID1-Zn²⁺ dimer. This construct was found to be quite active toward the hydrolysis of model substrates *p*-nitrophenyl acetate (pNPA) and *p*-nitrophenyl phosphate (pNPP), with rate accelerations of 7×10^5 and 1×10^4 and k_{cat}/K_m values of $630 \text{ M}^{-1} \text{ s}^{-1}$ and $14 \text{ M}^{-1} \text{ s}^{-1}$ for pNPA and pNPP, respectively.³⁴³ The finding that relatively high catalytic efficiencies could be obtained in a minimally designed interface suggested that metal-mediated protein interfaces could provide a powerful strategy for the rational engineering and evolution of new catalytic activities.

Following up on this concept, Song et al. developed an artificial metallo- β -lactamase that functioned *in vivo*.²⁰⁵ The researchers used the Zn-directed, D_2 -symmetric tetramer of the cyt *cb*₅₆₂ variant ^{C96}RIDC1 ($\text{Zn}_4\text{:}^{C96}\text{RIDC1}_4$) (Section 3.2.4, **Figure 25**) as a scaffold. While preserving the four, coordinatively saturated Zn sites to stabilize the tetrameric assembly, they designed four additional three-coordinate Zn²⁺ centers into one of the C_2 symmetric interfaces to generate a series of catalytically active protein complexes. One of the variants, $\text{Zn}_8\text{:}^{K104A}\text{AB3}_4$, displayed enzymatic activity not only for ester hydrolysis ($k_{\text{cat}} = 0.2 \text{ s}^{-1}$ and $k_{\text{cat}}/K_m = 120 \text{ M}^{-1} \text{ s}^{-1}$ for pNPA at pH 10), but also for the more challenging β -lactam hydrolysis reaction ($k_{\text{cat}}/K_m = 115 \text{ M}^{-1} \text{ min}^{-1}$ for ampicillin, a β -lactam antibiotic). Leveraging the previous observation that ^{C96}RIDC1 derivatives were capable of assembling into the designed tetramers by selectively binding Zn²⁺ in the periplasm of *E. coli* cells,^{210,367} the researchers then demonstrated that $\text{Zn}_8\text{:}^{K104A}\text{AB3}_4$ was also catalytically active in bacterial cells, allowing them to survive in the presence of moderate concentrations of ampicillin. This construct thus represented the first designed protein assembly to confer a functional benefit to a living organism.

The *in vivo* effect of $\text{Zn}_8\text{:}^{K104A}\text{AB3}_4$ on cell viability enabled the researchers to improve the artificial enzyme's β -lactamase activity by directed evolution. The optimized variant ^{K104A/E57G}AB3 had a three-fold higher catalytic efficiency (k_{cat}/K_m) $115 \text{ M}^{-1} \text{ min}^{-1}$, a catalytic proficiency ($k_{\text{cat}}/K_m/k_{\text{uncat}}$) of $2.3 \times 10^6 \text{ M}^{-1}$, and a four-fold enhanced selectivity over pNPA hydrolysis (**Figure 86a**). Unexpectedly, the crystal structure of $\text{Zn}_8\text{:}^{K104A/E57G}\text{AB3}_4$ revealed a substantially altered architecture compared to that of $\text{Zn}_8\text{:AB3}_4$. Notably, the E57G mutation, which emerged in the course of *in vivo* screening, was found to increase the mobility of an otherwise-ordered loop near the active site and enhance substrate binding through hydrophobic interactions, which is a key feature of natural β -lactamases.^{646,647} Follow-up studies using related tetrameric β -lactamases further established that the local structural flexibility/rigidity near the metal active sites can have

substantial effects on the evolvability of β -lactamase activities²⁰⁶ and that symmetry-related residues in protein-protein interfaces may have outsized effects on catalysis.⁶⁴⁶⁻⁶⁴⁸

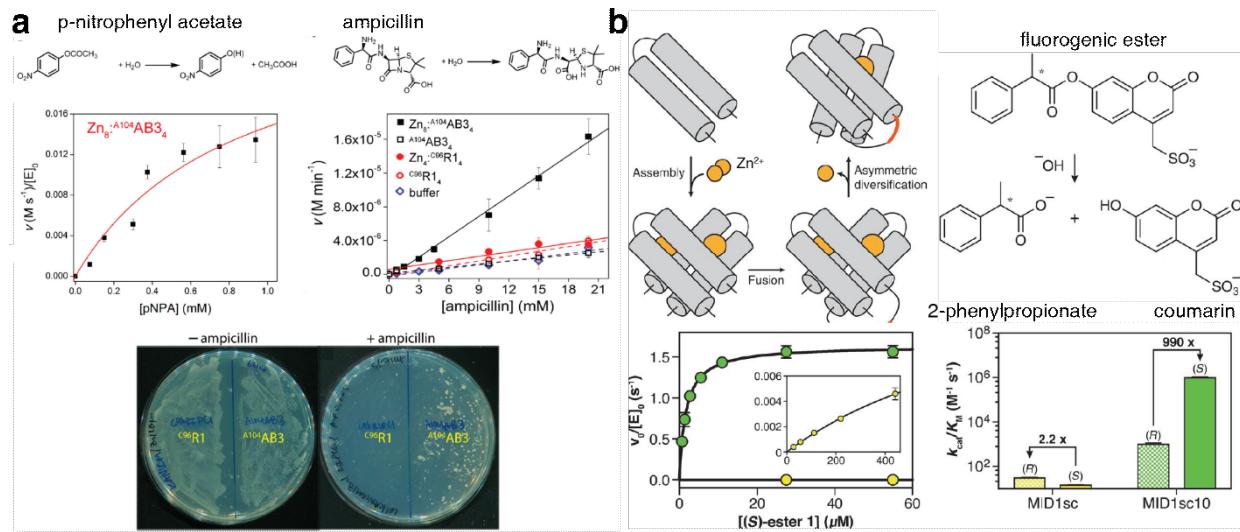


Figure 86. Designed protein assemblies as metalloenzymes. **a)** Studies of kinetic properties of cyt *cb562* variants for pNPA and ampicillin hydrolysis (top). Representative LB/agar plates in the absence and presence of 0.8 mg/L ampicillin streaked with cells expressing ^{c96}RIDC1 and ^{A104}AB3 (bottom). **b)** Emulating metalloenzyme biogenesis from Zn-mediated assembly and the hydrolysis of fluorogenic ester catalyzed by MID1sc (top). Michaelis-Menten plots and stereoselectivity of the hydrolysis reaction using MID1sc (yellow) and MID1sc10 (green). **(a)** Adapted with permission from Ref. ²⁰⁵. Copyright 2014 AAAS. **(b)** Adapted with permission from Ref. ²⁰⁷. Copyright 2018 AAAS.

Studer et al. recently adopted a similar strategy to convert the aforementioned MID1-Zn dimer into a highly efficient, single-chain esterase.²⁰⁷ Using the MID1-Zn dimer crystal structure as a guide, the MID1 monomers were appropriately linked through C-N terminus gene fusion. The resulting construct was subjected to an extensive series of redesign, rational and random mutagenesis and gene shuffling steps, which were coupled to an activity screen for the hydrolysis of a fluorogenic substrate. These directed-evolution steps led to a variant (MID1sc10) with a remarkably high catalytic performance ($k_{cat} = 1.6$ s⁻¹, $k_{cat}/K_m = 980,000$ M⁻¹ min⁻¹, $k_{cat}/K_m/k_{uncat} = 9.3 \times 10^{10}$ M⁻¹) and stereoselectivity (~1000-fold kinetic preference for S over R enantiomer), on par with highly evolved natural enzymes.⁶⁴⁹ (**Figure 86b**) In addition to hydrolysis reactions, Lewis acidic metal ions can catalyze a number of abiological reactions, including the Diels-Alder reaction.⁶⁵⁰ Basler et al. used the MID1-Zn dimer as a starting point to design and evolve an artificial Diels-Alderase.³⁴⁴ After multiple rounds of directed evolution, the researchers obtained an enzyme with >99% enantioselectivity and the highest reported catalytic proficiency of any Diels-Alderase, thereby demonstrating that interfacial metal active sites can be used to develop enzymes for reactions that are typically considered abiological.

Although the scope of designed metalloprotein assemblies with catalytic activities have thus far been limited to Zn-based hydrolases and Diels-Alderases, the examples summarized above demonstrate the immense potential to generate new and evolvable activities in nascent protein interfaces. Given the genetic economy with which metal ions can nucleate protein-protein

interactions and simultaneously provide inherent reactivities, these examples also implicate a potential role of metal-templated protein assembly in the evolution of modern metalloenzymes.^{368,369,651}

5. Future directions and challenges

Proteins are nature's premier building blocks for sophisticated molecular machines and materials, which has inspired scientists from diverse disciplines to try their own hand at designing protein assemblies with the structural and functional complexity of natural assemblies. In this review, we have summarized the remarkable progress that has been made in the rational design and construction of protein assemblies in the last two decades. To tackle the chemical and structural complexity of molecular interactions that guide protein self-assembly, scientists have devised many computational,³⁹² chemical⁹¹ and protein engineering⁸⁶ strategies to create an ever-expanding design toolkit. These advances have enabled the construction of complex protein architectures, ranging from small oligomers^{196,360} and cage-like assemblies^{127,195} to extended 1D structures^{202,224}, as well as 2D^{128,225} and 3D crystals.^{470,513} Many of these assemblies are not only stable, but also demonstrate dynamic, adaptive and stimuli-responsive behavior.

Importantly, the field of protein nanotechnology has made a swift transition from structure-building to property- and function-building. Protein materials with novel mechanical properties,⁶¹⁶ catalytically⁶⁵² and optically⁵⁹⁷ active systems, and assemblies that interface with cellular components^{489,637} and possess *bona fide* biological functions²⁰⁵ all reported within the last five years, aptly illustrate the field's transition toward function-building. The protein assembly design field has greatly benefited from contemporaneous technological advances in the characterization of static and dynamic structural protein assembly features,¹⁵⁷ the development of computational methodologies,¹⁵⁷ as well as new protein biochemistry and molecular biology tools.⁸⁶ Perhaps more importantly, these advances have been enabled by our growing understanding of molecular interactions and self-assembly, and the combination of knowledge from diverse disciplines such as protein design, supramolecular chemistry, inorganic chemistry, nanotechnology, polymer chemistry and physics.

What are some of the challenges and prospects of protein nanotechnology in the coming years? As usual, one needs to look no further than the structural and functional sophistication of natural protein assemblies or protein-based materials (e.g., multicomponent machines like photosystem II⁶⁵³ or dynamic materials like cytoskeletal filaments⁶⁵⁴) for inspiration. Despite our expanding design toolkit for controlling protein self-assembly (Section 2), essentially all the protein structures described in this review are symmetrical or periodic and consist of one, and at most two, protein components (Section 3). Often, these artificial assemblies are built for stability rather than dynamic or responsive behavior. Thus, an important technical challenge will be to improve the accuracy and efficiency of computational protein design methods, while also incorporating responsive elements/fluid bonds (e.g., metal coordination,²⁰⁰ electrostatic,²⁷² polar⁴⁸⁹ and water-mediated interactions¹⁹³) to enable the self-assembly of dynamic, multicomponent architectures via the simultaneous formation of multiple protein-protein interfaces. The presence of responsive elements will be instrumental in asserting control over self-assembly kinetics and for constructing reconfigurable and far-from-equilibrium systems.

Along these lines, it will be important to look beyond simple self-assembly. The formation of most natural protein-based machines and materials does not happen spontaneously from a one-pot mixture of building blocks. Instead, their assembly is often directed by templates (e.g., membranes,^{489,655} chaperones,⁶⁵⁶ polynucleotide sequences⁶⁵⁷) and takes place in a stepwise, spatially and temporally controlled fashion. Thus, it will be necessary to design/tailor protein building blocks for association not only with other proteins but also with external templates, and once again, to incorporate responsive/externally tunable interactions to access spatiotemporal control over protein assembly. Likewise, the construction of protein-based materials at the micrometer- or longer length scales will require us to think beyond the design of short-range, non-covalent interactions, which is the traditional realm of protein design. Considering other design elements, such as protein shape, charge, charge distribution, hydrophilic/hydrophobic patterning, etc. will enable the exploitation of other forces (electrostatic/dipole, magnetic, depletion, phase separation, etc.) and external fields to build hierarchical protein architectures at extended length scales.⁹¹

While the structure of a protein assembly is the obvious immediate target for design, it is hardly sufficient for obtaining a desired property or function. Indeed, the greater challenge in the design of functional protein assemblies does not stem from designing a particular structure with high accuracy, but rather from the fact that we possess an insufficient understanding of structure-property or structure-function relationships in proteins (except perhaps for simple functions like binding and recognition). For example, if we did not know the function of hemoglobin, we might still deduce from its globin fold and heme coordination environment that it is likely involved in small-molecule binding and not electron transfer. Yet, we would not be able to predict that it functions as a reversible O₂ transporter that displays positive cooperativity and allosteric modulation by CO₂ and BPG.⁶⁵⁸ Similarly, it would not be possible to describe the mechanical properties and nucleotide-dependent assembly/disassembly behavior of microtubules simply from 3D structures of α - and β -tubulin and the intermolecular interfaces between them.^{659,660}

Given this current knowledge gap between structure and property/function, we believe that it is impractical to solely rely on the conventional, deterministic protein design approach of “one design \rightarrow one structure \rightarrow one function”, which is challenging in its own right and inherently limited to one or few functional outcomes in the best-case scenario. More favorable, in our opinion, would be an alternative design pathway of “one design \rightarrow multiple structural outcomes \rightarrow multiple functional outcomes”. In this pathway, condition-dependent protein-protein interactions would enable one or few building blocks to yield a multitude of protein assemblies (for example, via switchable interactions), which in turn would provide a higher likelihood of obtaining the desired functions as well as the discovery of serendipitous ones. The key to the execution of this strategy would be modular protein assembly strategies (e.g., reconfigurable systems described in Section 4) coupled with efficient functional selection or screening methods, similar to those successfully used in the directed evolution of artificial enzymes and other functional proteins.

In closing, the structural and functional sophistication of natural protein assemblies and materials – which have evolved through functional selection over millions of years – may appear far out of reach for rational design. Yet, rational design is unbounded by the synthetic restrictions of living systems and has access to an unlimited source of building blocks, chemical tools and

assembly strategies. Indeed, as we have attempted to describe in this review, the field of protein nanotechnology has, in very short order, produced artificial protein assemblies and materials whose structural and functional attributes also lie beyond those of natural systems. Given the large momentum of this field and its expanding interdisciplinary scope, it would not be a big stretch to predict that there will soon be artificial protein assemblies used in real-life applications (e.g., as pharmaceutical agents,⁶⁶¹ catalysts in industrial processes,^{662,663} or device components⁶⁶⁴) and incorporated into the life cycles of organisms for synthetic biological purposes.

6. List of abbreviations

AFM	Atomic force microscopy
Ala	Alanine
Arg	Arginine
Asp	Aspartic acid
AUC	Analytical ultracentrifugation
AuNP	Gold nanoparticle
Bipy	Bipyridine
BPG	2,3-Bisphosphoglycerate
CCMV	Cowpea chlorotic mottle virus
CD	Circular dichroism
COF	Covalent-organic framework
Cys	Cysteine
Cyt	Cytochrome
DLS	Dynamic light scattering
dsDNA	Double-stranded DNA
cryo-EM	Cryogenic electron microscopy
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
Glu	Glutamic acid
GST	Glutathione-S-transferase
HA	Hydroxamic acid
Hcp1	Hemolysin-coregulated protein 1
His	Histidine
HRP	Horseradish peroxidase
HuHF	Human heavy chain ferritin
Lys	Lysine

MDPSA	Metal-directed protein self-assembly
MeTIR	Metal-templated interface redesign
MOF	Metal-organic framework
ns-TEM	Negative stain transmission electron microscopy
NHS	N-hydroxysuccinimide
NMR	Nuclear magnetic resonance
PAMAM	Polyamidoamine
Pc	Phthalocyanine
PDB	Protein Data Bank
pDEGMA	Poly(di(ethylene glycol) methyl ether methacrylate)
PEG	Polyethylene glycol
pl	Isoelectric point
pNPA	<i>p</i> -nitrophenyl acetate
pNPP	<i>p</i> -nitrophenyl phosphate
QD	Quantum dot
RhB	Rhodamine B
RhuA	L-rhamnulose-1-phosphate aldolase
rMeTIR	Reverse metal-templated interface redesign
SAXS	Small angle X-ray scattering
sfGFP	Superfolder green fluorescent protein
SP1	Stable protein 1
TMV	Tobacco mosaic virus
TMVCP	Tobacco mosaic virus coat protein
TRAP	Tryptophan RNA-binding attenuation protein
VLP	Virus-like particle
XRD	X-ray diffraction

7. Author biographies



Clockwise from top left: Chung-Jui Yu, Yiying Li, Nicole Avakyan, Zhiyin Zhang, Kenneth Han, Jie Zhu, F. Akif Tezcan, Youjeong (Yui) Na, Tae Su Choi, Alexander M. Hoffnagle and Albert A. Kakkis.

Jie Zhu received his B.S. degree (2011) in Chemistry from Sun Yat-sen University and his Ph.D. degree (2018) from Virginia Tech, where he worked with the late Prof. Karen Brewer and Prof. Amanda Morris on developing supramolecular complexes for photodynamic therapy and metal-organic frameworks for CO₂ chemical transformations. Currently, he is a postdoctoral scholar in the Tezcan Group at University of California, San Diego. His research interests focus on developing novel artificial protein-based materials and metal-organic frameworks.

Nicole Avakyan completed her B.Sc. (2011) and Ph.D. (2018) degrees in Chemistry at McGill University, under the supervision of Prof. Hanadi Sleiman. As an undergraduate, she worked on the synthesis and characterization of platinum(II)-based guanine quadruplex binders for chemotherapeutic applications. In her graduate work, she focused on DNA-based nanomaterials, developing methods to reprogram and manipulate DNA assembly using simple self-assembly principles. She joined the Tezcan Group at the University of California, San Diego

as an NSERC postdoctoral fellow in 2019, working on rational design of nucleoprotein architectures through hierarchical co-assembly of proteins and nucleic acids, as well as 2D protein assembly.

Albert A. Kakkis received his B.A. in Chemistry (2016) from Pomona College. As an undergraduate, he worked in the lab of Prof. Cynthia Selassie, performing computational studies to predict the impact of active site mutations on the thermal stability of lipases. He is currently a Ph.D. candidate in the Tezcan Group at the University of California, San Diego, where his research focuses on the *de novo* design of metalloproteins harboring redox-dependent functions.

Alexander M. Hoffnagle received his B.A. in Biochemistry, B.A. in Biophysics, and M.S. in Chemistry from the University of Pennsylvania in 2018, where he was a graduate of the Roy and Diana Vagelos Scholars Program in the Molecular Life Sciences. As an undergraduate, he worked in the laboratory of Prof. Jessica Anna on multi-dimensional spectroscopy of molecules in photosystem I. He is currently a Ph.D. candidate in the Tezcan Group at the University of California, San Diego, where his research focuses on the *de novo* design of functional metalloproteins and protein assemblies.

Kenneth Han received his B.S. degree in Chemical Biology from University of California, Berkeley in 2017. During his undergraduate career, he joined the laboratory of Prof. Matthew Francis and studied protein-drug conjugates for therapeutic purposes. He then joined the laboratory of Prof. F. Akif Tezcan to pursue a Ph.D. in Chemistry at the University of California, San Diego. His research focuses on the design, synthesis, and application of polymer-integrated protein crystals.

Yiying Li received her B.S. degree from the College of Chemistry and Molecular Sciences at Wuhan University in 2018. As an undergraduate, she worked on organic synthesis methodology and CO₂ electroreduction with gold-metal oxide nanoparticles. She then joined the Tezcan Group to pursue a Ph.D. at the University of California, San Diego. Her research interests center on the design of metal-mediated protein assemblies.

Zhiyin Zhang received her B.S. degree in Chemistry from Zhejiang University in 2018. As an undergraduate, she worked on transition metal catalyzed C-H activation. She then joined the Tezcan Group to pursue a Ph.D. in Chemistry at the University of California, San Diego. Currently, her research focuses on extended protein assembly mediated by supramolecular interactions or carried out at interfaces.

Tae Su Choi received his B.S. degree (2012) and Ph.D. (2017) in Chemistry at Pohang University of Science and Technology (POSTECH) under the supervision of Prof. Hugh Kim on the mechanism and modulation of amyloid self-assembly. Currently, he is a postdoctoral scholar in Prof. F. Akif Tezcan's research group at the University of California, San Diego. His research interest is *de novo* design of selective metal-binding proteins and development of artificial phosphatase activity.

Youjeong (Yui) Na received her B.S. degree (2019) in Pharmacological Chemistry at the University of California, San Diego. She worked as an undergraduate researcher in the Prof. Stanley J. Opella's laboratory, studying the structure of G-protein coupled receptors using NMR spectroscopy. She joined the Tezcan Group to pursue her Ph.D. in the following year. Her

research focuses on the biomedical applications of polymer-integrated protein crystals (PIX). As a student member of MRSEC, her goal is to design polymer-protein hybrid materials that harness the unique structural and functional features of both types of building blocks.

Chung-Jui Yu was born in Taipei, Taiwan. He received his B.S. in Chemistry from the University of California, Berkeley in 2015 and his Ph.D. in Chemistry from Northwestern University in 2020, where he worked under the supervision of Prof. Danna E. Freedman on understanding electron-electron and electron-nuclear spin dynamics in molecular quantum bits. He is currently a postdoctoral scholar in Prof. F. Akif Tezcan's research group at the University of California, San Diego. His research focuses on the development of new dynamic nuclear polarization methodologies with model metalloprotein assemblies and investigating novel magnetism in 3D protein architectures.

F. Akif Tezcan was born and raised in Istanbul, Turkey. He received his B.A. degree in Chemistry from Macalester College and his Ph.D. degree from the California Institute of Technology (Caltech), where he worked with Prof. Harry Gray and Prof. Jay Winkler on the folding and redox properties of metalloproteins. He continued his research career at Caltech as a postdoctoral fellow in Prof. Doug Rees' group, where he focused on structural investigations of nitrogenase complexes. He joined the University of California, San Diego in 2005, where he is currently Professor of Chemistry, Biochemistry and Materials Science and Leslie Orgel Faculty Scholar. His research interests revolve around the use of chemical tools and principles to address biological questions and to create new biological materials.

8. Acknowledgments

We thank L. Zhang, R. Subramanian, R. Alberstein and J. Esselborn for extensive discussions and assistance. Research by the Tezcan Lab described in this Review was supported by the following agencies: US Department of Energy (DOE) Basic Energy Sciences (BES) (Division of Materials Sciences, Biomolecular Materials, DE-SC0003844 for the design and construction of dynamic protein assemblies and protein MOFs), National Science Foundation (DMR-2004558 for the design of 2- and 3D protein assemblies; CHE-1607145 for metalloprotein/enzyme design), National Institutes of Health (R01-GM138884 for metalloprotein/enzyme design) and US Army Research Office (W911NF-19-1-0228 for polymer-integrated protein crystals). Additional support was provided by DOE-BES, as part of the Energy Frontier Research Centers program: CSSAS—The Center for the Science of Synthesis Across Scales—under Award Number DE-SC0019288 (for interfacial protein assembly) and NASA (80NSSC18M0093; ENIGMA: Evolution of Nanomachines in Geospheres and Microbial Ancestors (NASA Astrobiology Institute Cycle 8). The Tezcan Lab is additionally supported by the UC San Diego Materials Research Science and Engineering Center (UCSD MRSEC, DMR-2011924). We gratefully acknowledge support by a NSERC postdoctoral fellowship to N.A. and NIH training grants to A.K. (Chemistry Biology Interfaces; T32GM112584) and A.M.H. (Molecular Biophysics, T32 GM008326).

9. References

- (1) Urey, H. C. On the Early Chemical History of the Earth and the Origin of Life. *Proc. Natl. Acad. Sci. U. S. A.* **1952**, 38 (4), 351-363.
- (2) Miller, S. L. The Formation of Organic Compounds on the Primitive Earth. *Ann. N. Y. Acad. Sci.* **1957**, 69 (2), 260-275.
- (3) Miller, S. L.; Urey, H. C. Organic Compound Synthesis on the Primitive Earth. *Science* **1959**, 130 (3370), 245-251.
- (4) Miller, S. L.; Urey, H. C.; Oro, J. Origin of Organic Compounds on the Primitive Earth and in Meteorites. *J. Mol. Evol.* **1976**, 9 (1), 59-72.
- (5) Joyce, G. F. RNA Evolution and the Origins of Life. *Nature* **1989**, 338 (6212), 217-224.
- (6) Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, 319 (6055), 618-618.
- (7) Weis, W. I.; Drickamer, K. Structural Basis of Lectin-Carbohydrate Recognition. *Annu. Rev. Biochem.* **1996**, 65 (1), 441-473.
- (8) Dwek, R. A. Glycobiology: Toward Understanding the Function of Sugars. *Chem. Rev.* **1996**, 96 (2), 683-720.
- (9) Lis, H.; Sharon, N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chem. Rev.* **1998**, 98 (2), 637-674.
- (10) Bertozzi, C. R.; Kiessling, L. L. Chemical Glycobiology. *Science* **2001**, 291 (5512), 2357-2364.
- (11) Gabius, H.-J.; Siebert, H.-C.; André, S.; Jiménez-Barbero, J.; Rüdiger, H. Chemical Biology of the Sugar Code. *ChemBioChem* **2004**, 5 (6), 740-764.
- (12) Ochman, H.; Lawrence, J. G.; Groisman, E. A. Lateral Gene Transfer and the Nature of Bacterial Innovation. *Nature* **2000**, 405 (6784), 299-304.
- (13) Joyce, G. F. The Antiquity of RNA-Based Evolution. *Nature* **2002**, 418 (6894), 214-221.
- (14) Caprara, M. G.; Nilsen, T. W. RNA: Versatility in Form and Function. *Nat. Struct. Biol.* **2000**, 7 (10), 831-833.
- (15) Chen, I.; Christie, P. J.; Dubnau, D. The Ins and Outs of DNA Transfer in Bacteria. *Science* **2005**, 310 (5753), 1456-1460.
- (16) Stubbe, J.; van Der Donk, W. A. Protein Radicals in Enzyme Catalysis. *Chem. Rev.* **1998**, 98 (2), 705-762.
- (17) Benkovic, S. J.; Hammes-Schiffer, S. A Perspective on Enzyme Catalysis. *Science* **2003**, 301 (5637), 1196-1202.
- (18) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **2004**, 303 (5655), 186-195.
- (19) Ciccia, A.; Elledge, S. J. The DNA Damage Response: Making It Safe to Play with Knives. *Mol. Cell* **2010**, 40 (2), 179-204.
- (20) Lindahl, T.; Wood, R. D. Quality Control by DNA Repair. *Science* **1999**, 286 (5446), 1897-1905.
- (21) Kunkel, T. A.; Bebenek, K. DNA Replication Fidelity. *Annu. Rev. Biochem.* **2000**, 69 (1), 497-529.
- (22) Catterall, W. A. Structure and Function of Voltage-Gated Ion Channels. *Annu. Rev. Biochem.* **1995**, 64 (1), 493-531.
- (23) Long, S. B.; Campbell, E. B.; Mackinnon, R. Voltage Sensor of Kv1.2: Structural Basis of Electromechanical Coupling. *Science* **2005**, 309 (5736), 903-908.
- (24) Jiang, Y.; Ruta, V.; Chen, J.; Lee, A.; MacKinnon, R. The Principle of Gating Charge Movement in a Voltage-Dependent K⁺ Channel. *Nature* **2003**, 423 (6935), 42-48.
- (25) Chang, G.; Spencer, R. H.; Lee, A. T.; Barclay, M. T.; Rees, D. C. Structure of the Mscl Homolog from *Mycobacterium Tuberculosis*: A Gated Mechanosensitive Ion Channel. *Science* **1998**, 282 (5397), 2220-2226.

- (26) Perozo, E.; Kloda, A.; Cortes, D. M.; Martinac, B. Physical Principles Underlying the Transduction of Bilayer Deformation Forces During Mechanosensitive Channel Gating. *Nat. Struct. Biol.* **2002**, 9 (9), 696-703.
- (27) Perozo, E.; Cortes, D. M.; Sompornpisut, P.; Kloda, A.; Martinac, B. Open Channel Structure of Mscl and the Gating Mechanism of Mechanosensitive Channels. *Nature* **2002**, 418 (6901), 942-948.
- (28) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, 29 (31), 7133-7155.
- (29) Newberry, R. W.; Raines, R. T. Secondary Forces in Protein Folding. *ACS Chem. Biol.* **2019**, 14 (8), 1677-1686.
- (30) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997**, 48 (1), 545-600.
- (31) Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, 426 (6968), 884-890.
- (32) Dobson, C. M.; Sali, A.; Karplus, M. Protein Folding: A Perspective from Theory and Experiment. *Angew. Chem. Int. Ed. Engl.* **1998**, 37 (7), 868-893.
- (33) Jones, S.; Thornton, J. M. Principles of Protein-Protein Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, 93 (1), 13-20.
- (34) Lo Conte, L.; Chothia, C.; Janin, J. The Atomic Structure of Protein-Protein Recognition Sites. *J. Mol. Biol.* **1999**, 285 (5), 2177-2198.
- (35) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic Aspects of Protein-Protein Interactions. *Curr. Opin. Struct. Biol.* **2000**, 10 (2), 153-159.
- (36) Hung, A. Y.; Sheng, M. Pdz Domains: Structural Modules for Protein Complex Assembly. *J. Biol. Chem.* **2002**, 277 (8), 5699-5702.
- (37) Janin, J.; Bahadur, R. P.; Chakrabarti, P. Protein-Protein Interaction and Quaternary Structure. *Q. Rev. Biophys.* **2008**, 41 (2), 133-180.
- (38) Hamm, H. E. The Many Faces of G Protein Signaling. *J. Biol. Chem.* **1998**, 273 (2), 669-672.
- (39) Virkamaki, A.; Ueki, K.; Kahn, C. R. Protein-Protein Interaction in Insulin Signaling and the Molecular Mechanisms of Insulin Resistance. *J. Clin. Invest.* **1999**, 103 (7), 931-943.
- (40) Saltiel, A. R.; Kahn, C. R. Insulin Signalling and the Regulation of Glucose and Lipid Metabolism. *Nature* **2001**, 414 (6865), 799-806.
- (41) Nelson, N.; Ben-Shem, A. The Complex Architecture of Oxygenic Photosynthesis. *Nat. Rev. Mol. Cell Biol.* **2004**, 5 (12), 971-982.
- (42) Nelson, N.; Yocom, C. F. Structure and Function of Photosystems I and II. *Annu. Rev. Plant Biol.* **2006**, 57 (1), 521-565.
- (43) Howard, J. B.; Rees, D. C. Structural Basis of Biological Nitrogen Fixation. *Chem. Rev.* **1996**, 96 (7), 2965-2982.
- (44) Rutledge, H. L.; Tezcan, F. A. Electron Transfer in Nitrogenase. *Chem. Rev.* **2020**, 120 (12), 5158-5193.
- (45) Marsh, J. A.; Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, 84 (1), 551-575.
- (46) dos Remedios, C. G.; Chhabra, D.; Kekic, M.; Dedova, I. V.; Tsubakihara, M.; Berry, D. A.; Nosworthy, N. J. Actin Binding Proteins: Regulation of Cytoskeletal Microfilaments. *Physiol. Rev.* **2003**, 83 (2), 433-473.
- (47) Tojkander, S.; Gateva, G.; Lappalainen, P. Actin Stress Fibers--Assembly, Dynamics and Biological Roles. *J. Cell Sci.* **2012**, 125 (Pt 8), 1855-1864.
- (48) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L. et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **2018**, 28 (6), 420-435.
- (49) Chiesa, G.; Kiriakov, S.; Khalil, A. S. Protein Assembly Systems in Natural and Synthetic Biology. *BMC Biol.* **2020**, 18 (1), 35.

- (50) Bajaj, M.; Blundell, T. Evolution and the Tertiary Structure of Proteins. *Annu. Rev. Biophys. Bioeng.* **1984**, 13 (1), 453-492.
- (51) Quiocio, F. A. Carbohydrate-Binding Proteins: Tertiary Structures and Protein-Sugar Interactions. *Annu. Rev. Biochem.* **1986**, 55 (1), 287-315.
- (52) Gether, U. Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocr. Rev.* **2000**, 21 (1), 90-113.
- (53) Beratan, D. N.; Betts, J. N.; Onuchic, J. N. Protein Electron Transfer Rates Set by the Bridging Secondary and Tertiary Structure. *Science* **1991**, 252 (5010), 1285-1288.
- (54) Levy, E. D.; Teichmann, S. Structural, Evolutionary, and Assembly Principles of Protein Oligomerization. *Prog. Mol. Biol. Transl. Sci.* **2013**, 117, 25-51.
- (55) Levy, E. D.; Erba, E. B.; Robinson, C. V.; Teichmann, S. A. Assembly Reflects Evolution of Protein Complexes. *Nature* **2008**, 453 (7199), 1262-1265.
- (56) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The Protein Folding Problem. *Annu. Rev. Biophys.* **2008**, 37 (1), 289-316.
- (57) Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. *Science* **1973**, 181 (4096), 223-230.
- (58) Haber, E.; Anfinsen, C. B. Side-Chain Interactions Governing the Pairing of Half-Cystine Residues in Ribonuclease. *J. Biol. Chem.* **1962**, 237 (6), 1839-1844.
- (59) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, 234 (3), 779-815.
- (60) Eisenhaber, F.; Persson, B.; Argos, P. Protein Structure Prediction: Recognition of Primary, Secondary, and Tertiary Structural Features from Amino Acid Sequence. *Crit. Rev. Biochem. Mol. Biol.* **1995**, 30 (1), 1-94.
- (61) Jones, D. T. Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices. *J. Mol. Biol.* **1999**, 292 (2), 195-202.
- (62) Zhang, Y. I-Tasser Server for Protein 3d Structure Prediction. *BMC Bioinf.* **2008**, 9 (1), 40.
- (63) Kelley, L. A.; Sternberg, M. J. Protein Structure Prediction on the Web: A Case Study Using the Phyre Server. *Nat. Protoc.* **2009**, 4 (3), 363-371.
- (64) Hardin, C.; Pogorelov, T. V.; Luthey-Schulten, Z. Ab Initio Protein Structure Prediction. *Curr. Opin. Struct. Biol.* **2002**, 12 (2), 176-181.
- (65) Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S. High-Resolution Protein Design with Backbone Freedom. *Science* **1998**, 282 (5393), 1462-1467.
- (66) Dahiyat, B. I.; Mayo, S. L. De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997**, 278 (5335), 82-87.
- (67) Grigoryan, G.; Degrado, W. F. Probing Designability Via a Generalized Model of Helical Bundle Geometry. *J. Mol. Biol.* **2011**, 405 (4), 1079-1100.
- (68) Regan, L.; DeGrado, W. F. Characterization of a Helical Protein Designed from First Principles. *Science* **1988**, 241 (4868), 976-978.
- (69) Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. Principles for Designing Ideal Protein Structures. *Nature* **2012**, 491 (7423), 222-227.
- (70) Koepnick, B.; Flatten, J.; Husain, T.; Ford, A.; Silva, D.-A.; Bick, M. J.; Bauer, A.; Liu, G.; Ishida, Y.; Boykov, A. et al. De Novo Protein Design by Citizen Scientists. *Nature* **2019**, 570 (7761), 390-394.
- (71) Jaenicke, R.; Lilie, H. *Folding and Association of Oligomeric and Multimeric Proteins In Advances in Protein Chemistry*; Matthews, C. R., Ed.; Academic Press, 2000; Vol. 53.
- (72) Keskin, O.; Tuncbag, N.; Gursoy, A. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem. Rev.* **2016**, 116 (8), 4884-4909.
- (73) Kumari, N.; Yadav, S. Modulation of Protein Oligomerization: An Overview. *Prog. Biophys. Mol. Biol.* **2019**, 149, 99-113.
- (74) Minton, A. P. Implications of Macromolecular Crowding for Protein Assembly. *Curr. Opin. Struct. Biol.* **2000**, 10 (1), 34-39.

- (75) Bishop, K. J. M.; Wilmer, C. E.; Soh, S.; Grzybowski, B. A. Nanoscale Forces and Their Uses in Self-Assembly. *Small* **2009**, 5 (14), 1600-1630.
- (76) Gabrys, P. A.; Zornberg, L. Z.; Macfarlane, R. J. Programmable Atom Equivalents: Atomic Crystallization as a Framework for Synthesizing Nanoparticle Superlattices. *Small* **2019**, 15 (26), 1805424.
- (77) Grzelczak, M.; Vermant, J.; Furst, E. M.; Liz-Marzán, L. M. Directed Self-Assembly of Nanoparticles. *ACS Nano* **2010**, 4 (7), 3591-3605.
- (78) Kao, J.; Thorkelsson, K.; Bai, P.; Rancatore, B. J.; Xu, T. Toward Functional Nanocomposites: Taking the Best of Nanoparticles, Polymers, and Small Molecules. *Chem. Soc. Rev.* **2013**, 42 (7), 2654-2678.
- (79) Li, W.; Palis, H.; Merindol, R.; Majimel, J.; Ravaine, S.; Duguet, E. Colloidal Molecules and Patchy Particles: Complementary Concepts, Synthesis and Self-Assembly. *Chem. Soc. Rev.* **2020**, 49 (6), 1955-1976.
- (80) Zhang, W.-B.; Yu, X.; Wang, C.-L.; Sun, H.-J.; Hsieh, I. F.; Li, Y.; Dong, X.-H.; Yue, K.; Van Horn, R.; Cheng, S. Z. D. Molecular Nanoparticles Are Unique Elements for Macromolecular Science: From “Nanoatoms” to Giant Molecules. *Macromolecules* **2014**, 47 (4), 1221-1239.
- (81) Reedy, C. J.; Gibney, B. R. Heme Protein Assemblies. *Chem. Rev.* **2004**, 104 (2), 617-649.
- (82) Howorka, S. Rationally Engineering Natural Protein Assemblies in Nanobiotechnology. *Curr. Opin. Biotechnol.* **2011**, 22 (4), 485-491.
- (83) Lai, Y. T.; King, N. P.; Yeates, T. O. Principles for Designing Ordered Protein Assemblies. *Trends Cell Biol.* **2012**, 22 (12), 653-661.
- (84) Bailey, J. B.; Subramanian, R. H.; Churchfield, L. A.; Tezcan, F. A. *Chapter Eleven - Metal-Directed Design of Supramolecular Protein Assemblies In Method Enzymol.*; Pecoraro, V. L., Ed.; Academic Press, 2016; Vol. 580.
- (85) Pieters, B. J.; van Eldijk, M. B.; Nolte, R. J.; Mecinovic, J. Natural Supramolecular Protein Assemblies. *Chem. Soc. Rev.* **2016**, 45 (1), 24-39.
- (86) Nguyen, T. K.; Ueno, T. Engineering of Protein Assemblies within Cells. *Curr. Opin. Struct. Biol.* **2018**, 51, 1-8.
- (87) Cannon, K. A.; Ochoa, J. M.; Yeates, T. O. High-Symmetry Protein Assemblies: Patterns and Emerging Applications. *Curr. Opin. Struct. Biol.* **2019**, 55, 77-84.
- (88) Churchfield, L. A.; Tezcan, F. A. Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc. Chem. Res.* **2019**, 52 (2), 345-355.
- (89) Fegan, A.; White, B.; Carlson, J. C.; Wagner, C. R. Chemically Controlled Protein Assembly: Techniques and Applications. *Chem. Rev.* **2010**, 110 (6), 3315-3336.
- (90) Bai, Y.; Luo, Q.; Liu, J. Protein Self-Assembly Via Supramolecular Strategies. *Chem. Soc. Rev.* **2016**, 45 (10), 2756-2767.
- (91) Luo, Q.; Hou, C.; Bai, Y.; Wang, R.; Liu, J. Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures. *Chem. Rev.* **2016**, 116 (22), 13571-13632.
- (92) Wang, X.; Liu, X.; Huang, X. Bioinspired Protein-Based Assembling: Toward Advanced Life-Like Behaviors. *Adv. Mater.* **2020**, 32 (25), e2001436.
- (93) Norn, C. H.; Andre, I. Computational Design of Protein Self-Assembly. *Curr. Opin. Struct. Biol.* **2016**, 39, 39-45.
- (94) Aumiller, W. M.; Uchida, M.; Douglas, T. Protein Cage Assembly across Multiple Length Scales. *Chem. Soc. Rev.* **2018**, 47 (10), 3433-3469.
- (95) Kuan, S. L.; Bergamini, F. R. G.; Weil, T. Functional Protein Nanostructures: A Chemical Toolbox. *Chem. Soc. Rev.* **2018**, 47 (24), 9069-9105.
- (96) Hamley, I. W. Protein Assemblies: Nature-Inspired and Designed Nanostructures. *Biomacromolecules* **2019**, 20 (5), 1829-1848.

- (97) Solomonov, A.; Shimanovich, U. Self-Assembly in Protein-Based Bionanomaterials. *Isr. J. Chem.* **2020**, 60 (12), 1152-1170.
- (98) Beloqui, A.; Cortajarena, A. L. Protein-Based Functional Hybrid Bionanomaterials by Bottom-up Approaches. *Curr. Opin. Struct. Biol.* **2020**, 63, 74-81.
- (99) Hansen, W. A.; Khare, S. D. Recent Progress in Designing Protein-Based Supramolecular Assemblies. *Curr. Opin. Struct. Biol.* **2020**, 63, 106-114.
- (100) Majsterkiewicz, K.; Azuma, Y.; Heddle, J. G. Connectability of Protein Cages. *Nanoscale Adv.* **2020**, 2 (6), 2255-2264.
- (101) Korpi, A.; Anaya-Plaza, E.; Valimaki, S.; Kostiainen, M. Highly Ordered Protein Cage Assemblies: A Toolkit for New Materials. *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.* **2020**, 12 (1), e1578.
- (102) Dehsorkhi, A.; Castelletto, V.; Hamley, I. W. Self-Assembling Amphiphilic Peptides. *J. Pept. Sci.* **2014**, 20 (7), 453-467.
- (103) Lapenta, F.; Aupič, J.; Strmšek, Ž.; Jerala, R. Coiled Coil Protein Origami: From Modular Design Principles Towards Biotechnological Applications. *Chem. Soc. Rev.* **2018**, 47 (10), 3530-3542.
- (104) Katyal, P.; Meleties, M.; Montclare, J. K. Self-Assembled Protein- and Peptide-Based Nanomaterials. *ACS Biomater. Sci. Eng.* **2019**, 5 (9), 4132-4147.
- (105) Beesley, J. L.; Woolfson, D. N. The De Novo Design of Alpha-Helical Peptides for Supramolecular Self-Assembly. *Curr. Opin. Biotechnol.* **2019**, 58, 175-182.
- (106) Levin, A.; Hakala, T. A.; Schnaider, L.; Bernardes, G. J. L.; Gazit, E.; Knowles, T. P. J. Biomimetic Peptide Self-Assembly for Functional Materials. *Nat. Rev. Chem.* **2020**, 4 (11), 615-634.
- (107) Pizzi, A.; Pigliacelli, C.; Bergamaschi, G.; Gori, A.; Metrangolo, P. Biomimetic Engineering of the Molecular Recognition and Self-Assembly of Peptides and Proteins Via Halogenation. *Coordin. Chem. Rev.* **2020**, 411, 213242.
- (108) Song, Q.; Cheng, Z.; Kariuki, M.; Hall, S. C. L.; Hill, S. K.; Rho, J. Y.; Perrier, S. Molecular Self-Assembly and Supramolecular Chemistry of Cyclic Peptides. *Chem. Rev.* **2021**.
- (109) Willbold, D.; Strodel, B.; Schröder, G. F.; Hoyer, W.; Heise, H. Amyloid-Type Protein Aggregation and Prion-Like Properties of Amyloids. *Chem. Rev.* **2021**.
- (110) Goodsell, D. S.; Olson, A. J. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, 29 (1), 105-153.
- (111) Yuan, Y.; Tam, M. F.; Simplaceanu, V.; Ho, C. New Look at Hemoglobin Allostery. *Chem. Rev.* **2015**, 115 (4), 1702-1724.
- (112) Springer, B. A.; Sligar, S. G.; Olson, J. S.; Phillips, G. N. Mechanisms of Ligand Recognition in Myoglobin. *Chem. Rev.* **1994**, 94 (3), 699-714.
- (113) Nogales, E. Structural Insights into Microtubule Function. *Annu. Rev. Biochem.* **2000**, 69 (1), 277-302.
- (114) Janmey, P. A. Mechanical Properties of Cytoskeletal Polymers. *Curr. Opin. Cell Biol.* **1991**, 3 (1), 4-11.
- (115) Howard, J. Mechanical Signaling in Networks of Motor and Cytoskeletal Proteins. *Annu. Rev. Biophys.* **2009**, 38 (1), 217-234.
- (116) Fagan, R. P.; Fairweather, N. F. Biogenesis and Functions of Bacterial S-Layers. *Nat. Rev. Microbiol.* **2014**, 12 (3), 211-222.
- (117) Sleytr, U. B.; Schuster, B.; Egelseer, E. M.; Pum, D. S-Layers: Principles and Applications. *FEMS Microbiol. Rev.* **2014**, 38 (5), 823-864.
- (118) Lopez, M. G.; Diez, M.; Alfonso, V.; Taboga, O. Biotechnological Applications of Occlusion Bodies of Baculoviruses. *Appl. Microbiol. Biotechnol.* **2018**, 102 (16), 6765-6774.
- (119) Li, Y. V. Zinc and Insulin in Pancreatic Beta-Cells. *Endocrine* **2014**, 45 (2), 178-189.

- (120) Coulibaly, F.; Chiu, E.; Ikeda, K.; Gutmann, S.; Haebel, P. W.; Schulze-Briese, C.; Mori, H.; Metcalf, P. The Molecular Organization of Cypovirus Polyhedra. *Nature* **2007**, *446* (7131), 97-101.
- (121) Ozimek, P.; Veenhuis, M.; van der Klei, I. J. Alcohol Oxidase: A Complex Peroxisomal, Oligomeric Flavoprotein. *FEMS Yeast. Res.* **2005**, *5* (11), 975-983.
- (122) Mateu, M. G. Assembly, Stability and Dynamics of Virus Capsids. *Arch. Biochem. Biophys.* **2013**, *531* (1-2), 65-79.
- (123) He, D.; Marles-Wright, J. Ferritin Family Proteins and Their Use in Bionanotechnology. *New Biotechnol.* **2015**, *32* (6), 651-657.
- (124) Ahnert, S. E.; Marsh, J. A.; Hernandez, H.; Robinson, C. V.; Teichmann, S. A. Principles of Assembly Reveal a Periodic Table of Protein Complexes. *Science* **2015**, *350* (6266), aaa2245.
- (125) Shen, J.-R. *Photosystem II: Protein Components, Structure and Electron Transfer In Reference Module in Life Sciences*; Elsevier, 2020.
- (126) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. A Dissection of Specific and Non-Specific Protein-Protein Interfaces. *J. Mol. Biol.* **2004**, *336* (4), 943-955.
- (127) Padilla, J. E.; Colovos, C.; Yeates, T. O. Nanohedra: Using Symmetry to Design Self-Assembling Protein Cages, Layers, Crystals, and Filaments. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (5), 2217-2221.
- (128) Sinclair, J. C.; Davies, K. M.; Venien-Bryan, C.; Noble, M. E. Generation of Protein Lattices by Fusing Proteins with Matching Rotational Symmetry. *Nat Nanotechnol* **2011**, *6* (9), 558-562.
- (129) Sontz, P. A.; Bailey, J. B.; Ahn, S.; Tezcan, F. A. A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3d Protein Crystals. *J. Am. Chem. Soc.* **2015**, *137* (36), 11598-11601.
- (130) Linci, C. J.; MacDermaid, C. M.; Kang, S. G.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. Computational Design of a Protein Crystal. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (19), 7304-7309.
- (131) Laniado, J.; Yeates, T. O. A Complete Rule Set for Designing Symmetry Combination Materials from Protein Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (50), 31817-31823.
- (132) Yeates, T. O.; Liu, Y.; Laniado, J. The Design of Symmetric Protein Nanomaterials Comes of Age in Theory and Practice. *Curr. Opin. Struct. Biol.* **2016**, *39*, 134-143.
- (133) Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **2017**, *46* (1), 23-42.
- (134) Tsai, C. J.; Zheng, J.; Zanuy, D.; Haspel, N.; Wolfson, H.; Aleman, C.; Nussinov, R. Principles of Nanostructure Design with Protein Building Blocks. *Proteins* **2007**, *68* (1), 1-12.
- (135) Caulder, D. L.; Raymond, K. N. Supermolecules by Design. *Acc. Chem. Res.* **1999**, *32* (11), 975-982.
- (136) Holliday, B. J.; Mirkin, C. A. Strategies for the Construction of Supramolecular Compounds through Coordination Chemistry. *Angew. Chem. Int. Ed. Engl.* **2001**, *40* (11), 2022-2043.
- (137) Northrop, B. H.; Zheng, Y.-R.; Chi, K.-W.; Stang, P. J. Self-Organization in Coordination-Driven Self-Assembly. *Acc. Chem. Res.* **2009**, *42* (10), 1554-1563.
- (138) Yoshizawa, M.; Klosterman, J. K.; Fujita, M. Functional Molecular Flasks: New Properties and Reactions within Discrete, Self-Assembled Hosts. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (19), 3418-3438.
- (139) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. Reticular Synthesis and the Design of New Materials. *Nature* **2003**, *423* (6941), 705-714.

- (140) Sciore, A.; Su, M.; Koldewey, P.; Eschweiler, J. D.; Diffley, K. A.; Linhares, B. M.; Ruotolo, B. T.; Bardwell, J. C.; Skiniotis, G.; Marsh, E. N. Flexible, Symmetry-Directed Approach to Assembling Protein Cages. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113 (31), 8681-8686.
- (141) Badieyan, S.; Sciore, A.; Eschweiler, J. D.; Koldewey, P.; Cristie-David, A. S.; Ruotolo, B. T.; Bardwell, J. C. A.; Su, M.; Marsh, E. N. G. Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by De Novo-Designed Coiled Coils. *ChemBioChem* **2017**, 18 (19), 1888-1892.
- (142) Cristie-David, A. S.; Chen, J.; Nowak, D. B.; Bondy, A. L.; Sun, K.; Park, S. I.; Banaszak Holl, M. M.; Su, M.; Marsh, E. N. G. Coiled-Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability. *J. Am. Chem. Soc.* **2019**, 141 (23), 9207-9216.
- (143) Cannon, K. A.; Nguyen, V. N.; Morgan, C.; Yeates, T. O. Design and Characterization of an Icosahedral Protein Cage Formed by a Double-Fusion Protein Containing Three Distinct Symmetry Elements. *ACS Synth. Biol.* **2020**, 9 (3), 517-524.
- (144) Boas, F. E.; Harbury, P. B. Potential Energy Functions for Protein Design. *Curr. Opin. Struct. Biol.* **2007**, 17 (2), 199-204.
- (145) Samish, I.; MacDermaid, C. M.; Perez-Aguilar, J. M.; Saven, J. G. Theoretical and Computational Protein Design. *Annu. Rev. Phys. Chem.* **2011**, 62, 129-149.
- (146) Mackerell, A. D., Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, 25 (13), 1584-1604.
- (147) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984**, 106 (3), 765-784.
- (148) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, 4 (2), 187-217.
- (149) Hermans, J.; Berendsen, H. J. C.; Vangunsteren, W. F.; Postma, J. P. M. A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers* **1984**, 23 (8), 1513-1518.
- (150) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. Dreiding - a Generic Force-Field for Molecular Simulations. *J. Phys. Chem.* **1990**, 94 (26), 8897-8909.
- (151) Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102 (19), 6665-6670.
- (152) Lazaridis, T.; Karplus, M. Effective Energy Functions for Protein Structure Prediction. *Curr. Opin. Struct. Biol.* **2000**, 10 (2), 139-145.
- (153) Mohanty, D.; Dominy, B. N.; Kolinski, A.; Brooks, C. L.; Skolnick, J. Correlation between Knowledge-Based and Detailed Atomic Potentials: Application to the Unfolding of the Gcn4 Leucine Zipper. *Proteins: Struct. Funct. Bioinf.* **1999**, 35 (4), 447-452.
- (154) Dehouck, Y.; Gilis, D.; Roonan, M. A New Generation of Statistical Potentials for Proteins. *Biophys J* **2006**, 90 (11), 4010-4017.
- (155) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K. et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, 13 (6), 3031-3048.
- (156) Looger, L. L.; Dwyer, M. A.; Smith, J. J.; Hellinga, H. W. Computational Design of Receptor and Sensor Proteins with Novel Functions. *Nature* **2003**, 423 (6936), 185-190.
- (157) Kuhlman, B.; Bradley, P. Advances in Protein Structure Prediction and Design. *Nat. Rev. Mol. Cell Biol.* **2019**, 20 (11), 681-697.
- (158) Dunbrack, R. L., Jr. Rotamer Libraries in the 21st Century. *Curr. Opin. Struct. Biol.* **2002**, 12 (4), 431-440.

- (159) Dunbrack, R. L., Jr.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins. Application to Side-Chain Prediction. *J. Mol. Biol.* **1993**, *230* (2), 543-574.
- (160) Ponder, J. W.; Richards, F. M. Tertiary Templates for Proteins - Use of Packing Criteria in the Enumeration of Allowed Sequences for Different Structural Classes. *J. Mol. Biol.* **1987**, *193* (4), 775-791.
- (161) Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R. A New Approach to the Rapid Determination of Protein Side Chain Conformations. *J. Biomol. Struct. Dyn.* **1991**, *8* (6), 1267-1289.
- (162) Kono, H.; Doi, J. A New Method for Side-Chain Conformation Prediction Using a Hopfield Network and Reproduced Rotamers. *J. Comput. Chem.* **1996**, *17* (14), 1667-1683.
- (163) DeMaeyer, M.; Desmet, J.; Lasters, I. All in One: A Highly Detailed Rotamer Library Improves Both Accuracy and Speed in the Modelling of Sidechains by Dead-End Elimination. *Fold. Des.* **1997**, *2* (1), 53-66.
- (164) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The Penultimate Rotamer Library. *Proteins* **2000**, *40* (3), 389-408.
- (165) Holm, L.; Sander, C. Fast and Simple Monte-Carlo Algorithm for Side-Chain Optimization in Proteins - Application to Model-Building by Homology. *Proteins: Struct, Funct, Genet* **1992**, *14* (2), 213-223.
- (166) Hellinga, H. W.; Richards, F. M. Optimal Sequence Selection in Proteins of Known Structure by Simulated Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (13), 5803-5807.
- (167) Godzik, A. In Search of the Ideal Protein-Sequence. *Protein Eng.* **1995**, *8* (5), 409-416.
- (168) Dahiyat, B. I.; Mayo, S. L. Protein Design Automation. *Protein Sci.* **1996**, *5* (5), 895-903.
- (169) Voigt, C. A.; Gordon, D. B.; Mayo, S. L. Trading Accuracy for Speed: A Quantitative Comparison of Search Algorithms in Protein Sequence Design. *J. Mol. Biol.* **2000**, *299* (3), 789-803.
- (170) Kuhlman, B.; Baker, D. Native Protein Sequences Are Close to Optimal for Their Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (19), 10383-10388.
- (171) Desmet, J.; Demaeyer, M.; Hazes, B.; Lasters, I. The Dead-End Elimination Theorem and Its Use in Protein Side-Chain Positioning. *Nature* **1992**, *356* (6369), 539-542.
- (172) Goldstein, R. F. Efficient Rotamer Elimination Applied to Protein Side-Chains and Related Spin Glasses. *Biophys J* **1994**, *66* (5), 1335-1340.
- (173) Koehl, P.; Delarue, M. Application of a Self-Consistent Mean Field Theory to Predict Protein Side-Chains Conformation and Estimate Their Conformational Entropy. *J. Mol. Biol.* **1994**, *239* (2), 249-275.
- (174) Lee, C. Predicting Protein Mutant Energetics by Self-Consistent Ensemble Optimization. *J. Mol. Biol.* **1994**, *236* (3), 918-939.
- (175) Looger, L. L.; Hellinga, H. W. Generalized Dead-End Elimination Algorithms Make Large-Scale Protein Side-Chain Structure Prediction Tractable: Implications for Protein Design and Structural Genomics. *J. Mol. Biol.* **2001**, *307* (1), 429-445.
- (176) Mou, Y.; Huang, P. S.; Hsu, F. C.; Huang, S. J.; Mayo, S. L. Computational Design and Experimental Verification of a Symmetric Protein Homodimer. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (34), 10714-10719.
- (177) Vakser, I. A. Protein-Protein Docking: From Interaction to Interactome. *Biophys J* **2014**, *107* (8), 1785-1793.
- (178) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular Surface Recognition: Determination of Geometric Fit between Proteins and Their Ligands by Correlation Techniques. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (6), 2195-2199.
- (179) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein-Protein Docking with Simultaneous Optimization of Rigid-Body Displacement and Side-Chain Conformations. *J. Mol. Biol.* **2003**, *331* (1), 281-299.

- (180) Dominguez, C.; Boelens, R.; Bonvin, A. M. Haddock: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, 125 (7), 1731-1737.
- (181) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The Cluspro Web Server for Protein-Protein Docking. *Nat. Protoc.* **2017**, 12 (2), 255-278.
- (182) Yan, Y.; Tao, H.; He, J.; Huang, S. Y. The Hdock Server for Integrated Protein-Protein Docking. *Nat. Protoc.* **2020**, 15 (5), 1829-1852.
- (183) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W. et al. Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Method Enzymol.* **2011**, 487, 545-574.
- (184) Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. A.; Dym, O.; Peck, S. H.; Albeck, S.; Unger, T.; Hu, W.; Liu, G. et al. A De Novo Protein Binding Pair by Computational Design and Directed Evolution. *Mol. Cell* **2011**, 42 (2), 250-260.
- (185) Doyle, L.; Hallinan, J.; Bolduc, J.; Parmeggiani, F.; Baker, D.; Stoddard, B. L.; Bradley, P. Rational Design of Alpha-Helical Tandem Repeat Proteins with Closed Architectures. *Nature* **2015**, 528 (7583), 585-588.
- (186) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; Andre, I.; Gonen, T.; Yeates, T. O.; Baker, D. Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **2012**, 336 (6085), 1171-1174.
- (187) Shen, H.; Fallas, J. A.; Lynch, E.; Sheffler, W.; Parry, B.; Jannetty, N.; Decarreau, J.; Wagenbach, M.; Vicente, J. J.; Chen, J. et al. De Novo Design of Self-Assembling Helical Protein Filaments. *Science* **2018**, 362 (6415), 705-709.
- (188) Gonen, S.; DiMaio, F.; Gonen, T.; Baker, D. Design of Ordered Two-Dimensional Arrays Mediated by Noncovalent Protein-Protein Interfaces. *Science* **2015**, 348 (6241), 1365-1368.
- (189) Chen, Z.; Boyken, S. E.; Jia, M.; Busch, F.; Flores-Solis, D.; Bick, M. J.; Lu, P.; VanAernum, Z. L.; Sahasrabuddhe, A.; Langan, R. A. et al. Programmable Design of Orthogonal Protein Heterodimers. *Nature* **2019**, 565 (7737), 106-111.
- (190) Chen, Z.; Kibler, R. D.; Hunt, A.; Busch, F.; Pearl, J.; Jia, M.; VanAernum, Z. L.; Wicky, B. I. M.; Dods, G.; Liao, H. et al. De Novo Design of Protein Logic Gates. *Science* **2020**, 368 (6486), 78-84.
- (191) Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A. Metals in Protein-Protein Interfaces. *Annu. Rev. Biophys.* **2014**, 43 (1), 409-431.
- (192) Sontz, P. A.; Song, W. J.; Tezcan, F. A. Interfacial Metal Coordination in Engineered Protein and Peptide Assemblies. *Curr. Opin. Chem. Biol.* **2014**, 19, 42-49.
- (193) Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.; Kuhlman, B.; Tezcan, F. A. Metal Templated Design of Protein Interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107 (5), 1827-1832.
- (194) Salgado, E. N.; Lewis, R. A.; Mossin, S.; Rheingold, A. L.; Tezcan, F. A. Control of Protein Oligomerization Symmetry by Metal Coordination: C2 and C3 Symmetrical Assemblies through CuII and NiII Coordination. *Inorg. Chem.* **2009**, 48 (7), 2726-2728.
- (195) Golub, E.; Subramanian, R. H.; Esselborn, J.; Alberstein, R. G.; Bailey, J. B.; Chiong, J. A.; Yan, X.; Booth, T.; Baker, T. S.; Tezcan, F. A. Constructing Protein Polyhedra Via Orthogonal Chemical Interactions. *Nature* **2020**, 578 (7793), 172-176.
- (196) Kakkis, A.; Gagnon, D.; Esselborn, J.; Britt, R. D.; Tezcan, F. A. Metal-Templated Design of Chemically Switchable Protein Assemblies with High-Affinity Coordination Sites. *Angew. Chem. Int. Ed. Engl.* **2020**, 59 (49), 21940-21944.
- (197) Radford, R. J.; Nguyen, P. C.; Ditri, T. B.; Figueiroa, J. S.; Tezcan, F. A. Controlled Protein Dimerization through Hybrid Coordination Motifs. *Inorg. Chem.* **2010**, 49 (9), 4362-4369.

- (198) Salgado, E. N.; Radford, R. J.; Tezcan, F. A. Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* **2010**, 43 (5), 661-672.
- (199) Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A. Controlling Protein-Protein Interactions through Metal Coordination: Assembly of a 16-Helix Bundle Protein. *J. Am. Chem. Soc.* **2007**, 129 (44), 13374-13375.
- (200) Malay, A. D.; Miyazaki, N.; Biela, A.; Chakraborti, S.; Majsterkiewicz, K.; Stupka, I.; Kaplan, C. S.; Kowalczyk, A.; Piette, B.; Hochberg, G. K. A. et al. An Ultra-Stable Gold-Coordinated Protein Cage Displaying Reversible Assembly. *Nature* **2019**, 569 (7756), 438-442.
- (201) Sim, S.; Niwa, T.; Taguchi, H.; Aida, T. Supramolecular Nanotube of Chaperonin GroEL: Length Control for Cellular Uptake Using Single-Ring GroEL Mutant as End-Capper. *J. Am. Chem. Soc.* **2016**, 138 (35), 11152-11155.
- (202) Brodin, J. D.; Ambroggio, X. I.; Tang, C.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. Metal-Directed, Chemically Tunable Assembly of One-, Two- and Three-Dimensional Crystalline Protein Arrays. *Nat. Chem.* **2012**, 4 (5), 375-382.
- (203) Biswas, S.; Kinbara, K.; Niwa, T.; Taguchi, H.; Ishii, N.; Watanabe, S.; Miyata, K.; Kataoka, K.; Aida, T. Biomolecular Robotics for Chemomechanically Driven Guest Delivery Fuelled by Intracellular Atp. *Nat. Chem.* **2013**, 5 (7), 613-620.
- (204) Brodin, J. D.; Carr, J. R.; Sontz, P. A.; Tezcan, F. A. Exceptionally Stable, Redox-Active Supramolecular Protein Assemblies with Emergent Properties. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, 111 (8), 2897-2902.
- (205) Song, W. J.; Tezcan, F. A. A Designed Supramolecular Protein Assembly with in Vivo Enzymatic Activity. *Science* **2014**, 346 (6216), 1525-1528.
- (206) Song, W. J.; Yu, J.; Tezcan, F. A. Importance of Scaffold Flexibility/Rigidity in the Design and Directed Evolution of Artificial Metallo-Beta-Lactamases. *J. Am. Chem. Soc.* **2017**, 139 (46), 16772-16779.
- (207) Studer, S.; Hansen, D. A.; Pianowski, Z. L.; Mittl, P. R. E.; Debon, A.; Guffy, S. L.; Der, B. S.; Kuhlman, B.; Hilvert, D. Evolution of a Highly Active and Enantiospecific Metalloenzyme from Short Peptides. *Science* **2018**, 362 (6420), 1285-1288.
- (208) Radford, R. J.; Tezcan, F. A. A Superprotein Triangle Driven by Nickel(II) Coordination: Exploiting Non-Natural Metal Ligands in Protein Self-Assembly. *J. Am. Chem. Soc.* **2009**, 131 (26), 9136-9137.
- (209) Burazerovic, S.; Gradinaru, J.; Pierron, J.; Ward, T. R. Hierarchical Self-Assembly of One-Dimensional Streptavidin Bundles as a Collagen Mimetic for the Biominerization of Calcite. *Angew. Chem. Int. Ed. Engl.* **2007**, 46 (29), 5510-5514.
- (210) Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A. In Vitro and Cellular Self-Assembly of a Zn-Binding Protein Cryptand Via Templated Disulfide Bonds. *J. Am. Chem. Soc.* **2013**, 135 (32), 12013-12022.
- (211) Boyken, S. E.; Chen, Z.; Groves, B.; Langan, R. A.; Oberdorfer, G.; Ford, A.; Gilmore, J. M.; Xu, C.; DiMaio, F.; Pereira, J. H. et al. De Novo Design of Protein Homo-Oligomers with Modular Hydrogen-Bond Network-Mediated Specificity. *Science* **2016**, 352 (6286), 680-687.
- (212) Li, W.; Wang, J.; Zhang, J.; Wang, W. Molecular Simulations of Metal-Coupled Protein Folding. *Curr. Opin. Struct. Biol.* **2015**, 30, 25-31.
- (213) Der, B. S.; Jha, R. K.; Lewis, S. M.; Thompson, P. M.; Guntas, G.; Kuhlman, B. Combined Computational Design of a Zinc-Binding Site and a Protein-Protein Interaction: One Open Zinc Coordination Site Was Not a Robust Hotspot for De Novo Ubiquitin Binding. *Proteins* **2013**, 81 (7), 1245-1255.
- (214) Singh, R.; Whitesides, G. M. *Thiol-Disulfide Interchange In The Chemistry of Sulfur-Containing Functional Groups*; Patai, S., Ed.; J. Wiley and Sons, Ltd.: London, 1993; Vol. Supplement S.

- (215) Fass, D. Disulfide Bonding in Protein Biophysics. *Annu. Rev. Biophys.* **2012**, *41* (1), 63-79.
- (216) Liu, H.; May, K. Disulfide Bond Structures of IgG Molecules: Structural Variations, Chemical Modifications and Possible Impacts to Stability and Biological Function. *mAbs* **2012**, *4* (1), 17-23.
- (217) Pajvani, U. B.; Hawkins, M.; Combs, T. P.; Rajala, M. W.; Doepper, T.; Berger, J. P.; Wagner, J. A.; Wu, M.; Knopps, A.; Xiang, A. H. et al. Complex Distribution, Not Absolute Amount of Adiponectin, Correlates with Thiazolidinedione-Mediated Improvement in Insulin Sensitivity. *J. Biol. Chem.* **2004**, *279* (13), 12152-12162.
- (218) DiChiara, A. S.; Li, R. C.; Suen, P. H.; Hosseini, A. S.; Taylor, R. J.; Weickhardt, A. F.; Malhotra, D.; McCaslin, D. R.; Shoulders, M. D. A Cysteine-Based Molecular Code Informs Collagen C-Propeptide Assembly. *Nat. Commun.* **2018**, *9* (1), 4206.
- (219) Patel, S. D.; Rajala, M. W.; Rossetti, L.; Scherer, P. E.; Shapiro, L. Disulfide-Dependent Multimeric Assembly of Resistin Family Hormones. *Science* **2004**, *304* (5674), 1154-1158.
- (220) Ponnuswamy, N.; Cougnon, F. B. L.; Clough, J. M.; Pantoş, G. D.; Sanders, J. K. M. Discovery of an Organic Trefoil Knot. *Science* **2012**, *338* (6108), 783-785.
- (221) Corbett, P. T.; LeClaire, J.; Vial, L.; West, K. R.; Wietor, J. L.; Sanders, J. K.; Otto, S. Dynamic Combinatorial Chemistry. *Chem. Rev.* **2006**, *106* (9), 3652-3711.
- (222) Cougnon, F. B. L.; Sanders, J. K. M. Evolution of Dynamic Combinatorial Chemistry. *Acc. Chem. Res.* **2012**, *45* (12), 2211-2221.
- (223) Banatao, D. R.; Cascio, D.; Crowley, C. S.; Fleissner, M. R.; Tienson, H. L.; Yeates, T. O. An Approach to Crystallizing Proteins by Synthetic Symmetrization. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (44), 16230-16235.
- (224) Ballister, E. R.; Lai, A. H.; Zuckermann, R. N.; Cheng, Y.; Mougous, J. D. In Vitro Self-Assembly of Tailorable Nanotubes from a Simple Protein Building Block. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (10), 3733-3738.
- (225) Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. S.; Tezcan, F. A. Self-Assembly of Coherently Dynamic, Auxetic, Two-Dimensional Protein Crystals. *Nature* **2016**, *533* (7603), 369-373.
- (226) Alberstein, R.; Suzuki, Y.; Paesani, F.; Tezcan, F. A. Engineering the Entropy-Driven Free-Energy Landscape of a Dynamic Nanoporous Protein Assembly. *Nat. Chem.* **2018**, *10* (7), 732-739.
- (227) Zhang, S.; Alberstein, R. G.; De Yoreo, J. J.; Tezcan, F. A. Assembly of a Patchy Protein into Variable 2d Lattices Via Tunable Multiscale Interactions. *Nat. Commun.* **2020**, *11* (1), 3770.
- (228) Zhou, K.; Chen, H.; Zhang, S.; Wang, Y.; Zhao, G. Disulfide-Mediated Reversible Two-Dimensional Self-Assembly of Protein Nanocages. *Chem. Commun.* **2019**, *55* (52), 7510-7513.
- (229) Phillips, J. J.; Millership, C.; Main, E. R. Fibrous Nanostructures from the Self-Assembly of Designed Repeat Protein Modules. *Angew. Chem. Int. Ed. Engl.* **2012**, *51* (52), 13132-13135.
- (230) Zhao, L.; Zou, H.; Zhang, H.; Sun, H.; Wang, T.; Pan, T.; Li, X.; Bai, Y.; Qiao, S.; Luo, Q. et al. Enzyme-Triggered Defined Protein Nanoarrays: Efficient Light-Harvesting Systems to Mimic Chloroplasts. *ACS Nano* **2017**, *11* (1), 938-945.
- (231) Miao, L.; Fan, Q.; Zhao, L.; Qiao, Q.; Zhang, X.; Hou, C.; Xu, J.; Luo, Q.; Liu, J. The Construction of Functional Protein Nanotubes by Small Molecule-Induced Self-Assembly of Cricoid Proteins. *Chem. Commun.* **2016**, *52* (21), 4092-4095.
- (232) Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A. De Novo Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. *J. Am. Chem. Soc.* **2016**, *138* (40), 13163-13166.

- (233) Churchfield, L. A.; Alberstein, R. G.; Williamson, L. M.; Tezcan, F. A. Determining the Structural and Energetic Basis of Allostery in a De Novo Designed Metalloprotein Assembly. *J. Am. Chem. Soc.* **2018**, 140 (31), 10043-10053.
- (234) Chiou, N. R.; Epstein, A. J. Polyaniline Nanofibers Prepared by Dilute Polymerization. *Adv. Mater.* **2005**, 17 (13), 1679-1683.
- (235) Grill, L.; Dyer, M.; Laffermann, L.; Persson, M.; Peters, M. V.; Hecht, S. Nano-Architectures by Covalent Assembly of Molecular Building Blocks. *Nat Nanotechnol* **2007**, 2 (11), 687-691.
- (236) El-Kaderi, H. M.; Hunt, J. R.; Mendoza-Cortes, J. L.; Cote, A. P.; Taylor, R. E.; O'Keeffe, M.; Yaghi, O. M. Designed Synthesis of 3d Covalent Organic Frameworks. *Science* **2007**, 316 (5822), 268-272.
- (237) Waller, P. J.; Gandara, F.; Yaghi, O. M. Chemistry of Covalent Organic Frameworks. *Acc. Chem. Res.* **2015**, 48 (12), 3053-3063.
- (238) Rebek Jr, J. Host-Guest Chemistry of Calixarene Capsules. *Chem. Commun.* **2000**, (8), 637-643.
- (239) Chen, G.; Jiang, M. Cyclodextrin-Based Inclusion Complexation Bridging Supramolecular Chemistry and Macromolecular Self-Assembly. *Chem. Soc. Rev.* **2011**, 40 (5), 2254-2266.
- (240) Masson, E.; Ling, X. X.; Joseph, R.; Kyeremeh-Mensah, L.; Lu, X. Y. Cucurbituril Chemistry: A Tale of Supramolecular Success. *RSC Adv.* **2012**, 2 (4), 1213-1247.
- (241) Liu, Z.; Nalluri, S. K. M.; Stoddart, J. F. Surveying Macrocyclic Chemistry: From Flexible Crown Ethers to Rigid Cyclophanes. *Chem. Soc. Rev.* **2017**, 46 (9), 2459-2478.
- (242) Cram, D. J.; Cram, J. M. Host-Guest Chemistry: Complexes between Organic Compounds Simulate the Substrate Selectivity of Enzymes. *Science* **1974**, 183 (4127), 803-809.
- (243) Cram, D. The Design of Molecular Hosts, Guests, and Their Complexes. *Science* **1988**, 240 (4853), 760-767.
- (244) Lehn, J. M. Supramolecular Chemistry: Receptors, Catalysts, and Carriers. *Science* **1985**, 227 (4689), 849-856.
- (245) Qu, D. H.; Wang, Q. C.; Zhang, Q. W.; Ma, X.; Tian, H. Photoresponsive Host-Guest Functional Systems. *Chem. Rev.* **2015**, 115 (15), 7543-7588.
- (246) Yang, Y. W.; Sun, Y. L.; Song, N. Switchable Host-Guest Systems on Surfaces. *Acc. Chem. Res.* **2014**, 47 (7), 1950-1960.
- (247) Sasmal, R.; Das Saha, N.; Schueder, F.; Joshi, D.; Sheeba, V.; Jungmann, R.; Agasti, S. S. Dynamic Host-Guest Interaction Enables Autonomous Single Molecule Blinking and Super-Resolution Imaging. *Chem. Commun.* **2019**, 55 (96), 14430-14433.
- (248) Ooi, H. W.; Kocken, J. M. M.; Morgan, F. L. C.; Malheiro, A.; Zoetebier, B.; Karperien, M.; Wieringa, P. A.; Dijkstra, P. J.; Moroni, L.; Baker, M. B. Multivalency Enables Dynamic Supramolecular Host-Guest Hydrogel Formation. *Biomacromolecules* **2020**, 21 (6), 2208-2217.
- (249) Appel, E. A.; Biedermann, F.; Hoogland, D.; Del Barrio, J.; Driscoll, M. D.; Hay, S.; Wales, D. J.; Scherman, O. A. Decoupled Associative and Dissociative Processes in Strong yet Highly Dynamic Host-Guest Complexes. *J. Am. Chem. Soc.* **2017**, 139 (37), 12985-12993.
- (250) Kitagishi, H.; Kakikura, Y.; Yamaguchi, H.; Oohora, K.; Harada, A.; Hayashi, T. Self-Assembly of One- and Two-Dimensional Hemoprotein Systems by Polymerization through Heme-Heme Pocket Interactions. *Angew. Chem. Int. Ed. Engl.* **2009**, 48 (7), 1271-1274.
- (251) Oohora, K.; Onoda, A.; Hayashi, T. Supramolecular Assembling Systems Formed by Heme-Heme Pocket Interactions in Hemoproteins. *Chem. Commun.* **2012**, 48 (96), 11714-11726.
- (252) Mori, Y.; Minamihata, K.; Abe, H.; Goto, M.; Kamiya, N. Protein Assemblies by Site-Specific Avidin-Biotin Interactions. *Org. Biomol. Chem.* **2011**, 9 (16), 5641-5644.
- (253) Minamihata, K.; Goto, M.; Kamiya, N. Control of a Tyrosyl Radical Mediated Protein Cross-Linking Reaction by Electrostatic Interaction. *Bioconjug. Chem.* **2012**, 23 (8), 1600-1609.

- (254) Fairhead, M.; Veggiani, G.; Lever, M.; Yan, J.; Mesner, D.; Robinson, C. V.; Dushek, O.; van der Merwe, P. A.; Howarth, M. Spyavidin Hubs Enable Precise and Ultrastable Orthogonal Nanoassembly. *J. Am. Chem. Soc.* **2014**, 136 (35), 12355-12363.
- (255) Kurppa, K.; Hytonen, V. P.; Nakari-Setala, T.; Kulomaa, M. S.; Linder, M. B. Molecular Engineering of Avidin and Hydrophobin for Functional Self-Assembling Interfaces. *Colloids Surf. B* **2014**, 120, 102-109.
- (256) Singh, S.; Kluger, R. Self-Assembly of a Functional Triple Protein: Hemoglobin-Avidin-Hemoglobin Via Biotin-Avidin Interactions. *Biochemistry* **2016**, 55 (20), 2875-2882.
- (257) Guo, D.-S.; Uzunova, V. D.; Assaf, K. I.; Lazar, A. I.; Liu, Y.; Nau, W. M. Inclusion of Neutral Guests by Water-Soluble Macroyclic Hosts – a Comparative Thermodynamic Investigation with Cyclodextrins, Calixarenes and Cucurbiturils. *Supramol. Chem.* **2015**, 28 (5-6), 384-395.
- (258) Barrow, S. J.; Kasera, S.; Rowland, M. J.; del Barrio, J.; Scherman, O. A. Cucurbituril-Based Molecular Recognition. *Chem. Rev.* **2015**, 115 (22), 12320-12406.
- (259) Zhang, L.; Wu, Y.; Brunsved, L. A Synthetic Supramolecular Construct Modulating Protein Assembly in Cells. *Angew. Chem. Int. Ed. Engl.* **2007**, 46 (11), 1798-1802.
- (260) Zhang, W.; Luo, Q.; Miao, L.; Hou, C.; Bai, Y.; Dong, Z.; Xu, J.; Liu, J. Self-Assembly of Glutathione S-Transferase into Nanowires. *Nanoscale* **2012**, 4 (19), 5847-5851.
- (261) Uhlenheuer, D. A.; Young, J. F.; Nguyen, H. D.; Scheepstra, M.; Brunsved, L. Cucurbit[8]Uril Induced Heterodimerization of Methylviologen and Naphthalene Functionalized Proteins. *Chem. Commun.* **2011**, 47 (24), 6798-6800.
- (262) Hou, C.; Huang, Z.; Fang, Y.; Liu, J. Construction of Protein Assemblies by Host-Guest Interactions with Cucurbiturils. *Org. Biomol. Chem.* **2017**, 15 (20), 4272-4281.
- (263) Wang, R.; Qiao, S.; Zhao, L.; Hou, C.; Li, X.; Liu, Y.; Luo, Q.; Xu, J.; Li, H.; Liu, J. Dynamic Protein Self-Assembly Driven by Host-Guest Chemistry and the Folding-Unfolding Feature of a Mutually Exclusive Protein. *Chem. Commun.* **2017**, 53 (76), 10532-10535.
- (264) Si, C.; Li, J.; Luo, Q.; Hou, C.; Pan, T.; Li, H.; Liu, J. An Ion Signal Responsive Dynamic Protein Nano-Spring Constructed by High Ordered Host-Guest Recognition. *Chem. Commun.* **2016**, 52 (14), 2924-2927.
- (265) Dang, D. T.; Nguyen, H. D.; Merkx, M.; Brunsved, L. Supramolecular Control of Enzyme Activity through Cucurbit[8]Uril-Mediated Dimerization. *Angew. Chem. Int. Ed. Engl.* **2013**, 52 (10), 2915-2919.
- (266) McGovern, R. E.; Fernandes, H.; Khan, A. R.; Power, N. P.; Crowley, P. B. Protein Camouflage in Cytochrome C-Calixarene Complexes. *Nat. Chem.* **2012**, 4 (7), 527-533.
- (267) Rennie, M. L.; Doolan, A. M.; Raston, C. L.; Crowley, P. B. Protein Dimerization on a Phosphonated Calix[6]Arene Disc. *Angew. Chem. Int. Ed. Engl.* **2017**, 56 (20), 5517-5521.
- (268) Ramberg, K. O.; Engilberge, S.; Skorek, T.; Crowley, P. B. Facile Fabrication of Protein-Macrocyclic Frameworks. *J. Am. Chem. Soc.* **2021**, 143 (4), 1896-1907.
- (269) Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F. A Simple Calixarene Recognizes Post-Translationally Methylated Lysine. *ChemBioChem* **2010**, 11 (1), 63-66.
- (270) Yang, G.; Zhang, X.; Kochovski, Z.; Zhang, Y.; Dai, B.; Sakai, F.; Jiang, L.; Lu, Y.; Ballauff, M.; Li, X. et al. Precise and Reversible Protein-Microtubule-Like Structure with Helicity Driven by Dual Supramolecular Interactions. *J. Am. Chem. Soc.* **2016**, 138 (6), 1932-1937.
- (271) Kostiainen, M. A.; Hiekkataipale, P.; de la Torre, J. A.; Nolte, R. J. M.; Cornelissen, J. J. L. M. Electrostatic Self-Assembly of Virus-Polymer Complexes. *J. Mater. Chem.* **2011**, 21 (7), 2112-2117.
- (272) Kostiainen, M. A.; Hiekkataipale, P.; Laiho, A.; Lemieux, V.; Seitsonen, J.; Ruokolainen, J.; Ceci, P. Electrostatic Assembly of Binary Nanoparticle Superlattices Using Protein Cages. *Nat Nanotechnol* **2013**, 8 (1), 52-56.

- (273) Yang, R.; Chen, L.; Yang, S.; Lv, C.; Leng, X.; Zhao, G. 2d Square Arrays of Protein Nanocages through Channel-Directed Electrostatic Interactions with Poly(Alpha, L-Lysine). *Chem. Commun.* **2014**, 50 (22), 2879-2882.
- (274) Korpi, A.; Ma, C.; Liu, K.; Nonappa; Herrmann, A.; Ikkala, O.; Kostiainen, M. A. Self-Assembly of Electrostatic Cocrystals from Supercharged Fusion Peptides and Protein Cages. *ACS Macro Lett.* **2018**, 7 (3), 318-323.
- (275) Edwardson, T. G. W.; Mori, T.; Hilvert, D. Rational Engineering of a Designed Protein Cage for Sirna Delivery. *J. Am. Chem. Soc.* **2018**, 140 (33), 10439-10442.
- (276) Simon, A. J.; Zhou, Y.; Ramasubramani, V.; Glaser, J.; Pothukuchi, A.; Gollihar, J.; Gerberich, J. C.; Leggere, J. C.; Morrow, B. R.; Jung, C. et al. Supercharging Enables Organized Assembly of Synthetic Biomolecules. *Nat. Chem.* **2019**, 11 (3), 204-212.
- (277) Kunzle, M.; Eckert, T.; Beck, T. Binary Protein Crystals for the Assembly of Inorganic Nanoparticle Superlattices. *J. Am. Chem. Soc.* **2016**, 138 (39), 12731-12734.
- (278) Kostiainen, M. A.; Kasyutich, O.; Cornelissen, J. J.; Nolte, R. J. Self-Assembly and Optically Triggered Disassembly of Hierarchical Dendron-Virus Complexes. *Nat. Chem.* **2010**, 2 (5), 394-399.
- (279) Liljestrom, V.; Mikkila, J.; Kostiainen, M. A. Self-Assembly and Modular Functionalization of Three-Dimensional Crystals from Oppositely Charged Proteins. *Nat. Commun.* **2014**, 5, 4445.
- (280) Beck, T.; Tetter, S.; Kunzle, M.; Hilvert, D. Construction of Matryoshka-Type Structures from Supercharged Protein Nanocages. *Angew. Chem. Int. Ed. Engl.* **2015**, 54 (3), 937-940.
- (281) Sun, H.; Miao, L.; Li, J.; Fu, S.; Guo, A.; Si, C.; Dong, Z.; Luo, Q.; Yu, S.; Xu, J. et al. Self-Assembly of Cricoid Proteins Induced by "Soft Nanoparticles": An Approach to Design Multienzyme Cooperative Antioxidative Systems. *ACS Nano* **2015**, 9 (5), 5461-5469.
- (282) Wimley, W. C.; Gawrisch, K.; Creamer, T. P.; White, S. H. Direct Measurement of Salt-Bridge Solvation Energies Using a Peptide Model System: Implications for Protein Stability. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, 93 (7), 2985-2990.
- (283) Wilson, D. N.; Doudna Cate, J. H. The Structure and Function of the Eukaryotic Ribosome. *Cold Spring Harb. Perspect. Biol.* **2012**, 4 (5), a011536.
- (284) McGinty, R. K.; Tan, S. Nucleosome Structure and Function. *Chem. Rev.* **2015**, 115 (6), 2255-2573.
- (285) Saveleva, M. S.; Eftekhari, K.; Abalymov, A.; Douglas, T. E. L.; Volodkin, D.; Parakhonskiy, B. V.; Skirtach, A. G. Hierarchy of Hybrid Materials-the Place of Inorganics-in-Organics in It, Their Composition and Applications. *Front. Chem.* **2019**, 7 (179), 179.
- (286) Descalzo, A. B.; Martinez-Manez, R.; Sancenon, F.; Hoffmann, K.; Rurack, K. The Supramolecular Chemistry of Organic-Inorganic Hybrid Materials. *Angew. Chem. Int. Ed. Engl.* **2006**, 45 (36), 5924-5948.
- (287) Li, C.; Iscen, A.; Sai, H.; Sato, K.; Sather, N. A.; Chin, S. M.; Alvarez, Z.; Palmer, L. C.; Schatz, G. C.; Stupp, S. I. Supramolecular-Covalent Hybrid Polymers for Light-Activated Mechanical Actuation. *Nat. Mater.* **2020**, 19 (8), 900-909.
- (288) Chen, Y.; Phipps, M. L.; Werner, J. H.; Chakraborty, S.; Martinez, J. S. DNA Templatized Metal Nanoclusters: From Emergent Properties to Unique Applications. *Acc. Chem. Res.* **2018**, 51 (11), 2756-2763.
- (289) Taylor-Pashow, K. M.; Della Rocca, J.; Huxford, R. C.; Lin, W. Hybrid Nanomaterials for Biomedical Applications. *Chem. Commun.* **2010**, 46 (32), 5832-5849.
- (290) McMillan, J. R.; Brodin, J. D.; Millan, J. A.; Lee, B.; Olvera de la Cruz, M.; Mirkin, C. A. Modulating Nanoparticle Superlattice Structure Using Proteins with Tunable Bond Distributions. *J. Am. Chem. Soc.* **2017**, 139 (5), 1754-1757.

- (291) Miao, L.; Han, J.; Zhang, H.; Zhao, L.; Si, C.; Zhang, X.; Hou, C.; Luo, Q.; Xu, J.; Liu, J. Quantum-Dot-Induced Self-Assembly of Cricoid Protein for Light Harvesting. *ACS Nano* **2014**, 8 (4), 3743-3751.
- (292) Sim, S.; Miyajima, D.; Niwa, T.; Taguchi, H.; Aida, T. Tailoring Micrometer-Long High-Integrity 1d Array of Superparamagnetic Nanoparticles in a Nanotubular Protein Jacket and Its Lateral Magnetic Assembling Behavior. *J. Am. Chem. Soc.* **2015**, 137 (14), 4658-4661.
- (293) McCoy, K.; Uchida, M.; Lee, B.; Douglas, T. Templatized Assembly of a Functional Ordered Protein Macromolecular Framework from P22 Virus-Like Particles. *ACS Nano* **2018**, 12 (4), 3541-3550.
- (294) Praetorius, F.; Dietz, H. Self-Assembly of Genetically Encoded DNA-Protein Hybrid Nanoscale Shapes. *Science* **2017**, 355 (6331), eaam5488.
- (295) Zhang, L.; Bailey, J. B.; Subramanian, R. H.; Groisman, A.; Tezcan, F. A. Hyperexpandable, Self-Healing Macromolecular Crystals with Integrated Polymer Networks. *Nature* **2018**, 557 (7703), 86-91.
- (296) Liljestrom, V.; Ora, A.; Hassinen, J.; Rekola, H. T.; Nonappa; Heilala, M.; Hynninen, V.; Joensuu, J. J.; Ras, R. H. A.; Torma, P. et al. Cooperative Colloidal Self-Assembly of Metal-Protein Superlattice Wires. *Nat. Commun.* **2017**, 8 (1), 671.
- (297) Brodin, J. D.; Auyeung, E.; Mirkin, C. A. DNA-Mediated Engineering of Multicomponent Enzyme Crystals. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, 112 (15), 4564-4569.
- (298) Subramanian, R. H.; Smith, S. J.; Alberstein, R. G.; Bailey, J. B.; Zhang, L.; Cardone, G.; Suominen, L.; Chami, M.; Stahlberg, H.; Baker, T. S. et al. Self-Assembly of a Designed Nucleoprotein Architecture through Multimodal Interactions. *ACS Cent. Sci.* **2018**, 4 (11), 1578-1586.
- (299) Mou, Y.; Yu, J. Y.; Wannier, T. M.; Guo, C. L.; Mayo, S. L. Computational Design of Co-Assembling Protein-DNA Nanowires. *Nature* **2015**, 525 (7568), 230-233.
- (300) Park, J. I.; Nguyen, T. D.; de Queiros Silveira, G.; Bahng, J. H.; Srivastava, S.; Zhao, G.; Sun, K.; Zhang, P.; Glotzer, S. C.; Kotov, N. A. Terminal Supraparticle Assemblies from Similarly Charged Protein Molecules and Nanoparticles. *Nat. Commun.* **2014**, 5, 3593.
- (301) Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; DeGrado, W. F. Computational Design of Virus-Like Protein Assemblies on Carbon Nanotube Surfaces. *Science* **2011**, 332 (6033), 1071-1076.
- (302) Chen, X.; Wang, Y.; Wang, P. Peptide-Induced Affinity Binding of Carbonic Anhydrase to Carbon Nanotubes. *Langmuir* **2015**, 31 (1), 397-403.
- (303) Pinheiro, A. V.; Han, D.; Shih, W. M.; Yan, H. Challenges and Opportunities for Structural DNA Nanotechnology. *Nat Nanotechnol* **2011**, 6 (12), 763-772.
- (304) Linko, V.; Dietz, H. The Enabled State of DNA Nanotechnology. *Curr. Opin. Biotechnol.* **2013**, 24 (4), 555-561.
- (305) Hong, F.; Zhang, F.; Liu, Y.; Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.* **2017**, 117 (20), 12584-12640.
- (306) Seeman, N. C.; Sleiman, H. F. DNA Nanotechnology. *Nat. Rev. Mater.* **2017**, 3 (1), 17068.
- (307) Strable, E.; Johnson, J. E.; Finn, M. G. Natural Nanochemical Building Blocks: Icosahedral Virus Particles Organized by Attached Oligonucleotides. *Nano Lett.* **2004**, 4 (8), 1385-1389.
- (308) McMillan, J. R.; Hayes, O. G.; Winegar, P. H.; Mirkin, C. A. Protein Materials Engineering with DNA. *Acc. Chem. Res.* **2019**, 52 (7), 1939-1948.
- (309) Mikkila, J.; Anaya-Plaza, E.; Liljestrom, V.; Caston, J. R.; Torres, T.; Escosura Ade, L.; Kostiainen, M. A. Hierarchical Organization of Organic Dyes and Protein Cages into Photoactive Crystals. *ACS Nano* **2016**, 10 (1), 1565-1571.
- (310) Marianayagam, N. J.; Sunde, M.; Matthews, J. M. The Power of Two: Protein Dimerization in Biology. *Trends Biochem. Sci.* **2004**, 29 (11), 618-625.

- (311) Voss, S.; Klewer, L.; Wu, Y. W. Chemically Induced Dimerization: Reversible and Spatiotemporal Control of Protein Function in Cells. *Curr. Opin. Chem. Biol.* **2015**, 28, 194-201.
- (312) Milroy, L. G.; Grossmann, T. N.; Hennig, S.; Brunsved, L.; Ottmann, C. Modulators of Protein-Protein Interactions. *Chem. Rev.* **2014**, 114 (9), 4695-4748.
- (313) Bennett, M. J.; Choe, S.; Eisenberg, D. Domain Swapping: Entangling Alliances between Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, 91 (8), 3127-3131.
- (314) Liu, Y.; Eisenberg, D. 3d Domain Swapping: As Domains Continue to Swap. *Protein Sci.* **2002**, 11 (6), 1285-1299.
- (315) Mascarenhas, N. M.; Gosavi, S. Understanding Protein Domain-Swapping Using Structure-Based Models of Protein Folding. *Prog. Biophys. Mol. Biol.* **2017**, 128, 113-120.
- (316) Green, S. M.; Gittis, A. G.; Meeker, A. K.; Lattman, E. E. One-Step Evolution of a Dimer from a Monomeric Protein. *Nat. Struct. Biol.* **1995**, 2 (9), 746-751.
- (317) Murray, A. J.; Head, J. G.; Barker, J. J.; Brady, R. L. Engineering an Intertwined Form of Cd2 for Stability and Assembly. *Nat. Struct. Biol.* **1998**, 5 (9), 778-782.
- (318) Kuhlman, B.; O'Neill, J. W.; Kim, D. E.; Zhang, K. Y.; Baker, D. Conversion of Monomeric Protein L to an Obligate Dimer by Computational Protein Design. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98 (19), 10687-10691.
- (319) Assar, Z.; Nosseni, Z.; Wang, W.; Santos, E. M.; Kramer, K.; McCornack, C.; Vasileiou, C.; Borhan, B.; Geiger, J. H. Domain-Swapped Dimers of Intracellular Lipid-Binding Proteins: Evidence for Ordered Folding Intermediates. *Structure* **2016**, 24 (9), 1590-1598.
- (320) Ghanbarpour, A.; Pinger, C.; Esmatpour Salmani, R.; Assar, Z.; Santos, E. M.; Nosrati, M.; Pawlowski, K.; Spence, D.; Vasileiou, C.; Jin, X. et al. Engineering the HcrbpII Domain-Swapped Dimer into a New Class of Protein Switches. *J. Am. Chem. Soc.* **2019**, 141 (43), 17125-17132.
- (321) Ha, J. H.; Karchin, J. M.; Walker-Kopp, N.; Huang, L. S.; Berry, E. A.; Loh, S. N. Engineering Domain-Swapped Binding Interfaces by Mutually Exclusive Folding. *J. Mol. Biol.* **2012**, 416 (4), 495-502.
- (322) Ha, J. H.; Karchin, J. M.; Walker-Kopp, N.; Castaneda, C. A.; Loh, S. N. Engineered Domain Swapping as an on/Off Switch for Protein Function. *Chem. Biol.* **2015**, 22 (10), 1384-1393.
- (323) Karchin, J. M.; Ha, J. H.; Namitz, K. E.; Cosgrove, M. S.; Loh, S. N. Small Molecule-Induced Domain Swapping as a Mechanism for Controlling Protein Function and Assembly. *Sci. Rep.* **2017**, 7, 44388.
- (324) Nandwani, N.; Surana, P.; Negi, H.; Mascarenhas, N. M.; Udgaonkar, J. B.; Das, R.; Gosavi, S. A Five-Residue Motif for the Design of Domain Swapping in Proteins. *Nat. Commun.* **2019**, 10 (1), 452.
- (325) Correia, B. E.; Bates, J. T.; Loomis, R. J.; Baneyx, G.; Carrico, C.; Jardine, J. G.; Rupert, P.; Correnti, C.; Kalyuzhniy, O.; Vittal, V. et al. Proof of Principle for Epitope-Focused Vaccine Design. *Nature* **2014**, 507 (7491), 201-206.
- (326) Correia, B. E.; Ban, Y. E.; Holmes, M. A.; Xu, H.; Ellingson, K.; Kraft, Z.; Carrico, C.; Boni, E.; Sather, D. N.; Zenobia, C. et al. Computational Design of Epitope-Scaffolds Allows Induction of Antibodies Specific for a Poorly Immunogenic Hiv Vaccine Epitope. *Structure* **2010**, 18 (9), 1116-1126.
- (327) Ofek, G.; Guenaga, F. J.; Schief, W. R.; Skinner, J.; Baker, D.; Wyatt, R.; Kwong, P. D. Elicitation of Structure-Specific Antibodies by Epitope Scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107 (42), 17880-17887.
- (328) Azoitei, M. L.; Correia, B. E.; Ban, Y. E.; Carrico, C.; Kalyuzhniy, O.; Chen, L.; Schroeter, A.; Huang, P. S.; McLellan, J. S.; Kwong, P. D. et al. Computation-Guided Backbone Grafting of a Discontinuous Motif onto a Protein Scaffold. *Science* **2011**, 334 (6054), 373-376.

- (329) Azoitei, M. L.; Ban, Y. E.; Julien, J. P.; Bryson, S.; Schroeter, A.; Kalyuzhniy, O.; Porter, J. R.; Adachi, Y.; Baker, D.; Pai, E. F. et al. Computational Design of High-Affinity Epitope Scaffolds by Backbone Grafting of a Linear Epitope. *J. Mol. Biol.* **2012**, *415* (1), 175-192.
- (330) Procko, E.; Berguig, G. Y.; Shen, B. W.; Song, Y.; Frayo, S.; Convertine, A. J.; Margineantu, D.; Booth, G.; Correia, B. E.; Cheng, Y. et al. A Computationally Designed Inhibitor of an Epstein-Barr Viral Bcl-2 Protein Induces Apoptosis in Infected Cells. *Cell* **2014**, *157* (7), 1644-1656.
- (331) Berger, S.; Procko, E.; Margineantu, D.; Lee, E. F.; Shen, B. W.; Zelter, A.; Silva, D. A.; Chawla, K.; Herold, M. J.; Garnier, J. M. et al. Computationally Designed High Specificity Inhibitors Delineate the Roles of Bcl2 Family Proteins in Cancer. *eLife* **2016**, *5*, e20352.
- (332) Stranges, P. B.; Machius, M.; Miley, M. J.; Tripathy, A.; Kuhlman, B. Computational Design of a Symmetric Homodimer Using Beta-Strand Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (51), 20562-20567.
- (333) Huang, P. S.; Love, J. J.; Mayo, S. L. A De Novo Designed Protein Protein Interface. *Protein Sci.* **2007**, *16* (12), 2770-2774.
- (334) Jha, R. K.; Leaver-Fay, A.; Yin, S.; Wu, Y.; Butterfoss, G. L.; Szyperski, T.; Dokholyan, N. V.; Kuhlman, B. Computational Design of a Pak1 Binding Protein. *J. Mol. Biol.* **2010**, *400* (2), 257-270.
- (335) Dang, L. T.; Miao, Y.; Ha, A.; Yuki, K.; Park, K.; Janda, C. Y.; Jude, K. M.; Mohan, K.; Ha, N.; Vallon, M. et al. Receptor Subtype Discrimination Using Extensive Shape Complementary Designed Interfaces. *Nat. Struct. Mol. Biol.* **2019**, *26* (6), 407-414.
- (336) Foight, G. W.; Wang, Z.; Wei, C. T.; Jr Greisen, P.; Warner, K. M.; Cunningham-Bryant, D.; Park, K.; Brunette, T. J.; Sheffler, W.; Baker, D. et al. Multi-Input Chemical Control of Protein Dimerization for Programming Graded Cellular Responses. *Nat Biotechnol.* **2019**, *37* (10), 1209-1216.
- (337) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot Spots—a Review of the Protein–Protein Interface Determinant Amino-Acid Residues. *Proteins: Struct. Funct. Bioinf.* **2007**, *68* (4), 803-812.
- (338) Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E. M.; Wilson, I. A.; Baker, D. Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science* **2011**, *332* (6031), 816-821.
- (339) Procko, E.; Hedman, R.; Hamilton, K.; Seetharaman, J.; Fleishman, S. J.; Su, M.; Aramini, J.; Kornhaber, G.; Hunt, J. F.; Tong, L. et al. Computational Design of a Protein-Based Enzyme Inhibitor. *J. Mol. Biol.* **2013**, *425* (18), 3563-3575.
- (340) Radford, R. J.; Brodin, J. D.; Salgado, E. N.; Tezcan, F. A. Expanding the Utility of Proteins as Platforms for Coordination Chemistry. *Coordin. Chem. Rev.* **2011**, *255* (7-8), 790-803.
- (341) Churchfield, L. A.; George, A.; Tezcan, F. A. Repurposing Proteins for New Bioinorganic Functions. *Essays Biochem.* **2017**, *61* (2), 245-258.
- (342) Der, B. S.; Machius, M.; Miley, M. J.; Mills, J. L.; Szyperski, T.; Kuhlman, B. Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed Protein Homodimer. *J. Am. Chem. Soc.* **2012**, *134* (1), 375-385.
- (343) Der, B. S.; Edwards, D. R.; Kuhlman, B. Catalysis by a De Novo Zinc-Mediated Protein Interface: Implications for Natural Enzyme Evolution and Rational Enzyme Engineering. *Biochemistry* **2012**, *51* (18), 3933-3940.
- (344) Basler, S.; Studer, S.; Zou, Y.; Mori, T.; Ota, Y.; Camus, A.; Bunzel, H. A.; Helgeson, R. C.; Houk, K. N.; Jimenez-Oses, G. et al. Efficient Lewis Acid Catalysis of an Abiological Reaction in a De Novo Protein Scaffold. *Nat. Chem.* **2021**, *13* (3), 231-235.
- (345) Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M. Design of Functional Metalloproteins. *Nature* **2009**, *460* (7257), 855-862.
- (346) Cook, S. A.; Hill, E. A.; Borovik, A. S. Lessons from Nature: A Bio-Inspired Approach to Molecular Design. *Biochemistry* **2015**, *54* (27), 4167-8410.

- (347) Rittle, J.; Field, M. J.; Green, M. T.; Tezcan, F. A. An Efficient, Step-Economical Strategy for the Design of Functional Metalloproteins. *Nat. Chem.* **2019**, 11 (5), 434-441.
- (348) Vrancken, J. P. M.; Tame, J. R. H.; Voet, A. R. D. Development and Applications of Artificial Symmetrical Proteins. *Comput. Struct. Biotechnol. J.* **2020**, 18, 3959-3968.
- (349) Grueninger, D.; Treiber, N.; Ziegler, M. O.; Koetter, J. W.; Schulze, M. S.; Schulz, G. E. Designed Protein-Protein Association. *Science* **2008**, 319 (5860), 206-209.
- (350) Wolynes, P. G. Symmetry and the Energy Landscapes of Biomolecules. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, 93 (25), 14249-14255.
- (351) Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. Structural Evidence for Evolution of the Beta/Alpha Barrel Scaffold by Gene Duplication and Fusion. *Science* **2000**, 289 (5484), 1546-1550.
- (352) Fortenberry, C.; Bowman, E. A.; Proffitt, W.; Dorr, B.; Combs, S.; Harp, J.; Mizoue, L.; Meiler, J. Exploring Symmetry as an Avenue to the Computational Design of Large Protein Domains. *J. Am. Chem. Soc.* **2011**, 133 (45), 18026-18029.
- (353) Voet, A. R.; Noguchi, H.; Addy, C.; Simoncini, D.; Terada, D.; Unzai, S.; Park, S. Y.; Zhang, K. Y.; Tame, J. R. Computational Design of a Self-Assembling Symmetrical Beta-Propeller Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, 111 (42), 15102-15107.
- (354) Lin, Y. R.; Koga, N.; Vorobiev, S. M.; Baker, D. Cyclic Oligomer Design with De Novo Alphabeta-Proteins. *Protein Sci.* **2017**, 26 (11), 2187-2194.
- (355) Fallas, J. A.; Ueda, G.; Sheffler, W.; Nguyen, V.; McNamara, D. E.; Sankaran, B.; Pereira, J. H.; Parmeggiani, F.; Brunette, T. J.; Cascio, D. et al. Computational Design of Self-Assembling Cyclic Protein Homo-Oligomers. *Nat. Chem.* **2017**, 9 (4), 353-360.
- (356) Mohan, K.; Ueda, G.; Kim, A. R.; Jude, K. M.; Fallas, J. A.; Guo, Y.; Hafer, M.; Miao, Y.; Saxton, R. A.; Piehler, J. et al. Topological Control of Cytokine Receptor Signaling Induces Differential Effects in Hematopoiesis. *Science* **2019**, 364 (6442), eaav7532.
- (357) Boyken, S. E.; Benhaim, M. A.; Busch, F.; Jia, M.; Bick, M. J.; Choi, H.; Klima, J. C.; Chen, Z.; Walkey, C.; Mileant, A. et al. De Novo Design of Tunable, Ph-Driven Conformational Changes. *Science* **2019**, 364 (6441), 658-664.
- (358) Lu, P.; Min, D.; DiMaio, F.; Wei, K. Y.; Vahey, M. D.; Boyken, S. E.; Chen, Z.; Fallas, J. A.; Ueda, G.; Sheffler, W. et al. Accurate Computational Design of Multipass Transmembrane Proteins. *Science* **2018**, 359 (6379), 1042-1046.
- (359) Xu, C.; Lu, P.; Gamal El-Din, T. M.; Pei, X. Y.; Johnson, M. C.; Uyeda, A.; Bick, M. J.; Xu, Q.; Jiang, D.; Bai, H. et al. Computational Design of Transmembrane Pores. *Nature* **2020**, 585 (7823), 129-134.
- (360) Kobayashi, N.; Yanase, K.; Sato, T.; Unzai, S.; Hecht, M. H.; Arai, R. Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric De Novo Protein. *J. Am. Chem. Soc.* **2015**, 137 (35), 11285-11293.
- (361) Kobayashi, N.; Inano, K.; Sasahara, K.; Sato, T.; Miyazawa, K.; Fukuma, T.; Hecht, M. H.; Song, C.; Murata, K.; Arai, R. Self-Assembling Supramolecular Nanostructures Constructed from De Novo Extender Protein Nanobuilding Blocks. *ACS Synth. Biol.* **2018**, 7 (5), 1381-1394.
- (362) Kim, Y. E.; Kim, Y. N.; Kim, J. A.; Kim, H. M.; Jung, Y. Green Fluorescent Protein Nanopolymers as Monodisperse Supramolecular Assemblies of Functional Proteins with Defined Valency. *Nat. Commun.* **2015**, 6 (1), 7134.
- (363) Campos, L. A.; Sharma, R.; Alvira, S.; Ruiz, F. M.; Ibarra-Molero, B.; Sadqi, M.; Alfonso, C.; Rivas, G.; Sanchez-Ruiz, J. M.; Romero Garrido, A. et al. Engineering Protein Assemblies with Allosteric Control Via Monomer Fold-Switching. *Nat. Commun.* **2019**, 10 (1), 5703.
- (364) Tolbert, A. E.; Ervin, C. S.; Ruckthong, L.; Paul, T. J.; Jayasinghe-Arachchige, V. M.; Neupane, K. P.; Stuckey, J. A.; Prabhakar, R.; Pecoraro, V. L. Heteromeric Three-

- Stranded Coiled Coils Designed Using a Pb(II)(Cys)3 Template Mediated Strategy. *Nat. Chem.* **2020**, 12 (4), 405-411.
- (365) Mills, J. H.; Sheffler, W.; Ener, M. E.; Almhjell, P. J.; Oberdorfer, G.; Pereira, J. H.; Parmeggiani, F.; Sankaran, B.; Zwart, P. H.; Baker, D. Computational Design of a Homotrimeric Metalloprotein with a Trisbipyridyl Core. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113 (52), 15012-15017.
- (366) Salgado, E. N.; Lewis, R. A.; Faraone-Mennella, J.; Tezcan, F. A. Metal-Mediated Self-Assembly of Protein Superstructures: Influence of Secondary Interactions on Protein Oligomerization and Aggregation. *J. Am. Chem. Soc.* **2008**, 130 (19), 6082-6084.
- (367) Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A. Evolution of Metal Selectivity in Templated Protein Interfaces. *J. Am. Chem. Soc.* **2010**, 132 (25), 8610-8617.
- (368) Armstrong, R. N. Mechanistic Diversity in a Metalloenzyme Superfamily. *Biochemistry* **2000**, 39 (45), 13625-13632.
- (369) Bergdoll, M.; Eltis, L. D.; Cameron, A. D.; Dumas, P.; Bolin, J. T. All in the Family: Structural and Evolutionary Relationships among Three Modular Proteins with Diverse Functions and Variable Assembly. *Protein Sci.* **1998**, 7 (8), 1661-1670.
- (370) Volbeda, A.; Hol, W. G. Pseudo 2-Fold Symmetry in the Copper-Binding Domain of Arthropodan Haemocyanins. Possible Implications for the Evolution of Oxygen Transport Proteins. *J. Mol. Biol.* **1989**, 206 (3), 531-546.
- (371) Salgado, E. N.; Brodin, J. D.; To, M. M.; Tezcan, F. A. Templated Construction of a Zn-Selective Protein Dimerization Motif. *Inorg. Chem.* **2011**, 50 (13), 6323-6329.
- (372) Chaikuad, A.; Pilka, E. S.; De Riso, A.; von Delft, F.; Kavanagh, K. L.; Venien-Bryan, C.; Oppermann, U.; Yue, W. W. Structure of Human Aspartyl Aminopeptidase Complexed with Substrate Analogue: Insight into Catalytic Mechanism, Substrate Specificity and M18 Peptidase Family. *BMC Struct Biol* **2012**, 12, 14.
- (373) Douglas, T.; Young, M. Viruses: Making Friends with Old Foes. *Science* **2006**, 312 (5775), 873-875.
- (374) Young, M.; Willits, D.; Uchida, M.; Douglas, T. Plant Viruses as Biotemplates for Materials and Their Use in Nanotechnology. *Annu. Rev. Phytopathol.* **2008**, 46 (1), 361-384.
- (375) Wang, Z.; Gao, H.; Zhang, Y.; Liu, G.; Niu, G.; Chen, X. Functional Ferritin Nanoparticles for Biomedical Applications. *Front. Chem. Sci. Eng.* **2017**, 11 (4), 633-646.
- (376) Lai, Y. T.; Cascio, D.; Yeates, T. O. Structure of a 16-Nm Cage Designed by Using Protein Oligomers. *Science* **2012**, 336 (6085), 1129-1129.
- (377) Lai, Y. T.; Tsai, K. L.; Sawaya, M. R.; Asturias, F. J.; Yeates, T. O. Structure and Flexibility of Nanoscale Protein Cages Designed by Symmetric Self-Assembly. *J. Am. Chem. Soc.* **2013**, 135 (20), 7738-7743.
- (378) Lai, Y. T.; Reading, E.; Hura, G. L.; Tsai, K. L.; Laganowsky, A.; Asturias, F. J.; Tainer, J. A.; Robinson, C. V.; Yeates, T. O. Structure of a Designed Protein Cage That Self-Assembles into a Highly Porous Cube. *Nat. Chem.* **2014**, 6 (12), 1065-1071.
- (379) Lai, Y. T.; Hura, G. L.; Dyer, K. N.; Tang, H. Y.; Tainer, J. A.; Yeates, T. O. Designing and Defining Dynamic Protein Cage Nanoassemblies in Solution. *Sci. Adv.* **2016**, 2 (12), e1501855.
- (380) Patterson, D. P.; Desai, A. M.; Holl, M. M.; Marsh, E. N. Evaluation of a Symmetry-Based Strategy for Assembling Protein Complexes. *RSC Adv.* **2011**, 1 (6), 1004-1012.
- (381) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N. et al. Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* **2013**, 340 (6132), 595-599.
- (382) Woolfson, D. N. *The Design of Coiled-Coil Structures and Assemblies In Advances in Protein Chemistry*; Parry, D. A. D. and Squire, J. M., Eds.; Academic Press, 2005; Vol. 70.

- (383) Huang, P.-S.; Oberdorfer, G.; Xu, C.; Pei, X. Y.; Nannenga, B. L.; Rogers, J. M.; DiMaio, F.; Gonen, T.; Luisi, B.; Baker, D. High Thermodynamic Stability of Parametrically Designed Helical Bundles. *Science* **2014**, 346 (6208), 481-485.
- (384) Thomson, A. R.; Wood, C. W.; Burton, A. J.; Bartlett, G. J.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N. Computational Design of Water-Soluble A-Helical Barrels. *Science* **2014**, 346 (6208), 485-488.
- (385) Dawson, W. M.; Martin, F. J. O.; Rhys, G. G.; Shelley, K. L.; Brady, R. L.; Woolfson, D. N. Coiled Coils 9-to-5: Rational De Novo Design of A-Helical Barrels with Tunable Oligomeric States. *Chem. Sci.* **2021**, 12 (20), 6923-6928.
- (386) Gradišar, H.; Božič, S.; Doles, T.; Vengust, D.; Hafner-Bratkovič, I.; Mertelj, A.; Webb, B.; Šali, A.; Klavžar, S.; Jerala, R. Design of a Single-Chain Polypeptide Tetrahedron Assembled from Coiled-Coil Segments. *Nat. Chem. Biol.* **2013**, 9 (6), 362-366.
- (387) Patterson, D. P.; Su, M.; Franzmann, T. M.; Sciore, A.; Skiniotis, G.; Marsh, E. N. Characterization of a Highly Flexible Self-Assembling Protein System Designed to Form Nanocages. *Protein Sci.* **2014**, 23 (2), 190-199.
- (388) Cristie-David, A. S.; Koldewey, P.; Meinen, B. A.; Bardwell, J. C. A.; Marsh, E. N. G. Elaborating a Coiled-Coil-Assembled Octahedral Protein Cage with Additional Protein Domains. *Protein Sci.* **2018**, 27 (11), 1893-1900.
- (389) King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D. Accurate Design of Co-Assembling Multi-Component Protein Nanomaterials. *Nature* **2014**, 510 (7503), 103-108.
- (390) Cannon, K. A.; Park, R. U.; Boyken, S. E.; Nattermann, U.; Yi, S.; Baker, D.; King, N. P.; Yeates, T. O. Design and Structure of Two New Protein Cages Illustrate Successes and Ongoing Challenges in Protein Engineering. *Protein Sci.* **2020**, 29 (4), 919-929.
- (391) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P. et al. Accurate Design of Megadalton-Scale Two-Component Icosahedral Protein Complexes. *Science* **2016**, 353 (6297), 389-394.
- (392) Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P. S.; Ravichandran, R. et al. Design of a Hyperstable 60-Subunit Protein Dodecahedron. *Nature* **2016**, 535 (7610), 136-139.
- (393) Wargacki, A. J.; Worner, T. P.; van de Waterbeemd, M.; Ellis, D.; Heck, A. J. R.; King, N. P. Complete and Cooperative in Vitro Assembly of Computationally Designed Self-Assembling Protein Nanomaterials. *Nat. Commun.* **2021**, 12 (1), 883.
- (394) Divine, R.; Dang, H. V.; Ueda, G.; Fallas, J. A.; Vulovic, I.; Sheffler, W.; Saini, S.; Zhao, Y. T.; Raj, I. X.; Morawski, P. A. et al. Designed Proteins Assemble Antibodies into Modular Nanocages. *Science* **2021**, 372 (6537), eabd9994.
- (395) Ni, T. W.; Tezcan, F. A. Structural Characterization of a Microperoxidase inside a Metal-Directed Protein Cage. *Angew. Chem. Int. Ed. Engl.* **2010**, 49 (39), 7014-7018.
- (396) Miyamoto, T.; Kurabayashi, M.; Nagao, S.; Shomura, Y.; Higuchi, Y.; Hirota, S. Domain-Swapped Cytochrome Cb(562) Dimer and Its Nanocage Encapsulating a Zn-So₄ Cluster in the Internal Cavity. *Chem. Sci.* **2015**, 6 (12), 7336-7342.
- (397) Cristie-David, A. S.; Marsh, E. N. G. Metal-Dependent Assembly of a Protein Nano-Cage. *Protein Sci.* **2019**, 28 (9), 1620-1629.
- (398) Pearson, R. G. Hard and Soft Acids and Bases. *J. Am. Chem. Soc.* **1963**, 85 (22), 3533-3539.
- (399) Wong, G. B.; Kappel, M. J.; Raymond, K. N.; Matzanke, B.; Winkelmann, G. Coordination Chemistry of Microbial Iron Transport Compounds. 24. Characterization of Coprogen and Ferricrocin, Two Ferric Hydroxamate Siderophores. *J. Am. Chem. Soc.* **1983**, 105 (4), 810-815.
- (400) Crumbliss, A. L. Iron Bioavailability and the Coordination Chemistry of Hydroxamic Acids. *Coordin. Chem. Rev.* **1990**, 105, 155-179.

- (401) Huard, D. J.; Kane, K. M.; Tezcan, F. A. Re-Engineering Protein Interfaces Yields Copper-Inducible Ferritin Cage Assembly. *Nat. Chem. Biol.* **2013**, 9 (3), 169-176.
- (402) Zhang, S. L.; Zang, J. C.; Zhang, X. R.; Chen, H.; Mikami, B.; Zhao, G. H. "Silent" Amino Acid Residues at Key Subunit Interfaces Regulate the Geometry of Protein Nanocages. *ACS Nano* **2016**, 10 (11), 10382-10388.
- (403) Zhang, S.; Zang, J.; Wang, W.; Chen, H.; Zhang, X.; Wang, F.; Wang, H.; Zhao, G. Conversion of the Native 24-Mer Ferritin Nanocage into Its Non-Native 16-Mer Analogue by Insertion of Extra Amino Acid Residues. *Angew. Chem. Int. Ed. Engl.* **2016**, 55 (52), 16064-16070.
- (404) Wang, W. M.; Wang, L. L.; Chen, H.; Zang, J. C.; Zhao, X.; Zhao, G. H.; Wang, H. F. Selective Elimination of the Key Subunit Interfaces Facilitates Conversion of Native 24-Mer Protein Nanocage into 8-Mer Nanorings. *J. Am. Chem. Soc.* **2018**, 140 (43), 14078-14081.
- (405) Zang, J. C.; Chen, H.; Zhang, X. R.; Zhang, C. X.; Guo, J.; Du, M.; Zhao, G. H. Disulfide-Mediated Conversion of 8-Mer Bowl-Like Protein Architecture into Three Different Nanocages. *Nat. Commun.* **2019**, 10, 778.
- (406) Yu, Y.; Lutz, S. Circular Permutation: A Different Way to Engineer Enzyme Structure and Function. *Trends Biotechnol.* **2011**, 29 (1), 18-25.
- (407) Jorda, J.; Leibly, D. J.; Thompson, M. C.; Yeates, T. O. Structure of a Novel 13 Nm Dodecahedral Nanocage Assembled from a Redesigned Bacterial Microcompartment Shell Protein. *Chem. Commun.* **2016**, 52 (28), 5041-5044.
- (408) Azuma, Y.; Herger, M.; Hilvert, D. Diversification of Protein Cage Structure Using Circularly Permuted Subunits. *J. Am. Chem. Soc.* **2018**, 140 (2), 558-561.
- (409) Azuma, Y.; Edwardson, T. G. W.; Hilvert, D. Tailoring Lumazine Synthase Assemblies for Bionanotechnology. *Chem. Soc. Rev.* **2018**, 47 (10), 3543-3557.
- (410) Wagstaff, J.; Lowe, J. Prokaryotic Cytoskeletons: Protein Filaments Organizing Small Cells. *Nat. Rev. Microbiol.* **2018**, 16 (4), 187-201.
- (411) Fletcher, D. A.; Mullins, R. D. Cell Mechanics and the Cytoskeleton. *Nature* **2010**, 463 (7280), 485-492.
- (412) Knowles, T. P.; Vendruscolo, M.; Dobson, C. M. The Amyloid State and Its Association with Protein Misfolding Diseases. *Nat. Rev. Mol. Cell Biol.* **2014**, 15 (6), 384-396.
- (413) Hou, C.; Li, J.; Zhao, L.; Zhang, W.; Luo, Q.; Dong, Z.; Xu, J.; Liu, J. Construction of Protein Nanowires through Cucurbit[8]Uril-Based Highly Specific Host-Guest Interactions: An Approach to the Assembly of Functional Proteins. *Angew. Chem. Int. Ed. Engl.* **2013**, 52 (21), 5590-5593.
- (414) Kitagishi, H.; Oohora, K.; Yamaguchi, H.; Sato, H.; Matsuo, T.; Harada, A.; Hayashi, T. Supramolecular Hemoprotein Linear Assembly by Successive Interprotein Heme-Heme Pocket Interactions. *J. Am. Chem. Soc.* **2007**, 129 (34), 10326-10327.
- (415) Onoda, A.; Takahashi, A.; Oohora, K.; Onuma, Y.; Hayashi, T. Fibrous Supramolecular Hemoprotein Assemblies Connected with Synthetic Heme Dimer and Apohemoprotein Dimer. *Chem. Biodivers.* **2012**, 9 (9), 1684-1692.
- (416) Abe, S.; Pham, T. T.; Negishi, H.; Yamashita, K.; Hirata, K.; Ueno, T. Design of an in-Cell Protein Crystal for the Environmentally Responsive Construction of a Supramolecular Filament. *Angew. Chem. Int. Ed. Engl.* **2021**, 60 (22), 12341-12345.
- (417) Biswas, S.; Kinbara, K.; Oya, N.; Ishii, N.; Taguchi, H.; Aida, T. A Tubular Biocontainer: Metal Ion-Induced 1d Assembly of a Molecularly Engineered Chaperonin. *J. Am. Chem. Soc.* **2009**, 131 (22), 7556-7557.
- (418) Miranda, F. F.; Iwasaki, K.; Akashi, S.; Sumitomo, K.; Kobayashi, M.; Yamashita, I.; Tame, J. R.; Heddle, J. G. A Self-Assembled Protein Nanotube with High Aspect Ratio. *Small* **2009**, 5 (18), 2077-2084.

- (419) Brodin, J. D.; Smith, S. J.; Carr, J. R.; Tezcan, F. A. Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. *J. Am. Chem. Soc.* **2015**, 137 (33), 10468-10471.
- (420) Hyman, P.; Denyes, J. *Bacteriophages in Nanotechnology: History and Future* In *Bacteriophages*; Harper, D. R., Abedon, S. T., Burrowes, B. H. and McConville, M. L., Eds.; Springer: Cham, 2018.
- (421) Czapar, A. E.; Steinmetz, N. F. Plant Viruses and Bacteriophages for Drug Delivery in Medicine and Biotechnology. *Curr. Opin. Chem. Biol.* **2017**, 38, 108-116.
- (422) De Santis, E.; Ryadnov, M. G. Peptide Self-Assembly for Nanomaterials: The Old New Kid on the Block. *Chem. Soc. Rev.* **2015**, 44 (22), 8288-8300.
- (423) Wei, G.; Su, Z.; Reynolds, N. P.; Arosio, P.; Hamley, I. W.; Gazit, E.; Mezzenga, R. Self-Assembling Peptide and Protein Amyloids: From Structure to Tailored Function in Nanotechnology. *Chem. Soc. Rev.* **2017**, 46 (15), 4661-4708.
- (424) Smith, A. M.; Acquah, S. F.; Bone, N.; Kroto, H. W.; Ryadnov, M. G.; Stevens, M. S.; Walton, D. R.; Woolfson, D. N. Polar Assembly in a Designed Protein Fiber. *Angew. Chem. Int. Ed. Engl.* **2004**, 44 (2), 325-328.
- (425) O'Leary, L. E.; Fallas, J. A.; Bakota, E. L.; Kang, M. K.; Hartgerink, J. D. Multi-Hierarchical Self-Assembly of a Collagen Mimetic Peptide from Triple Helix to Nanofibre and Hydrogel. *Nat. Chem.* **2011**, 3 (10), 821-828.
- (426) Bowerman, C. J.; Nilsson, B. L. Self-Assembly of Amphipathic Beta-Sheet Peptides: Insights and Applications. *Biopolymers* **2012**, 98 (3), 169-184.
- (427) Egelman, E. H.; Xu, C.; DiMaio, F.; Magnotti, E.; Modlin, C.; Yu, X.; Wright, E.; Baker, D.; Conticello, V. P. Structural Plasticity of Helical Nanotubes Based on Coiled-Coil Assemblies. *Structure* **2015**, 23 (2), 280-289.
- (428) Burgess, N. C.; Sharp, T. H.; Thomas, F.; Wood, C. W.; Thomson, A. R.; Zaccai, N. R.; Brady, R. L.; Serpell, L. C.; Woolfson, D. N. Modular Design of Self-Assembling Peptide-Based Nanotubes. *J. Am. Chem. Soc.* **2015**, 137 (33), 10554-10562.
- (429) Xu, C.; Liu, R.; Mehta, A. K.; Guerrero-Ferreira, R. C.; Wright, E. R.; Dunin-Horkawicz, S.; Morris, K.; Serpell, L. C.; Zuo, X.; Wall, J. S. et al. Rational Design of Helical Nanotubes from Self-Assembly of Coiled-Coil Lock Washers. *J. Am. Chem. Soc.* **2013**, 135 (41), 15565-15578.
- (430) Lee, M. J.; Mantell, J.; Hodgson, L.; Alibhai, D.; Fletcher, J. M.; Brown, I. R.; Frank, S.; Xue, W.-F.; Verkade, P.; Woolfson, D. N. et al. Engineered Synthetic Scaffolds for Organizing Proteins within the Bacterial Cytoplasm. *Nat. Chem. Biol.* **2018**, 14 (2), 142-147.
- (431) McPartland, L.; Heller, D. M.; Eisenberg, D. S.; Hochschild, A.; Sawaya, M. R. Atomic Insights into the Genesis of Cellular Filaments by Globular Proteins. *Nat. Struct. Mol. Biol.* **2018**, 25 (8), 705-714.
- (432) Nagarkar, R. P.; Hule, R. A.; Pochan, D. J.; Schneider, J. P. Domain Swapping in Materials Design. *Biopolymers* **2010**, 94 (1), 141-155.
- (433) Kitagishi, H.; Oohora, K.; Hayashi, T. Thermodynamically Controlled Supramolecular Polymerization of Cytochrome B 562. *Biopolymers* **2009**, 91 (3), 194-200.
- (434) Oohora, K.; Onoda, A.; Kitagishi, H.; Yamaguchi, H.; Harada, A.; Hayashi, T. A Chemically-Controlled Supramolecular Protein Polymer Formed by a Myoglobin-Based Self-Assembly System. *Chem. Sci.* **2011**, 2 (6), 1033-1038.
- (435) Oohora, K.; Burazerovic, S.; Onoda, A.; Wilson, Y. M.; Ward, T. R.; Hayashi, T. Chemically Programmed Supramolecular Assembly of Hemoprotein and Streptavidin with Alternating Alignment. *Angew. Chem. Int. Ed. Engl.* **2012**, 51 (16), 3818-3821.
- (436) Nagano, S.; Banwell, E. F.; Iwasaki, K.; Michalak, M.; Palka, R.; Zhang, K. Y. J.; Voet, A. R. D.; Heddle, J. G. Understanding the Assembly of an Artificial Protein Nanotube. *Adv Mater Interfaces* **2016**, 3 (24), 1600846.

- (437) Nguyen, T. K.; Negishi, H.; Abe, S.; Ueno, T. Construction of Supramolecular Nanotubes from Protein Crystals. *Chem. Sci.* **2019**, *10* (4), 1046-1051.
- (438) Bai, Y.; Luo, Q.; Zhang, W.; Miao, L.; Xu, J.; Li, H.; Liu, J. Highly Ordered Protein Nanorings Designed by Accurate Control of Glutathione S-Transferase Self-Assembly. *J. Am. Chem. Soc.* **2013**, *135* (30), 10966-10969.
- (439) Sendai, T.; Biswas, S.; Aida, T. Photoreconfigurable Supramolecular Nanotube. *J. Am. Chem. Soc.* **2013**, *135* (31), 11509-11512.
- (440) Kashiwagi, D.; Sim, S.; Niwa, T.; Taguchi, H.; Aida, T. Protein Nanotube Selectively Cleavable with DNA: Supramolecular Polymerization of "DNA-Appended Molecular Chaperones". *J. Am. Chem. Soc.* **2018**, *140* (1), 26-29.
- (441) Weissman, J. S.; Hohl, C. M.; Kovalenko, O.; Kashi, Y.; Chen, S.; Braig, K.; Saibil, H. R.; Fenton, W. A.; Horwich, A. L. Mechanism of GroEL Action: Productive Release of Polypeptide from a Sequestered Position under GroES. *Cell* **1995**, *83* (4), 577-587.
- (442) Kashiwagi, D.; Shen, H. K.; Sim, S.; Sano, K.; Ishida, Y.; Kimura, A.; Niwa, T.; Taguchi, H.; Aida, T. Molecularly Engineered "Janus GroEL": Application to Supramolecular Copolymerization with a Higher Level of Sequence Control. *J. Am. Chem. Soc.* **2020**, *142* (31), 13310-13315.
- (443) Wang, W. X.; Dgany, O.; Wolf, S. G.; Levy, I.; Algom, R.; Pouny, Y.; Wolf, A.; Marton, I.; Altman, A.; Shoseyov, O. Aspen Sp1, an Exceptional Thermal, Protease and Detergent-Resistant Self-Assembled Nano-Particle. *Biotechnol. Bioeng.* **2006**, *95* (1), 161-168.
- (444) Heyman, A.; Levy, I.; Altman, A.; Shoseyov, O. Sp1 as a Novel Scaffold Building Block for Self-Assembly Nanofabrication of Submicron Enzymatic Structures. *Nano Lett.* **2007**, *7* (6), 1575-1579.
- (445) Medalsky, I.; Dgany, O.; Sowwan, M.; Cohen, H.; Yukashevskaya, A.; Wolf, S. G.; Wolf, A.; Koster, A.; Almog, O.; Marton, I. et al. Sp1 Protein-Based Nanostructures and Arrays. *Nano Lett.* **2008**, *8* (2), 473-477.
- (446) Sun, H.; Zhao, L.; Wang, T.; An, G.; Fu, S.; Li, X.; Deng, X.; Liu, J. Photocontrolled Reversible Morphology Conversion of Protein Nanowires Mediated by an Azobenzene-Cored Dendrimer. *Chem. Commun.* **2016**, *52* (35), 6001-6004.
- (447) Sun, H.; Zhang, X.; Miao, L.; Zhao, L.; Luo, Q.; Xu, J.; Liu, J. Micelle-Induced Self-Assembling Protein Nanowires: Versatile Supramolecular Scaffolds for Designing the Light-Harvesting System. *ACS Nano* **2016**, *10* (1), 421-428.
- (448) Schreiber, A.; Huber, M. C.; Colfen, H.; Schiller, S. M. Molecular Protein Adaptor with Genetically Encoded Interaction Sites Guiding the Hierarchical Assembly of Plasmonically Active Nanoparticle Architectures. *Nat. Commun.* **2015**, *6* (1), 6705.
- (449) Cate, J. H.; Yusupov, M. M.; Yusupova, G. Z.; Earnest, T. N.; Noller, H. F. X-Ray Crystal Structures of 70s Ribosome Functional Complexes. *Science* **1999**, *285* (5436), 2095-2104.
- (450) Lin, J.; Ly, H.; Hussain, A.; Abraham, M.; Pearl, S.; Tzfati, Y.; Parslow, T. G.; Blackburn, E. H. A Universal Telomerase RNA Core Structure Includes Structured Motifs Required for Binding the Telomerase Reverse Transcriptase Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (41), 14713-14718.
- (451) Eitoku, M.; Sato, L.; Senda, T.; Horikoshi, M. Histone Chaperones: 30 Years from Isolation to Elucidation of the Mechanisms of Nucleosome Assembly and Disassembly. *Cell Mol. Life Sci.* **2008**, *65* (3), 414-444.
- (452) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816-821.
- (453) Pum, D.; Toca-Herrera, J. L.; Sleytr, U. B. S-Layer Protein Self-Assembly. *Int. J. Mol. Sci.* **2013**, *14* (2), 2484-2501.

- (454) Hampp, N. Bacteriorhodopsin as a Photochromic Retinal Protein for Optical Memories. *Chem. Rev.* **2000**, 100 (5), 1755-1776.
- (455) Caspar, D. L.; Goodenough, D. A.; Makowski, L.; Phillips, W. C. Gap Junction Structures. I. Correlated Electron Microscopy and X-Ray Diffraction. *J. Cell Biol.* **1977**, 74 (2), 605-628.
- (456) Scheuring, S.; Ringler, P.; Borgnia, M.; Stahlberg, H.; Muller, D. J.; Agre, P.; Engel, A. High Resolution Afm Topographs of the Escherichia Coli Water Channel Aquaporin Z. *EMBO J.* **1999**, 18 (18), 4981-4987.
- (457) Nannenga, B. L.; Iadanza, M. G.; Vollmar, B. S.; Gonen, T. Overview of Electron Crystallography of Membrane Proteins: Crystallization and Screening Strategies Using Negative Stain Electron Microscopy. *Curr. Protoc. Protein Sci.* **2013**, Chapter 17 (1), 17.15.11-17.15.11.
- (458) Wukovitz, S. W.; Yeates, T. O. Why Protein Crystals Favour Some Space-Groups over Others. *Nat. Struct. Biol.* **1995**, 2 (12), 1062-1067.
- (459) Kim, T.; Park, J. Y.; Hwang, J.; Seo, G.; Kim, Y. Supramolecular Two-Dimensional Systems and Their Biological Applications. *Adv. Mater.* **2020**, 32 (51), e2002405.
- (460) Huang, G.; Mei, Y. Assembly and Self-Assembly of Nanomembrane Materials - from 2d to 3d. *Small* **2018**, 14 (14), e1703665.
- (461) Gopalakrishnan, S.; Xu, J.; Zhong, F.; Rotello, V. M. Strategies for Fabricating Protein Films for Biomaterials Applications. *Adv. Sustain. Syst.* **2021**, 5 (1), 2000167.
- (462) Jiang, T.; Xu, C.; Liu, Y.; Liu, Z.; Wall, J. S.; Zuo, X.; Lian, T.; Salaita, K.; Ni, C.; Pochan, D. et al. Structurally Defined Nanoscale Sheets from Self-Assembly of Collagen-Mimetic Peptides. *J. Am. Chem. Soc.* **2014**, 136 (11), 4300-4308.
- (463) Magnotti, E. L.; Hughes, S. A.; Dillard, R. S.; Wang, S.; Hough, L.; Karumbamkandathil, A.; Lian, T.; Wall, J. S.; Zuo, X.; Wright, E. R. et al. Self-Assembly of an Alpha-Helical Peptide into a Crystalline Two-Dimensional Nanoporous Framework. *J. Am. Chem. Soc.* **2016**, 138 (50), 16274-16282.
- (464) Poulos, S.; Agah, S.; Jallah, N.; Faham, S. Symmetry Based Assembly of a 2 Dimensional Protein Lattice. *PLoS One* **2017**, 12 (4), e0174485.
- (465) Nauli, S.; Farr, S.; Lee, Y. J.; Kim, H. Y.; Faham, S.; Bowie, J. U. Polymer-Driven Crystallization. *Protein Sci.* **2007**, 16 (11), 2542-2551.
- (466) Ringler, P.; Schulz, G. E. Self-Assembly of Proteins into Designed Networks. *Science* **2003**, 302 (5642), 106-109.
- (467) Heyman, A.; Medalsy, I.; Dgany, O.; Porath, D.; Markovich, G.; Shoseyov, O. Float and Compress: Honeycomb-Like Array of a Highly Stable Protein Scaffold. *Langmuir* **2009**, 25 (9), 5226-5229.
- (468) Zhao, L.; Li, Y.; Wang, T.; Qiao, S.; Li, X.; Wang, R.; Luo, Q.; Hou, C.; Xu, J.; Liu, J. Photocontrolled Protein Assembly for Constructing Programmed Two-Dimensional Nanomaterials. *J. Mater. Chem. B* **2018**, 6 (1), 75-83.
- (469) Li, X.; Qiao, S.; Zhao, L.; Liu, S.; Li, F.; Yang, F.; Luo, Q.; Hou, C.; Xu, J.; Liu, J. Template-Free Construction of Highly Ordered Monolayered Fluorescent Protein Nanosheets: A Bioinspired Artificial Light-Harvesting System. *ACS Nano* **2019**, 13 (2), 1861-1869.
- (470) Zhou, K.; Zang, J.; Chen, H.; Wang, W.; Wang, H.; Zhao, G. On-Axis Alignment of Protein Nanocage Assemblies from 2d to 3d through the Aromatic Stacking Interactions of Amino Acid Residues. *ACS Nano* **2018**, 12 (11), 11323-11332.
- (471) Zheng, B.; Zhou, K.; Zhang, T.; Lv, C.; Zhao, G. Designed Two- and Three-Dimensional Protein Nanocage Networks Driven by Hydrophobic Interactions Contributed by Amyloidogenic Motifs. *Nano Lett.* **2019**, 19 (6), 4023-4028.
- (472) Butler, P. J. Assembly of Tobacco Mosaic Virus Particle. *Nature* **1971**, 233 (5314), 25-27.
- (473) Durham, A. C.; Klug, A. Polymerization of Tobacco Mosaic Virus Protein and Its Control. *Nat. New Biol.* **1971**, 229 (2), 42-46.

- (474) Durham, A. C.; Finch, J. T.; Klug, A. States of Aggregation of Tobacco Mosaic Virus Protein. *Nat. New Biol.* **1971**, 229 (2), 37-42.
- (475) Bruckman, M. A.; Soto, C. M.; McDowell, H.; Liu, J. L.; Ratna, B. R.; Korpany, K. V.; Zahr, O. K.; Blum, A. S. Role of Hexahistidine in Directed Nanoassemblies of Tobacco Mosaic Virus Coat Protein. *ACS Nano* **2011**, 5 (3), 1606-1616.
- (476) Zhang, J.; Zhou, K.; Wang, Q. Tailoring the Self-Assembly Behaviors of Recombinant Tobacco Mosaic Virus by Rationally Introducing Covalent Bonding at the Protein-Protein Interface. *Small* **2016**, 12 (36), 4955-4959.
- (477) Zhang, S.; Zhang, J.; Fang, W.; Zhang, Y.; Wang, Q.; Jin, J. Ultralarge Single-Layer Porous Protein Nanosheet for Precise Nanosize Separation. *Nano Lett.* **2018**, 18 (10), 6563-6569.
- (478) Zhang, J.; Wang, X.; Zhou, K.; Chen, G.; Wang, Q. Self-Assembly of Protein Crystals with Different Crystal Structures Using Tobacco Mosaic Virus Coat Protein as a Building Block. *ACS Nano* **2018**, 12 (2), 1673-1679.
- (479) Zhang, J.; Zhou, K.; Zhang, Y.; Du, M.; Wang, Q. Precise Self-Assembly of Nanoparticles into Ordered Nanoarchitectures Directed by Tobacco Mosaic Virus Coat Protein. *Adv. Mater.* **2019**, 31 (23), e1901485.
- (480) Qiao, S. P.; Lang, C.; Wang, R. D.; Li, X. M.; Yan, T. F.; Pan, T. Z.; Zhao, L. L.; Fan, X. T.; Zhang, X.; Hou, C. X. et al. Metal Induced Self-Assembly of Designed V-Shape Protein into 2d Wavy Supramolecular Nanostructure. *Nanoscale* **2016**, 8 (1), 333-341.
- (481) Yang, M.; Song, W. J. Diverse Protein Assembly Driven by Metal and Chelating Amino Acids with Selectivity and Tunability. *Nat. Commun.* **2019**, 10 (1), 5545.
- (482) Sakai, F.; Yang, G.; Weiss, M. S.; Liu, Y.; Chen, G.; Jiang, M. Protein Crystalline Frameworks with Controllable Interpenetration Directed by Dual Supramolecular Interactions. *Nat. Commun.* **2014**, 5, 4634.
- (483) Yang, G.; Ding, H. M.; Kochovski, Z.; Hu, R.; Lu, Y.; Ma, Y. Q.; Chen, G.; Jiang, M. Highly Ordered Self-Assembly of Native Proteins into 1d, 2d, and 3d Structures Modulated by the Tether Length of Assembly-Inducing Ligands. *Angew. Chem. Int. Ed. Engl.* **2017**, 56 (36), 10691-10695.
- (484) Liu, R.; Kochovski, Z.; Li, L.; Yin, Y. W.; Yang, J.; Yang, G.; Tao, G.; Xu, A.; Zhang, E.; Ding, H. M. et al. Fabrication of Pascal-Triangle Lattice of Proteins by Inducing Ligand Strategy. *Angew. Chem. Int. Ed. Engl.* **2020**, 59 (24), 9617-9623.
- (485) Xu, M.; Zeng, R.; Xiang, J.; Yan, Q. Shaping Protein Amphiphilic Assemblies Via Allosteric Effect: From 1d Nanofilament to 2d Rectangular Nanosheet. *J. Am. Chem. Soc.* **2019**, 141 (35), 13724-13728.
- (486) Matthaei, J. F.; DiMaio, F.; Richards, J. J.; Pozzo, L. D.; Baker, D.; Baneyx, F. Designing Two-Dimensional Protein Arrays through Fusion of Multimers and Interface Mutations. *Nano Lett.* **2015**, 15 (8), 5235-5239.
- (487) Thomas, A.; Matthaei, J. F.; Baneyx, F. A Self-Assembling Two-Dimensional Protein Array Is a Versatile Platform for the Assembly of Multicomponent Nanostructures. *Biotechnol. J.* **2018**, 13 (12), e1800141.
- (488) Chen, Z.; Johnson, M. C.; Chen, J.; Bick, M. J.; Boyken, S. E.; Lin, B.; De Yoreo, J. J.; Kollman, J. M.; Baker, D.; DiMaio, F. Self-Assembling 2d Arrays with De Novo Protein Building Blocks. *J. Am. Chem. Soc.* **2019**, 141 (22), 8891-8895.
- (489) Ben-Sasson, A. J.; Watson, J. L.; Sheffler, W.; Johnson, M. C.; Bittleston, A.; Somasundaram, L.; Decarreau, J.; Jiao, F.; Chen, J.; Mela, I. et al. Design of Biologically Active Binary Protein 2d Materials. *Nature* **2021**, 589 (7842), 468-473.
- (490) Fromherz, P. Electron Microscopic Studies of Lipid Protein Films. *Nature* **1971**, 231 (5300), 267-268.

- (491) Uzgiris, E. E.; Kornberg, R. D. Two-Dimensional Crystallization Technique for Imaging Macromolecules, with Application to Antigen--Antibody--Complement Complexes. *Nature* **1983**, *301* (5896), 125-129.
- (492) Engel, A.; Hoenger, A.; Hefti, A.; Henn, C.; Ford, R. C.; Kistler, J.; Zulauf, M. Assembly of 2-D Membrane Protein Crystals: Dynamics, Crystal Order, and Fidelity of Structure Analysis by Electron Microscopy. *J. Struct. Biol.* **1992**, *109* (3), 219-234.
- (493) Stahlberg, H.; Fotiadis, D.; Scheuring, S.; Rémy, H.; Braun, T.; Mitsuoka, K.; Fujiyoshi, Y.; Engel, A. Two-Dimensional Crystals: A Powerful Approach to Assess Structure, Function and Dynamics of Membrane Proteins. *FEBS Lett.* **2001**, *504* (3), 166-172.
- (494) Brisson, A.; Olofsson, A.; Ringler, P.; Schmutz, M.; Stoylova, S. Two-Dimensional Crystallization of Proteins on Planar Lipid Films and Structure Determination by Electron Crystallography. *Biol. Cell* **1994**, *80* (2-3), 221-228.
- (495) Bourg, I. C.; Lee, S. S.; Fenter, P.; Tournassat, C. Stern Layer Structure and Energetics at Mica-Water Interfaces. *J. Phys. Chem. C* **2017**, *121* (17), 9402-9412.
- (496) Pyles, H.; Zhang, S.; De Yoreo, J. J.; Baker, D. Controlling Protein Assembly on Inorganic Crystals through Designed Protein Interfaces. *Nature* **2019**, *571* (7764), 251-256.
- (497) McPherson, A. Introduction to Protein Crystallization. *Methods* **2004**, *34* (3), 254-265.
- (498) Derewenda, Z. Application of Protein Engineering to Enhance Crystallizability and Improve Crystal Properties. *Acta Crystallogr, Sect. D* **2010**, *66* (5), 604-615.
- (499) Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G. et al. Solving the Structure of Human H Ferritin by Genetically Engineering Intermolecular Crystal Contacts. *Nature* **1991**, *349* (6309), 541-544.
- (500) Banyard, S. H.; Stammers, D. K.; Harrison, P. M. Electron Density Map of Apoferritin at 2.8- \AA Resolution. *Nature* **1978**, *271* (5642), 282-284.
- (501) Rypniewski, W. R.; Holden, H. M.; Rayment, I. Structural Consequences of Reductive Methylation of Lysine Residues in Hen Egg White Lysozyme: An X-Ray Analysis at 1.8- \AA Resolution. *Biochemistry* **1993**, *32* (37), 9851-9858.
- (502) Rayment, I.; Rypniewski, W. R.; Schmidt-Base, K.; Smith, R.; Tomchick, D. R.; Benning, M. M.; Winkelmann, D. A.; Wesenberg, G.; Holden, H. M. Three-Dimensional Structure of Myosin Subfragment-1: A Molecular Motor. *Science* **1993**, *261* (5117), 50-58.
- (503) Cooper, D. R.; Boczek, T.; Grelewska, K.; Pinkowska, M.; Sikorska, M.; Zawadzki, M.; Derewenda, Z. Protein Crystallization by Surface Entropy Reduction: Optimization of the Ser Strategy. *Acta Crystallogr, Sect. D: Biol. Crystallogr.* **2007**, *63* (Pt 5), 636-645.
- (504) Derewenda, Z. S.; Vekilov, P. G. Entropy and Surface Engineering in Protein Crystallization. *Acta Crystallogr, Sect. D: Biol. Crystallogr.* **2006**, *62* (Pt 1), 116-124.
- (505) Yamada, H.; Tamada, T.; Kosaka, M.; Miyata, K.; Fujiki, S.; Tano, M.; Moriya, M.; Yamanishi, M.; Honjo, E.; Tada, H. et al. 'Crystal Lattice Engineering', an Approach to Engineer Protein Crystal Contacts by Creating Intermolecular Symmetry: Crystallization and Structure Determination of a Mutant Human RNase 1 with a Hydrophobic Interface of Leucines. *Protein Sci.* **2007**, *16* (7), 1389-1397.
- (506) Du, M.; Zhou, K.; Yu, R.; Zhai, Y.; Chen, G.; Wang, Q. Noncovalent Self-Assembly of Protein Crystals with Tunable Structures. *Nano Lett.* **2021**, *21* (4), 1749-1757.
- (507) Kunzle, M.; Eckert, T.; Beck, T. Metal-Assisted Assembly of Protein Containers Loaded with Inorganic Nanoparticles. *Inorg. Chem.* **2018**, *57* (21), 13431-13436.
- (508) Forse, G. J.; Ram, N.; Banatao, D. R.; Cascio, D.; Sawaya, M. R.; Klock, H. E.; Lesley, S. A.; Yeates, T. O. Synthetic Symmetrization in the Crystallization and Structure Determination of Cela from Thermotoga Maritima. *Protein Sci.* **2011**, *20* (1), 168-178.
- (509) Laganowsky, A.; Zhao, M.; Soriaga, A. B.; Sawaya, M. R.; Cascio, D.; Yeates, T. O. An Approach to Crystallizing Proteins by Metal-Mediated Synthetic Symmetrization. *Protein Sci.* **2011**, *20* (11), 1876-1890.

- (510) Radford, R. J.; Lawrenz, M.; Nguyen, P. C.; McCammon, J. A.; Tezcan, F. A. Porous Protein Frameworks with Unsaturated Metal Centers in Sterically Encumbered Coordination Sites. *Chem. Commun.* **2011**, 47 (1), 313-315.
- (511) Broomell, C. C.; Birkedal, H.; Oliveira, C. L. P.; Pedersen, J. S.; Gertenbach, J. A.; Young, M.; Douglas, T. Protein Cage Nanoparticles as Secondary Building Units for the Synthesis of 3-Dimensional Coordination Polymers. *Soft Matter* **2010**, 6 (14), 3167-3171.
- (512) Bailey, J. B.; Zhang, L.; Chiong, J. A.; Ahn, S.; Tezcan, F. A. Synthetic Modularity of Protein-Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2017**, 139 (24), 8160-8166.
- (513) Bailey, J. B.; Tezcan, F. A. Tunable and Cooperative Thermomechanical Properties of Protein-Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2020**, 142 (41), 17265-17270.
- (514) Olsen, L. R.; Dessen, A.; Gupta, D.; Sabesan, S.; Sacchettini, J. C.; Brewer, C. F. X-Ray Crystallographic Studies of Unique Cross-Linked Lattices between Four Isomeric Biantennary Oligosaccharides and Soybean Agglutinin. *Biochemistry* **1997**, 36 (49), 15073-15080.
- (515) Dessen, A.; Gupta, D.; Sabesan, S.; Brewer, C. F.; Sacchettini, J. C. X-Ray Crystal Structure of the Soybean Agglutinin Cross-Linked with a Biantennary Analog of the Blood Group I Carbohydrate Antigen. *Biochemistry* **1995**, 34 (15), 4933-4942.
- (516) Dotan, N.; Arad, D.; Frolov, F.; Freeman, A. Self-Assembly of a Tetrahedral Lectin into Predesigned Diamondlike Protein Crystals. *Angew. Chem. Int. Ed. Engl.* **1999**, 38 (16), 2363-2366.
- (517) Beyeh, N. K.; Nonappa; Liljestrom, V.; Mikkila, J.; Korpi, A.; Bochicchio, D.; Pavan, G. M.; Ikkala, O.; Ras, R. H. A.; Kostiainen, M. A. Crystalline Cyclophane-Protein Cage Frameworks. *ACS Nano* **2018**, 12 (8), 8029-8036.
- (518) Rennie, M. L.; Fox, G. C.; Perez, J.; Crowley, P. B. Auto-Regulated Protein Assembly on a Supramolecular Scaffold. *Angew. Chem. Int. Ed. Engl.* **2018**, 57 (42), 13764-13769.
- (519) Engilberge, S.; Rennie, M. L.; Dumont, E.; Crowley, P. B. Tuning Protein Frameworks Via Auxiliary Supramolecular Interactions. *ACS Nano* **2019**, 13 (9), 10343-10350.
- (520) Guagnini, F.; Huber, A.; Alex, J. M.; Marx, F.; Crowley, P. B. Porous Assembly of an Antifungal Protein Mediated by Zinc and Sulfonato-Calix[8]Arene. *J. Struct. Biol.* **2021**, 213 (2), 107711.
- (521) Vandebroek, L.; Noguchi, H.; Kamata, K.; Tame, J. R. H.; Van Meervelt, L.; Parac-Vogt, T. N.; Voet, A. R. D. Shape and Size Complementarity-Induced Formation of Supramolecular Protein Assemblies with Metal-Oxo Clusters. *Cryst. Growth Des.* **2021**, 21 (2), 1307-1313.
- (522) Vandebroek, L.; Noguchi, H.; Kamata, K.; Tame, J. R. H.; Van Meervelt, L.; Parac-Vogt, T. N.; Voet, A. R. D. Hybrid Assemblies of a Symmetric Designer Protein and Polyoxometalates with Matching Symmetry. *Chem. Commun.* **2020**, 56 (78), 11601-11604.
- (523) Liljestrom, V.; Seitsonen, J.; Kostiainen, M. A. Electrostatic Self-Assembly of Soft Matter Nanoparticle Cocrystals with Tunable Lattice Parameters. *ACS Nano* **2015**, 9 (11), 11278-11285.
- (524) Winfree, E.; Liu, F.; Wenzler, L. A.; Seeman, N. C. Design and Self-Assembly of Two-Dimensional DNA Crystals. *Nature* **1998**, 394 (6693), 539-544.
- (525) Seeman, N. C. Nucleic Acid Junctions and Lattices. *J. Theor. Biol.* **1982**, 99 (2), 237-247.
- (526) Yang, Y. R.; Liu, Y.; Yan, H. DNA Nanostructures as Programmable Biomolecular Scaffolds. *Bioconjug. Chem.* **2015**, 26 (8), 1381-1395.
- (527) Liu, N.; Liedl, T. DNA-Assembled Advanced Plasmonic Architectures. *Chem. Rev.* **2018**, 118 (6), 3032-3053.
- (528) Cigler, P.; Lytton-Jean, A. K.; Anderson, D. G.; Finn, M. G.; Park, S. Y. DNA-Controlled Assembly of a Natl Lattice Structure from Gold Nanoparticles and Protein Nanoparticles. *Nat. Mater.* **2010**, 9 (11), 918-922.

- (529) Mirkin, C. A.; Letsinger, R. L.; Mucic, R. C.; Storhoff, J. J. A DNA-Based Method for Rationally Assembling Nanoparticles into Macroscopic Materials. *Nature* **1996**, 382 (6592), 607-609.
- (530) Park, S. Y.; Lytton-Jean, A. K.; Lee, B.; Weigand, S.; Schatz, G. C.; Mirkin, C. A. DNA-Programmable Nanoparticle Crystallization. *Nature* **2008**, 451 (7178), 553-556.
- (531) Wang, M. X.; Brodin, J. D.; Millan, J. A.; Seo, S. E.; Girard, M.; Olvera de la Cruz, M.; Lee, B.; Mirkin, C. A. Altering DNA-Programmable Colloidal Crystallization Paths by Modulating Particle Repulsion. *Nano Lett.* **2017**, 17 (8), 5126-5132.
- (532) Hayes, O. G.; McMillan, J. R.; Lee, B.; Mirkin, C. A. DNA-Encoded Protein Janus Nanoparticles. *J. Am. Chem. Soc.* **2018**, 140 (29), 9269-9274.
- (533) Winegar, P. H.; Hayes, O. G.; McMillan, J. R.; Figg, C. A.; Focia, P. J.; Mirkin, C. A. DNA-Directed Protein Packing within Single Crystals. *Chem* **2020**, 6 (4), 1007-1017.
- (534) Partridge, B. E.; Winegar, P. H.; Han, Z.; Mirkin, C. A. Redefining Protein Interfaces within Protein Single Crystals with DNA. *J. Am. Chem. Soc.* **2021**, 143 (23), 8925-8934.
- (535) Stephanopoulos, N. Hybrid Nanostructures from the Self-Assembly of Proteins and DNA. *Chem* **2020**, 6 (2), 364-405.
- (536) Bellomo, E. G.; Wyrsta, M. D.; Pakstis, L.; Pochan, D. J.; Deming, T. J. Stimuli-Responsive Polypeptide Vesicles by Conformation-Specific Assembly. *Nat. Mater.* **2004**, 3 (4), 244-248.
- (537) Anzini, P.; Xu, C.; Hughes, S.; Magnotti, E.; Jiang, T.; Hemmingsen, L.; Demeler, B.; Conticello, V. P. Controlling Self-Assembly of a Peptide-Based Material Via Metal-Ion Induced Registry Shift. *J. Am. Chem. Soc.* **2013**, 135 (28), 10278-10281.
- (538) Yan, C.; Pochan, D. J. Rheological Properties of Peptide-Based Hydrogels for Biomedical and Other Applications. *Chem. Soc. Rev.* **2010**, 39 (9), 3528-3540.
- (539) Chockalingam, K.; Blenner, M.; Banta, S. Design and Application of Stimulus-Responsive Peptide Systems. *Protein Eng. Des. Sel.* **2007**, 20 (4), 155-161.
- (540) Iyer, P. V.; Ananthanarayan, L. Enzyme Stability and Stabilization - Aqueous and Non-Aqueous Environment. *Process Biochem.* **2008**, 43 (10), 1019-1032.
- (541) Jiang, X.; Bishop, E. J.; Farid, R. S. A De Novo Designed Protein with Properties That Characterize Natural Hyperthermophilic Proteins. *J. Am. Chem. Soc.* **1997**, 119 (4), 838-839.
- (542) Korkegian, A.; Black, M. E.; Baker, D.; Stoddard, B. L. Computational Thermostabilization of an Enzyme. *Science* **2005**, 308 (5723), 857-860.
- (543) Malakauskas, S. M.; Mayo, S. L. Design, Structure and Stability of a Hyperthermophilic Protein Variant. *Nat. Struct. Biol.* **1998**, 5 (6), 470-475.
- (544) Eijsink, V. G.; Bjork, A.; Gaseidnes, S.; Sirevag, R.; Synstad, B.; van den Burg, B.; Vriend, G. Rational Engineering of Enzyme Stability. *J. Biotechnol.* **2004**, 113 (1-3), 105-120.
- (545) Hao, J.; Berry, A. A Thermostable Variant of Fructose Bisphosphate Aldolase Constructed by Directed Evolution Also Shows Increased Stability in Organic Solvents. *Protein Eng. Des. Sel.* **2004**, 17 (9), 689-697.
- (546) Lele, B. S.; Murata, H.; Matyjaszewski, K.; Russell, A. J. Synthesis of Uniform Protein-Polymer Conjugates. *Biomacromolecules* **2005**, 6 (6), 3380-3387.
- (547) Heredia, K. L.; Maynard, H. D. Synthesis of Protein-Polymer Conjugates. *Org. Biomol. Chem.* **2007**, 5 (1), 45-53.
- (548) Ge, J.; Lu, D. N.; Liu, Z. X.; Liu, Z. Recent Advances in Nanostructured Biocatalysts. *Biochem. Eng. J.* **2009**, 44 (1), 53-59.
- (549) Margolin, A. L.; Navia, M. A. Protein Crystals as Novel Catalytic Materials. *Angew. Chem. Int. Ed. Engl.* **2001**, 40 (12), 2204-2222.
- (550) St. Clair, N. L.; Navia, M. A. Cross-Linked Enzyme Crystals as Robust Biocatalysts. *J. Am. Chem. Soc.* **2002**, 114 (18), 7314-7316.

- (551) Gupta, M. N.; Perwez, M.; Sardar, M. Protein Crosslinking: Uses in Chemistry, Biology and Biotechnology. *Biocatal. Biotransform.* **2020**, 38 (3), 178-201.
- (552) Govardhan, C. P. Crosslinking of Enzymes for Improved Stability and Performance. *Curr. Opin. Biotechnol.* **1999**, 10 (4), 331-335.
- (553) Liu, Q.; Xun, G.; Feng, Y. The State-of-the-Art Strategies of Protein Engineering for Enzyme Stabilization. *Biotechnol. Adv.* **2019**, 37 (4), 530-537.
- (554) Ó'Fágáin, C. Enzyme Stabilization—Recent Experimental Progress. *Enzyme Microb. Technol.* **2003**, 33 (2-3), 137-149.
- (555) Mines, G. A.; Pascher, T.; Lee, S. C.; Winkler, J. R.; Gray, H. B. Cytochrome C Folding Triggered by Electron Transfer. *Chem. Biol.* **1996**, 3 (6), 491-497.
- (556) Perlmutter, J. D.; Hagan, M. F. Mechanisms of Virus Assembly. *Annu. Rev. Phys. Chem.* **2015**, 66 (1), 217-239.
- (557) Roos, W. H.; Bruinsma, R.; Wuite, G. J. L. Physical Virology. *Nat. Phys.* **2010**, 6 (10), 733-743.
- (558) Roos, W. H.; Gertsman, I.; May, E. R.; Brooks, C. L., 3rd; Johnson, J. E.; Wuite, G. J. Mechanics of Bacteriophage Maturation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109 (7), 2342-2347.
- (559) Perkins, J. R.; Diboun, I.; Dessailly, B. H.; Lees, J. G.; Orengo, C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. *Structure* **2010**, 18 (10), 1233-1243.
- (560) Kirschner, M.; Mitchison, T. Beyond Self-Assembly: From Microtubules to Morphogenesis. *Cell* **1986**, 45 (3), 329-342.
- (561) Zlotnick, A. Theoretical Aspects of Virus Capsid Assembly. *J. Mol. Recognit.* **2005**, 18 (6), 479-490.
- (562) Antonini, E.; Brunori, M. Hemoglobin. *Annu. Rev. Biochem.* **1970**, 39 (1), 977-1042.
- (563) Lippow, S. M.; Tidor, B. Progress in Computational Protein Design. *Curr. Opin. Biotechnol.* **2007**, 18 (4), 305-311.
- (564) Fleishman, S. J.; Baker, D. Role of the Biomolecular Energy Gap in Protein Design, Structure, and Evolution. *Cell* **2012**, 149 (2), 262-273.
- (565) Dawson, W. M.; Rhys, G. G.; Woolfson, D. N. Towards Functional De Novo Designed Proteins. *Curr. Opin. Chem. Biol.* **2019**, 52, 102-111.
- (566) Gibaud, T.; Barry, E.; Zakhary, M. J.; Henglin, M.; Ward, A.; Yang, Y.; Berciu, C.; Oldenbourg, R.; Hagan, M. F.; Nicastro, D. et al. Reconfigurable Self-Assembly through Chiral Control of Interfacial Tension. *Nature* **2012**, 481 (7381), 348-351.
- (567) Li, T.; Zan, X.; Winans, R. E.; Wang, Q.; Lee, B. Biomolecular Assembly of Thermoresponsive Superlattices of the Tobacco Mosaic Virus with Large Tunable Interparticle Distances. *Angew. Chem. Int. Ed. Engl.* **2013**, 52 (26), 6638-6642.
- (568) Heo, K.; Jin, H. E.; Kim, H.; Lee, J. H.; Wang, E.; Lee, S. W. Transient Self-Templating Assembly of M13 Bacteriophage for Enhanced Biopiezoelectric Devices. *Nano Energy* **2019**, 56, 716-723.
- (569) Ross, B.; Mehta, S.; Zhang, J. Molecular Tools for Acute Spatiotemporal Manipulation of Signal Transduction. *Curr. Opin. Chem. Biol.* **2016**, 34, 135-142.
- (570) Lella, M.; Mahalakshmi, R. Metamorphic Proteins: Emergence of Dual Protein Folds from One Primary Sequence. *Biochemistry* **2017**, 56 (24), 2971-2984.
- (571) Xu, M.; Zeng, R.; Xiang, J.; Yan, Q. Self-Assembly of Switchable Protein Nanocages Via Allosteric Effect. *CCS Chem* **2020**, 2, 2223-2232.
- (572) Xu, M.; Liu, L.; Yan, Q. Allosterically Activated Protein Self-Assembly for the Construction of Helical Microfilaments with Tunable Helicity. *Angew. Chem. Int. Ed. Engl.* **2018**, 57 (18), 5029-5032.

- (573) Han, K.; Bailey, J. B.; Zhang, L.; Tezcan, F. A. Anisotropic Dynamics and Mechanics of Macromolecular Crystals Containing Lattice-Patterned Polymer Networks. *J. Am. Chem. Soc.* **2020**, 142 (45), 19402-19410.
- (574) Cui, Q.; Karplus, M. Allostery and Cooperativity Revisited. *Protein Sci.* **2008**, 17 (8), 1295-1307.
- (575) Changeux, J. P. Allostery and the Monod-Wyman-Changeux Model after 50 Years. *Annu. Rev. Biophys.* **2012**, 41 (1), 103-133.
- (576) Strickland, D.; Moffat, K.; Sosnick, T. R. Light-Activated DNA Binding in a Designed Allosteric Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105 (31), 10709-10714.
- (577) Chen, R. P.; Blackstock, D.; Sun, Q.; Chen, W. Dynamic Protein Assembly by Programmable DNA Strand Displacement. *Nat. Chem.* **2018**, 10 (4), 474-481.
- (578) Kowalski, A. E.; Johnson, L. B.; Dierl, H. K.; Park, S.; Huber, T. R.; Snow, C. D. Porous Protein Crystals as Scaffolds for Enzyme Immobilization. *Biomater. Sci.* **2019**, 7 (5), 1898-1904.
- (579) Huber, T. R.; Hartje, L. F.; McPherson, E. C.; Kowalski, A. E.; Snow, C. D. Programmed Assembly of Host-Guest Protein Crystals. *Small* **2017**, 13 (7), 1602703.
- (580) Uchida, M.; McCoy, K.; Fukuto, M.; Yang, L.; Yoshimura, H.; Miettinen, H. M.; LaFrance, B.; Patterson, D. P.; Schwarz, B.; Karty, J. A. et al. Modular Self-Assembly of Protein Cage Lattices for Multistep Catalysis. *ACS Nano* **2018**, 12 (2), 942-953.
- (581) Ward, A. R.; Snow, C. D. Porous Crystals as Scaffolds for Structural Biology. *Curr. Opin. Struct. Biol.* **2020**, 60, 85-92.
- (582) Butterfield, G. L.; Lajoie, M. J.; Gustafson, H. H.; Sellers, D. L.; Nattermann, U.; Ellis, D.; Bale, J. B.; Ke, S.; Lenz, G. H.; Yehdego, A. et al. Evolution of a Designed Protein Assembly Encapsulating Its Own RNA Genome. *Nature* **2017**, 552 (7685), 415-420.
- (583) Edwardson, T. G. W.; Tetter, S.; Hilvert, D. Two-Tier Supramolecular Encapsulation of Small Molecules in a Protein Cage. *Nat. Commun.* **2020**, 11 (1), 5410.
- (584) Maita, N. Crystal Structure Determination of Ubiquitin by Fusion to a Protein That Forms a Highly Porous Crystal Lattice. *J. Am. Chem. Soc.* **2018**, 140 (42), 13546-13549.
- (585) Yeates, T. O.; Agdanowski, M. P.; Liu, Y. Development of Imaging Scaffolds for Cryo-Electron Microscopy. *Curr. Opin. Struct. Biol.* **2020**, 60, 142-149.
- (586) Koshiyama, T.; Kawaba, N.; Hikage, T.; Shirai, M.; Miura, Y.; Huang, C. Y.; Tanaka, K.; Watanabe, Y.; Ueno, T. Modification of Porous Protein Crystals in Development of Biohybrid Materials. *Bioconjug. Chem.* **2010**, 21 (2), 264-269.
- (587) Huber, T. R.; McPherson, E. C.; Keating, C. E.; Snow, C. D. Installing Guest Molecules at Specific Sites within Scaffold Protein Crystals. *Bioconjug. Chem.* **2018**, 29 (1), 17-22.
- (588) Liu, Y. X.; Huynh, D. T.; Yeates, T. O. A 3.8 Ångstrom Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold. *Nat. Commun.* **2019**, 10, 1864.
- (589) Challenges for Cryo-EM. *Nat. Methods* **2018**, 15 (12), 985.
- (590) Renaud, J.-P.; Chari, A.; Ciferri, C.; Liu, W.-t.; Rémigy, H.-W.; Stark, H.; Wiesmann, C. Cryo-EM in Drug Discovery: Achievements, Limitations and Prospects. *Nat. Rev. Drug Discovery* **2018**, 17 (7), 471-492.
- (591) Bale, J. B.; Park, R. U.; Liu, Y.; Gonen, S.; Gonen, T.; Cascio, D.; King, N. P.; Yeates, T. O.; Baker, D. Structure of a Designed Tetrahedral Protein Assembly Variant Engineered to Have Improved Soluble Expression. *Protein Sci.* **2015**, 24 (10), 1695-1701.
- (592) Nguyen, T. K.; Abe, S.; Kasamatsu, M.; Maity, B.; Yamashita, K.; Hirata, K.; Kojima, M.; Ueno, T. In-Cell Engineering of Protein Crystals with Nanoporous Structures for Promoting Cascade Reactions. *ACS Appl. Nano. Mater.* **2021**, 4 (2), 1672-1681.
- (593) McConnell, S. A.; Cannon, K. A.; Morgan, C.; McAllister, R.; Amer, B. R.; Clubb, R. T.; Yeates, T. O. Designed Protein Cages as Scaffolds for Building Multienzyme Materials. *ACS Synth. Biol.* **2020**, 9 (2), 381-391.

- (594) Subramanian, R. H.; Suzuki, Y.; Tallorin, L.; Sahu, S.; Thompson, M.; Gianneschi, N. C.; Burkart, M. D.; Tezcan, F. A. Enzyme-Directed Functionalization of Designed, Two-Dimensional Protein Lattices. *Biochemistry* **2021**, *60* (13), 1050-1062.
- (595) Correnti, C. E.; Hallinan, J. P.; Doyle, L. A.; Ruff, R. O.; Jaeger-Ruckstuhl, C. A.; Xu, Y.; Shen, B. W.; Qu, A.; Polkinghorn, C.; Friend, D. J. et al. Engineering and Functionalization of Large Circular Tandem Repeat Protein Nanoparticles. *Nat. Struct. Mol. Biol.* **2020**, *27* (4), 342-350.
- (596) Saif, B.; Yang, P. Metal-Protein Hybrid Materials with Desired Functions and Potential Applications. *ACS Appl. Bio. Mater.* **2021**, *4* (2), 1156-1177.
- (597) Kostiainen, M. A.; Ceci, P.; Fornara, M.; Hiekkataipale, P.; Kasyutich, O.; Nolte, R. J.; Cornelissen, J. J.; Desautels, R. D.; van Lierop, J. Hierarchical Self-Assembly and Optical Disassembly for Controlled Switching of Magnetoferitin Nanoparticle Magnetism. *ACS Nano* **2011**, *5* (8), 6394-6402.
- (598) Okuda, M.; Suzumoto, Y.; Yamashita, I. Bioinspired Synthesis of Homogenous Cerium Oxide Nanoparticles and Two- or Three-Dimensional Nanoparticle Arrays Using Protein Supramolecules. *Cryst. Growth Des.* **2011**, *11* (6), 2540-2545.
- (599) Maity, B.; Abe, S.; Ueno, T. Observation of Gold Sub-Nanocluster Nucleation within a Crystalline Protein Cage. *Nat. Commun.* **2017**, *8* (1), 14820.
- (600) Chakraborti, S.; Korpi, A.; Kumar, M.; Stepien, P.; Kostiainen, M. A.; Heddle, J. G. Three-Dimensional Protein Cage Array Capable of Active Enzyme Capture and Artificial Chaperone Activity. *Nano Lett.* **2019**, *19* (6), 3918-3924.
- (601) Jutz, G.; van Rijn, P.; Santos Miranda, B.; Böker, A. Ferritin: A Versatile Building Block for Bionanotechnology. *Chem. Rev.* **2015**, *115* (4), 1653-1701.
- (602) Lach, M.; Kunzle, M.; Beck, T. Free-Standing Metal Oxide Nanoparticle Superlattices Constructed with Engineered Protein Containers Show in Crystallo Catalytic Activity. *Chemistry* **2017**, *23* (69), 17482-17486.
- (603) Ardini, M.; Giansanti, F.; Di Leandro, L.; Pitari, G.; Cimini, A.; Ottaviano, L.; Donarelli, M.; Santucci, S.; Angelucci, F.; Ippoliti, R. Metal-Induced Self-Assembly of Peroxiredoxin as a Tool for Sorting Ultrasmall Gold Nanoparticles into One-Dimensional Clusters. *Nanoscale* **2014**, *6* (14), 8052-8061.
- (604) Manuguri, S.; Webster, K.; Yewdall, N. A.; An, Y.; Venugopal, H.; Bhugra, V.; Turner, A.; Domigan, L. J.; Gerrard, J. A.; Williams, D. E. et al. Assembly of Protein Stacks with in Situ Synthesized Nanoparticle Cargo. *Nano Lett.* **2018**, *18* (8), 5138-5145.
- (605) Anaya-Plaza, E.; Aljarilla, A.; Beaune, G.; Nonappa; Timonen, J. V. I.; de la Escosura, A.; Torres, T.; Kostiainen, M. A. Phthalocyanine-Virus Nanofibers as Heterogeneous Catalysts for Continuous-Flow Photo-Oxidation Processes. *Adv. Mater.* **2019**, *31* (39), e1902582.
- (606) McMillan, R. A.; Paavola, C. D.; Howard, J.; Chan, S. L.; Zaluzec, N. J.; Trent, J. D. Ordered Nanoparticle Arrays Formed on Engineered Chaperonin Protein Templates. *Nat. Mater.* **2002**, *1* (4), 247-252.
- (607) McMillan, R. A.; Howard, J.; Zaluzec, N. J.; Kagawa, H. K.; Mogul, R.; Li, Y. F.; Paavola, C. D.; Trent, J. D. A Self-Assembling Protein Template for Constrained Synthesis and Patterning of Nanoparticle Arrays. *J. Am. Chem. Soc.* **2005**, *127* (9), 2800-2801.
- (608) Du, M.; Zhou, K.; Wang, X.; Zhang, J.; Zhang, Y.; Dong, J.; Wu, L.; Qiao, Z.; Chen, G.; Wang, Q. Precise Fabrication of De Novo Nanoparticle Lattices on Dynamic 2d Protein Crystalline Lattices. *Nano Lett.* **2020**, *20* (2), 1154-1160.
- (609) Coste, B.; Xiao, B.; Santos, J. S.; Syeda, R.; Grandl, J.; Spencer, K. S.; Kim, S. E.; Schmidt, M.; Mathur, J.; Dubin, A. E. et al. Piezo Proteins Are Pore-Forming Subunits of Mechanically Activated Channels. *Nature* **2012**, *483* (7388), 176-181.
- (610) Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. L. Silk-Based Biomaterials. *Biomaterials* **2003**, *24* (3), 401-416.

- (611) Almine, J. F.; Bax, D. V.; Mithieux, S. M.; Nivison-Smith, L.; Rnjak, J.; Waterhouse, A.; Wise, S. G.; Weiss, A. S. Elastin-Based Materials. *Chem. Soc. Rev.* **2010**, 39 (9), 3371-3379.
- (612) Chattopadhyay, S.; Raines, R. T. Review Collagen-Based Biomaterials for Wound Healing. *Biopolymers* **2014**, 101 (8), 821-833.
- (613) van Hest, J. C.; Tirrell, D. A. Protein-Based Materials, toward a New Level of Structural Control. *Chem. Commun.* **2001**, (19), 1897-1904.
- (614) Buehler, M. J.; Yung, Y. C. Deformation and Failure of Protein Materials in Physiologically Extreme Conditions and Disease. *Nat. Mater.* **2009**, 8 (3), 175-188.
- (615) Hu, X.; Cebe, P.; Weiss, A. S.; Omenetto, F.; Kaplan, D. L. Protein-Based Composite Materials. *Mater. Today* **2012**, 15 (5), 208-215.
- (616) Sun, J.; Su, J.; Ma, C.; Gostl, R.; Herrmann, A.; Liu, K.; Zhang, H. Fabrication and Mechanical Properties of Engineered Protein-Based Adhesives and Fibers. *Adv. Mater.* **2020**, 32 (6), e1906360.
- (617) Greaves, G. N.; Greer, A. L.; Lakes, R. S.; Rouxel, T. Poisson's Ratio and Modern Materials. *Nat. Mater.* **2011**, 10 (11), 823-837.
- (618) Evans, K. E.; Alderson, A. Auxetic Materials: Functional Materials and Structures from Lateral Thinking! *Adv. Mater.* **2000**, 12 (9), 617-628.
- (619) Baughman, R. H. Auxetic Materials: Avoiding the Shrink. *Nature* **2003**, 425 (6959), 667.
- (620) Valimaki, S.; Mikkila, J.; Liljestrom, V.; Rosilo, H.; Ora, A.; Kostiainen, M. A. Hierarchically Ordered Supramolecular Protein-Polymer Composites with Thermoresponsive Properties. *Int. J. Mol. Sci.* **2015**, 16 (5), 10201-10213.
- (621) Okada, Y.; Tanaka, F. Cooperative Hydration, Chain Collapse, and Flat Lcst Behavior in Aqueous Poly(N-Isopropylacrylamide) Solutions. *Macromolecules* **2005**, 38 (10), 4465-4471.
- (622) Li, P.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F. et al. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, 483 (7389), 336-340.
- (623) Cule, D.; Hwa, T. Denaturation of Heterogeneous DNA. *Phys. Rev. Lett.* **1997**, 79 (12), 2375-2378.
- (624) Sato, O. Dynamic Molecular Crystals with Switchable Physical Properties. *Nat. Chem.* **2016**, 8 (7), 644-656.
- (625) Shao, Z. W.; Cao, X.; Luo, H. J.; Jin, P. Recent Progress in the Phase-Transition Mechanism and Modulation of Vanadium Dioxide Materials. *NPG Asia Mater.* **2018**, 10 (7), 581-605.
- (626) Fratzl, P.; Weinkamer, R. Nature's Hierarchical Materials. *Prog. Mater. Sci.* **2007**, 52 (8), 1263-1334.
- (627) Meyers, M. A.; Chen, P.-Y.; Lin, A. Y.-M.; Seki, Y. Biological Materials: Structure and Mechanical Properties. *Prog. Mater. Sci.* **2008**, 53 (1), 1-206.
- (628) Carlson, J. C.; Jena, S. S.; Flenniken, M.; Chou, T. F.; Siegel, R. A.; Wagner, C. R. Chemically Controlled Self-Assembly of Protein Nanorings. *J. Am. Chem. Soc.* **2006**, 128 (23), 7630-7638.
- (629) Li, Q.; So, C. R.; Fegan, A.; Cody, V.; Sarikaya, M.; Vallera, D. A.; Wagner, C. R. Chemically Self-Assembled Antibody Nanorings (Csans): Design and Characterization of an Anti-Cd3 IgM Biomimetic. *J. Am. Chem. Soc.* **2010**, 132 (48), 17247-17257.
- (630) Boyoglu-Barnum, S.; Ellis, D.; Gillespie, R. A.; Hutchinson, G. B.; Park, Y. J.; Moin, S. M.; Acton, O. J.; Ravichandran, R.; Murphy, M.; Pettie, D. et al. Quadrivalent Influenza Nanoparticle Vaccines Induce Broad Protection. *Nature* **2021**, 592 (7855), 623-628.
- (631) Petersburg, J. R.; Shen, J.; Csizmar, C. M.; Murphy, K. A.; Spanier, J.; Gabrielse, K.; Griffith, T. S.; Fife, B.; Wagner, C. R. Eradication of Established Tumors by Chemically Self-Assembled Nanoring Labeled T Cells. *ACS Nano* **2018**, 12 (7), 6563-6576.

- (632) Marcandalli, J.; Fiala, B.; Ols, S.; Perotti, M.; de van der Schueren, W.; Snijder, J.; Hodge, E.; Benhaim, M.; Ravichandran, R.; Carter, L. et al. Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell* **2019**, *176* (6), 1420-1431
- (633) Ueda, G.; Antanasićević, A.; Fallas, J. A.; Sheffler, W.; Coppers, J.; Ellis, D.; Hutchinson, G. B.; Moyer, A.; Yasmeen, A.; Tsybovsky, Y. et al. Tailored Design of Protein Nanoparticle Scaffolds for Multivalent Presentation of Viral Glycoprotein Antigens. *eLife* **2020**, *9*, e57659.
- (634) Brouwer, P. J. M.; Antanasićević, A.; Berndsen, Z.; Yasmeen, A.; Fiala, B.; Bijl, T. P. L.; Bontjer, I.; Bale, J. B.; Sheffler, W.; Allen, J. D. et al. Enhancing and Shaping the Immunogenicity of Native-Like Hiv-1 Envelope Trimers with a Two-Component Protein Nanoparticle. *Nat. Commun.* **2019**, *10* (1), 4272.
- (635) Walls, A. C.; Fiala, B.; Schafer, A.; Wrenn, S.; Pham, M. N.; Murphy, M.; Tse, L. V.; Shehata, L.; O'Connor, M. A.; Chen, C. et al. Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for Sars-Cov-2. *Cell* **2020**, *183* (5), 1367-1382
- (636) Park, J.; Selvaraj, B.; McShan, A. C.; Boyken, S. E.; Wei, K. Y.; Oberdorfer, G.; DeGrado, W.; Sgourakis, N. G.; Cuneo, M. J.; Myles, D. A. A. et al. De Novo Design of a Homo-Trimeric Amantadine-Binding Protein. *eLife* **2019**, *8*, e47839.
- (637) Votteler, J.; Ogohara, C.; Yi, S.; Hsia, Y.; Nattermann, U.; Belnap, D. M.; King, N. P.; Sundquist, W. I. Designed Proteins Induce the Formation of Nanocage-Containing Extracellular Vesicles. *Nature* **2016**, *540* (7632), 292-295.
- (638) Schwizer, F.; Okamoto, Y.; Heinisch, T.; Gu, Y.; Pellizzoni, M. M.; Lebrun, V.; Reuter, R.; Kohler, V.; Lewis, J. C.; Ward, T. R. Artificial Metalloenzymes: Reaction Scope and Optimization Strategies. *Chem. Rev.* **2018**, *118* (1), 142-231.
- (639) Thomas, C. M.; Ward, T. R. Artificial Metalloenzymes: Proteins as Hosts for Enantioselective Catalysis. *Chem. Soc. Rev.* **2005**, *34* (4), 337-346.
- (640) Chen, K.; Arnold, F. H. Engineering New Catalytic Activities in Enzymes. *Nat. Catal.* **2020**, *3* (3), 203-213.
- (641) Lewis, J. C.; Coelho, P. S.; Arnold, F. H. Enzymatic Functionalization of Carbon-Hydrogen Bonds. *Chem. Soc. Rev.* **2011**, *40* (4), 2003-2021.
- (642) Nastri, F.; Chino, M.; Maglio, O.; Bhagi-Damodaran, A.; Lu, Y.; Lombardi, A. Design and Engineering of Artificial Oxygen-Activating Metalloenzymes. *Chem. Soc. Rev.* **2016**, *45* (18), 5020-5054.
- (643) Reed, C. J.; Lam, Q. N.; Mirts, E. N.; Lu, Y. Molecular Understanding of Heteronuclear Active Sites in Heme-Copper Oxidases, Nitric Oxide Reductases, and Sulfite Reductases through Biomimetic Modelling. *Chem. Soc. Rev.* **2021**, *50* (4), 2486-2539.
- (644) Yu, F.; Cangelosi, V. M.; Zastrow, M. L.; Tegoni, M.; Plegaria, J. S.; Tebo, A. G.; Mocny, C. S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V. L. Protein Design: Toward Functional Metalloenzymes. *Chem. Rev.* **2014**, *114* (7), 3495-3578.
- (645) Zastrow, M. L.; Pecoraro, V. L. Designing Functional Metalloproteins: From Structural to Catalytic Metal Sites. *Coordin. Chem. Rev.* **2013**, *257* (17-18), 2565-2588.
- (646) Scrofani, S. D.; Chung, J.; Huntley, J. J.; Benkovic, S. J.; Wright, P. E.; Dyson, H. J. Nmr Characterization of the Metallo-Beta-Lactamase from *Bacteroides fragilis* and Its Interaction with a Tight-Binding Inhibitor: Role of an Active-Site Loop. *Biochemistry* **1999**, *38* (44), 14507-14514.
- (647) Tomatis, P. E.; Fabiane, S. M.; Simona, F.; Carloni, P.; Sutton, B. J.; Vila, A. J. Adaptive Protein Evolution Grants Organismal Fitness by Improving Catalysis and Flexibility. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (52), 20605-20610.
- (648) Yu, J.; Yang, J.; Seok, C.; Song, W. J. Symmetry-Related Residues as Promising Hotspots for the Evolution of De Novo Oligomeric Enzymes. *Chem. Sci.* **2021**.

- (649) Bar-Even, A.; Noor, E.; Savir, Y.; Liebermeister, W.; Davidi, D.; Tawfik, D. S.; Milo, R. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **2011**, *50* (21), 4402-4410.
- (650) Pindur, U.; Lutz, G.; Otto, C. Acceleration and Selectivity Enhancement of Diels-Alder Reactions by Special and Catalytic Methods. *Chem. Rev.* **1993**, *93* (2), 741-761.
- (651) Eck, R. V.; Dayhoff, M. O. Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences. *Science* **1966**, *152* (3720), 363-366.
- (652) DiMarco, R. L.; Heilshorn, S. C. Multifunctional Materials through Modular Protein Engineering. *Adv. Mater.* **2012**, *24* (29), 3923-3940.
- (653) Zabret, J.; Bohn, S.; Schuller, S. K.; Arnolds, O.; Moller, M.; Meier-Credo, J.; Liauw, P.; Chan, A.; Tajkhorshid, E.; Langer, J. D. et al. Structural Insights into Photosystem II Assembly. *Nat. Plants* **2021**.
- (654) Mattila, P. K.; Lappalainen, P. Filopodia: Molecular Architecture and Cellular Functions. *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (6), 446-454.
- (655) White, S. H.; von Heijne, G. The Machinery of Membrane Protein Assembly. *Curr. Opin. Struct. Biol.* **2004**, *14* (4), 397-404.
- (656) Ellis, R. J. Molecular Chaperones: Assisting Assembly in Addition to Folding. *Trends Biochem. Sci.* **2006**, *31* (7), 395-401.
- (657) Klug, A. The Tobacco Mosaic Virus Particle: Structure and Assembly. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1999**, *354* (1383), 531-535.
- (658) Ackers, G. K.; Doyle, M. L.; Myers, D.; Daugherty, M. A. Molecular Code for Cooperativity in Hemoglobin. *Science* **1992**, *255* (5040), 54-63.
- (659) Janke, C.; Magiera, M. M. The Tubulin Code and Its Role in Controlling Microtubule Properties and Functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (6), 307-326.
- (660) Weisenberg, R. C.; Deery, W. J.; Dickinson, P. J. Tubulin-Nucleotide Interactions During the Polymerization and Depolymerization of Microtubules. *Biochemistry* **1976**, *15* (19), 4248-4254.
- (661) Dimitrov, D. S. *Therapeutic Proteins In Therapeutic Proteins: Methods and Protocols*; Voynov, V. and Caravella, J. A., Eds.; Humana Press: Totowa, NJ, 2012.
- (662) Arbige, M. V.; Shetty, J. K.; Chotani, G. K. Industrial Enzymology: The Next Chapter. *Trends Biotechnol.* **2019**, *37* (12), 1355-1366.
- (663) Kirk, O.; Borchert, T. V.; Fuglsang, C. C. Industrial Enzyme Applications. *Curr. Opin. Biotechnol.* **2002**, *13* (4), 345-351.
- (664) Torculas, M.; Medina, J.; Xue, W.; Hu, X. Protein-Based Bioelectronics. *ACS Biomater. Sci. Eng.* **2016**, *2* (8), 1211-1223.