Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Recent progress in designing protein-based supramolecular assemblies

William A Hansen[1] and Sagar D Khare[1,2]

The design of protein-based assemblies is an emerging area in bionanotechnology with wide ranging applications, from vaccines to smart biomaterials. Design approaches have sought to mimic both the topologies of assemblies observed in nature, as well as their functionally relevant properties, such as being responsive to external cues. In the last few years, diverse design approaches have been used to construct assemblies with integer-dimensional (e.g. filaments, layers, lattices and polyhedra) and non-integer-dimensional (fractal) topologies. Supramolecular structures that assemble/disassemble in response to chemical and physical stimuli have also been built. Hybrid protein-DNA assemblies have expanded the set of building blocks used for generating supramolecular architectures. While still far from reproducing the sophistication of natural assemblies, these exciting results represent important steps towards the design of responsive and functional biomaterials built from the bottom up. As the complexity of topologies and diversity of building blocks increases, considerations of both thermodynamics and kinetics of assembly formation will play crucial roles in making the design of protein-based assemblies robust and useful.

Addresses
[1] Institute for Quantitative Biomedicine, Rutgers – The State University of New Jersey, NJ, USA
[2] Department of Chemistry and Chemical Biology, Rutgers – The State University of New Jersey, NJ, USA

Corresponding author: Khare, Sagar D (sagar.khare@rutgers.edu)

Current Opinion in Structural Biology 2020, **63**:106–114

This review comes from a themed issue on **Engineering and design**

Edited by **Tijana Z Grove** and **Thomas J Magliery**

For a complete overview see the Issue and the Editorial

Available online 20th June 2020

https://doi.org/10.1016/j.sbi.2020.05.001

Introduction

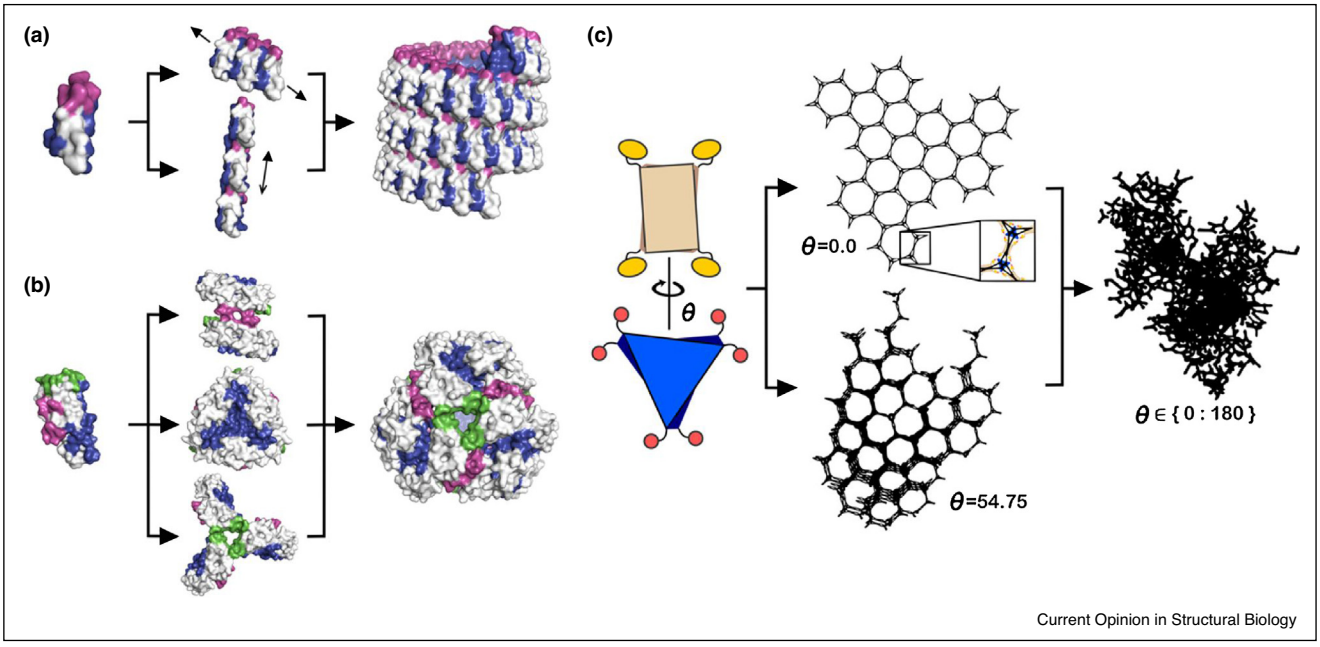
Life is sustained by the controlled self-assembly and disassembly of nanometer-sized biomolecules into mesoscale (defined as 100 nm–10 μm) objects in response to external stimuli. For example, cells move using mesoscale scaffolding provided by the cytoskeleton as a result of dynamic assembly/disassembly of tubulin proteins in response to chemical gradients [1]. The ability to programmatically and robustly produce mesoscale biomolecular assemblies promises to enable diverse applications in bionanotechnology, from tissue engineering to drug delivery to responsive biomaterials. Inspired by the diversity of natural assemblies and potential applications, biomolecular designers have sought over the years to use diverse building blocks (e.g. nucleic acids, proteins, peptides) to construct synthetic self-assembled systems with properties that mimic, complement or even surpass those produced by evolution [2–5].

Advancements in the field of computational design have made available a variety of *de novo* designed proteins, thereby significantly expanding the palette of building blocks available to a protein assembly designer [6]. Concomitantly, assemblies built by taking advantage of known switchable interactions between proteins and/or other biomolecules such as nucleic acids, and small molecule ligands, have enabled some measure of external control over assembly formation. In this review we highlight some recent achievements (2017 onwards) in designing new assembly topologies, and the use of various external stimuli including metal ions, light, and post-translational modifications which give rise to reversibility and topological tunability.

Topologies of designed assemblies

Protein assemblies in nature feature a variety of topologies including shell-like and cage-like structures (e.g. virus capsids [7]), three-dimensional crystals (e.g. occlusion bodies [8]), two-dimensional layers (e.g. bacterial S-layers [9]), fractional-dimensional topologies (e.g. intermediates in silicatein formation [10]), and one-dimensional wire-like topologies (e.g. microtubules [11]). These naturally observed protein-based structures have inspired the design of synthetic assemblies featuring these topologies using diverse strategies. The first wave of designed protein assemblies used symmetry as a key element of the design approach to generate a variety of shapes — an idea pioneered by Yeates [12]. The approach is to arrange building blocks with internal symmetry (e.g. trimers, dimers) at the vertices of target architectures (e.g. lattices, polyhedra) and either use protein fusions or computational design to optimize the stability of the single, new protein–protein interface generated as a result of this placement. While remarkably successful, the requirement for the use of symmetric building blocks (as opposed to asymmetric monomers) limits the range of designable topologies using this approach. Similarly, going beyond the stabilization of a single target inter-component interaction topology has the potential to yield novel supramolecular topologies. Exciting developments in the last few years have led to an expansion of the type of approaches used for design — asymmetric

Figure 1



Overview of recently demonstrated new design approaches and novel topologies in protein-based assembly design.
Success achieved in designing multiple interfaces (pink, blue and green surface colors) with a single asymmetric monomer has led to the construction of **(a)** protein-based filaments: lateral and axial interactions are highlighted with arrows, and **(b)** metal-ion stabilized polyhedra: constitutent C_2 and C_3-symmetric oligomers of the polyhedron are shown (middle). **(c)** Fractional-dimensional topologies were constructed by using dihedral symmetric building blocks and designing multiple high-affinity anisotropic binding modes between components such that stochastic sampling of interaction geometry leads to an emergent fractal-like structure. As edge cases, adoption of a uniform interaction geometry between components would lead to two-dimensional or three-dimensional lattices.

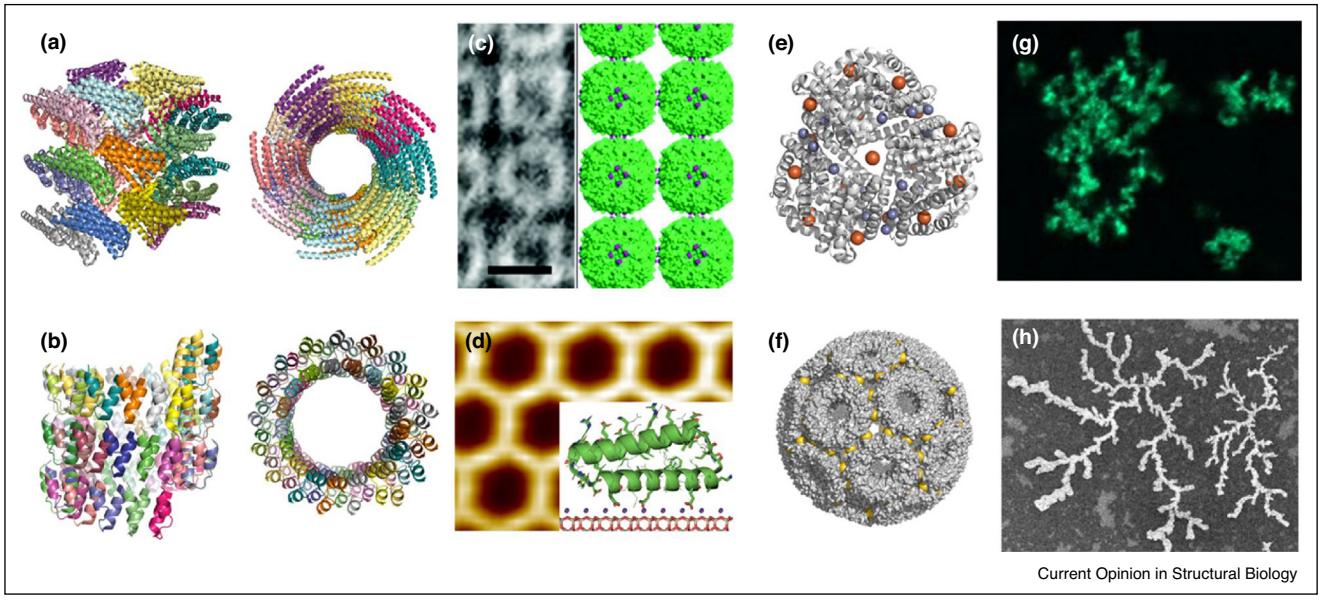
monomer building blocks have been to generate assemblies (Figure 1a,b) — as well as the range of designable topologies — fractional-dimensional topologies have been designed (Figure 1c).

One-dimensional (filaments and fibrils) assemblies

In filaments, building block proteins are generally asymmetric and make two types of interactions — lateral interactions along the perimeter of the filament cross section, and axial interactions parallel to the filament axis (Figure 1a). Therefore, the design of protein-based filaments requires the simultaneous design of at least two interfaces between asymmetric building blocks. Shen *et al.* reasoned that a large set of filamentous topologies can be generated by sampling a combination of cyclic and helical symmetric transforms starting from an asymmetric protein monomer [13••]. Rosetta-based interface design was used to generate sequences that stabilized candidate fibrillar morphologies. Cryo–electron microscopy-derived structures of six designs showed excellent agreement with the computational models (Figure 2a). The choice of highly stable and architecturally modular building blocks — *de novo* designed helical repeat proteins — which can tolerate multiple substitutions due to their high stability and can be shortened or expanded by removing or adding repeat units enabled the design of diverse native-like filamentous topologies with controllable geometries (filament diameter and superhelicity parameters). In a complementary approach, Hughes *et al.* used as building blocks variants of the consensus sequence of naturally occuring tandem repeat proteins that already form lateral interactions with each other leading to a ring-like architecture with partial superhelicity [14••]. Careful choice of amino acid substitutions along axial contacts then led to the formation of filamentous nanotubes (Figure 2b). The advantage of this approach is the ease of synthesis of building blocks (small peptides) but unlike the approach by Shen *et al.*, the parameters of the resulting filaments cannot be controlled *a priori*.

In nature, fibrillar protein architectures with one-dimensional order typically arise from beta-sheet hydrogen bonding and sidechain intercalation, for example, amyloid fibrils. In contrast, alpha helices rarely form fibrillar structures. Zhang *et al.* built upon their serendipitous discovery of a novel cross-alpha structure in a membrane protein crystal structure to design soluble peptides with this fibrillar topology [15•]. The rules of heptad-repeat packing derived from coiled-coils were

Figure 2



Examples of diverse topologies in designed protein assemblies.
Superhelical filaments (repeating monomers colored in pastel rainbow) by Shen *et al.* **(a)** and Hughes *et al.* **(b)**; two-dimensional layers of ferritin nanocages **(c)** and *de novo* designed protein layers on a mica surface **(d)**; metal-mediated polyhedra **(e)** and **(f)**, (Fe - gold, Zn - grey, and Au - yellow, highlighted as spheres); fractal-like topologies **(g)** and **(h)**.

used to engineer contacts between 4-helix bundles that form superhelical structures featuring lateral contacts between helices. A small (15° to 20°) left-handed crossing angle gives rise to a progressive left-handed screw that generates spiralling amyloid-like fibrillar structures.

Two-dimensional and three-dimensional lattices

Previous reviews have exhaustively covered 2D/3D self-assembled structures formed with homo-oligomeric protein building blocks using a variety of strategies: computational interface design, domain fusion, metal coordination and disulfide bonds [2–5]. Recently, Zhou *et al.* extended this approach to hierarchical supramolecular 2D layer formation by constructing disulfide-mediated 2D superlattices of ferritin nanocages induced by a single point mutation at the C_4 interface of the ferritin nanocage [16]. In a similar clever design approach for hierarchical 3D crystal design, ferritin nanocage building blocks that were engineered to form Ca^{2+}-bridged crystals were embedded in polymer hydrogel that occupied the void volume of the crystals [17•]. Hydration/dehydration of the gel led to reversible swelling/contraction of the crystalline superlattice for multiple cycles — a property described as 'self healing'. The chemical constitution of the gel is key for determining the self-healing properties of the superlattices, suggesting that explicit consideration of the chemical interactions between proteins and polymeric matrix should enable greater control over the emergent topology. Indeed, the computational design of two-dimensional arrays on inorganic surfaces was impressively demonstrated by explicit consideration of the chemical character of the surface in the design process [18••]. Inspired by ice-binding proteins which present patterned arrays of hydrogen bonding residues matched to the ice lattice, *de novo* designed proteins were modeled to expose arrays of carboxylate sidechains with geometries matched to the potassium ion (K+) sublattice on muscovite mica (Figure 2d). Depending on K+ ion concentration, liquid crystal-like arrays ordered over tens of millimetres were experimentally detected.

Success with these design approaches promises to extend the reach of the principles of protein assembly design to the bottom up construction of synthetic hybrid materials with abiological (organic/inorganic) and biological (protein) building blocks, leading to biosensing, and light and energy harvesting as some potential downstream applications. However, the design problem also becomes more complex when considering multiple types of organic/inorganic constituent building blocks and their interactions with each other. Careful matching of interactions to avoid kinetic traps and/or explicit modeling of the non-protein components will be key for these approaches to become robust.

Protein-based polyhedral (zero-dimensional) assemblies

Previously designed polyhedral architectures were generated using the computational design of a single interface between oligomeric proteins (typically featuring C_3 or C_2

symmetry) building blocks. A complementary approach is to generate the required new interface by introducing metal ion-binding residues at appropriate locations on the protein surface. Metal chelation-dependent polyhedron formation was demonstrated using divalent cation-mediated coiled-coil formation where coiled-coil forming monomers were fused to a symmetric trimer building block [19]. In a related elegant approach that forgoes the requirement for symmetric oligomer building blocks, two novel interfaces were simultaneously designed using metal-ion chelation (Figures 1b, 2 e). Golub *et al.* used the different metal co-ordination geometry of zinc and iron binding (the latter achieved via a covalently attached hydroxamate moiety) to simultaneously create novel C_3 and C_2 interfaces between copies of a single monomeric protein [20••]. Remarkably, in spite of a small number (4–5) of substitutions made in the monomeric protein, the resulting hetero-bi-metallic shell is tightly packed and the largest opening measures less than 4 Å across. Different shell topologies and protein-metal stoichiometries could be obtained by varying the number and locations of metal-co-ordinating sites. It will be interesting to see how general and tunable this design approach is by using differently shaped monomeric building blocks and diverse metal-chelation geometries.

The strategy of utilizing metal chelation for assembly formation has led to the realization of a novel supramolecular topology, the snub cube, which belongs to a group of polyhedra known as Archimedean solids (Figure 2f) [21]. These differ from Platonic solids in that all inter-building block interfaces are not identical, thereby breaking symmetry. Malay *et al.* incorporated a surface-exposed cysteine residue that can effectively co-ordinate metal ions (Au or Hg) on each monomer of a 11-mer ring-shaped oligomeric protein building block. Cryo-electron microscopy of the resulting polyhedral assembly revealed the topology to be a snub cube. Symmetry-breaking was achieved at the molecular level by incomplete saturation of cysteine-metal ion interactions: only 10 out of the 11 cysteine residues co-ordinated metal ions. The design of well known mathematical space tessellations including other Archimedean solids and a variety of tilings (e.g. Archimedean, Penrose) represents a fundamental future challenge for the field.

Fractional-dimensional assemblies

Fractal topologies are ubiquitous at all length scales in nature — the shapes of mountain ranges, trees, ice formations, lungs, neuronal networks in brains are all fractal [22]. These shapes are characterized by a high surface area:volume ratio, which in turn allows efficient molecular capture and energy dispersal, for example in objects such as trees and lungs [23]. Fractal patterns arise from stochastic and directional self-association of building blocks under kinetic control, for example, ice crystals or window frost arise from hydrogen bonding-driven nucleated self-association of water molecules [22]. Although the design of surface-mediated fractals has been well documented for non-protein nanomaterials [24,25], these topologies can emerge in protein assemblies under certain conditions where self-association is under kinetic control [26]. For example, in their design protein oligomers mediated by supercharging of interfaces, Simon *et al.* observed the emergence of fractal-like geometry due to random association of oppositely charged proteins (Figure 2g) [27].

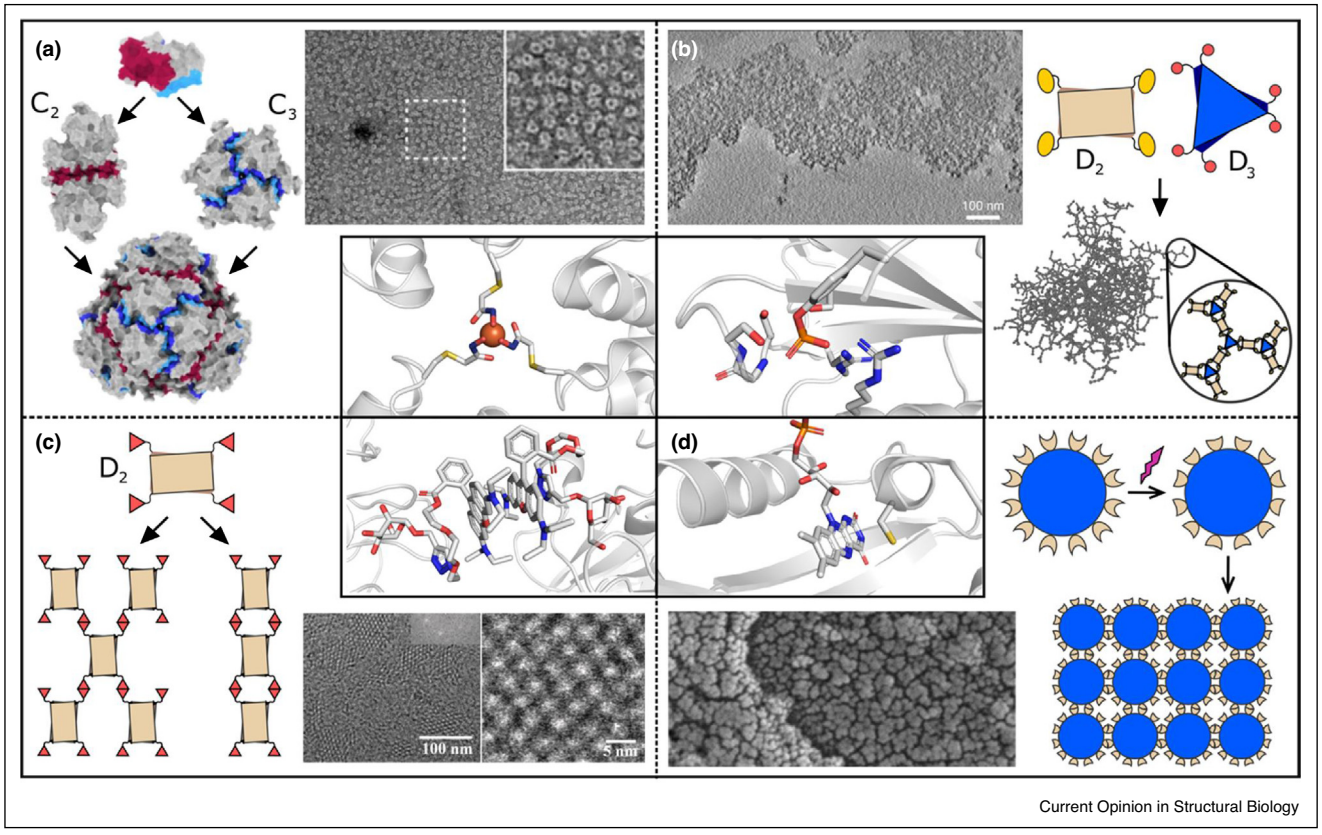
In our own work, we explored the design principles for generating fractals with protein building blocks [28••]. Based on theories of colloidal patchy particle self-association, we reasoned that multivalent association of building blocks with strong inter-particle interactions and limited conformational flexibility would enable stochastic but directional propagation, leading to fractal-like geometries [29,30]. We used dihedral symmetric building blocks (D_2 and D_3 symmetry; Figure 1c) and an engineered SH2 domain-phosphopeptide interaction [31] to encode strong inter-particle interactions. We encoded anisotropy by restricting growth along a single C2 axis of symmetry and engineering the sequences of protein to allow for multiple binding modes (Figure 1c). The resulting assemblies were fractal-like (Figure 2h) and extended over several micrometers, on surfaces and in solution, where they were efficient at macromolecular capture due to their highly 'holey' nature. This general design method can be applied to any pair of D-symmetric oligomers and should allow generation of a range of emergent topologies characterized by unique fractal dimensions, and controllable surface:volume ratios. Applications of fractal assemblies include molecular capture and filtration, and novel hierarchically organized biomaterials.

Designing stimulus-responsive assemblies

While some protein-based assemblies in nature are relatively static, for example, collagen, many, for example, cytoskeletal supramolecular polymers, are dynamically formed and dissolved in response to external stimuli. Naturally evolved cages are also often conformationally flexible and chemically tunable, allowing them to undergo motions for influx or egress of chemicals or dynamically respond to changes in environment [32]. These observations have inspired the design of assemblies that can be controlled by external environmental cues.

Several groups have utilized designed metal-protein bonds to construct dynamic assemblies that can be formed and dissolved in response to metal co-ordination (Figure 3a) [5,19,20••,21]. Treatment with chelating agents such as EDTA leads to the facile dissolution of the assembly, demonstrating its stimulus-responsive properties. Metal ion-responsive assemblies are excellent model systems as they allow detailed delineation of the mechanism of assembly formation [33••]. However, the tightly regulated homeostasis of metal ions in biological systems limits the ability to create and control these assemblies in biological environments.

Figure 3



Design of stimulus-responsive assemblies.
(a) Dimeric and trimeric metal-mediated interfaces (colored red and blue, respectively) were simultaneously designed in an asymmetric monomeric protein, leading to the reversible formation of polyhedral architecture, *top inset*. The trimeric interface was encoded by a covalently attached hydroxamate moiety, leading to preferential binding of Fe^{3+} (orange sphere) *bottom inset*, over divalent Zn^{2+} which was used to create the dimeric interface (not shown). **(b)** We generated fractal shapes, *top inset*, in a phosphorylation-dependent manner, *bottom inset*, by using an engineered SH2 domain-phosphopeptide interaction to drive self-assembly of dihedral proteins (D_2 and D_3 represented as tan rectangle and blue triangle). Kinase and phosphatase enzymes can be used to control assembly/disassembly for multiple cycles. **(c)** Small molecue interactions (orange triangles, pi-pi stacking, *top inset*) and stoichiometry was used to generate multiple supramolecular topologies, *bottom inset*, from a single protein, Lec A, represented as tan rectangle. **(d)** Engineered photoreceptors (LOV domain, *top inset*) were used to generate photocontrolled assemblies of an enzyme, *bottom inset*. All molecular interactions shown as sticks.

Phosphorylation is a common biologically utilized stimulus for controlling signal transduction. Phosphate groups are covalently attached to Tyr, Ser and Thr sidechains by kinase enzymes, and removed by the action of phosphatase enzymes [34]. These enzymes, thus, provide an opportunity to toggle between phosphorylation states of designed building block proteins. Naturally occurring binding modules such as SH2 domains that selectively bind to phosphorylated tyrosine-containing peptide motifs can be used as fusion partners to engineer protein–protein interactions that propagate to form assemblies. We utilized an engineered Fyn-SH2 binding domain that binds to its cognate phosphorylated peptide with high affinity to reversibly assemble a uniform fractal topology with high surface-area:volume ratio (Figure 3b) [28••]. This phosphorylation-dependent assembly design strategy is yet to be demonstrated for integer-dimensional topologies and in cells but promises to enable several *in vivo* in applications for dynamic, designed assemblies that respond to or report on signal transduction pathways using phosphorylation.

Protein-small molecule binding has also been used as a stimulus to control diverse assembly structures. The interactions between lectins and sugars modified with aromatic groups (to introduce pi–pi interactions between sugars) were used by Yang *et al.* to create diverse protein assemblies (Figure 3c) [35]. Impressively, with a single protein LecA as the building block, five kinds of protein assembly structures (1D nanoribbons, nanowires, 2D crystalline nanosheets, and 3D layered structures) were obtained by systematically varying the molecular structure of the assembly inducing ligands and salt concentration, demonstrating the potential of this approach

to generate diverse topologies with minimal protein engineering. In principle, these assemblies can be dissolved by adding excess ligands, however, this control was not directly demonstrated.

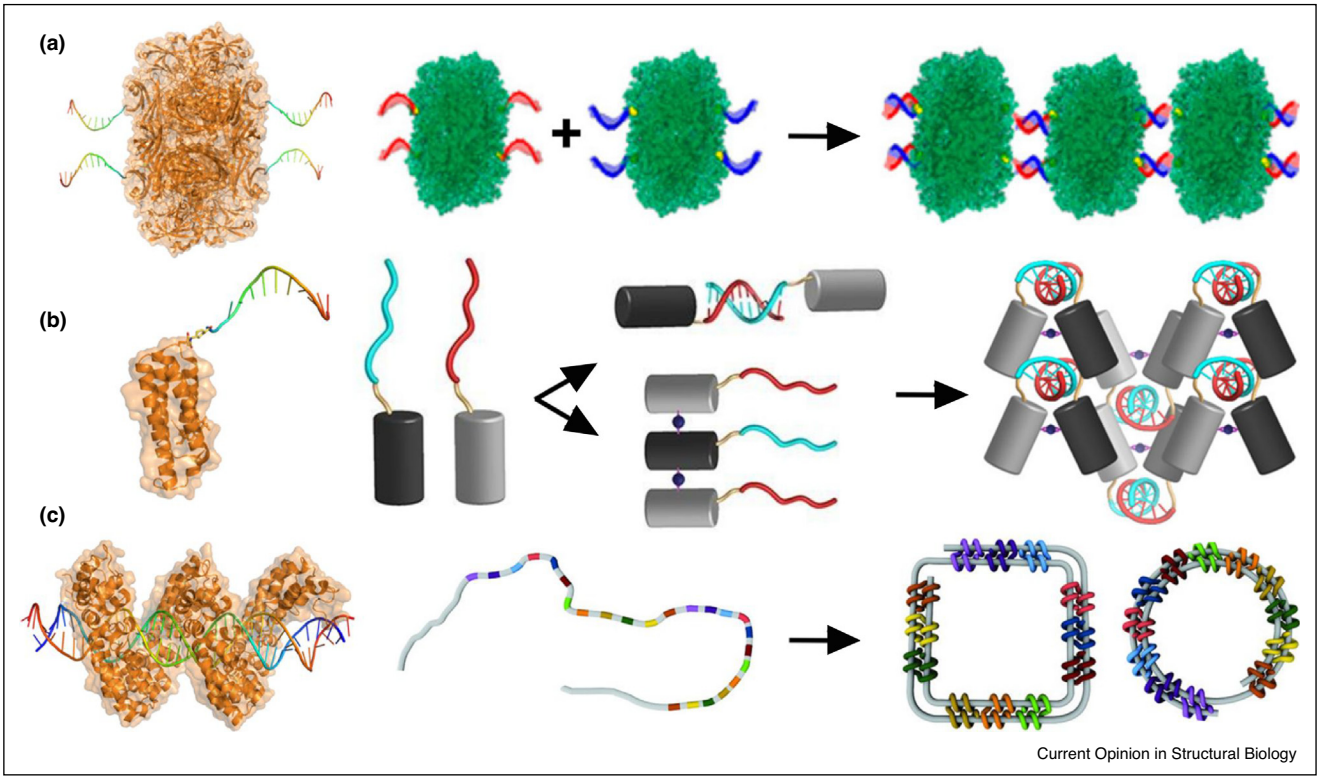
Optogenetics provides a powerful tool for controlling biological processes through the utilization of several naturally occurring photoreceptors [36]. Intense engineering efforts have led to the generation of several photocontrolled protein–protein interaction pairs with tunable affinities under dark conditions and light [37]. Taking advantage of these developments, Yu *et al.* used fusions of an oligomeric enzyme with a designed LOV domain [38] to reversibly induce enzyme assembly formation both *in vitro* and *in vivo*, although no high-resolution structures of assemblies were determined (Figure 3d) [39•]. Multiple cycles of assembly disassembly could be induced by irradiation with appropriate light wavelengths. The photocontrolled enzyme assembly strategy may find use in organizing enzymatic cascades to realize precise and reversible control of metabolic processes using optical stimuli. It should also be possible to use these or other light-sensitive moieties [40], for example, azobenzenes [41], for photocontrolling other protein-based assemblies, such as functionalized nanocages.

Thus, while much work remains in building functionally useful switchable biomaterials from the bottom up, the last few years have seen exciting successes in proof-of-principle studies for building stimulus-responsiveness into designed assemblies.

Hybrid assemblies built with protein and DNA building blocks

Some of the most complex and sophisticated assemblies in nature, for example, the ribosome, nucleosomes, feature a combination of proteins and nucleic acid building blocks. These naturally occurring assemblies have inspired efforts to build synthetic protein-DNA assemblies with complex higher order structures. McMillan and Mirkin created protein nanowires using DNA hybridization [42]. A single cysteine residues was introduced on each protomer of a tetrameric protein, and complementary DNA strands were covalently attached to two different samples of the protein (Figure 4a). Upon mixing, hybridization of the DNA

Figure 4



Hybrid protein-DNA nanostructures by design.
(a) Covalent attachement of complementart single-stranded DNA to an oligomeric protein was used to contruct DNA nanowires, **(b)** Simultaneous DNA base pairing and metal-binding was used to generate crystalline nucleoprotein lattices, **(c)** DNA-binding proteins were used in lieu of 'staple strands' to organize the supramolecular structure of DNA origami.

strands led to the formation of a supramolecular assembly with one-dimensional order. The tunability of double-stranded DNA length, binding strength, and flexibility should allow modularity in design. Another promise of this approach is the possibility of engineering multiple orthogonal DNA interaction pairs on protein surfaces by covalent attachment via different residue types to access higher-order, hierarchically assembled materials.

Subramanian *et al.* used similar covalent ssDNA-protein hybrids as building blocks for design of crystalline lattices [17•]. The protein component was a previously engineered monomeric protein that forms 2D planar assemblies using metal-chelation [43], and the introduction of DNA hybridization can be expected to help array protein layers on top of each other (Figure 4b). Ordered protein—DNA conjugate arrays were observed but only in a very narrow window of conditions; outside these conditions disordered aggregates resulted. Structural analyses of the arrays revealed not only DNA–DNA and protein–protein contacts but also pH-dependent protein-DNA contacts. These results highlight the complex phase diagrams of nucleoprotein assemblies that arise from the distinct structural and chemical properties of proteins and DNA, and the challenges associated with prospective prediction of the emergent topologies of hybrid nanostructures.

The DNA–protein interactions observed in the above hybrid assemblies are fortuitous, as the intention is to use protein–protein and/or DNA–DNA interactions to generate positional order. A complementary approach is to use protein–DNA interactions to control the emergent shape of the assembly. DNA origami has been used *in vitro* to develop a vast array of shapes, but the reliance of stapling single-stranded DNA pieces makes it incompatible with the cellular environment [44,45]. Praeotrius and Dietz used engineered and highly specific DNA-binding proteins, TALENs [46], as stapling elements to generate diverse shapes with DNA origami (Figure 4c) [47]. By their use of proteins instead of single-stranded DNA staples promises to significantly enhance the biocompatibility of DNA-based nanostructures, and lead to a plethora of novel self-assembled shapes inside cells.

Thermodynamics and kinetics in assembly design

A key consideration in assembly design is thermodynamic versus kinetic control of assembly formation. In most design approaches, the target topology (e.g. a polyhedron) is chosen *a priori* and stabilized by an appropriate choice of protein sequence by computation, domain fusion and/or metal chelation. The implicit assumption is that by sufficiently stabilizing the target state using multivalency, that is, multiple copies of individually weak protein–protein interfaces (encoding co-operativity), assembly formation will be under thermodynamic control: alternative states, that is, kinetic traps, will be disfavored because the interactions between constituents are individually weak [48]. Success in computationally designing polyhedral architectures bears this assumption out strikingly: 120 copies of two different proteins are required to correctly assemble for generating a designed icosahedron [49]. For placing assembly formation under thermodynamic control, the designed inter-subunit interactions require high directional specificity. On the other hand, if sufficiently strong and directionally flexible (yet anisotropic) inter-monomer interactions are used, kinetic traps (e.g., fractal-like structures) can themselves be stabilized by design [50].

Consideration of kinetics is likely to be crucial when multiple types of building blocks, for example, in hybrid protein-DNA assemblies, and systems with high conformational flexibility are involved in assembly formation. In these systems, many different types (protein–protein, protein–DNA, DNA–DNA) of interactions of varying strength have to be simultaneously satisfied to ensure target geometry formation. In other words, the interaction heterogeneity makes the encoding of thermodynamic control considerably more difficult when compared with pure protein-based assemblies. Indeed, assembly formation is likely under kinetic control when using metal chelation as the driving force for assembly formation: Yang and Song elegantly demonstrate how, under conditions of excess metal ions relative to chelating protein building blocks, metal unbinding is a required step before the building blocks can correctly assemble into supramolecular assembly [33••]. The explicit consideration of kinetics, although experimentally challenging, may help reduce the high degree of optimization or serendipity required for realization of target geometry in many assembly design studies.

Conclusions and future outlook

With several proof-of-principle demonstrations of protein-based supramolecular assemblies, the field is poised to build upon these successes for the generation of many functionally useful supramolecular complexes with proteins. One direction that has already seen exciting progress is the development of custom-designed vaccines by presenting antigens on designed nanocages [51••]. The bottom up design of cages enables unprecedented and precise control over epitope density, presentation geometry and accessibility. Other potential applications of designed assemblies include encapsulation of biomolecules [52] for synthetic biology [53] and structural biology studies [54], custom designed nanoreactors for controlled colocalization of metabolic processes [55,56], drug delivery [57], and molecular capture [28••] and filtration.

Key challenges for assembly design include increasing success rates in computational design (typically ~10%), successfully deploying and optimizing the functions of stimulus-responsive assemblies in the complex

environments *in vivo*, and engineering subtle topological changes beyond complete assembly disassembly, for example engineering opening-closing motions in existing cages, and generating topologically variable materials, which would require designing a system that can reversibly change dimensionality (e.g. 3D to 2D) in response to an external cue. The conformational flexibility and well known allosteric properties of proteins make them attractive building block candidates for generating topologically tunable biomaterials.

Finally, for protein assembly design to become useful for downstream applications, the ability to generate specifications from a target mesoscale topological description will need to be developed, for example, a user might desire a 100 nm cage with 10 A pores that undergo open-close motions in response to red light. Building hybrid protein-based stimulus-responsive materials that involve interfaces with supramolecular organics and polymers also represents a key growth area for the field. With the developments highlighted above (and several that we could not include due to space restrictions), the field of protein-based supramolecular assembly design is poised to see many exciting successes in the coming years.

Conflict of interest statement

Nothing declared.

Acknowledgement

We thank NSF (MCB1716623 to SDK) for funding.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Fletcher DA, Mullins RD: **Cell mechanics and the cytoskeleton**. *Nature* 2010, **463**:485-492.
2. Yeates TO: **Geometric principles for designing highly symmetric self-assembling protein nanomaterials**. *Annu Rev Biophys* 2017, **46**:23-42.
3. King NP, Lai YT: **Practical approaches to designing novel protein assemblies**. *Curr Opin Struct Biol* 2013, **23**:632-638.
4. Zhang J, Zheng F, Grigoryan G: **Design and designability of protein-based assemblies**. *Curr Opin Struct Biol* 2014, **27**:79-86.
5. Churchfield LA, Tezcan FA: **Design and construction of functional supramolecular metalloprotein assemblies**. *Acc Chem Res* 2019, **52**:345-355.
6. Huang PS, Boyken SE, Baker D: **The coming of age of de novo protein design**. *Nature* 2016, **537**:320-327.
7. Mateu MG: **Assembly, stability and dynamics of virus capsids**. *Arch Biochem Biophys* 2013, **531**:65-79.
8. Lopez MG, Diez M, Alfonso V, Taboga O: **Biotechnological applications of occlusion bodies of Baculoviruses**. *Appl Microbiol Biotechnol* 2018, **102**:6765-6774.
9. Fagan RP, Fairweather NF: **Biogenesis and functions of bacterial S-layers**. *Nat Rev Microbiol* 2014, **12**:211-222.
10. Murr MM, Morse DE: **Fractal intermediates in the self-assembly of silicatein filaments**. *Proc Natl Acad Sci U S A* 2005, **102**:11657-11662.
11. Nogales E: **Structural insights into microtubule function**. *Annu Rev Biochem* 2000, **69**:277-302.
12. Norn CH, Andre I: **Computational design of protein self-assembly**. *Curr Opin Struct Biol* 2016, **39**:39-45.
13. •• Shen H, Fallas JA, Lynch E, Sheffler W, Parry B, Jannetty N, Decarreau J, Wagenbach M, Vicente JJ, Chen J *et al.*: **De novo design of self-assembling helical protein filaments**. *Science* 2018, **362**:705-709

Two protein–protein interfaces are simultaneously designed to demonstrate the formation of helical filaments with monomeric protein building blocks.

14. •• Hughes SA, Wang F, Wang S, Kreutzberger MAB, Osinski T, Orlova A, Wall JS, Zuo X, Egelman EH, Conticello VP: **Ambidextrous helical nanotubes from self-assembly of designed helical hairpin motifs**. *Proc Natl Acad Sci U S A* 2019, **116**:14456-14464

Variants of tandem repeat proteins were engineered to assemble into helical filaments.

15. • Zhang SQ, Huang H, Yang J, Kratochvil HT, Lolicato M, Liu Y, Shu X, Liu L, DeGrado WF: **Designed peptides that assemble into cross-alpha amyloid-like structures**. *Nat Chem Biol* 2018, **14**:870-875

This study demonstrates the formation of a novel topology (cross-alpha fibers) using the rules of coiled-coil packing.

16. Zhou K, Chen H, Zhang S, Wang Y, Zhao G: **Disulfide-mediated reversible two-dimensional self-assembly of protein nanocages**. *Chem Commun (Camb)* 2019, **55**:7510-7513.
17. • Zhang L, Bailey JB, Subramanian RH, Groisman A, Tezcan FA: **Hyperexpandable, self-healing macromolecular crystals with integrated polymer networks**. *Nature* 2018, **557**:86-91

Large, stimulus-responsive crystals were engineered by arraying ferritin nanocages in a matrix of expandable polymers, leading to the property of 'self-healing' depending on solvent (water) and Ca^{2+} content.

18. •• Pyles H, Zhang S, De Yoreo JJ, Baker D: **Controlling protein assembly on inorganic crystals through designed protein interfaces**. *Nature* 2019, **571**:251-256

Proteins were designed to form nanoscale patterns on an inorganic mica surface. Carboxylate groups of amino acid sidechains were strategically placed to make electrostatic interactions with K+ ions.

19. Cristie-David AS, Marsh ENG: **Metal-dependent assembly of a protein nano-cage**. *Protein Sci* 2019, **28**:1620-1629.
20. •• Golub E, Subramanian RH, Esselborn J, Alberstein RG, Bailey JB, Chiong JA, Yan X, Booth T, Baker TS, Tezcan FA: **Constructing protein polyhedra via orthogonal chemical interactions**. *Nature* 2020, **578**:172-176 http://dx.doi.org/10.1038/s41586-019-1928-2

Metal chelation in multiple geometries was enforced by conjugating small molecules to a monomeric building block protein, leading to the formation of polyhdera dependent on two metal ions.

21. Malay AD, Miyazaki N, Biela A, Chakraborti S, Majsterkiewicz K, Stupka I, Kaplan CS, Kowalczyk A, Piette B, Hochberg GKA *et al.*: **An ultra-stable gold-coordinated protein cage displaying reversible assembly**. *Nature* 2019, **569**:438-442.
22. Stanley HE, Meakin P: **Multifractal phenomena in physics and chemistry**. *Nature* 1988, **335**:405-409.
23. Fairbanks MS, McCarthy DN, Scott SA, Brown SA, Taylor RP: **Fractal electronic devices: simulation and implementation**. *Nanotechnology* 2011, **22**:365304.
24. Tikhomirov G, Petersen P, Qian L: **Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns**. *Nature* 2017, **552**:67-71.
25. Shin S, Gu ML, Yu CY, Jeon J, Lee E, Choi TL: **Polymer self-assembly into unique fractal nanostructures in solution by a one-shot synthetic procedure**. *J Am Chem Soc* 2018, **140**:475-482.
26. Lomander A, Hwang W, Zhang S: **Hierarchical self-assembly of a coiled-coil peptide into fractal structure**. *Nano Lett* 2005, **5**:1255-1260.

27. Simon AJ, Zhou Y, Ramasubramani V, Glaser J, Pothukuchy A, Gollihar J, Gerberich JC, Leggere JC, Morrow BR, Jung C *et al.*: **Supercharging enables organized assembly of synthetic biomolecules**. *Nat Chem* 2019, **11**:204-212.

28. •• Hernandez NE, Hansen WA, Zhu D, Shea ME, Khalid M, Manichev V, Putnins M, Chen M, Dodge AG, Yang L *et al.*: **Stimulus-responsive self-assembly of protein-based fractals by computational design**. *Nat Chem* 2019, **11**:605-614

Hyperbranched fractal-like patterns that are phosphorylation dependent were constructed from the bottom up by utilizing a combination of symmetric building blocks, and SH2 domain-peptide interactions. Control over fractal dimension and other topological features could be achieved, and designed assemblies served as molecular sponges for capturing protein cargo.

29. Yan Y, Huang J, Tang BZ: **Kinetic trapping - a strategy for directing the self-assembly of unique functional nanostructures**. *Chem Commun (Camb)* 2016, **52**:11870-11884.

30. Nicolas-Carlock JR, Carrillo-Estrada JL, Dossetti V: **Fractality a la carte: a general particle aggregation model**. *Sci Rep* 2016, **6**:19505.

31. Kaneko T, Huang H, Cao X, Li X, Li C, Voss C, Sidhu SS, Li SS: **Superbinder SH2 domains act as antagonists of cell signaling**. *Sci Signal* 2012, **5**:ra68.

32. Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM: **Protein-based organelles in bacteria: carboxysomes and related microcompartments**. *Nat Rev Microbiol* 2008, **6**:681-691.

33. •• Yang M, Song WJ: **Diverse protein assembly driven by metal and chelating amino acids with selectivity and tunability**. *Nat Commun* 2019, **10**:5545

Genetic incorporation of non-natural amino acid incorporation was used to guide the formation of diverse topologies mediated by metal chelation. Elegant studies of assembly formation kinetics revealed the intricate interplay of kinetics and thermodynamics in metal-mediated assembly formation.

34. Pawson T, Scott JD: **Protein phosphorylation in signaling–50 years and counting**. *Trends Biochem Sci* 2005, **30**:286-290.

35. Yang G, Ding HM, Kochovski Z, Hu R, Lu Y, Ma YQ, 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**:10691-10695.

36. Fenno L, Yizhar O, Deisseroth K: **The development and application of optogenetics**. *Annu Rev Neurosci* 2011, **34**:389-412.

37. Ziegler T, Moglich A: **Photoreceptor engineering**. *Front Mol Biosci* 2015, **2**:30.

38. Guntas G, Hallett RA, Zimmerman SP, Williams T, Yumerefendi H, Bear JE, Kuhlman B: **Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins**. *Proc Natl Acad Sci U S A* 2015, **112**:112-117.

39. • Yu Q, Wang Y, Zhao S, Ren Y: **Photocontrolled reversible self-assembly of dodecamer nitrilase**. *Bioresour Bioprocess* 2017, **4**:36

Assembly disassembly of an enzyme complex was shown to be controllable by light using fusions of light-dependent protein–protein interaction domains.

40. Pianowski ZL: **Recent implementations of molecular photoswitches into smart materials and biological systems**. *Chemistry* 2019, **25**:5128-5144.

41. Blacklock KM, Yachnin BJ, Woolley GA, Khare SD: **Computational design of a photocontrolled cytosine deaminase**. *J Am Chem Soc* 2018, **140**:14-17.

42. McMillan JR, Mirkin CA: **DNA-functionalized, bivalent proteins**. *J Am Chem Soc* 2018, **140**:6776-6779.

43. Brodin JD, Ambroggio XI, Tang C, Parent KN, Baker TS, Tezcan FA: **Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays**. *Nat Chem* 2012, **4**:375-382.

44. Rothemund PW: **Folding DNA to create nanoscale shapes and patterns**. *Nature* 2006, **440**:297-302.

45. Hong F, Zhang F, Liu Y, Yan H: **DNA origami: scaffolds for creating higher order structures**. *Chem Rev* 2017, **117**:12584-12640.

46. Joung JK, Sander JD: **TALENs: a widely applicable technology for targeted genome editing**. *Nat Rev Mol Cell Biol* 2013, **14**:49-55.

47. Praetorius F, Dietz H: **Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes**. *Science* 2017, **355**.

48. Baker D: **What has de novo protein design taught us about protein folding and biophysics?** *Protein Sci* 2019, **28**:678-683.

49. Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP *et al.*: **Accurate design of megadalton-scale two-component icosahedral protein complexes**. *Science* 2016, **353**:389-394.

50. Lomakin A, Asherie N, Benedek GB: **Aeolotopic interactions of globular proteins**. *Proc Natl Acad Sci U S A* 1999, **96**:9465-9468.

51. •• 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**:1420-1431.e17

Designed protein nanocages were engineered to display a a surface antigen from a virus. High-density and uniform antigen presentation led to the induction of a robust immune response, highlighting a key application of designed protein assemblies.

52. Sasaki E, Bohringer D, van de Waterbeemd M, Leibundgut M, Zschoche R, Heck AJ, Ban N, Hilvert D: **Structure and assembly of scalable porous protein cages**. *Nat Commun* 2017, **8**:14663.

53. Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, Bale JB, Ke S, Lenz GH, Yehdego A *et al.*: **Evolution of a designed protein assembly encapsulating its own RNA genome**. *Nature* 2017, **552**:415-420.

54. Liu Y, Gonen S, Gonen T, Yeates TO: **Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system**. *Proc Natl Acad Sci U S A* 2018, **115**:3362-3367.

55. Chen RP, Blackstock D, Sun Q, Chen W: **Dynamic protein assembly by programmable DNA strand displacement**. *Nat Chem* 2018, **10**:474-481.

56. Jordan PC, Patterson DP, Saboda KN, Edwards EJ, Miettinen HM, Basu G, Thielges MC, Douglas T: **Self-assembling biomolecular catalysts for hydrogen production**. *Nat Chem* 2016, **8**:179-185.

57. Schoonen L, van Hest JC: **Functionalization of protein-based nanocages for drug delivery applications**. *Nanoscale* 2014, **6**:7124-7141.