Review

Special Issue – Synthetic Cell Biology

Principles for designing ordered protein assemblies

Yen-Ting Lai[1], Neil P. King[2], and Todd O. Yeates[3,4]

[1] Biomedical Engineering Interdepartmental Degree Program, University of California, Los Angeles, CA, USA
[2] Department of Biochemistry, University of Washington, Seattle, WA, USA
[3] Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA
[4] University of California Los Angeles (UCLA)–Department of Energy Institute for Genomics and Proteomics, University of California, Los Angeles, CA, USA

In nature, many proteins have evolved to have self-complementary shapes. This drives them to assemble into supramolecular structures, sometimes of great complexity, and often carrying out sophisticated cellular functions. Designing novel proteins that can self-assemble into similarly complex structures is a longstanding goal in bioengineering. New ideas, combined with continually improving computer algorithms, are making it possible to advance on that goal, bringing wide-ranging applications in synthetic biology within reach. Prospective applications range from vaccine design to molecular delivery to bioactive materials. Recent strategies and examples of successfully designed protein cages, layers, and crystals are reviewed.

Protein assemblies as a synthetic biology goal

The emerging research area of synthetic biology seeks to recreate various complex phenomena exhibited by biological systems, especially at the molecular level [1]. The phenomena of interest are often characterized by a high degree of order – either in time or space. The emergent behavior or 'output' of synthetic systems can be considerably more complex than the behavior of the individual components [2]. Often, this arises from the introduction of non-trivial or self-referring interactions between components. For example, if rather than simply A affecting B, instead B also affects C, and C affects A, then an output behavior may arise where molecular concentrations oscillate in time [3]. Likewise, if rather than simply A binding to B, instead A binds to itself in multiple ways, surprisingly large and complex molecular assemblies can arise, leading to spatial organization of various types, such as compartmentalization or long-range propagation of forces by rigid structures. We concern ourselves here with strategies for designing protein molecules that self-associate to produce large, complex assemblies with potential synthetic biology applications.

Diverse efforts in the area of designing protein-based assemblies and materials can be divided into two groups: stochastic and deterministic [4,5]. The stochastic group encompasses several design strategies where the self-assembling protein molecule is highly flexible [4,6]. When a highly flexible molecule self-associates to form a higher-order assembly, the result is typically an extended and geometrically irregular material. Such network or mesh-like materials have interesting bulk properties, which can sometimes be modulated in useful ways [7]. The second group, where the assembly behavior is intended to be deterministic, encompasses those strategies aimed at producing specific 3D structures [5]. These structures, which may be finite in size (e.g., clusters or cages) or indefinite in extent (e.g., arrays or crystals), can be built with atomic level features in mind.

Early work on designing geometrically specific protein assemblies focused on filamentous structures as design targets. That early focus reflects the relatively simple design requirements for filamentous assemblies: a single self-associating interface can produce end-to-end polymerization. Specific well-studied self-associating protein motifs have been a rich source of building blocks for designing filamentous structures. Helical coiled-coils have been especially useful [8,9]. Cyclic polypeptides composed of β-strand-preferring amino acids have provided another self-associating motif, in this case leading to rigid tubular assemblies [10]. Variations on filamentous designs have sought to produce more complex patterns, such as branching [11], but the end-to-end polymerization strategies central to filament design do not extend easily to the problem of creating highly specific 3D architectures.

In this review we focus on strategies for designing proteins that self-assemble to give defined structures with complex architectures, including cages and extended 2D and 3D crystalline arrays.

Underlying principles

In nature, wherever supramolecular structures are built up by the assembly of multiple copies of the same subunit (or similar subunits), the subunits are nearly always assembled in a symmetrical fashion [12,13]. The reason for this was anticipated as early as 1956 by Crick and Watson [14]: symmetric assemblies require fewer distinct kinds of specific interaction interfaces compared to asymmetric assemblies. It is natural then that efforts to design ordered protein assemblies should rely on principles of symmetry. We articulate three connected ideas that we believe are important to permit full exploitation of symmetry-based

Corresponding author: Yeates, T.O. (yeates@mbi.ucla.edu).
Keywords: protein design; symmetry; cages; biomaterials; nanomaterials; assembly.

0962-8924/$ – see front matter http://dx.doi.org/10.1016/j.tcb.2012.08.004 Trends in Cell Biology, December 2012, Vol. 22, No. 12

approaches for designing self-assembling molecular systems:

(i) A symmetric molecular assembly, whether it is a finite structure or an extended array, is characterized by its symmetry group. A symmetry group is an exact mathematical idea that expresses the complete set of spatial operations that interrelate a set of individual components, in this case a set of structurally identical protein subunits in a 3D assembly. For example, a 24-subunit assembly built with the symmetry of a cube (referred to as octahedral symmetry) is described by a symmetry group with 24 spatial operations. Every pair of subunits is related by some operation in the group. For chiral biological molecules, such as proteins, these operations must be rotations, potentially combined with translations, but excluding reflections.

(ii) Any interface formed between two subunits in a symmetric assembly corresponds to one of the operations of the symmetry group (*vide supra*).

(iii) For a hypothetical, symmetric constellation of subunits (whether finite or indefinite) to constitute a plausible, physically connected assembly – in other words, to not be two or more disjoint sub-complexes – the following condition is both necessary and sufficient. Taking any individual subunit as a reference, the distinct interface types it uses to contact its neighbors comprise a subset of the operations of the complete symmetry group, and this subset of operations must be capable of generating the full symmetry group. That is, repeated combinations of this typically small subset of operations must be able to reform the whole symmetry group. In physical terms, this is equivalent to stating that it must be possible to trace a path from any one molecule to any other molecule in the assembly through the contacts between molecules in the assembly (Box 1). Otherwise, the collection of molecules would be disjoint.

The first two ideas are relatively well-known to those familiar with crystallography or molecular symmetry, but the third is less obvious. It was first articulated in the context of crystal symmetries [15], and then in the context of designed protein assemblies [16]. Particularly surprising, and important from a design perspective, was the realization that large, complex symmetries (including some 3D crystal symmetries) could be generated using only two distinct symmetry elements [16]. This translates to the idea that if one can design two distinct, geometrically specific, self-associating interfaces into a single molecule, then a wide range of assembly architectures can be realized. Although more than the minimum number of distinct contact types can be introduced in a design – and large natural assemblies such as viral capsids almost always exhibit more than the minimum number of distinct contact types [17] – the minimum contact number (which is just two for many cases) establishes an important design principle.

Design strategies and successes

Owing largely to the complexity of protein molecules, and our incomplete understanding of the rules by which they fold and recognize each other, designing complex protein assemblies has been a difficult challenge. However, efforts along multiple lines are beginning to bear fruit. Recalling the discussions above on the design requirement of introducing two (or more) modes of self-association into a single protein molecule, varied approaches to the problem of designing self-assembling proteins can be grouped according to the strategy used to satisfy this central requirement. Different strategies rely to different degrees on natural (or native-like) protein–protein interfaces versus novel interfaces created by amino acid sequence design (Figure 1).

Fusion of natural oligomers

An early idea for introducing two self-associating interfaces into a single protein molecule emerged at a time when the prospects for designing novel protein–protein interfaces by computational methods were limited. In 2001, we proposed that genetically fusing two different, naturally oligomeric domains into a single protein chain could satisfy the design requirement of combining two self-associating motifs [16]. The problem was how to dictate, or at least predict in advance, what the relative orientation would be between two genetically fused domains; free backbone rotations occur at a point of fusion, and this would make the final geometry unpredictable. The solution for how to predict the relative orientation in advance was to use oligomeric domains that began or ended in an α-helix, such that the protein backbone (and any α-helix-preferring amino acids introduced as a linker) might adopt an unbroken α-helix running from within one oligomeric domain into the next. In this way, pairs of oligomeric domains of known structure would be combined in hypothetical, predicted arrangements, and a search could be made for pairs that would satisfy specific geometric rules for constructing different architectures such as cages or crystals. The same paper laid out geometric rules for how the symmetry axes of the component domains would have to be oriented relative to each other to obtain various architectures ([16], Box 1). That initial set of geometric rules only covered combinations of dimers and trimers because the protein structure database at that time contained relatively few proteins with both higher oligomeric symmetry and subunits ending in helices [16]. A tremendous range of assembly architectures can be achieved using higher symmetry building blocks; those possibilities have not been completely enumerated yet.

The oligomeric fusion method was first used in the design of protein filaments – a relatively easy design target – and a 12-subunit molecular cage with a tetrahedral shape, representing the first protein construction of its kind. However, that initially designed protein sequence formed cage-like assemblies whose sizes were too heterogeneous to characterize in atomic detail; crystals could not be obtained [16]. This barrier was surmounted in recent work. When the original design was revisited, and two amino acid changes were made based on a visual identification of potential steric conflicts, a homogenous 12-subunit cage was obtained and crystallized [18]. The designed cage is roughly 16 nm in diameter and contains a central opening about 5 nm in diameter (Figure 2a,b). Interestingly, despite an overall match to the design, the observed assembly exhibited significant deviations (about 8 Å root-mean-square deviation, RMSD) from perfect symmetry [18].

Box 1. Principles for designing ordered protein assemblies

A general requirement for engineering complex self-assembling structures is to create a protein molecule bearing multiple distinct self-associating interfaces (Figure I, top). This can be achieved through the use of natural oligomeric proteins or computational sequence design, separately or in combination. The multiple interfaces must be combined according to specific geometric rules [16] if defined structures are to be created (Figure I, bottom). The geometric design requirements are given for some example assemblies. In each case, the two types of symmetry elements are noted in bold, together with the angle they must form. The symmetries of the resulting assemblies are given (T: tetrahedral; O: octahedral; I: icosahedral). For 2D layers and 3D crystals (not shown), rules for constructing the full range of possible symmetries have not been articulated yet [16], though specific instances of 2D layers and 3D crystals have been successfully engineered [22,35,36].

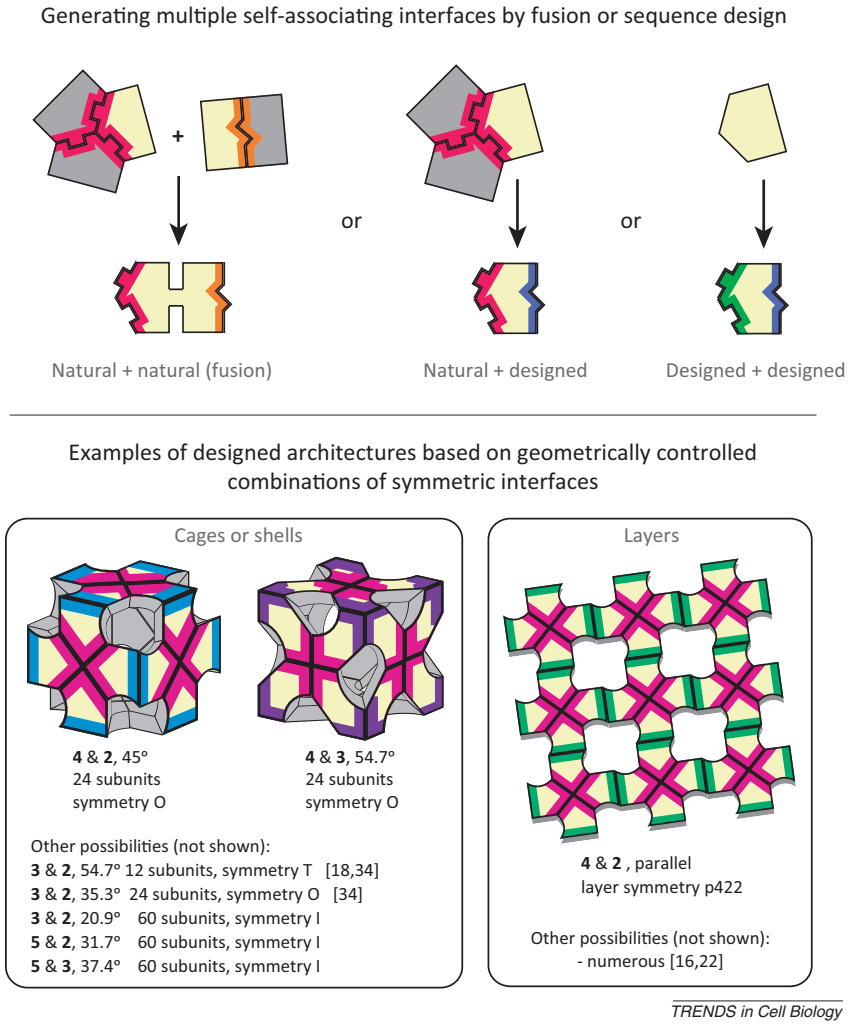

Figure I. Strategies and geometric rules for combining multiple protein interfaces.

Variations on the fusion approach have been introduced by others. In one study [19], natural oligomers were used to supply the two self-associating interfaces required for an extended 2D array of protein molecules, but in this case the components were connected by binding to biotin instead of by genetic fusion; one of the oligomers was streptavidin, a D_2 tetramer, and the other was aldolase (RhuA), a C_4 tetramer. Arrays of only limited extent were obtained, probably owing to flexibility of the biotin linker and imperfect control over the relative orientation of the component oligomers. In another study, two types of building blocks were generated by fusing PDZ domains or PDZ-binding peptides to the termini of a tetrameric (D_2 symmetry) superoxide reductase [20]. These building blocks formed linear filaments upon mixing. Smaller protein motifs have also been used in oligomeric fusion approaches. One study connected a trimeric coiled-coil to a pentameric coiled-coil with the goal of generating 60-subunit icosahedral particles [21]. A disulfide bond was introduced to force a bend between the two helical components; an angle of precisely 37.4° is required to satisfy icosahedral symmetry. The particles formed by this design appeared polymorphic but were roughly the intended size of 16 nm.

A notable extension to the original oligomer fusion strategy was recently introduced [22] (Figure 2c,d). These authors noted that, if higher-order oligomers are used as

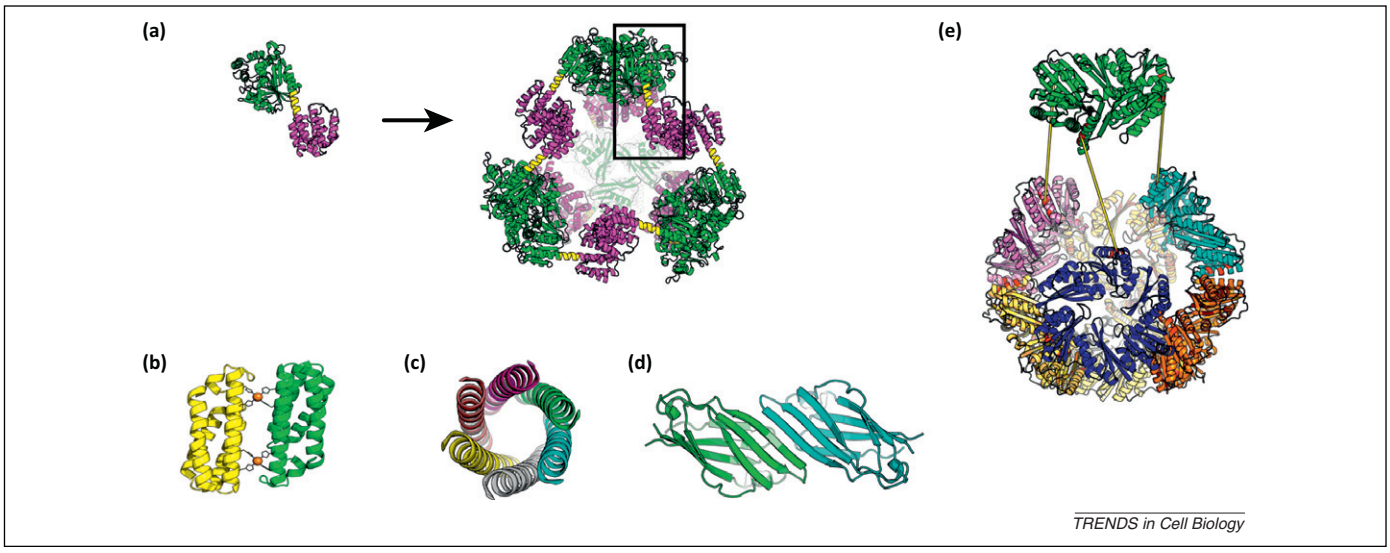

Figure 1. Strategies for introducing a new oligomerization interface. **(a)** Oligomeric fusion strategy. Two different oligomeric proteins can be fused together to generate two oligomerization interfaces within a single protein subunit, thereby driving the assembly of complex structures [16]. A single protein chain within the complex is shown in the black rectangle. In the example shown, the green domain derives from a natural trimer and the magenta domain derives from a natural dimer. Linking residues are yellow. **(b)** New interfaces generated through metal-binding. Multiple histidine residues incorporated into the side of an exposed α-helix can constitute a new dimeric interface [30]. Two different chains are shown in green and yellow, histidine residues are shown in black lines, and metal ions are shown as orange spheres. **(c)** α-Helix-based oligomerization. α-Helices can be designed to form different oligomerization states based on well-studied coiled-coil motifs [32]. Different chains are shown in different colors. **(d)** β-Sheet-based oligomerization. The open edge of a β-sheet can be used as a site for designing a dimeric interaction [31]. Two different chains are colored in green and cyan. **(e)** *De novo* design of a new interface. A new interface can be introduced into a natural oligomer of relatively low initial symmetry to generate higher-order assemblies [34]. In the case illustrated, multiple copies of a natural trimer are shown in different colors. One trimer, shown in green, is lifted from its assembled position. The *de novo* designed interaction patches are shown in red. The designed interactions between the green subunit and other subunits are indicated by thin lines.

the components, the two oligomers to be genetically fused can be aligned at an axis of symmetry that they both share. In such a scenario, there will be two (or more) polypeptide linkers instead of one running between adjacent oligomers. The requirement for a rigid α-helical linker could then be removed, with the necessary orientational control resulting instead from the multiple chain connections. This symmetry-matching fusion protein strategy successfully generated a linear filament and 2D arrays. A notable success rate of 40% was achieved when constructing 2D arrays, and one of the arrays exhibited exceptionally good long-range order. The design of 3D crystals based on this strategy led to large solid aggregates having crystal-like morphology, but the long-range order required for crystallographic analysis has not been reported yet.

In theory, the shared symmetry axis method can be used to create 2D and 3D arrays obeying a variety of different symmetries, the full range of which has not been articulated yet. However, an unavoidable constraint of the method is that only extended materials (e.g., 2D arrays and 3D crystals) can be generated, but not finite structures such as molecular cages.

Interface design

Computer programs for designing protein–protein interactions are becoming increasingly powerful [23–25], making it possible to design sequence mutations that drive specific modes of symmetric self-association. This has opened up more direct strategies for designing large protein assemblies and extended materials. The most conservative approaches for designing such large assemblies, including most of the successful experiments reported so far, rely on a natural (or native-like) oligomerization motif to provide one of the modes of self-association, and computer algorithms are then used to introduce an additional interface (or interfaces) (Box 1). This approach minimizes the number of novel interfaces that must be successfully designed computationally. Strategies that combine a natural oligomeric interface with the computational design of an additional interaction were presaged in experiments [26] in which assemblies such as double-ring structures were generated by introducing relatively simple interfacial features such as a hydrophobic patch into simpler, single-ring, natural protein assemblies. The success rate of that strategy was higher when the starting structures contained more subunits (for example, C4 compared to C2), presumably owing to the higher multiplicity of the newly introduced interaction sites. In addition, the observed assemblies typically showed substantial deviations from the intended structures, presumably reflecting the limited geometric precision provided by the sequence design strategy.

Numerous sequence design strategies for promoting self-association have been demonstrated, with strategies that deliver the highest geometric specificity placing the greatest demand on the sequence design process (Figure 1b–e). At one end of the spectrum, a simple sequence element such as two histidines at positions i and i+4 on the exposed surface of an α-helix has proved to be a straightforward approach for driving protein self-association in the presence of metals [27–29]. However, the specific geometries of those associations have been hard to predict [30]. Higher levels of geometric specificity have been demonstrated in the design of self-associating interactions involving an exposed β-strand in one case [31], and α-helical bundles in several other studies [32,33]. At the

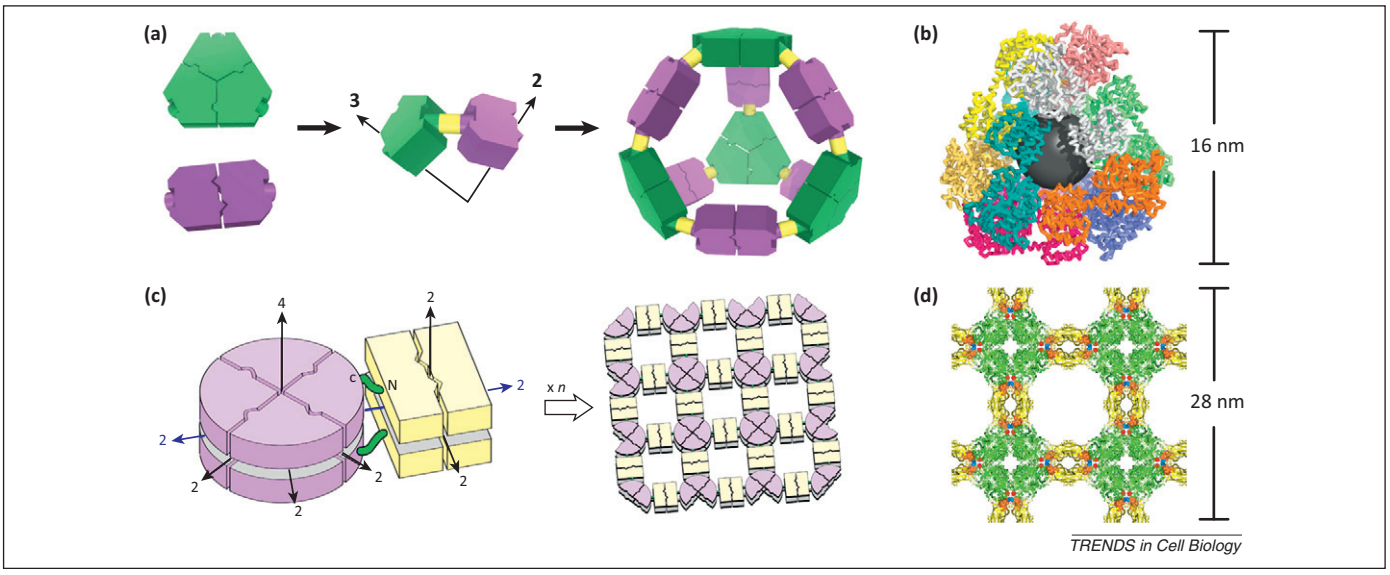

Figure 2. Designed assemblies based on two types of oligomer-fusion strategies. **(a)** Fusion through a semi-rigid helix linker. Two natural oligomeric domains (green and purple; left) are held in a predetermined orientation by genetic fusion of the two domains and an α-helical linker (yellow; middle) [16]. Multiple copies of the fusion protein can then self-assemble into a well-defined shape (right). **(b)** Crystal structure of a successful example by using the helix-linker fusion strategy [18]. Twelve different chains in the assembly are colored differently; a central sphere of about 5 nm is shown in black. **(c)** Fusion by common symmetry axis. Short and relatively flexible linkers (green; left) are used to fuse natural oligomeric domains (pink and yellow) such that the shared symmetry axes tend to align with each other (blue arrow; left) [22]. Multiple subunits can then self-assemble into extended arrays (right). **(d)** Molecular model of a layer assembled by the common-symmetry-axis fusion strategy, which was consistent with electron micrographs of the resulting material. Panels (a) and (b) adapted, with permission, from [18]. Panel (c) adapted, with permission, from [55]. Panel (d) adapted, with permission, from [22].

extreme end, the sequence design of entirely novel, geometrically specific self-associating interfaces (i.e., not modeled on well-characterized natural motifs) has been demonstrated in a newly published study [34]. Together, these varied modes of interface design have been exploited in a recent flurry of experiments demonstrating the creation of complex assemblies and extended materials.

Metal site design, combined with additional interface design, was used to create a small protein that self-assembles into a series of different architectures [35]. A cytochrome protein that had been engineered by the introduction of coordinating histidine residues to form a metal-induced dimer was stabilized in that specific geometrical configuration by further mutations using computer-aided interface design. In the resulting configuration, the two symmetric metal-binding motifs each had one open coordination site to promote a further twofold association between dimers. Typically, a tail-to-tail association between head-to-head dimers tends to produce a filament or strip of molecules. In this study [35] it appears that different experimental conditions promoted different arrangements and associations of such strips, such as side-by-side associations to give sheets, or helical wrappings to give tubular structures (Figure 3). Assemblies produced by the metal-mediated strategy have the potential to be fine-tuned through the adjustment of the protein:metal ratio or pH. At the present stage, the assembly outcomes in metal-mediated designs must be established experimentally, and the generation of predictable structures constitutes an important future challenge.

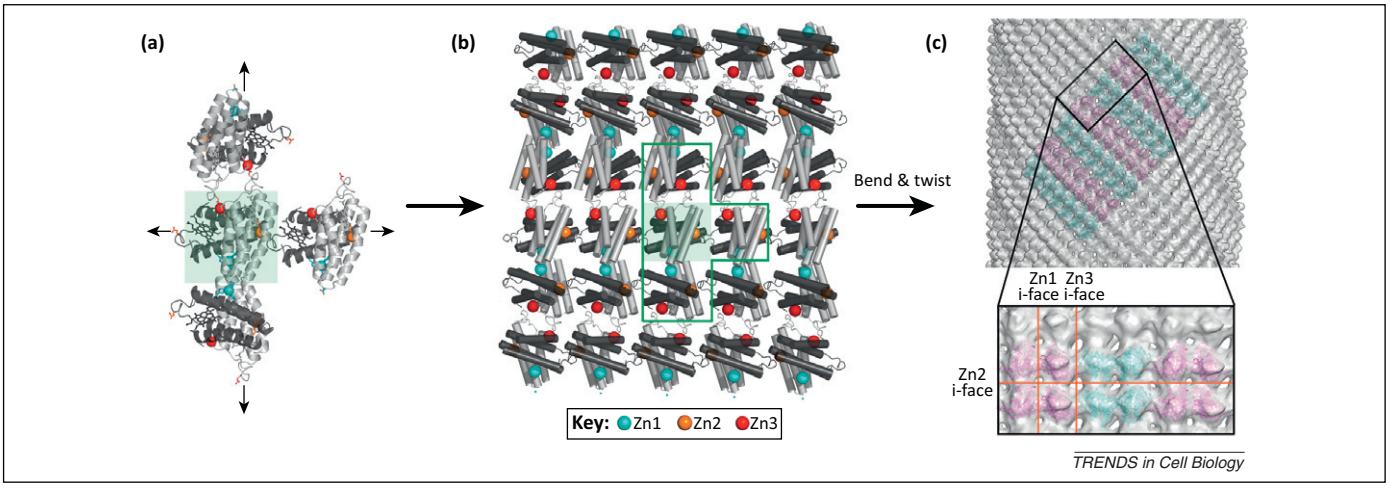

Figure 3. Layers and nanotubes by metal-mediated assembly. **(a)** Zinc was used to trigger the assembly of protein subunits with designed di-histidine motifs in two perpendicular directions into a layer, as shown in **(b)** [35]. The equivalent part of (a) is shown in the green box. Three differently located zinc ions are colored in cyan, orange, and red. **(c)** Bending and twisting around the binding site for zinc #3 caused the layer to curl into a nanotube. Adapted, with permission, from [35].

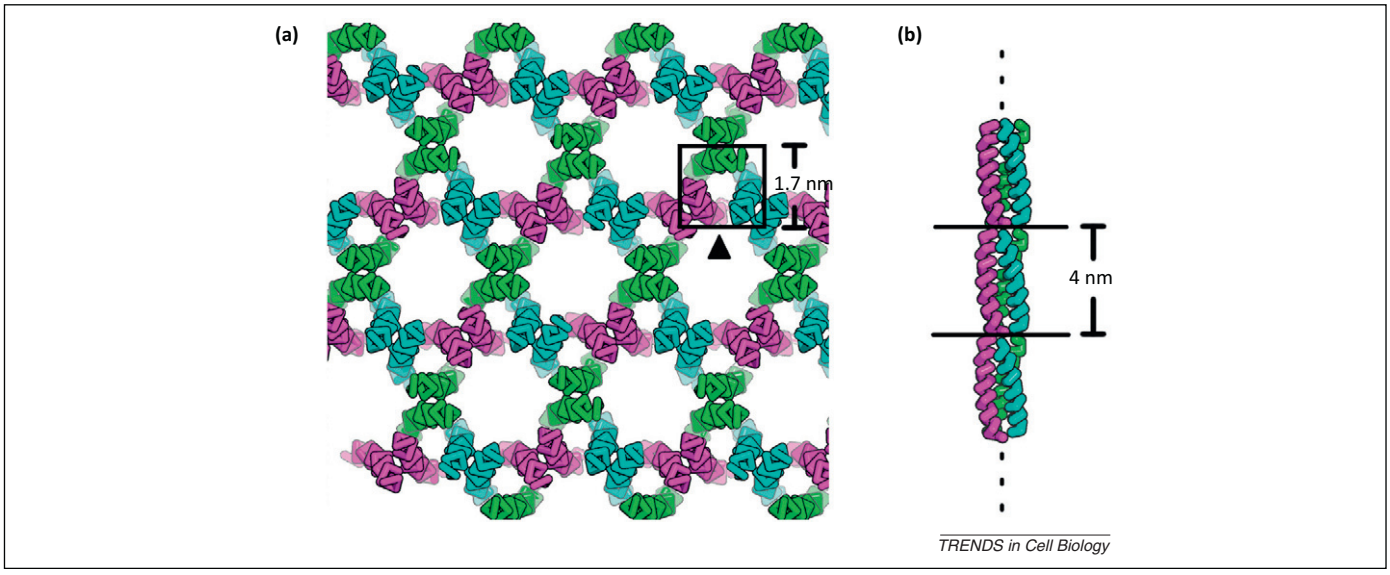

Figure 4. A designed 3D protein crystal using coiled-coil building blocks. **(a)** An assembly obeying P6 crystal symmetry was designed using a parallel 3-helix coiled-coil as the starting motif [36]. Computational sequence design was used to promote lateral and vertical contacts between trimers. Distinct chains in the trimer are colored in green, cyan, and magenta. Minor breakage of the intended P6 symmetry is visible. **(b)** The vertical stacking of layers is illustrated in a perpendicular view from the position of the black triangle in (a). The ends of the helices in one layer form hydrogen bonds with helices in neighboring layers, forming semi-continuous helices along the longitudinal crystal dimension. The layer shown in (a) is located between the two solid lines.

The successful design of a self-assembling 3D protein crystal was recently demonstrated [36]. First, a protein backbone conformation based on a well-studied protein motif, a three-helix bundle, was placed in a chosen crystal lattice. In this case, the template trimer was placed on the threefold symmetry axis of a P6 crystal unit cell (Figure 4). Favorable unit cell spacings and trimer orientations were then sampled, after which the sequence of the protein was optimized based on its interactions with symmetry-based neighboring protein molecules. Following production of the designed protein, two different protein crystal forms were obtained, one of which agreed very closely with the designed crystal model, with deviations in the range of about 1 Å.

Two different test cases recently demonstrated the ability to design a large, novel self-associating interface into a naturally oligomeric protein subunit to produce large molecular cages and clusters [34] (Figure 5). Novel homodimeric interfaces were designed into two different naturally trimeric proteins in geometries intended to yield a 24-subunit cage with octahedral symmetry in one case and a 12-subunit assembly with tetrahedral symmetry in the other. This design strategy relied on a computational docking procedure that restrained the rigid body and conformational sampling to configurations obeying the target symmetry [37,38]. Multiple crystal structures of the designed proteins revealed that the resulting assemblies matched the design models at atomic-level resolution, exhibiting backbone RMSDs of about 1 Å or less. It should be noted that the initially designed sequence for the tetrahedral case yielded an assembly that was slightly less accurate (backbone RMSD of 2.7 Å); introducing three additional mutations yielded the more accurate assembly. This study demonstrated that designing interfaces based on many weak interactions over large interface areas – similar to the interfaces found in natural protein assemblies – can result in highly ordered and accurately assembled structures. However, the two successes were accompanied by 39 other designed proteins that either expressed insolubly or failed to assemble, emphasizing the need for improved methods for designing protein–protein interactions if this approach is to become widely used.

Strategic variations

The basic design ideas discussed above admit numerous variations, some of which have been demonstrated already. For extended protein assemblies, one immediate obstacle is the production of recombinant proteins in soluble form during bacterial protein expression. The building blocks for such assemblies tend to associate as soon as they are produced inside the host cell, leading to large aggregates and inclusion bodies that seriously complicate the problem of protein purification. Several groups have overcome this problem by splitting the assemblies into two or more distinct components that can be produced in soluble forms in separate bacterial strains [19,20,22]. Purified, soluble components can then be mixed to form the intended assemblies *in situ*. Other methods to trigger protein assemblies rely on the addition of metals or other ligands into the otherwise soluble protein building blocks [35,39,40]. These approaches offer the additional advantage that the metals or ligands can be chelated or scavenged to revert the protein assemblies to their soluble building blocks.

The surfaces of protein building blocks can also be decorated with various functional groups to generate desirable properties. In one case, a rhodamine chemical group was conjugated to building blocks to promote association between protein layers [35]. In another example, protein sidechains in a repeating pattern along a helical peptide were chosen to adhere to the periodic pattern of carbon atoms in single-walled carbon nanotubes [41]. The helical peptides wrapping around the carbon nanotubes further served as a periodic registry for gold nanoparticle attachment, highlighting the potential for generating hybrid bio-nanomaterials.

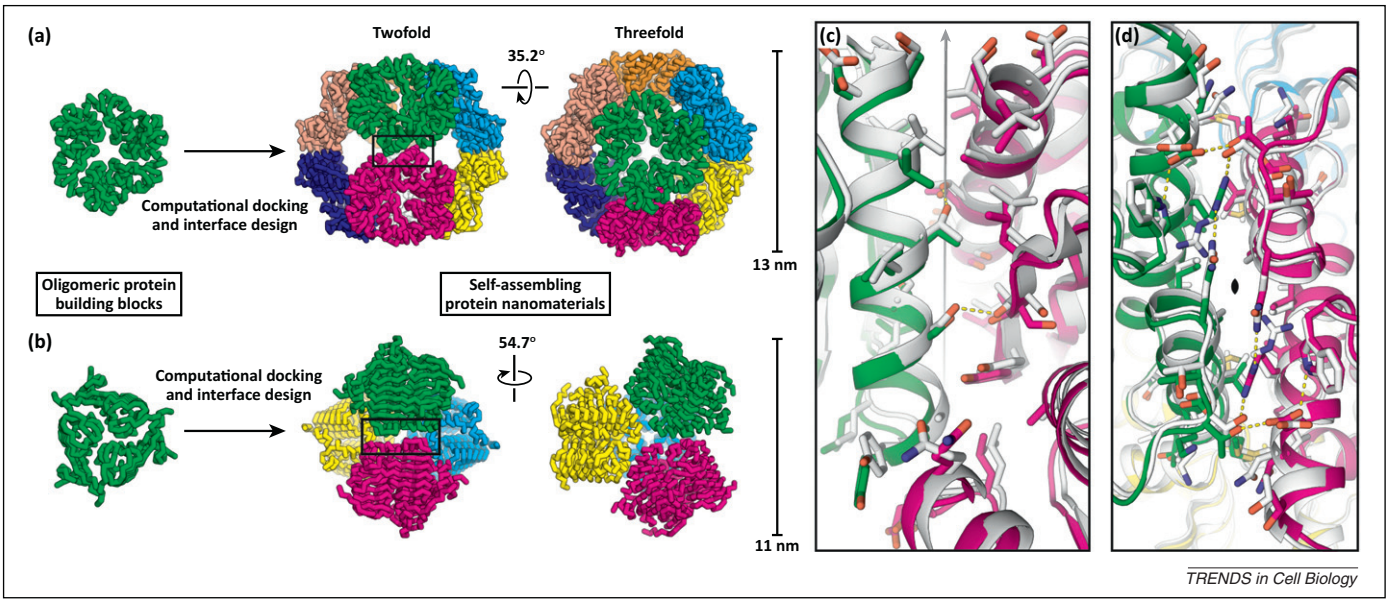

Figure 5. Accurate design of large symmetric protein assemblies based on novel interface design. Among the wide variety of symmetric architectures that can be obtained by interface design, two large cubic assemblies have been demonstrated [34]. Crystal structures of designed **(a)** 24-subunit octahedral and **(b)** 12-subunit tetrahedral assemblies are shown, viewed along their twofold and threefold symmetry axes. Within each assembly, individual trimers are depicted in different colors. The designed interfaces [regions boxed in (a) and (b)] in the crystal structures (green/magenta) of the **(c)** octahedral and **(d)** tetrahedral assemblies closely agree with the designed models (white). Backbone RMSDs between the observed structures and the designed models were 1.1 Å and 0.6 Å over all subunits for the octahedral and tetrahedral assemblies, respectively.

Further variations for designing protein assemblies have been conceptualized. In one example, the building blocks comprise smaller fragments of natural proteins, which would then be stitched together to produce molecules having specific self-association properties [42]. Although this approach has not been realized yet in practice, it highlights the variety of creative ideas being brought to bear on the problem of designing ordered assemblies.

Different design strategies offer their own advantages and disadvantages. Recent studies emphasize key distinctions between the two main strategies discussed here – oligomeric fusion and interface design. The former approach avoids the introduction of mutations within naturally occurring protein domains, and hence minimizes the chance that the natively folded structure will be disrupted. On the other hand, *de novo* interface design should ultimately enable the use of arbitrary protein domains as building blocks, without being limited to those natural oligomers offered by nature.

Concluding remarks

If engineered protein assemblies are to become a reliable technology for real-life applications, the success rate of the strategies discussed here will need to be improved. For the oligomeric fusion strategies, the major challenge is to control better the relative orientation of the fusion partners. This is especially true for the helix-fusion strategy given that the recent crystal structure of the designed cage revealed considerable bending and twisting of the helical linkers [18]. For the symmetry-matching fusion strategy, although the presence of multiple linkers between joined oligomers promotes alignment of the shared symmetric axes, twisting about this axis may be harder to control [22]. Such rotational flexibility may need to be dealt with if a high success rate is to be achieved, particularly in the challenging area of extended 3D materials. For strategies that rely on creating new interfaces by computational design, the key challenge is in achieving geometrically precise interfaces. The successful design of a protein crystal based on an α-helical polypeptide [36] suggests that this may be relatively straightforward for protein building blocks consisting of simple motifs. By comparison, designing precise interfaces into larger proteins of arbitrary shape appears to present more complex problems. Early experiments showed that single rings of protein oligomers could be reliably converted to double rings by incorporating a few nonpolar residues on the surface, but the assemblies obtained generally deviated from the designs in detail [26]. King *et al.* [34] twice succeeded in designing extensive, geometrically-accurate interfaces into globular protein subunits, which consequently gave rise to large, symmetric cubic assemblies in very close agreement with the designed models. However, the successful designs were only a small fraction of the total designs attempted, and in one of the two successful designs a second round of sequence mutations was necessary to fine-tune the initial design. The successes (and accompanying failures) of various recent design efforts suggest that further improvements in interface design algorithms could have an important impact.

An ultimate challenge will be to create a large, complex assembly without relying on existing oligomerization interfaces or well-known self-associating motifs. This will require designing at least two distinct interfaces into a naturally monomeric protein. In addition to the high accuracy required, the need to introduce multiple interfaces compounds the success-rate problem, and also increases the likelihood that the sequence changes will be so many as to compromise protein folding and stability. The latter issue may call for more careful consideration of which

protein folds might be the most suitable as starting templates for design.

The success-rate problem might be mitigated if efficient screening methods can be developed. King *et al.* implemented a multistep screening method to identify which of their many computationally designed protein sequences were soluble and compatible with self-assembly [34]. Although their screening protocol, which involved gel electrophoresis or liquid chromatography, was suitable for the number of designs examined (~50), design strategies that exploit much larger sequence pools could benefit from the development of higher-throughput screening strategies [43].

The next stage in developing designed protein assemblies will need to focus on applications [44,45]. The possibilities are myriad, arising from different combinations of functional elements (such as catalytic sites or recognition motifs) with various types of spatial organization, such as encapsulation or ordered display [46,47]. Finite structures such as cages and shells invite the idea of encapsulation, as well as targeted delivery based on externally displayed motifs [48,49]. Motifs displayed on the surface of large assemblies could be useful in their own right, for example in the design of synthetic vaccines [50]. Extended materials offer additional modes of spatial organization [51]. Numerous areas of synthetic biology could be advanced by materials presenting ordered arrangements of enzymes or receptors [52,53]. Across all the possible architectures that might be explored, the ability to trigger assembly or disassembly should enable a broad range of responsive materials [54].

Acknowledgments

The authors thank Jennifer Padilla for helpful discussions and critical reading of the manuscript.

References

1 Schwille, P. (2011) Bottom-up synthetic biology: engineering in a tinkerer's world. *Science* 333, 1252–1254
2 Tan, C. *et al.* (2009) Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* 5, 842–848
3 Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338
4 Zhang, S.G. *et al.* (2002) Design of nanostructured biological materials through self-assembly of peptides and proteins. *Curr. Opin. Chem. Biol.* 6, 865–871
5 Yeates, T.O. and Padilla, J.E. (2002) Designing supramolecular protein assemblies. *Curr. Opin. Struct. Biol.* 12, 464–470
6 Zhang, S.G. (2003) Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* 21, 1171–1178
7 Petka, W.A. *et al.* (1998) Reversible hydrogels from self-assembling artificial proteins. *Science* 281, 389–392
8 Ogihara, N.L. *et al.* (2001) Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1404–1409
9 Woolfson, D.N. (2010) Building fibrous biomaterials from alpha-helical and collagen-like coiled-coil peptides. *Biopolymers* 94, 118–127
10 Ghadiri, M.R. *et al.* (1993) Self-assembling organic nanotubes based on a cyclic peptide architecture. *Nature* 366, 324–327
11 Ryadnov, M.G. and Woolfson, D.N. (2003) Introducing branches into a self-assembling peptide fiber. *Angew. Chem. Int. Ed. Engl.* 42, 3021–3023
12 Levy, E.D. *et al.* (2008) Assembly reflects evolution of protein complexes. *Nature* 453, U1262–U1266
13 Goodsell, D.S. and Olson, A.J. (2000) Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* 29, 105–153
14 Crick, F.H.C. and Watson, J.D. (1956) Structure of small viruses. *Nature* 177, 473–475
15 Wukovitz, S.W. and Yeates, T.O. (1995) why protein crystals favor some space-groups over others. *Nat. Struct. Biol.* 2, 1062–1067
16 Padilla, J.E. *et al.* (2001) Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U. S. A.* 98, 2217–2221
17 Plevka, P. *et al.* (2012) Crystal structure of human enterovirus 71. *Science* 336, 1274
18 Lai, Y.T. *et al.* (2012) Structure of a 16-nm cage designed by using protein oligomers. *Science* 336, 1129
19 Ringler, P. and Schulz, G.E. (2003) Self-assembly of proteins into designed networks. *Science* 302, 106–109
20 Usui, K. *et al.* (2009) Nanoscale elongating control of the self-assembled protein filament with the cysteine-introduced building blocks. *Protein Sci.* 18, 960–969
21 Raman, S. *et al.* (2006) Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. *Nanomedicine* 2, 95–102
22 Sinclair, J.C. *et al.* (2011) Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* 6, 558–562
23 Leaver-Fay, A. *et al.* (2011) Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487, 545–574
24 Fleishman, S.J. *et al.* (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332, 816–821
25 Huang, P.S. *et al.* (2007) A de novo designed protein-protein interface. *Protein Sci.* 16, 2770–2774
26 Grueninger, D. *et al.* (2008) Designed protein–protein association. *Science* 319, 206–209
27 Radford, R.J. and Tezcan, F.A. (2009) A superprotein triangle driven by nickel(II) coordination: exploiting non-natural metal ligands in protein self-assembly. *J. Am. Chem. Soc.* 131, 9136–9137
28 Salgado, E.N. *et al.* (2010) Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1827–1832
29 Der, B.S. *et al.* (2012) Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* 134, 375–385
30 Salgado, E.N. *et al.* (2010) Metal-directed protein self-assembly. *Acc. Chem. Res.* 43, 661–672
31 Stranges, P.B. *et al.* (2011) Computational design of a symmetric homodimer using beta-strand assembly. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20562–20567
32 Zaccai, N.R. *et al.* (2011) A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.* 7, 935–941
33 Korendovych, I.V. *et al.* (2010) De novo design and molecular assembly of a transmembrane diporphyrin-binding protein complex. *J. Am. Chem. Soc.* 132, 15516–15518
34 King, N.P. *et al.* (2012) Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* 336, 1171–1174
35 Brodin, J.D. *et al.* (2012) Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* 4, 375–382
36 Lanci, C.J. *et al.* (2012) Computational design of a protein crystal. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7304–7309
37 Andre, I. *et al.* (2007) Prediction of the structure of symmetrical protein assemblies. *Proc. Natl. Acad. Sci. U. S. A.* 104, 17656–17661
38 Wang, C. *et al.* (2007) Protein-protein docking with backbone flexibility. *J. Mol. Biol.* 373, 503–519
39 Carlson, J.C.T. *et al.* (2006) Chemically controlled self-assembly of protein nanorings. *J. Am. Chem. Soc.* 128, 7630–7638
40 Fegan, A. *et al.* (2010) Chemically controlled protein assembly: techniques and applications. *Chem. Rev.* 110, 3315–3336
41 Grigoryan, G. *et al.* (2011) Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* 332, 1071–1076
42 Tsai, C.J. *et al.* (2007) Principles of nanostructure design with protein building blocks. *Proteins* 68, 1–12
43 Worsdorfer, B. *et al.* (2011) Directed evolution of a protein container. *Science* 331, 589–592

44 Woolfson, D.N. and Mahmoud, Z.N. (2010) More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. *Chem. Soc. Rev.* 39, 3464–3479
45 Howorka, S. (2011) Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* 22, 485–491
46 Tschiggerl, H. *et al.* (2008) Exploitation of the S-layer self-assembly system for site directed immobilization of enzymes demonstrated for an extremophilic laminarinase from Pyrococcus furiosus. *J. Biotechnol.* 133, 403–411
47 Minten, I.J. *et al.* (2009) Controlled encapsulation of multiple proteins in virus capsids. *J. Am. Chem. Soc.* 131, 17771–17773
48 Tong, G.J. *et al.* (2009) Viral capsid dna aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* 131, 11174–11178
49 Han, M. *et al.* (2011) Targeted vault nanoparticles engineered with an endosomolytic peptide deliver biomolecules to the cytoplasm. *ACS Nano* 5, 6128–6137
50 Smith, M.L. *et al.* (2006) Modified tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* 348, 475–488
51 Ilk, N. *et al.* (2011) S-layer fusion proteins – construction principles and applications. *Curr. Opin. Biotechnol.* 22, 824–831
52 Delebecque, C.J. *et al.* (2011) Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333, 470–474
53 Dueber, J.E. *et al.* (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27, 753–759
54 de la Rica, R. and Matsui, H. (2010) Applications of peptide and protein-based materials in bionanotechnology. *Chem. Soc. Rev.* 39, 3499–3509
55 Yeates, T.O. (2011) Nanobiotechnology: protein arrays made to order. *Nat. Nanotechnol.* 6, 541–542