

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/318373380>

# Advances in design of protein folds and assemblies

Article in *Current opinion in chemical biology* · October 2017

DOI: 10.1016/j.cbpa.2017.06.020

CITATIONS

0

READS

61

3 authors, including:



[Ajasja Ljubetič](#)

National Institute of Chemistry

20 PUBLICATIONS 40 CITATIONS

[SEE PROFILE](#)



[Roman Jerala](#)

National Institute of Chemistry

224 PUBLICATIONS 5,120 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Innate immunity [View project](#)



Protein origami [View project](#)



# Advances in design of protein folds and assemblies

Ajasja Ljubetič<sup>1</sup>, Helena Gradišar<sup>1,2</sup> and Roman Jerala<sup>1,2</sup>

Conceptual and computational advances triggered an explosion of designed protein structures in the recent years. Various protein fold geometries have been robustly designed with atomic accuracy, including protein folds unseen in nature. The same principles and tools have been extended to design multi-chain assemblies. By exploiting symmetry, mega-Dalton structures have been created with exciting potential applications for synthetic biology. In this review we focus on design of single chain and multi polypeptide chain assemblies of defined size and composition. Several innovative strategies have been developed to create *de novo* protein assemblies, with the two main approaches to the design of multi-chain assemblies being genetic fusion of interacting modules and engineering of novel protein–protein interfaces.

## Addresses

<sup>1</sup> Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia

<sup>2</sup> EN-FIST Centre of Excellence, Ljubljana, Slovenia

Corresponding author: Jerala, Roman ([roman.jerala@ki.si](mailto:roman.jerala@ki.si))

**Current Opinion in Chemical Biology** 2017, **40**:65–71

This review comes from a themed issue on **Synthetic biomolecules**

Edited by **Peter H Seeberger** and **Beate Kokscher**

<http://dx.doi.org/10.1016/j.cbpa.2017.06.020>

1367-5931/© 2017 Published by Elsevier Ltd.

## Introduction

Proteins are the most versatile type of biomacromolecules. They adopt the wide-ranging 3D shapes and functions based on the defined complex tertiary structures that are encoded by the linear sequence of amino-acid residues. Proteins mediate most of the processes that are essential to all living organism. While natural proteins have been selected for function within the context of the cell, they have the huge untapped potential to adopt many new shapes and functions unknown in nature [1] which can be applied to solve many technological, medical, scientific and other problems [2].

To accurately design the fold of engineered proteins still remains an exciting challenge for protein scientists. A variety of approaches for designing novel protein structures have been developed in recent years: from modification of the existing natural domains to more advanced

strategies that employ computational tools and are able to design *de novo* proteins from scratch based on rational engineering modular principles. The ability to design and construct proteins with arbitrary size, shape and function could be a milestone comparable to the industrial revolution, as the proteins represent smart nanostructures and molecular machines [3\*].

## Scope of the review

In this review we discuss the recent advances in the *de novo* design of protein assemblies with defined size and composition. In the last four years since our previous review [4], the field of protein design witnessed important milestones in *de novo* design of single chain protein structures, which can be used as building blocks for larger and more complex assemblies as well as new types of assemblies. The main driving forces for this advance have been the conceptual leap in new design principles (where particularly the Rosetta based approach needs to be singled out), wider availability of synthetic DNA and advances in structural characterizations of macromolecules, from small angle X-ray scattering (SAXS), X-ray crystallography to cryo-electron microscopy. The above factors facilitated flourishing of designed proteins and assemblies once deemed beyond our capability.

## *De novo* design of single-chain proteins

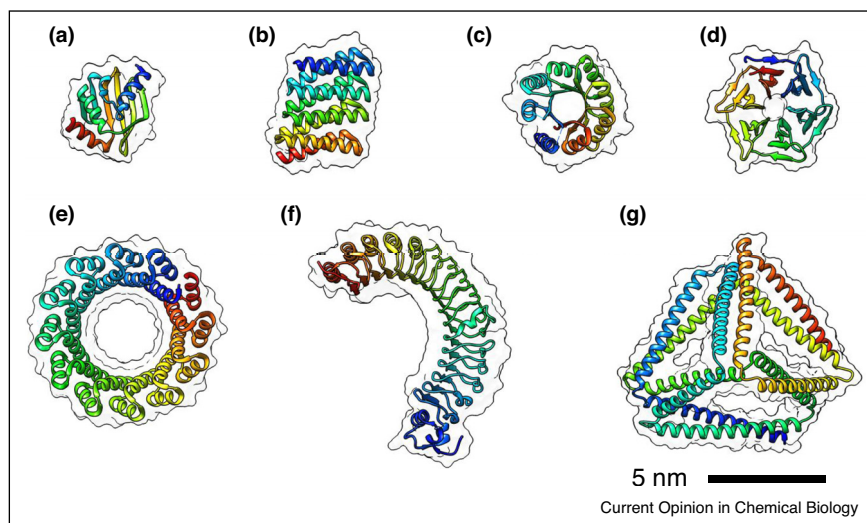
Protein design can be viewed as the inverse protein folding problem [5]: the goal is to find a sequence of amino acids that is compatible with a given backbone fold [6]. *De novo* design involves solving an even more complex problem, where neither the backbone nor the sequence is completely specified [3\*]. Instead architectural features such as desired size or type of fold, number of secondary structure elements are specified (this is sometimes termed the topology of the protein). The advantages are numerous: proteins can be designed with atomic accuracy for specific functions, such as scaffold support, specific binding [7] or catalytic activity [8].

The current hurdles in protein design are inaccuracies in the energy functions and problems with exhaustive sampling. This is also where the most progress is being made. Consequently the design efficiency is increasing [9]. The sampling of different backbones is usually achieved either by assembling them from fragments or by describing backbones parametrically; both approaches are presented below. Examples of several *de novo* designed single-chain proteins are depicted in Figure 1.

## $\alpha/\beta$ folds

One of the first *de novo* globular proteins consisted of 2  $\alpha$ -helices and 4  $\beta$ -strands [10]. Lin *et al.* [11] expanded the

Figure 1



Single-chain *de novo* proteins. Examples of *de novo* designed proteins. (a)  $\alpha/\beta$  fold with controlled curvature [12\*\*] (PDB: 5TPJ). (b) Repeat protein with controlled curvature [17\*\*] (PDB: 5CWG). (c) First *de novo* TIM-barrel [22] (PDB: 5BVL). (d) Designed  $\beta$ -propeller with 6 identical repeats [21] (PDB: 3WW9). (e) Closed  $\alpha$ -helical solenoid [20] (PDB: 5BYO). (f) Designed leucine rich repeat protein with varying curvature [18] (PDB: 4R6G). (g) Coiled-coil protein origami tetrahedron [26\*\*]. Scale bar denotes 5 nm. Ribbon representation is colored by amino acid position from blue (N-terminal) to red (C-terminal). Solvent accessible surface is represented by an outline. All images were generated using UCSF Chimera [47,48].

range of possible topologies, as well as elucidated the rules needed to efficiently link various secondary structural motives. Marcos *et al.* [12\*\*] (Figure 1a) extended the principles to non-ideal  $\beta$ -sheets that contain kinks. Insertion of kinks can regulate local curvature enabling the researchers to install cavities into some of the designs. Larger  $\alpha/\beta$  constructs can be obtained by fusing of two  $\alpha/\beta$  designs. King *et al.* [13] created a 147 residue fusion construct containing 4  $\alpha$ -helices and 6  $\beta$ -strands.

### Repeat proteins

Current *de novo* designs of compact protein folds are typically limited to  $\sim 150$  residues [14]. One possibility to design larger structures is to design proteins with internal symmetry [15]. Parmeggiani *et al.* [16] devised a general computational approach to designing repeat proteins, based on making modifications simultaneously on all repeats. Six different types of repeat proteins were successfully designed. Repeat proteins can be characterized using helical parameters (radius, axial displacement and twist). Brunette *et al.* [17\*\*] (Figure 1b) have designed helix-loop-helix repeat proteins of various geometries, including many that have not been found among natural proteins, indicating that the design space of proteins is larger than what nature has so far explored. Park *et al.* [18] (Figure 1f) have developed a method to chain repeat building blocks of different curvatures, thereby enabling design of proteins with controlled local curvature. Repeat proteins with closed architectures (zero axial displacement) have also been designed [19,20] (Figure 1e)

including a  $\beta$ -propeller [21] (Figure 1d) and a TIM-barrel [22] (Figure 1c).

### Coiled-coil protein origami (CCPO)

A completely different innovative approach to designing self-assembled nanostructures is based on concatenating dimeric coiled-coil building elements, which have orthogonal binding preferences, into a single polypeptide chain [23,24]. The structure is formed by ensuring that the polypeptide chain traverses each edge of a designed polyhedron exactly twice. It has been proven mathematically that any polyhedra can be constructed in this way. From the topological point of view it is not important whether the CC pairs are heterodimers or homodimers, however it is important to have available both parallel and antiparallel CC dimers, since only their combination guarantees that a selected single chain polyhedron can be constructed. The modular design allows a construction of new polyhedral cage-like folds, unseen in nature. In many aspects this strategy resembles the principles of DNA nanostructures where polyhedral structures can be designed based on the complementary DNA segments [25]. However proteins have several advantages over DNA, such as greater chemical variability, availability of parallel and antiparallel building blocks and easier *in vivo* production.

Gradišar *et al.* [26\*\*] demonstrated the feasibility of the coiled-coil origami strategy by designing a 50 kDa tetrahedral cage, with 476 amino acid residues. Correct folding

of designed polypeptide chain is based on the interactions between twelve concatenated coiled coil-forming modules, far apart in the residue sequence. The coiled coil-forming segments dimerize independently and form the edges of the tetrahedral cage with a cavity in the center. Two antiparallel and four parallel CC pairs were used. Three of the parallel pairs were heterodimers and three were homodimers. Using another topological solution, a tetrahedron can also be constructed from three antiparallel and three parallel CC pairs.

Longer coiled-coil building elements are twisted. Since the structures composed of long coiled-coil dimers would contain topological knots, their folding needs to be kinetically defined and the order of self-assembly guided. This issue has been recently resolved using the single chain assembly of DNA pyramid, which established the rules for the fast folding design [25], which awaits its implementation in polypeptide-based folds.

The platform CoCoPOD ([https://github.com/NIC-SBI/CC\\_protein\\_origami](https://github.com/NIC-SBI/CC_protein_origami)) was developed by the same group and offers many tools for design and building of polyhedral nano-cages. The main advantage of such modular coiled coil-based nano-cages is their excellent potential for variation in shape, size, complexity and functionalization for medical and biotechnological applications.

### De novo design of $\alpha$ -helical bundles

One approach to successfully sampling *de novo* backbones is to describe the backbones parametrically. This is most easily accomplished for  $\alpha$ -helical coiled-coils [27], which are dimers or higher oligomers composed of  $\alpha$ -helices and stabilized by characteristic ‘knobs-into-holes’ packing [28]. New tools have been developed to facilitate the design of such structures [29]. Examples of designed bundles are shown in Figure 2.

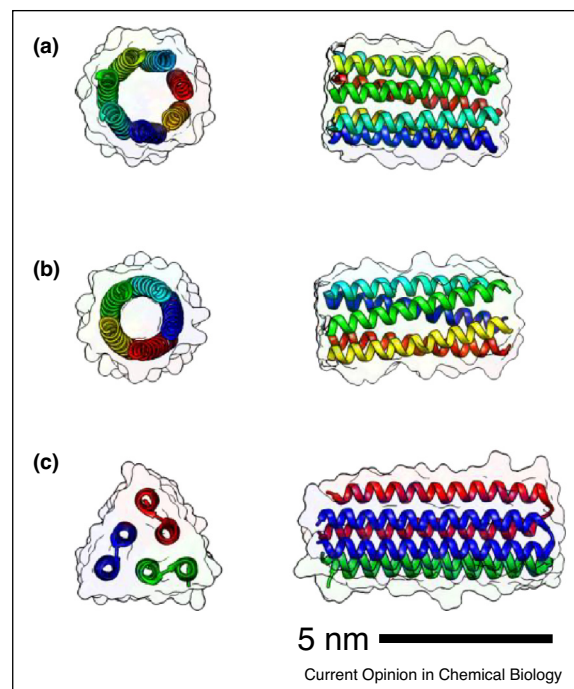
#### Coiled-coil $\alpha$ -helical bundles

Huang *et al.* [30] (Figure 2a) have designed extremely stable trimeric and tetrameric and pentameric  $\alpha$ -helical bundles. Thomson *et al.* [31] (Figure 2b) have demonstrated the utility of this approach by designing pentameric, hexameric and heptameric  $\alpha$ -helical barrels. Catalytic activity has been successfully grafted onto the heptameric bundle [32]. A four helix membrane-spanning  $\text{Zn}^{2+}$  transporter has also been designed [33].

#### Design of hydrogen bond networks at protein–protein interfaces

Historically protein design focused on the hydrophobic protein cores. Therefore it is not surprising that most designed interfaces have a more hydrophobic character compared to natural counterparts [9]. However progress is being made in understanding polar interactions as exemplified by the design of a protein with a polar core [34].

Figure 2



Parametric  $\alpha$ -helical bundles. Examples of *de novo* parametrically designed  $\alpha$ -helical bundles. (a) A designed heptameric coiled coil [31] (PDB: 4PNA). (b) Designed hyperstable pentameric coiled-coil [30] (PDB: 4UOT). (c) Helical bundle containing two buried hydrogen networks at the binding interface [35] (PDB: 5IZS). Scale bar denotes 5 nm. Each chain is colored uniquely. Solvent accessible surface is represented by an outline. All images were generated using UCSF Chimera [47,48].

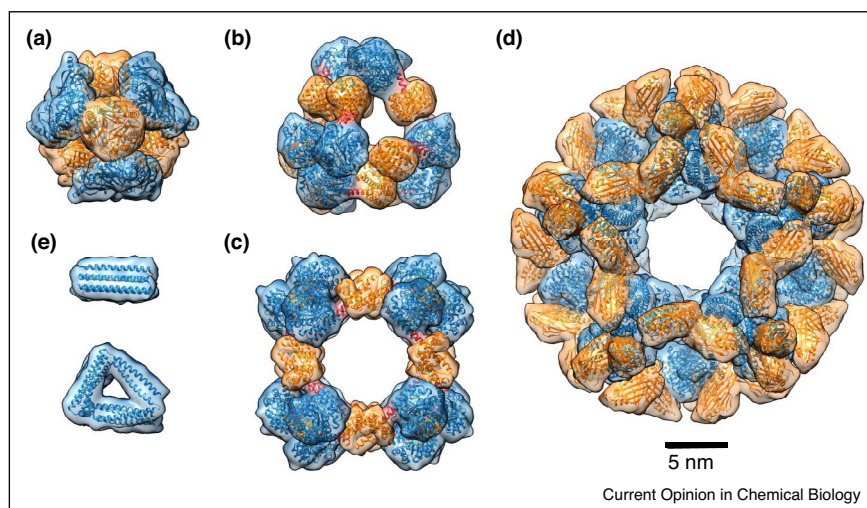
Boyken *et al.* [35] (Figure 2c) designed polypeptide chain interfaces with buried hydrogen bond networks. Eight different  $\alpha$ -helical bundles topologies with C2, C3 or C4 symmetry and various levels of supercoiling were successfully designed and tested by SAXS and crystallography. Protein–protein interfaces (PPIs) based on buried hydrogen bonding offer elegant solution to specificity and solubility problems faced by current state of the art design methods and are reminiscent of how nature solved the specificity problem in DNA binding.

### De novo design of multi-chain protein assemblies

The first developed strategy for designed protein assemblies employed natural oligomerization domains that have been genetically or chemically fused. Another strategy that has been developed involves design of novel protein–protein interfaces that can be grafted onto existing oligomers. Over last two decades, the formation of well-ordered assemblies, lattices or cage-like assemblies where symmetry plays an important role have been constructed. Both types of assemblies are depicted in Figure 3.



Figure 3



Multi-chain *de novo* assemblies. Multi-chain *de novo* assemblies based either on interface design ((a) and (d)) or genetic fusion ((b) and (c)). (a) A 24-subunit two-component (4 trimers + 4 trimers) assembly with tetrahedral symmetry based on protein interface design. The different trimers are colored orange or blue [43\*\*] (PDB: 4NWR). (b) A 12-subunit one component assembly with tetrahedral symmetry. The dimers (orange) and trimers (blue) are linked via an alpha-helix (red) [39\*] (PDB: 4QF0). (c) A 24-subunit cube with octahedral symmetry [38] (PDB: 4QCC). (d) A 120-subunit two-component (20 trimers + 30 dimers) icosahedral assembly [45] (PDB: 5IM6). The dimers are colored orange and the trimers blue. (e) Two examples from Figures 1 and 2 are included for scale. Top: helical bundle with hydrogen networks [35\*] (PDB: 5IZS). Bottom: coiled-coil protein origami tetrahedron [26\*\*]. All images were generated using UCSF Chimera [47,48].

### Genetic fusion of natural oligomerization domains

One of the earliest strategies for creating designed protein nanostructures relied on genetic fusion of two distinct protein domains, typically comprised of 100–200 amino acid residues. The building blocks composed of linked oligomerization domains, which can non-covalently self-assemble into larger, highly symmetric assemblies. In the pioneering study of this approach the fusion of dimeric and trimeric domain led to formation of a tetrahedral cage, while fusion of two different dimeric components led to filaments [36]. Later, refinement of the original sequence for the tetrahedron improved control of the relative orientation over subunit interaction which resulted in homogeneous 16-nm cage-like particles, confirmed by X-ray crystal structure determination [37]. The extension of the oligomer fusion method enabled the construction of larger polyhedral assemblies, including highly porous, 24-subunit cube with perfect octahedral symmetry [38] (Figure 3c). A crystal structure of the cage which had a diameter of 22.5 nm showed atomic-level agreement with the designed model. Further, the same group created and validated an advanced design of a 600-kDa protein homododecamer that self-assembles into a cage with tetrahedral symmetry [39\*] (Figure 3b). The solution properties of designed variants were measured and the effects of amino acid sequence on structural robustness defined by applying new algorithms for systematic analyses by HT-SAXS experiments. This approach enables to reveal cage symmetry, flexibility, multimeric stability, and

conformational transitions as a function of solution environment. In summary this represents a new tool for effective feedback on experimental constructs relative to design and has general applicability in analyzing the solution behavior of heterogeneous nanosystems.

Kobayashi *et al.* [40] constructed novel assemblies by genetically fusing a dimeric (WA20) and trimeric (foldon) building block. The resulting building block (PN Block) was shown to simultaneously form several stable homo-oligomeric forms: small (S), middle (M) and large (L). SAXS analyses suggested that S is a hexameric barrel and M a tetrahedron-shaped dodecamer.

### Design of protein–protein interfaces (PPIs)

Similar principles and tools used in design of single-chain assemblies can also be used to design novel PPIs. However additional challenges are present: the interface must be sufficiently soluble, which requires control of ion pairing and hydrogen bonding in addition to hydrophobic interactions [41]. To make the design problem more tractable symmetry is often employed in the design. Symmetric designs have a better tolerance for non-optimal interactions and higher specificity as reviewed in Zhang *et al.* [42\*].

In a ground-breaking study King *et al.* [43\*\*] (Figure 3a) have created multiple two-component 24-subunit assemblies with tetrahedral symmetry. Two architectures T33 (4 trimers + 4 trimers) or T32 (4 trimers + 6 dimers) were

tested. The authors first searched the structures in the PDB for suitable dimers and trimers. All combinations of dimers and trimers (over 250 000 pairs) were rigidly docked in the pre-described symmetry orientation. Because of symmetry constraints there are only two rigid body degrees of freedom (one rotation and one translation) for each distinct subunit that need to be systematically searched. Only backbone and C $\beta$  atoms were used at this stage. The minimized docked complexes were evaluated for designability using several metrics (for example shape complementarity). The best candidates were then used for all-atom interface design. Four designs were crystallized and found to match their design blueprints with atomic resolution. The designs were also confirmed using electron microscopy.

By using existing dimers and trimers only one additional PPI needs to be designed, which makes the design problem more tractable. Two-component assemblies also offer the option of controlled assembly, by mixing the separate components, which can be used for packaging and encapsulation.

Hsia *et al.* [44] designed a one-component 60-subunit assembly with icosahedral symmetry. The design was checked using electron microscopy. Superfolder GFP was genetically fused either to one or both termini and the measured intensity was consistent with either 60-subunits or 120-subunits of GFP. Such molecules could serve as an intensity reference for (single molecule) fluorescence microscopy.

Bale *et al.* [45] (Figure 3c) have extended the approach of symmetric assemblies by designing two-component

120-subunit assemblies with icosahedral symmetry. Three architectures I52 (12 pentamers + 30 dimers), I53 (12 pentamers + 20 trimers) and I32 (20 trimers + 30 dimers) were tested. The designs were verified using SAXS, electron microscopy and crystallography. The best designs show an impressive 0.8 Å RMSD over all backbone atoms. Negatively supercharged GFP was packaged into a positively charged variant of one of the designs. 7–11 molecules of GFP associated with one 120-subunit design.

In an attempt to design more negatively charged 60-subunit assemblies with icosahedral symmetry Sasaki *et al.* [46] discovered that symmetric PPI can be robust and flexible. Increasing the negative net charge in the lumen resulted in larger particles. Two larger assemblies were characterized by Cryo-EM, one with tetrahedral symmetry composed of 180 subunits and one with icosahedral symmetry composed of 360 subunits. The assemblies are porous, which may be an advantage for packaging applications.

## Challenges and outlook

The most difficult technical challenges in computational protein design are inaccuracies in the energy functions and problems with exhaustive sampling, including the sampling of alternate unwanted states. Inaccuracies in energy calculations may lead to incorrect ranking of these unwanted conformations that might therefore be significantly populated in solution. Consequently this leads to incorrect protein folding and aggregation, which is reflected in higher failure rates of multi-chain design approaches (Table 1). The presented data outline the current difficulties in PPI design, although as was the case

**Table 1**

### Success rates of single-chain versus multi-chain designed protein assemblies

Study	Year	Type	Tested	Soluble	Designed oligomeric state/co-purification	Validated	Atomic resolution structure
Marcos <sup>d</sup> [12**]	2017	Single-chain	37	33 (89%)	32 (86%)	24 (65%) <sup>a</sup>	7 (19%)
Brunette [17**]	2015	Single-chain	83	79 (95%)	74 (89%)	43 (52%) <sup>b</sup>	15 (18%)
Boyken [35*]	2016	$\alpha$ -Helical bundles	114	101 (89%)	66 (58%)	25 (22%) <sup>b</sup>	10 (9%)
King [43**]	2014	Multi-chain	57	Several	5 (9%)	5 (9%) <sup>b,c</sup>	4 (7%)
Hsia [44]	2016	Multi-chain	17	3 (18%)	1 (6%)	1 (6%) <sup>c</sup>	0 (0%)
Bale [45]	2016	Multi-chain	183	Several	12 (7%)	10 (5%) <sup>b,c</sup>	3 (2%)

A comparison of design efficiencies across several studies. The success rates are typically lower for multi-chain assemblies. Tested — denotes the number of experimentally tested constructs (i.e. the number of ordered genes). Soluble — the number of soluble constructs (values in parenthesis are percentage of tested constructs). Designed oligomeric state/co-purification — number of designs with correct oligomeric state (either monomeric or oligomeric). For multi-chain types this number reflects the number of designs that co-purified.

Validated — number of designs validated by experimental methods. Superscripts denote the used method:

<sup>a</sup> HSQC.

<sup>b</sup> SAXS.

<sup>c</sup> EM or Cryo-EM.

Atomic resolution structure — number of designs with determined crystal, NMR or Cryo-EM structure with coordinates deposited in the PDB. The real success rates are likely higher, since not all validated designs were necessarily used for structure determination.

<sup>d</sup> Only the basic A to F folds were taken into account (and not the disulfide, dimeric or cavity mutants).

for single-chain designs, the success rate will most likely increase in the future.

A fundamental property of amino acid interactions is the complexity of cooperative interactions, which can still decrease the success of functionalizing the designed proteins. For example, introducing a cavity into a designed protein usually decreases its overall stability [12<sup>••</sup>]. The work done by the group of Yeates also demonstrates the difficulty of predicting effects of amino acid interactions: a few mutations in the linker region had a large impact on the global structure of obtained assemblies [39<sup>•</sup>].

One solution to this problem involves constructing assemblies from orthogonal elements as demonstrated by the design of repeat proteins [15] and coiled-coiled protein origami [26<sup>••</sup>]. The other option is to re-design the larger regions of the protein (not just the binding site) and extensively test the new design *in silico*, for example by folding simulations [14]. Several of the designed protein structures and assemblies are extremely stable [12<sup>••</sup>,30,44], demonstrating that protein design can improve the properties of natural proteins.

Designing new protein structures opens many new challenging possibilities, such as, for example, dynamic assemblies. Most *de novo* proteins are designed to have only a single global minimum. But dynamic assemblies must have several distinct stable conformations (and therefore several distinct free energy minima), which will require design strategies for targeting multiple minima.

The design of proteins and protein assemblies unseen in nature opens new perspectives for synthetic biology. We can expect in the coming years additional advances in design of new scaffolds and also demonstration of some applications in the areas of materials, sensing, industrial, medical and many others.

## Acknowledgements

We acknowledge the financial support of the Slovenian Research Agency (program P4-0176, projects N4-0037 and J4-5528), the ERANET SynBio project BioOrigami (ERASYNBIO1-006) and grant from the ICGEB (CRP/SLO14-03).

## 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. Woolfson DN, Bartlett GJ, Burton AJ, Heal JW, Niitsu A, Thomson AR, Wood CW: **De novo protein design: how do we expand into the universe of possible protein structures?** *Curr Opin Struct Biol* 2015, **33**:16-26.
  2. Gerrard JA (Ed): *Protein Nanotechnology: Protocols, Instrumentation, and Applications*, edn 2. Humana Press; 2013.
  3. Huang P-S, Boyken SE, Baker D: **The coming of age of de novo protein design.** *Nature* 2016, **537**:320-327.  
An excellent review article describing the recent progress of *de novo* protein design. Most emphasis is on the Rosetta framework.
  4. Božič S, Doles T, Gradišar H, Jerala R: **New designed protein assemblies.** *Curr Opin Chem Biol* 2013, **17**:940-945.
  5. Regan L, Caballero D, Hinrichsen MR, Virrueta A, Williams DM, O'Hern CS: **Protein design: past, present, and future.** *Pept Sci* 2015, **104**:334-350.
  6. MacDonald JT, Freemont PS: **Computational protein design with backbone plasticity.** *Biochem Soc Trans* 2016, **44**:1523-1529.
  7. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL *et al.*: **Computational design of ligand-binding proteins with high affinity and selectivity.** *Nature* 2013, **501**:212-216.
  8. Siegel JB, Smith AL, Poust S, Wargacki AJ, Bar-Even A, Louw C, Shen BW, Eiben CB, Tran HM, Noor E *et al.*: **Computational protein design enables a novel one-carbon assimilation pathway.** *Proc Natl Acad Sci U S A* 2015, **112**:3704-3709.
  9. Stranges PB, Kuhlman B: **A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds.** *Protein Sci Publ Protein Soc* 2013, **22**:74-82.
  10. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D: **Design of a novel globular protein fold with atomic-level accuracy.** *Science* 2003, **302**:1364-1368.
  11. Lin Y-R, Koga N, Tatsumi-Koga R, Liu G, Clouser AF, Montelione GT, Baker D: **Control over overall shape and size in de novo designed proteins.** *Proc Natl Acad Sci U S A* 2015, **112**:E5478-E5485.
  12. Marcos E, Basanta B, Chidyausiku TM, Tang Y, Oberdorfer G, Liu G, Swapna GVT, Guan R, Silva D-A, Dou J *et al.*: **Principles for designing proteins with cavities formed by curved  $\beta$  sheets.** *Science* 2017, **355**:201-206.  
•• *De novo* design of  $\alpha/\beta$  proteins with various topologies. Kinks were inserted into the  $\beta$  sheet, demonstrating control of curvature and also design of proteins with cavities.
  13. King IC, Gleixner J, Doyle L, Kuzin A, Hunt JF, Xiao R, Montelione GT, Stoddard BL, DiMaio F, Baker D: **Precise assembly of complex beta sheet topologies from de novo designed building blocks.** *eLife* 2015, **4**:e11012.
  14. Bender BJ, Cisneros A, Duran AM, Finn JA, Fu D, Lokits AD, Mueller BK, Sangha AK, Sauer MF, Sevy AM *et al.*: **Protocols for molecular modeling with Rosetta3 and RosettaScripts.** *Biochemistry (Mosc)* 2016, **55**:4748-4763.
  15. Parmeggiani F, Huang P-S: **Designing repeat proteins: a modular approach to protein design.** *Curr Opin Struct Biol* 2017, **45**:116-123.
  16. Parmeggiani F, Huang P-S, Vorobiev S, Xiao R, Park K, Caprari S, Su M, Seetharaman J, Mao L, Janjua H *et al.*: **A general computational approach for repeat protein design.** *J Mol Biol* 2015, **427**:563-575.
  17. Brunette TJ, Parmeggiani F, Huang P-S, Bhabha G, Ekiert DC, Tsutakawa SE, Hura GL, Tainer JA, Baker D: **Exploring the repeat protein universe through computational protein design.** *Nature* 2015, **528**:580-584.  
•• A variety of helical repeat proteins are design and crystallized, some with curvatures unseen in natural helical repeat proteins, demonstrating that the protein design space is larger than what nature has explored.
  18. Park K, Shen BW, Parmeggiani F, Huang P-S, Stoddard BL, Baker D: **Control of repeat-protein curvature by computational protein design.** *Nat Struct Mol Biol* 2015, **22**:167-174.
  19. Rämisch S, Weininger U, Martinsson J, Akke M, André I: **Computational design of a leucine-rich repeat protein with a predefined geometry.** *Proc Natl Acad Sci U S A* 2014, **111**:17875-17880.
  20. Doyle L, Hallinan J, Bolduc J, Parmeggiani F, Baker D, Stoddard BL, Bradley P: **Rational design of  $\alpha$ -helical tandem**

- repeat proteins with closed architectures. *Nature* 2015, **528**:585-588.
21. Voet ARD, Noguchi H, Addy C, Simoncini D, Terada D, Unzai S, Park S-Y, Zhang KYJ, Tame JRH: **Computational design of a self-assembling symmetrical  $\beta$ -propeller protein**. *Proc Natl Acad Sci U S A* 2014, **111**:15102-15107.
  22. Huang P-S, Feldmeier K, Parmeggiani F, Fernandez Velasco DA, Höcker B, Baker D: **De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy**. *Nat Chem Biol* 2016, **12**:29-34.
  23. Drobna I, Ljubetič A, Gradišar H, Pisanski T, Jerala R: **Designed protein origami [Internet]**. In *Protein-based Engineered Nanostructures*. Edited by Cortajarena AL, Grove TZ. Springer International Publishing; 2016:7-27.
  24. Ljubetič A, Drobna I, Gradišar H, Jerala R: **Designing the structure and folding pathway of modular topological bionanostructures**. *Chem Commun* 2016, **52**:5220-5229.
  25. Kočar V, Schreck JS, Čeru S, Gradišar H, Bašič N, Pisanski T, Doye JPK, Jerala R: **Design principles for rapid folding of knotted DNA nanostructures**. *Nat Commun* 2016, **7**:10803.
  26. Gradišar H, Božič S, Doles T, Vengust D, Hafner-Bratkovič I, Mertelj A, Webb B, Sali A, Klavžar S, Jerala R: **Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments**. *Nat Chem Biol* 2013, **9**:362-366.
- The authors introduced a new method to design polyhedral protein folds based on modular coiled-coil forming modules that were arranged in a sequence, so that the polypeptide chain crosses each polyhedron edge exactly twice.
27. Lupas AN, Bassler J: **Coiled coils — a model system for the 21st century**. *Trends Biochem Sci* 2017, **42**:130-140.
  28. Walshaw J, Woolfson DN: **SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures**. *J Mol Biol* 2001, **307**:1427-1450.
  29. Wood CW, Bruning M, Ibarra AA, Bartlett GJ, Thomson AR, Sessions RB, Brady RL, Woolfson DN: **CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies**. *Bioinformatics* 2014, **30**:3029-3035.
  30. Huang P-S, Oberdorfer G, Xu C, Pei XY, Nannenga BL, Rogers JM, DiMaio F, Gonen T, Luisi B, Baker D: **High thermodynamic stability of parametrically designed helical bundles**. *Science* 2014, **346**:481-485.
  31. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN: **Computational design of water-soluble  $\alpha$ -helical barrels**. *Science* 2014, **346**:485-488.
  32. Burton AJ, Thomson AR, Dawson WM, Brady RL, Woolfson DN: **Installing hydrolytic activity into a completely de novo protein framework [Internet]**. *Nat Chem* 2016. advance online publication.
  33. Joh NH, Wang T, Bhate MP, Acharya R, Wu Y, Grabe M, Hong M, Grigoryan G, DeGrado WF: **De novo design of a transmembrane  $Zn^{2+}$ -transporting four-helix bundle**. *Science* 2014, **346**:1520-1524.
  34. Basanta B, Chan KK, Barth P, King T, Sosnick TR, Hinshaw JR, Liu G, Everett JK, Xiao R, Montelione GT *et al.*: **Introduction of a polar core into the de novo designed protein Top7: Polar Core in De Novo Designed Protein Top7**. *Protein Sci* 2016, **25**:1299-1307.
  35. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH *et al.*: **De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity**. *Science* 2016, **352**:680-687.
- The authors explored a method to develop more polar and specific interfaces, by designing buried hydrogen networks. Several such interfaces were successfully designed between  $\alpha$ -helical bundles.
36. Padilla JE, Colovos C, Yeates TO: **Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments**. *Proc Natl Acad Sci U S A* 2001, **98**:2217-2221.
  37. Lai Y-T, Tsai K-L, Sawaya MR, Asturias FJ, Yeates TO: **Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly**. *J Am Chem Soc* 2013, **135**:7738-7743.
  38. Lai Y-T, Reading E, Hura GL, Tsai K-L, Laganowsky A, Asturias FJ, Tainer JA, Robinson CV, Yeates TO: **Structure of a designed protein cage that self-assembles into a highly porous cube**. *Nat Chem* 2014, **6**:1065-1071.
  39. Lai Y-T, Hura GL, Dyer KN, Tang HYH, Tainer JA, Yeates TO: **Designing and defining dynamic protein cage nanoassemblies in solution**. *Sci Adv* 2016, **2**:e1501855.
- A tetrahedron is created by genetically fusing a dimer and trimer. The constructs are extensively characterized via solution SAXS and crystallization. Comparison to previous constructs demonstrates that structures are very sensitive to mutations in the linker region.
40. Kobayashi N, Yasase K, Sato T, Unzai S, Hecht MH, 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 <http://dx.doi.org/10.1021/jacs.5b03593>.
  41. Aupič J, Lapenta F, Strmšek Ž, Jerala R: **Towards designing new nano-scale protein architectures**. *Essays Biochem* 2016, **60**:315-324.
  42. Zhang J, Zheng F, Grigoryan G: **Design and designability of protein-based assemblies**. *Curr Opin Struct Biol* 2014, **27**:79-86.
- A review that discusses the problem of designing assemblies with specific binding and avoiding alternate conformations.
43. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D: **Accurate design of co-assembling multi-component protein nanomaterials**. *Nature* 2014, **510**:103-108.
- A method for constructing symmetric multicomponent assemblies is developed. Several multicomponent assemblies with tetrahedral symmetry are designed and characterized by X-ray crystallography.
44. Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang P-S, Ravichandran R *et al.*: **Design of a hyperstable 60-subunit protein icosahedron**. *Nature* 2016, **535**:136-139.
  45. 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.
  46. Sasaki E, Böhlinger D, van de Waterbeemd M, Leibundgut M, Zschoche R, Heck AJR, Ban N, Hilvert D: **Structure and assembly of scalable porous protein cages**. *Nat Commun* 2017, **8**:14663.
  47. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera — a visualization system for exploratory research and analysis**. *J Comput Chem* 2004, **25**:1605-1612.
  48. Goddard TD, Huang CC, Ferrin TE: **Software extensions to UCSF Chimera for interactive visualization of large molecular assemblies**. *Structure* 2005, **13**:473-482.