Towards designing new nano-scale protein architectures

Jana Aupič*[1], Fabio Lapenta*[1], Žiga Strmšek*[1] and Roman Jerala[1,2‡]

[1]Department of synthetic biology and immunology, National institute of chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia; [2]EN-FIST Centre of excellence, Trg OF 13, 1000 Ljubljana, Slovenia

*Shared equal contribution; ‡To whom the correspondence should be addressed. E-mail: roman.jerala@ki.si.

Abstract

The complexity of designed bionano-scale architectures is rapidly increasing mainly due to the expanding field of DNA-origami technology and accurate protein design approaches. The major advantage offered by polypeptide nanostructures compared to most other polymers resides in their highly programmable complexity. Proteins allow *in vivo* formation of well-defined structures with a precise spatial arrangement of functional groups, providing extremely versatile nano-scale scaffolds. Extending beyond existing proteins that perform a wide range of functions in biological systems, it became possible in the last few decades to engineer and predict properties of completely novel protein folds, opening the field of protein nanostructure design. This review offers an overview on rational and computational design approaches focusing on the main achievements of novel protein nanostructures design.

Introduction

Predicting the tertiary structure of proteins from their sequence is still a challenge, particularly for large and non-natural folds [1]. Designing novel protein structures as an inverse approach to solving the folding problem consists of finding polypeptide sequences compatible to a target structure [2]. However, novel protein nanostructures can also be developed by combining and reengineering already existing protein domains through rational or computational design. Protein design is especially attractive as it could lead to protein structures with completely new functionality, since known protein folds represent only a fraction of the conformational space potentially accessible to polypeptide chains [3]. This review provides an overview of different strategies for protein nanostructure design, briefly describing a few of the numerous examples provided in the recent decades.

Nano-assemblies through protein fusion

Symmetry is widely utilised in nature as a way of building larger protein assemblies such as viruses, actin filaments, clathrin scaffold and S-layer proteins. Viral capsids form highly organised, stable and symmetrical protein nanostructures, where viruses manage to achieve this with small genomes and very limited number of building blocks. Therefore, viral capsids offer a perfect platform to learn from and adapt for our use. One of the strategies to design protein nanostructures is represented by fusion of discrete polypeptide domains. The most straightforward approach is to redesign already existing

nanostructures. An example of this approach is the design of an Epstein-Barr virus vaccine using ferritin and encapsulin fused with part of the gp350 ectodomain [4]. In this case individual building blocks self-assembled into virus-like particles with ferritin or encapsulin determining shape and size, while fused gp350 ectodomain was presented at the surface. The designed virus like particle retained its form and self-assembly properties, successfully presenting gp350 ectodomain to the immune system. Another example of using similar approach is presented in ref. [5] where authors report a design of ball-and-spike protein that is built from a dimer of DNA binding Dps protein and cell-punctuating needles of the T-4 bacteriophage fused together with a flexible linker.

Use of naturally existing nanostructures for design is not limited to 3D motives – S-layers offer a robust starting point for design of protein lattices. S-layers are crystalline protein arrays present at the surface of prokaryotes, forming the outermost envelope of the cell [6]. In an example of creating S-layer based nanostructures [7] researchers reported a fusion of heavy chain camel antibody with truncated form of *Bacillus sphaericus* S-layer through a short linker. Fusion resulted in a square lattice structure that self-assembled on the solid support. Moreover, obtained nanostructure retained the ability of heavy camel antibody to bind prostate-specific antigen, which was detected with surface plasmon resonance, thus proving that fusions of antigen binding domains with S-layer could be used as biosensors.

New shapes of nanostructures can also be obtained by fusing naturally oligomerizing domains to make e.g. polyhedra. In the pioneering example of this approach [8] dimerization and trimerization domains were fused resulting in formation of a tetrahedral protein cage and filaments. To achieve the desired nanostructure it was necessary to identify building blocks that obey geometric restrictions, such as the angle between two symmetry elements, imposed by the selected architecture; an appropriate helical linker, which meets both rigidity and directional requirements and di- and trimerizing proteins in the Protein Data Bank that start or end with an α-helix. After these requirements were met, fused dimer-trimer and dimer-dimer proteins were designed in order to form a tetrahedral protein cage and filaments, respectively. While the initial cages were heterogeneous in size and shape, their redesign [9] differed in the structure of the linker, which offered higher helicity and thus locked the subunits in the appropriate position with less freedom of movement resulting in the formation of monodisperse well-defined tetrahedral cages (Fig. 1A).

Higher polyhedral nanocages designed based on the principle of protein domain fusion were reported, including a highly porous cube [10]. Various fusion constructs using dimeric and trimeric protein domains joined together by a continuous α-helix were designed *in silico,* followed by screening for a defined angle between the two- and three-fold symmetry axes (35.3° for the 24-subunit cube). A fusion of trimeric *E. coli* 2-keto-3-deoxy-6-phosphogalactonate aldolase31 and the dimeric N-terminal domain of *E. coli* FkpA protein 32 linked with a four-residue α-helical linker assembled into three porous nanocages – a 12-, 18- and 24-mer, which represent a tetrahedron, a trigonal prism and a cube, respectively. The porous cube formed a nanostructure with 100 Å openings to a central cavity with a diameter of 130 Å, while the whole structure had a diameter of 225 Å (Fig. 1B).

Nano-assemblies through protein interface design

An alternative but similar approach to building nanostructures based on the self-assembly of multiple domains is by engineering novel protein-protein interactions (PPIs). This strategy offers various advantages in comparison to using naturally oligomerizing protein domains. An immediately evident benefit is the removal of limitation of using only naturally associating subunits, extending the range of potential building blocks. Furthermore, this approach reduces the intrinsic conformational flexibility of designed nanostructures, which often accompanies the use of linked oligomerizing domains as associating units and renders the isolated structures somewhat polydisperse [11–14].

At the root of successes in designing novel protein folds through interface design lie tremendous advances in computational protein design [15–18] and lessons learned from previous attempts to redesign existing PPIs for greater selectivity or binding affinity as well as to engineer completely novel PPIs [19–22]. Computational tools used for designing interacting surfaces are in principle the same as those used in *de novo* protein design. They must be able of generating as many conformations as possible and adequately scoring them to find tightly packed interfaces that form hydrogen bonds and have a favourable solvation energy [18, 23]. For examining conformational space different algorithms such as dead-end elimination, linear integer programming and Monte Carlo optimization are being applied [24–26]. Scoring functions are usually a mix of terms based on physical and empirical considerations [23]. Among computer programs that can carry out interface design, Orbit [27] and Rosetta [28] are by far the most popular.

Protein interfaces are considerably more polar than protein cores. In natural dimers around 40% of interfacial amino acids have polar side chains [29]. On the other hand, a comparison of successful and failed interface designs attempted with Rosetta modelling suite revealed that the former are mainly hydrophobic. This shows that while Rosetta scoring function faithfully recreates packing interactions it still somewhat struggles with describing polar interactions [29], failing to accurately describe the trade-off between desolvation and formation of hydrogen bonds as interface contacts form. Since water is usually only implicitly considered in protein design calculations, the fact that interfacial hydrogen bonds, which are important for high binding affinity [30], are often water-mediated poses an additional problem. This issue has been addressed by development of different algorithms that place individual water molecules at appropriate positions along the interface [31]. Despite these drawbacks, recent advances show that Rosetta might be able to tackle the challenge of designing polar interfaces rather soon [32]. The accuracy of scoring functions is not the only hurdle facing the design of protein interfaces as well as protein design in general. Another limitation is speed, restricting the size of conformational space that can be inspected in reasonable time, consequently imposing constraints on the size of designable interfaces [17]. Different approaches for reducing computational costs connected with calculating interaction terms and searching conformational space are constantly being developed [33–37]. Currently, interface design is feasible for surfaces consisting of approximately 100 residues.

Nanostructures composed of multiple subunits can be achieved through symmetry constraints. Rosetta has been used to design protein complexes with octahedral and tetrahedral point group symmetry that self-assemble from C3 symmetric trimers. Their design process consisted of two stages. Firstly, the trimeric subunit was rigidly docked along the symmetry axes of the target structure in order to find designable interfaces with a high-density of contacting residues in well-anchored regions. After the

appropriate position of the building blocks was found, sequence design was performed. The designed structures were confirmed by X-ray crystallography and electron microscopy. An interesting result was the revelation that very small changes can notably affect the final outcome. For example, a single replacement of Serine with Alanine residue significantly impaired self-assembly [38]. Recently this strategy was advanced even further as King et al. [11] established a method for designing nanostructures from multiple copies of two distinct subunits. To showcase their developments they designed a protein with a dual tetrahedral architecture (Fig. 2A). X-ray crystallography showed that interfaces were designed with high accuracy with a backbone root mean square deviations between 0.5 and 1.2 Å. Moreover, in structures where resolution was high enough to allow a detailed analysis of side chain configurations it was discovered that 87 of 113 side chains were positioned as predicted. The main benefit of using multiple subunits for nanostructure construction is that it increases the range of accessible target structures. As this strategy is advanced even further it will be necessary to take into account also the specificity of designed interactions. While algorithms performing explicit optimisation of the energy difference between target and off-target interactions have been developed, it is becoming clear that mutations promoting the desired interaction are typically also destabilising off-target interactions [16, 39].

Protein domains can assemble also through metal coordination. Metal binding can be achieved through covalent modification of proteins with non-natural metal chelates, though structures of this type usually exhibit pronounced flexibility around the metal centre [40]. To achieve structurally well-defined self-assemblies, the metal ion should be coordinated by the interfacial amino acids. However, coordination bonds are relatively weak, leading to protein assemblies with low stability. Furthermore, it is hard to precisely predict the orientation between the individual building blocks due to the presence of multiple metal-coordinating interfacial amino acids like Asp, Glu, Cys and His. So far only helical bundles, 1D nanotubes and 2D arrays have been assembled through metal binding. A particularly attractive feature of using metal coordination as the driving force for self-assembly is that it can be easily controlled through environmental conditions, like concentration of metal ions, competing ligand and pH [15, 40].

A clever way for designing protein interfaces is the so-called MeTIR approach [41]. Here monomeric protein units are firstly brought together through metal coordination by rationally placing coordinating amino acids at appropriate positions. Once a crystal structure is determined for metal-mediated complexes the interface area is redesigned so that binding occurs even in the absence of metal ions. In such a manner cytochrome cb_{562} was engineered to form a dimer (Fig. 2B), though the complex induced by metal coordination was a tetramer. Design can also be undertaken in reverse. Interfaces between domains of ferritin, a cage-like protein for iron-storage, was redesigned so the assembly could occur only in the presence of divalent copper ions [42].

Design of PPIs has also been extended to yield 3D protein crystals and 2D arrays. In a pioneering attempt Saven et al. designed an α-helical coiled-coil trimer that formed 3D crystals with a P6 space group symmetry [43]. The final structure deviated from the design only by ~1 Å. Their achievement is especially noteworthy as interface design is usually carried out on naturally occurring protein structures, while here the trimeric building block was *de novo* designed. Recently, Rosetta was used to create 2D protein arrays [44]. They sought to achieve layer groups that contain only two unique interfaces and belong to a cyclic

rather than to a dihedral point groups. Out of 6 layer groups that fit this criteria, regular lattices were realised for representatives of three different layer groups (Fig. 2C).

De novo protein design

In addition to using naturally occurring protein domains, protein nanostructures can also be designed completely from scratch. In protein science, the *de novo* design concept emerged gradually in the 1980s. The knowledge of short consensus metal-binding sites, amino acid rotamers and conserved sequence-structure relation for small protein folds stimulated early attempts to test the validity of an inverse approach to the folding problem, establishing the foundation of protein *de novo* design [45–48].

Helix bundle proteins revealed to be a solid template-structure to test the potentiality of a design approach from first principles. First attempts at *de novo* design relied on this type of fold mainly due to the low complexity of its topology, the presence of only one type of secondary structure and the intrinsic periodicity of α-helices [49, 50]. Computational approaches slowly assumed a considerable importance in protein design, mostly thanks to more accurate force field calculations, implementation of energy minimisation through Monte Carlo and larger rotamer libraries [51]. The first example of a computational protocol written to find the optimal sequences for a given backbone conformation resulted in the design of a zinc finger protein [27]. Another important landmark in this field is the computational design of Top7, a 93 residue protein assuming a non-natural topology, which has been accomplished with Rosetta modelling suite [52].

De novo multiple chain peptide-assembly design

Controlled and highly precise intermolecular self-assembling of individual building blocks seems a promising avenue for the construction of sophisticated biomaterials [53]. Multi-chain design is mainly used to compose symmetrical complexes by combining a small number of the same or a small population of peptide building blocks. Larger protein complex designs are based on the juxtaposition of oligomeric protein domains, which primarily relies on the energy minimisation of protein-protein interfaces, extensively described in another section of this review. Coiled coils possess properties such as periodicity, rigidity, specificity and auto-stabilisation which make them a popular choice to build self-assembling multi-chain nanostructures [54]. Rational design of coiled coil units is based on the heptad register rationalisation and is assisted by parametric modelling based on Crick equations [55, 56]. Several examples of peptides forming helix bundles in different oligomeric states have been described [57–59]. A useful property of helical bundles is an internal channel offering positioning of specific functional groups in a controlled environment. An example of such a multi-chain assembly has been reported for a homo-hexameric channel where reduced cysteine residues could be selectively alkylated [60], proving that such structures could be used as a catalytic platform. Coiled coils have also been used to assemble larger multi-chain nanostructures, such as cages [61] and nanotubes [62–66] (Fig. 3A).

De novo design of repetitive and non-repetitive single-chain protein nanostructures

Since protein design requires an extensive optimization of individual structural units, most of protein designs, be it rationally or computationally designed, are based on a modular strategy, which prioritizes

the stability of the building elements. These modular architectures allow a relatively high degree of flexibility in the design process, establishing innovative non-globular folds based more on repetitive and topologically isolated interactions rather than a shared hydrophobic core [67, 68].

Repeat-proteins represent an interesting scaffold for *de novo* design; the regularity of their inner structure, assessed by repetitive interactions makes this kind of structures an interesting choice for nanostructure design. Libraries of randomized repeat-proteins are already widely used in detecting protein-protein interactions [69, 70] and as an aid in protein crystallization [71], showing the remarkable flexibility, variability and functionality of this kind of folds. A common feature of these nanostructures is the encoding of structural units in tandem arrays characterized by repetition of short-range interactions. For instance, a dimeric toroidal nanostructure with a precise curvature allowing for a specific target binding was designed by using Leucine-rich repeat motifs (ribonuclease inhibitor subtype) as modular scaffolding units [72]. Rosetta has been intensively used to design protein nanostructure based on repetitive units and has proven to possess an energy function well-suited for designing such structures [73, 74] (Fig. 3B). The complexity of folds achievable by tandem repeats is however limited as their fold is defined by short and medium range interactions that can define the curvature and pitch of the assemblies.

To create complex folds long range interactions are required [75]. Non-repetitive single-chain polypeptide designs based on helix bundles featuring a localized long-range interaction network were the target of the first *de novo* design attempts [49, 50, 76]. Recently, the helix bundle motif has been revisited and redesigned through a computational approach employing Rosetta modelling suite. In this case, parametric design of different helix bundle proteins has been coupled to sequence optimization, providing non-repetitive sequences able to fold in a three, four and five helix-bundle structure [77]. Despite major progresses in *de novo* design, obtaining a non-repetitive protein scaffold with a long-range network of interactions and complex topologies is still a challenge. Although non-repetitive single chain nanostructures still require modular structural units, the design of long-range intramolecular interactions necessitates a different strategy, which consists of finding a specific arrangement of distinct and separate structural elements that will promote folding into the target structure [75]. A polypeptide requires topological constraints to fold into a complex intertwined architecture and these constraints consist of long-range interactions rather than structurally individual units as in repeat-proteins [78]. These properties, utilizing concepts similar to that of DNA origami technology [68], were applied in order to design a single-chain polypeptide which assumes a tetrahedral fold not found among natural proteins [79] (Fig. 3C). In the case of this peculiar fold, the edges of the polyhedral cage were formed by pairwise long-range interactions between different dimeric coiled coil units. The main advantage offered by a modular coiled coil-based structures lays in the possibility of individually designing each orthogonal pair and assigning specific properties to them, since each dimeric coiled coil provides a distinct structural unit and may be subjected to specific and localized modification. The required unique network of topological constraints that leads the protein to assume its final fold is imposed by the orthogonality of each coiled-coil dimer-forming segment, which are concatenated within the same polypeptide chain. These constraints are designed by positioning segments along the polypeptide chain, leaving the possibility to vary the shape, complexity, length and the topology of such nanocages. This topological-design approach

allows the design of single chain proteins able of forming hollow polyhedral scaffolds characterized by the presence of a modifiable cavity providing a basis for designing modular nano-architectures with an intrinsic stability determined by each structural unit, which control the final shape of the internal cavity and the overall folding state. Overall, foldable nanocages represent an interesting novel scaffold for biomaterial engineering and *in vivo* applications due to their peculiar fold and modularity.

Conclusion

Design of protein nanostructures is in an exciting period of rapid development, employing several different design strategies. Polypeptide design can be achieved by either using existing protein domains and combining them through regulated linking or through specific protein-protein interactions or *de novo* design. While DNA nanostructures offer an easily achievable orthogonality, which underlies the complexity of designed nanostructures, protein nanotechnology is a fast developing field whose boundaries are constantly being expanded [80]. Developments are going in the direction of achieving larger and more complex nanostructures and also towards bestowing function to the designed structures as they show a great potential for a vast array of biotechnological applications.

References

1. R. Shenoy, S. and Jayaram, B. (2010) Proteins: Sequence to Structure and Function – Current Status. Curr. Protein Pept. Sci. **11**, 498–514

2. Shakhnovich, E.I. (1998) Protein design: a perspective from simple tractable models. Fold. Des. **3**, R45–58

3. Taylor, W.R., Chelliah, V., Hollup, S.M., MacDonald, J.T. and Jonassen, I. (2009) Probing the "Dark Matter" of Protein Fold Space. Structure **17**, 1244–52

4. Kanekiyo, M., Bu, W., Joyce, M.G., Meng, G., Whittle, J.R.R., Baxa, U., Yamamoto, T., Narpala, S., Todd, J.-P., Rao, S.S., McDermott, A.B., Koup, R.A., Rossmann, M.G., Mascola, J.R., Graham, B.S., Cohen, J.I. and Nabel, G.J. (2015) Rational Design of an Epstein-Barr Virus Vaccine Targeting the Receptor-Binding Site. Cell **162**, 1090–100

5. Sugimoto, K., Kanamaru, S., Iwasaki, K., Arisaka, F. and Yamashita, I. (2006) Construction of a ball-and-spike protein supramolecule. Angew. Chemie - Int. Ed. **45**, 2725–8

6. Sleytr, U.B. and Beveridge, T.J. (1999) Bacterial S-layers. Trends Microbiol. **7**, 253–60

7. Pleschberger, M., Saerens, D., Weigert, S., Sleytr, U.B., Muyldermans, S., Sára, M. and Egelseer, E.M. (2004) An S-layer heavy chain camel antibody fusion protein for generation of a nanopatterned sensing layer to detect the prostate-specific antigen by surface plasmon resonance technology. Bioconjug. Chem. **15**, 664–71

8. Padilla, J.E., Colovos, C. and Yeates, T.O. (2001) Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc. Natl. Acad. Sci. U. S. A. **98**, 2217–21

9. Lai, Y.-T., Cascio, D. and Yeates, T.O. (2012) Structure of a 16-nm Cage Designed by Using Protein Oligomers. Science **336**, 1129–1129

10. Lai, Y.-T., Reading, E., Hura, G.L., Tsai, K.-L., Laganowsky, A., Asturias, F.J., Tainer, J.A., Robinson, C.

V. and Yeates, T.O. (2014) Structure of a designed protein cage that self-assembles into a highly porous cube. Nat. Chem. **6**, 1065–71

11. King, N.P., Bale, J.B., Sheffler, W., McNamara, D.E., Gonen, S., Gonen, T., Yeates, T.O. and Baker, D. (2014) Accurate design of co-assembling multi-component protein nanomaterials. Nature **510**, 103–8

12. Zhang, J., Zheng, F. and Grigoryan, G. (2014) Design and designability of protein-based assemblies. Curr. Opin. Struct. Biol. **27**, 79–86

13. Božič, S., Doles, T., Gradišar, H. and Jerala, R. (2013) New designed protein assemblies. Curr. Opin. Chem. Biol. **17**, 940–5

14. Lai, Y.T., Tsai, K.L., Sawaya, M.R., Asturias, F.J. and Yeates, T.O. (2013) Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. J. Am. Chem. Soc. **135**, 7738–43

15. Khare, S.D. and Fleishman, S.J. (2013) Emerging themes in the computational design of novel enzymes and protein-protein interfaces. FEBS Lett. **587**, 1147–54

16. Karanicolas, J. and Kuhlman, B. (2009) Computational design of affinity and specificity at protein-protein interfaces. Curr. Opin. Struct. Biol. **19**, 458–63

17. Das, R. and Baker, D. (2008) Macromolecular modeling with rosetta. Annu. Rev. Biochem. **77**, 363–82

18. Lippow, S.M. and Tidor, B. (2007) Progress in computational protein design. Curr. Opin. Biotechnol. **18**, 305–11

19. Huang, P.-S., Love, J.J. and Mayo, S.L. (2007) A de novo designed protein protein interface. Protein Sci. **16**, 2770–4

20. Jha, R.K., Leaver-Fay, A., Yin, S., Wu, Y., Butterfoss, G.L., Szyperski, T., Dokholyan, N. V. and Kuhlman, B. (2010) Computational Design of a PAK1 Binding Protein. J. Mol. Biol. **400**, 257–70

21. Karanicolas, J., Corn, J.E., Chen, I., Joachimiak, L.A., Dym, O., Peck, S.H., Albeck, S., Unger, T., Hu, W., Liu, G., Delbecq, S., T. Montelione, G., P. Spiegel, C., Liu, D.R. and Baker, D. (2011) A De Novo Protein Binding Pair By Computational Design and Directed Evolution. Mol. Cell **42**, 250–60

22. Stranges, P.B., Machius, M., Miley, M.J., Tripathy, A. and Kuhlman, B. (2011) Computational design of a symmetric homodimer using B-strand assembly. Proc. Natl. Acad. Sci. **108**, 20562–7

23. Boas, F.E. and Harbury, P.B. (2007) Potential energy functions for protein design. Curr. Opin. Struct. Biol. **17**, 199–204

24. Kingsford, C.L., Chazelle, B. and Singh, M. (2005) Solving and analyzing side-chain positioning problems using linear and integer programming. Bioinformatics **21**, 1028–36

25. Desmet, J., De Maeyer, M., Hazes, B. and Lasters, I. (1992) The dead-end elimination theorem and its use in protein side-chain positioning. Nature **356**, 539–42

26. Voigt, C.A., Gordon, D.B. and Mayo, S.L. (2000) Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. J. Mol. Biol. **299**, 789–803

27. Dahiyat, B.I. and Mayo, S.L. (1997) De Novo Protein Design: Fully Automated Sequence Selection. Science **278**, 82–7

28. Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W., Davis, I.W., Cooper, S., Treuille, A., Mandell, D.J., Richter, F., Ban, Y.E.A., Fleishman, S.J., Corn, J.E., Kim, D.E., Lyskov, S., Berrondo, M., Mentzer, S., Popović, Z., Havranek, J.J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J.J., Kuhlman, B., Baker, D. and Bradley, P. (2011) Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. **487**, 545–74

29. Stranges, P.B. and Kuhlman, B. (2013) A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. Protein Sci. **22**, 74–82

30. Reichmann, D., Phillip, Y., Carmi, A. and Schreiber, G. (2008) On the contribution of water-mediated interactions to protein-complex stability. Biochemistry **47**, 1051–60

31. Jiang, L., Kuhlman, B., Kortemme, T. and Baker, D. (2005) A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. Proteins Struct. Funct. Genet. **58**, 893–904

32. Basanta, B., Chan, K.K., Barth, P., King, T., Hinshaw, J.R., Sosnick, T.R., Liu, G., Everett, J., Xiao, R., Montelione, G.T. and Baker, D. (2016) Introduction of a polar core into the de novo designed protein Top7. Protein Sci.

33. Traoré, S., Roberts, K.E., Allouche, D., Donald, B.R., André, I., Schiex, T. and Barbe, S. (2016) Fast search algorithms for computational protein design. J. Comput. Chem. , 1048 – 58

34. Yang, X. and Saven, J.G. (2005) Computational methods for protein design and protein sequence variability: Biased Monte Carlo and replica exchange. Chem. Phys. Lett. **401**, 205–10

35. Allen, B.D. and Mayo, S.L. (2006) Dramatic performance enhancements for the FASTER optimization algorithm. J. Comput. Chem. **27**, 1071–5

36. Xie, W. and Sahinidis, N. V. (2006) Residue-rotamer-reduction algorithm for the protein side-chain conformation problem. Bioinformatics **22**, 188–94

37. Georgiev, I., Lilien, R.H. and Donald, B.R. (2008) The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. J. Comput. Chem. **29**, 1527–42

38. King, N.P., Sheffler, W., Sawaya, M.R., Vollmar, B.S., Sumida, J.P., André, I., Gonen, T., Yeates, T.O. and Baker, D. (2012) Computational design of self-assembling protein nanomaterials with atomic level accuracy. Science **336**, 1171–4

39. Havranek, J.J. and Harbury, P.B. (2003) Automated design of specificity in molecular recognition. Nat. Struct. Biol. **10**, 45–52

40. Sontz, P.A., Song, W.J. and Tezcan, F.A. (2014) Interfacial metal coordination in engineered protein and peptide assemblies. Curr. Opin. Chem. Biol. **19**, 42–9

41. Salgado, E.N., Ambroggio, X.I., Brodin, J.D., Lewis, R.A., Kuhlman, B. and Tezcan, F.A. (2010) Metal templated design of protein interfaces. Proc. Natl. Acad. Sci. U. S. A. **107**, 1827–32

42. Huard, D.J.E., Kane, K.M. and Tezcan, F.A. (2013) Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. Nat. Chem. Biol. **9**, 169–76

43. Lanci, C.J., MacDermaid, C.M., Kang, S. -g., Acharya, R., North, B., Yang, X., Qiu, X.J., DeGrado,

W.F. and Saven, J.G. (2012) Computational design of a protein crystal. Proc. Natl. Acad. Sci. **109**, 7304–9

44. Gonen, S., DiMaio, F., Gonen, T. and Baker, D. (2015) Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. Science **348**, 1365–8

45. Drexler, K.E. (1981) Molecular engineering: An approach to the development of general capabilities for molecular manipulation. Proc. Natl. Acad. Sci. U. S. A. **78**, 5275–8

46. Pabo, C. (1983) Molecular technology. Designing proteins and peptides. Nature **301**, 200

47. Ponder, J.W. and Richards, F.M. (1987) Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J. Mol. Biol. **193**, 775–91

48. Bowie, J.U., Lüthy, R. and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science **253**, 164–70

49. Regan, L. and DeGrado, W.F. (1988) Characterization of a helical protein designed from first principles. Science **241**, 976–8

50. Hecht, M.H., Richardson, J.H., Richardson, D.C. and Ogden, R.C. (1990) De Novo Design, Expression and Characterization of Felix: A Four-Helix Bundle Protein of Native-Like Sequence. Science **249**, 973–973

51. Kuhlman, B. and Baker, D. (2000) Native protein sequences are close to optimal for their structures. Proc. Natl. Acad. Sci. **97**, 10383–8

52. Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L. and Baker, D. (2003) Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. Science **302**, 1364–8

53. Busseron, E., Ruff, Y., Moulin, E. and Giuseppone, N. (2013) Supramolecular self-assemblies as functional nanomaterials. Nanoscale **5**, 7098

54. Kohn, W.D. and Hodges, R.S. (1998) De novo design of α-helical coiled coils and bundles: models for the development of protein-design principles. Trends Biotechnol. **16**, 379–89

55. Wood, C.W., Bruning, M., Ibarra, A.A., Bartlett, G.J., Thomson, A.R., Sessions, R.B., Brady, R.L. and Woolfson, D.N. (2014) CCBuilder: An interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. Bioinformatics **30**, 3029–35

56. Crick, F.H.C. (1953) The packing of α-helices: simple coiled-coils. Acta Crystallogr. **6**, 689–97

57. Liu, J., Zheng, Q., Deng, Y., Cheng, C.-S., Kallenbach, N.R. and Lu, M. (2006) A seven-helix coiled coil. Proc. Natl. Acad. Sci. **103** , 15457–62

58. Zaccai, N.R., Chi, B., Thomson, A.R., Boyle, A.L., Bartlett, G.J., Bruning, M., Linden, N., Sessions, R.B., Booth, P.J., Brady, R.L. and Woolfson, D.N. (2011) A de novo peptide hexamer with a mutable channel. Nat. Chem. Biol. **7**, 935–41

59. Thomson, A.R., Wood, C.W., Burton, A.J., Bartlett, G.J., Sessions, R.B., Brady, R.L. and Woolfson, D.N. (2014) Computational design of water-soluble α-helical barrels. Science **346**, 485–8

60. Burton, A.J., Thomas, F., Agnew, C., Hudson, K.L., Halford, S.E., Brady, R.L. and Woolfson, D.N. (2013) Accessibility, reactivity, and selectivity of side chains within a channel of de novo peptide assembly. J. Am. Chem. Soc. **135**, 12524–7

61. Fletcher, J.M., Harniman, R.L., Barnes, F.R.H., Boyle, A.L., Collins, A., Mantell, J., Sharp, T.H., Antognozzi, M., Booth, P.J., Linden, N., Miles, M.J., Sessions, R.B., Verkade, P. and Woolfson, D.N. (2013) Self-assembling cages from coiled-coil peptide modules. Science **340**, 595–9

62. Hume, J., Sun, J., Jacquet, R., Renfrew, P.D., Martin, J.A., Bonneau, R., Gilchrist, M.L. and Montclare, J.K. (2014) Engineered coiled-coil protein microfibers. Biomacromolecules **15**, 3503–10

63. Potekhin, S.A., Melnik, T.N., Popov, V., Lanina, N.F., Vazina, A.A., Rigler, P., Verdini, A.S., Corradin, G. and Kajava, A. V. (2001) De novo design of fibrils made of short α-helical coiled coil peptides. Chem. Biol. **8**, 1025–32

64. Egelman, E.H., Xu, C., Dimaio, F., Magnotti, E., Modlin, C., Yu, X., Wright, E., Baker, D. and Conticello, V.P. (2015) Structural plasticity of helical nanotubes based on coiled-coil assemblies. Structure **23**, 280–9

65. Burgess, N.C., Sharp, T.H., Thomas, F., Wood, C.W., Thomson, A.R., Zaccai, N.R., Brady, R.L., Serpell, L.C. and Woolfson, D.N. (2015) Modular Design of Self-Assembling Peptide-Based Nanotubes. J. Am. Chem. Soc. **137**, 10554–62

66. Xu, C., Liu, R., Mehta, A.K., Guerrero-Ferreira, R.C., Wright, E.R., Dunin-Horkawicz, S., Morris, K., Serpell, L.C., Zuo, X., Wall, J.S. and Conticello, V.P. (2013) Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. J. Am. Chem. Soc. **135**, 15565–78

67. Brunette, T., Parmeggiani, F., Huang, P.-S., Bhabha, G., Ekiert, D.C., Tsutakawa, S.E., Hura, G.L., Tainer, J.A. and Baker, D. (2015) Exploring the repeat protein universe through computational protein design. Nature **528**, 580–4

68. Kočar, V., Božič Abram, S., Doles, T., Bašić, N., Gradišar, H., Pisanski, T. and Jerala, R. (2015) TOPOFOLD, the designed modular biomolecular folds: polypeptide-based molecular origami nanostructures following the footsteps of DNA. Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology **7**, 218–37

69. Varadamsetty, G., Tremmel, D., Hansen, S., Parmeggiani, F. and Plückthun, A. (2012) Designed armadillo repeat proteins: Library generation, characterization and selection of peptide binders with high specificity. J. Mol. Biol. **424**, 68–87

70. Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P. and Plückthun, A. (2003) Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J. Mol. Biol. **332**, 489–503

71. Valerio-Lepiniec, M., Urvoas, A., Chevrel, A., Guellouz, A., Ferrandez, Y., Mesneau, A., de la Sierra-Gallay, I.L., Aumont-Nicaise, M., Desmadril, M., van Tilbeurgh, H. and Minard, P. (2015) The αRep artificial repeat protein scaffold: a new tool for crystallization and live cell applications. Biochem. Soc. Trans. **43**, 819–24

72. Rämisch, S., Weininger, U., Martinsson, J., Akke, M. and André, I. (2014) Computational design of a leucine-rich repeat protein with a predefined geometry. Proc. Natl. Acad. Sci. U. S. A. **111**, 17875–80

73. Parmeggiani, F., Huang, P.-S., Vorobiev, S., Xiao, R., Park, K., Caprari, S., Su, M., Seetharaman, J., Mao, L., Janjua, H., Montelione, G., Hunt, J. and Baker, D. (2015) A general computational approach for repeat protein design. J. Mol. Biol. **427**, 563–75

74. Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B.L. and Bradley, P. (2015) Rational design of α-helical tandem repeat proteins with closed architectures. Nature **528**, 585–8

75. Gromiha, M.M. and Selvaraj, S. (1999) Importance of long-range interactions in protein folding. Biophys. Chem. **77**, 49–68

76. Kamtekar, S., Schiffer, J., Xiong, H., Babik, J. and Hecht, M. (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science **262**, 1680–5

77. Huang, P.-S., Oberdorfer, G., Xu, C., Pei, X.Y., Nannenga, B.L., Rogers, J.M., DiMaio, F., Gonen, T., Luisi, B. and Baker, D. (2014) High thermodynamic stability of parametrically designed helical bundles. Science **346**, 481–5

78. Ljubetič, A., Drobnak, I., Gradišar, H. and Jerala, R. (2016) Designing the structure and folding pathway of modular topological bionanostructures. Chem. Commun. **52**, 5220–9

79. Gradišar, H., Božič, S., Doles, T., Vengust, D., Hafner-Bratkovič, I., Mertelj, A., Webb, B., Šali, A., Klavžar, S. and Jerala, R. (2013) Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. Nat. Chem. Biol. **9**, 362–6

80. Hsia, Y., Bale, J.B., Gonen, S., Shi, D., Sheffler, W., Fong, K.K., Nattermann, U., Xu, C., Huang, P.-S., Ravichandran, R., Yi, S., Davis, T.N., Gonen, T., King, N.P. and Baker, D. (2016) Design of a hyperstable 60-subunit protein icosahedron. Nature **in press**

Figures

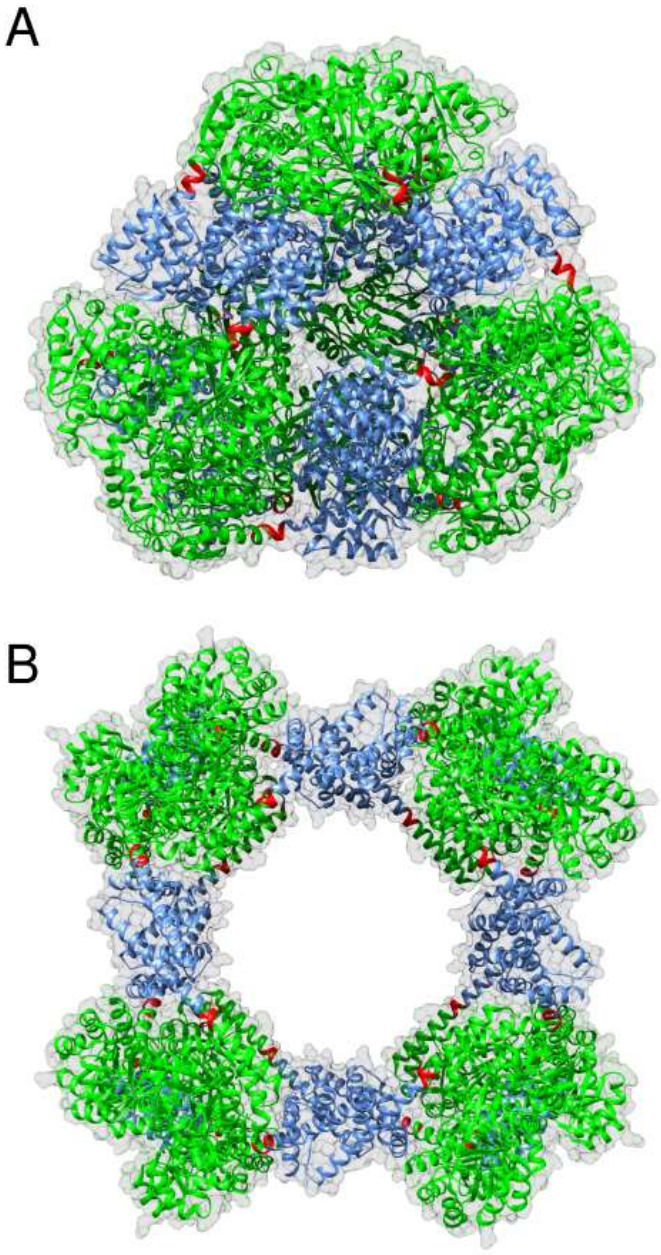

Figure 1 – Examples of nanocages designed based on the principle of protein domain fusion. (A) Crystal structure of a tetrahedral nanocage, self-assembled from a fusion of di- and trimer (PDB: 3VDX) [9]. (B) A crystal structure of a nanocage in the shape of a cube also self-assembled from a fusion of di- and trimer (PDB: 4QCC) [10]. Hydrophilic core was observed in both examples.

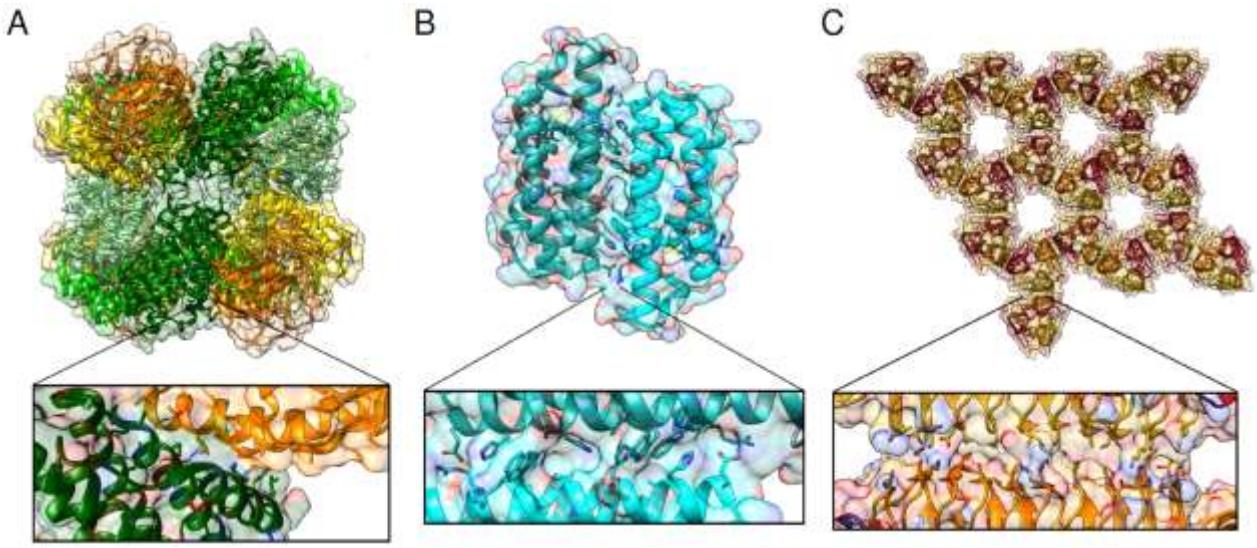

Figure 2 – Nano assemblies achieved through designing PPIs. Upper panels show structures of three different examples of nanocages constructed through designing PPIs, while the lower panels provide a close-up look at the designed interface. (A) Crystal structure of a protein complex with a tetrahedral symmetry, achieved by designing protein interactions between two distinct protein units (PDB: 4NWQ). 87 out of 113 interfacial amino acids assumed the predicted conformation [11]. (B) Crystal structure of a redesigned cb_{562} dimer obtained by using MeTIR approach (PDB: 3HNK). PPIs are mainly hydrophobic [41]. (C) Designed structure of a 2D protein crystal belonging to layer group P 3 2 1. The designed interface is characterized by zipper-like hydrophobic packing and hydrogen bonds at the edges of the interface [44].

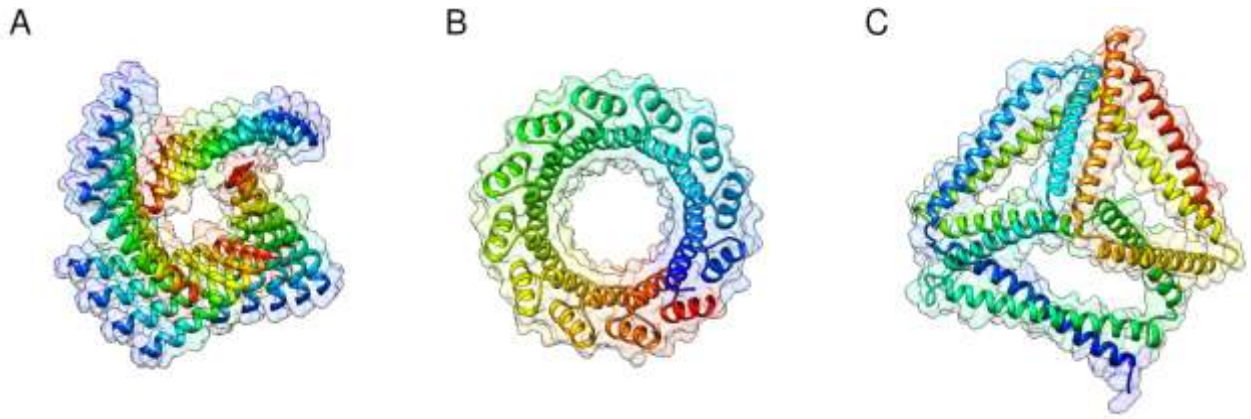

Figure 3 – *De novo* protein design. (A) Self-assembling 29-residue coiled coil peptides forming nanotubes [64] (PDB: 3J89). (B) Toroidal tandem-repeat *de novo* protein designed with Rosetta (PDB: 5BYO) [73]. (C) Protein tetrahedron coiled coil-based nanocage rationally designed [79].

Acknowledgments:

We acknowledge the financial support of the Slovenian Research Agency (program P4-0176, projects N4-0037, L4-6812, J4-5528), the ERANET SynBio project Bioorigami, and TOLLerant project (H2020 Marie Sklodowska-Curie grant agreement No 642157).