PORTLAND PRESS

Review Article

Better together: building protein oligomers naturally and by design

Rebecca E.A. Gwyther[1], D. Dafydd Jones[1] and Harley L. Worthy[2]

[1]Molecular Biosciences, School of Biosciences, Cardiff University, Cardiff, U.K.; [2]School of Biosciences, University of Exeter, Exeter, U.K.

Correspondence: Harley L. Worthy (H.Worthy@exeter.ac.uk) or D. Dafydd Jones (jonesdd@cardiff.ac.uk)

OPEN ACCESS

Protein oligomers are more common in nature than monomers, with dimers being the most prevalent final structural state observed in known structures. From a biological perspective, this makes sense as it conserves vital molecular resources that may be wasted simply by generating larger single polypeptide units, and allows new features such as cooperativity to emerge. Taking inspiration from nature, protein designers and engineers are now building artificial oligomeric complexes using a variety of approaches to generate new and useful supramolecular protein structures. Oligomerisation is thus offering a new approach to sample structure and function space not accessible through simply tinkering with monomeric proteins.

Introduction

The vast expanse of protein sequence space has been the driving force behind molecular evolution, with the quaternary structure adding another level of diversity so further expanding the variation observed. Individual polypeptide chains amalgamate to produce a macromolecular complex through a process of oligomerisation (Figure 1a). The molecular coupling of individual subunits leads to new structural and functional properties beyond those available in a simple monomeric system so offering evolutionary benefits, especially when starting from a limited monomer repertoire. This is apparent from the dominance of oligomers over monomers as a final structural form for proteins; with symmetrical homodimers representing the most commonly observed form in the Protein Databank [1,2].

The fundamental basis behind protein oligomerisation is the network of intricate interactions holding the individual units together. Protein oligomers are normally formed and stabilised by a variety of mechanisms, the most common of which is the formation of a mutually compatible non-covalent interface (Figure 1b). The interface is stabilised by hydrophobic interactions, H-bonds and electrostatic attraction [3,4]. Other less common mechanisms include domain swapping (e.g. β-strand exchange, as seen in dimeric cytokines and some cytochromes [5]), binding of shared metal cofactors (e.g. insulin [3,6]) and formation of inter-protein disulfide bridges to connect monomers, with linking of the light and heavy chains in antibodies being an excellent example [1–3,7,8]. Crucially, the total sum of interaction between monomers dictate its strength, which is varied advantageously to provide temporal and spatial regulation over protein complex interactions. Examples of this are evident in large complex structures like the cytoskeleton and collagen [2,9], and multienzyme complexes such as RNA polymerase and pyruvate dehydrogenase [3], to small functional dimers such as cell receptors and transcription factors [3,9,10]. However, looking beyond the interactions that drive oligomerisation, there are several reasons behind the triumph of the protein oligomer; and these are something synthetic biologists need to consider when designing novel proteins and genomes.
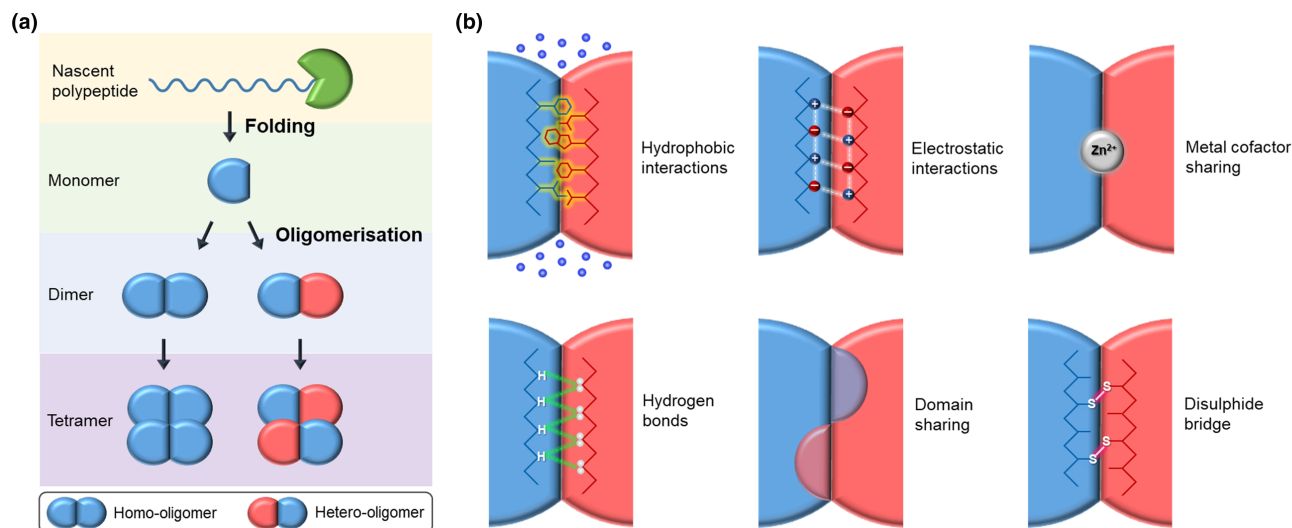
Figure 1. Oligomerisation of proteins.

(**a**) Oligomer formation. Nascent polypeptides enter the endoplasmic reticulum and undergo folding to form their monomeric structure. Monomers can then partake in a process of oligomerisation to form dimers, tetramers and other higher order oligomers. Association of the same monomer forms homo-oligomers, whilst interaction with a different monomer will form hetero-oligomers. (**b**) Mechanisms of interaction at the oligomeric interface to form a mutually compatible interface. Non-polar residues can mediate hydrophobic interactions, whilst polar residues can facilitate hydrogen bonds and, to a lesser extent, electrostatic interaction. Less common approaches utilise domain and metal cofactor sharing to bring monomers together. Disulfide bridges are the main if relatively rare covalent mechanism mediating oligomerisation, and this is often exploited artificially to oligomerise two polypeptides.

Fundamentals of oligomeric success

Given the abundance of oligomeric proteins, a key question is what makes oligomeric proteins so evolutionarily successful and why are the majority symmetrical? [3,9] Numerous people have attempted to answer this including Klotz [11] and Monod et al. [12] in the 1960s, and more recently by both Goodsell and Olson [3], and Ali and Imperiali [9], who reviewed the topic at the turn of the century. General conclusions included that larger proteins were seen as preferential to smaller proteins, because they are more resistant to denaturation and degradation through reduced solvent to surface area exposure [3,4,11,12]. There is also the advantage of having multiple active sites allowing for cooperative functionality. For example, large multienzyme complexes, such as RNA polymerase have an increased turnover rate compared with the subunits acting independently [3]. Whilst these are benefits that may have been achieved through larger single proteins, nature has favoured the use of smaller subunits to generate the same effect.

Multiple proteins subunits offer a safety net for translation errors; by creating a large protein complex of monomers, subunits with an error can be quickly discarded without great strain on the cells' resources [3]. In prokaryotes, ∼25% of proteins of 500 amino acids or more contain an amino acid substitution and $1/7$ of proteins are released from the ribosome before the full-length protein is created [13,14]. This means that proteins greater than 2000 amino acids are rarely fully translated and when they are translated the protein contains at least one error [13,14]. There is also an increased risk of misfolding with longer polypeptide chains having a more complex folding energy landscape and as such may require chaperone proteins [13,14]. A final advantage of utilising multiple subunits is increased coding efficiency at the DNA level. For example, a protein with 1000 amino acids could be coded by a single gene 3 kb long (not including regulatory DNA) or could be made up of four identical subunits 250 amino acids in length, requiring a gene of only 750 bp. Not only does this save genetic space and space within the cell, energy is conserved through more efficient replication and transcription of the subunits [2,3,15]. This, combined with monomeric interactions, are key factors to consider when designing novel oligomeric proteins.

Designing artificial oligomers

Designing and producing artificial self-assembling protein complexes is currently of great interest to protein engineering, and a variety of inventive techniques [5,16,17] have been used to mediate this, including fusion

proteins/domains [18] (e.g. Nanohedra [19] and protein nanobuilding blocks [20,21]), split proteins/domains (e.g. Spycatcher [22] and split luciferase domains [23]), helix-helix interactions [24–26], metal ion bridging [27–29], cofactor bridging [30], and disulfide bridging [7] — to name a few [31–34].

The primary reason for developing artificial oligomers is to explore the new structure and function space not currently present in nature and to use these new designs to try and solve technological, medical and scientific problems [16,35,36]. For example, using domain insertion of split luciferase domains has been used to create Rho GTPase biosensors, where one half of the luciferase is genetically incorporated into the GTPase and the second part of luciferase is attached to potential ligands [23]. If the ligand binds to the GTPase then the two fragments of luciferase combine to form a functionally active bioluminescent protein [23]. Using the same split-domain technique, the Spycatcher–Spytag system was developed and used for localisation and assisted purification of proteins of interest [22,37], Meanwhile, Bailey et al. [38] pursued a different a strategy: metal-coordination chemistry to direct assembly of protein oligomers. This resulted in the creation of protein nano-tubes and lattices with variable diameters [28,39], whilst Song and Tezcan [27] used the strategy to engineer rudimentary β-lactamase activity into a cytochrome-based scaffold.

Most of the techniques mentioned above rely on using structural information gleaned from the protein databank based on oligomers found in nature [28,33,36]. This information is then used to engineer proteins of interest to conform to a set of rules to generate dimerisation interfaces [10,25,40]. These interfaces multimerise because they form hydrophobic patches that associate to escape solvent, hydrogen bonding networks where many weak interactions create a stable interface [2], or the halves of an interface co-ordinate a metal ion [11,12] or cofactors like heme [2,9,30]. A potential downside to these methodologies is the requirement for an extensive design process and significant engineering of proteins to incorporate oligomerisation into normally monomeric systems. Relying solely on symmetry can also limit the shapes and lattices that can be explored [41].

A potential alternative to the empirical-based dimer design is disulfide cross-linking via engineered cysteine residues. As mentioned above, engineering cysteines into proteins to form inter-protein cross-links is an approach used by nature and popular amongst researchers, requiring very little modification of the target protein (usually only a single residue). Functional covalent dimers of azurin were created using this method, mutating residue Asn42 to cysteine and allowing the formation of a disulfide linkage between two monomers [42]. However, the dimers showed a decrease in electron transfer, which was attributed to reduced flexibility from the short disulfide linker [42]. The short side chain of cysteine ($-CH_2-SH$) will make many interaction interfaces sterically impossible and when generating hetero-complexes or beyond dimers, off-pathway disulfide bonds will generate a mix of oligomeric states. Other risks accompanying this method include the increased likelihood of misfolded proteins from the newly mutated cysteines forming incorrect disulfide bridges with existing cysteine residues [43], and the unsuitability for *in vivo* work, due to the reducing environment of most living cells [43].

Use of non-canonical amino acids

Through the use of non-canonical amino acids (ncAAs) coupled with codon reprogramming new chemistries have been introduced into proteins by design which overcome the site specificity issues [44–47]. An example of these bioorthogonal chemistries is strain promoted azide-alkyne cycloaddition (SPAAC) [48]. Incorporation of ncAAs and the SPAAC reaction has been extensively reviewed and studied elsewhere [47,49–55]. In theory, incorporating an azide moiety (e.g. p-azido-ʟ-phenylalanine, AzF) into one protein and an alkyne (e.g. s-cyclooctyne-ʟ-lysine, SCO) into another (Figure 2a), it is possible to create defined covalent protein dimers in defined orientations. Other examples of generating protein oligomers using ncAAs include the synthesis of antibody dimers [56], ubiquitin [57] dimers, metal-chelating homotrimers [29] (Figure 2b), and multifunctional enzymatic complexes [58,59], which use a variety different ncAAs: p-acetylphenylalanine [56], azidohomoalanine [57,59], (2,2′-bipyridin-5yl)alanine [29], and other aliphatic and aromatic alkyne derivatives [60,61].

The Jones group has successfully demonstrated the use of genetically encoded SPAAC by linking monomers of fluorescent proteins together to create functionally enhanced dimers of super-folder green fluorescent protein (sfGFP) (Figure 2a) [62]. The fluorescence mechanism and properties of sfGFP and other *Aequorea victoria* derived FPs is well defined [63–66] with the central active component being the solvent-shielded chromophore, buried within the β-barrel structure. Composed of residues 65 (variable in variants of GFP and Thr in sfGFP), Tyr66 and Gly67, the chromophore can exist in two protonation states: the less populated CRO A, featuring a neutral protonated phenol group of Tyr66, or the more fluorescent and highly populated CRO B with
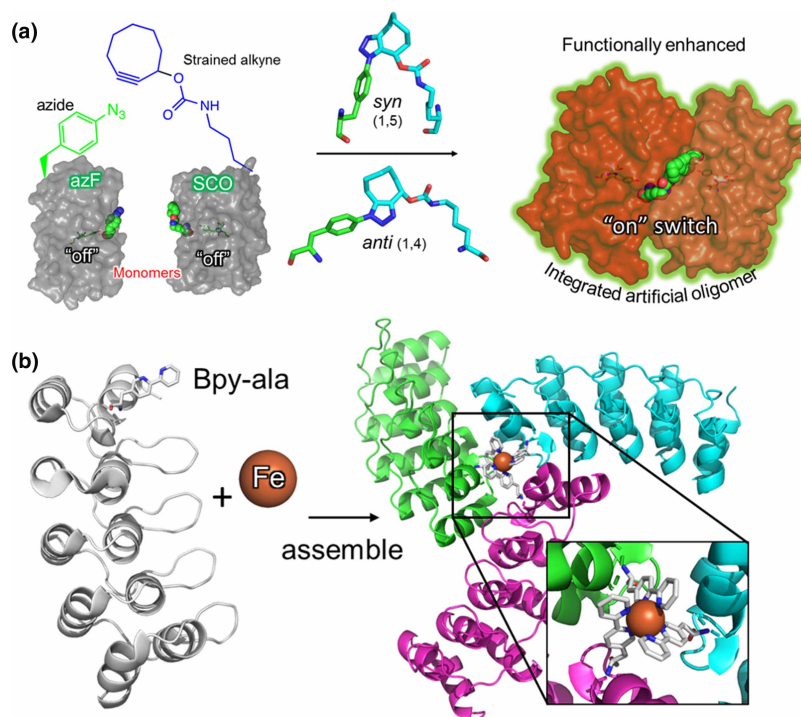
Figure 2. Oligomerisation of proteins via designed ncAA incorporation.

(**a**) Dimerisation of sfGFP^{ncAA} monomers. A scheme depicting the strain promoted azide-alkyne cycloaddition between two non-fluorescent monomers containing azF and SCO (left), forming either 'syn' or 'anti' triazole linkages (centre) to generate a fluorescent dimer (dimer). In the case above, the original monomeric proteins are largely inactive that are switched on when dimerised through linkage via residue 148 (green spheres, PDB 5NHN). Figure is copied with permission from Worthy et al. [62] under a creative commons license. (**b**) Scheme showing the oligomerisation of a synthetic ankyrin repeat protein (grey), containing the metal-chelating ncAA 2,2′-bipyridin-5yl-alanine (Bpy-ala, shown as sticks). Briefly, upon addition of iron ions (orange sphere, Fe), three protein monomers assemble to form a spiral pattern. The iron atom forms the nucleus of the trimer co-ordinated by three Bpy-ala residues. (PDB 5EIL [20]).

a charged phenolate; switching between these two states gives rise to its characteristic spectral properties [64,65]. Residue His148 plays a crucial role in the deprotonation of Tyr66 [63]. Mutation of H148 to a ncAA results in the breakage of this key H-bond causing the CRO A chromophore to predominate [51,54]. The formation of sfGFP homodimers using SPAAC compatible residues at 148 not only reverses this protonation state so switching on CRO B, but enhances brightness over threefold above wild type sfGFP — indicative of functional synergy [62]. Comparison of the 400 nm : 485 nm excitation peaks would thus allow a ratiometric estimation of the CRO A monomer to CRO B dimer population. The study of the structures arising from these artificial protein dimers suggests that the improved fluorescence is due to the formation of extended hydrogen bonding networks between both chromophores. This work paves the way for not just linking monomeric proteins together but shows how generating intimate interactions can lead to new emergent properties.

Using protein dimerisation to monitor protein–protein interactions

The archetypal technique for monitoring protein–protein interactions (PPIs) is fluorescent biosensors, which transduce real-time ligand-binding events into a measurable fluorescence signal [67]. These proximity-based biosensors have numerous advantages over alternative strategies, including their selectivity and sensitivity in spectral analysis, temporal and spatial resolution in biomolecular imaging and relative low cost [68,69]. However, these properties vary inherently between different subtypes of the fluorescent biosensor, bringing selective advantages and disadvantages to each application.

Fluorescence resonance energy transfer (FRET) [70,71], utilises the overlapping emission and excitation spectra of two different fluorophores to stimulate a change in fluorescence when their proximity is <10 nm [72]. This becomes a useful experimental tool when fusing the fluorophores to two potential interaction partners/domains, as the fluorescence output should correlate with their proximity, and thus interaction. Limitations to this, however, include the low signal-to-noise ratio (SNR) from background autofluorescence and the sensitivity of fluorescent proteins (FPs) to changes in their microenvironment [73]. Plus, the most abundant oligomerisation event, homo-dimerisation cannot be easily monitored. Biomolecular fluorescent complementation overcomes the background autofluorescence of FRET by physically splitting the FPs and attaching the two halves to putative interacting proteins, restoring emission only when an interaction occurs [74,75]. Nevertheless, limitations here are often temporally linked: slow off-rates between the split fragments prevent time-dependent studies, delays in fluorescent readouts arise from protein folding and chromophore maturation and false-positives arising from non-specific self-assembly [76]. The final biosensing approach involves engineering single FPs to respond to analytes directly by incorporating receptor elements into FP design [77–80]. This approach effectively increases the temporal perception but is hampered by the complex design process; with a prerequisite for precise structural knowledge and conformational change modelling to ensure correct protein folding upon analyte binding [81].

In an attempt to expand this repertoire of proximity-based biosensors, dimerisation-dependent FP (ddFP) biosensors have become a new focus for the scientific community [82–84]. This strategy typically involves the formation of a fluorescent heterodimer from two non-fluorescent counterparts: a quenched monomer and a monomer lacking a chromophore [84]. Attaching these FPs to separate interacting proteins brings the complex together to form a FP dimer and so enhancing the fluorescence output and thus the SNR. However, considering the aforementioned developments in mutations centred on residue 148 of sfGFP mutants [51,62], it is exciting to note the potential homodimers may have used as a novel suite of ddFP biosensors. Not only will these be easier to engineer than heterodimers, homodimers are likely to be more stable to fluctuations in the microenvironment. This is because monomers should respond in a near-enough identical manner; whereas heterodimers will experience inherent variation and as such, unpredictably skew fluorescent measurements. It also allows ratiometric sensing which removes concentration-dependent measurement issues. Additionally, it opens up the ability to monitor homo-oligomeric complexes as only one engineered FP is needed to achieve a response. Overall, oligomerisation is shown to be a fantastic tool for monitoring PPIs, and future developments in this area will likely have a widespread impact across biological research.

Conclusion

Focusing on converting perceived 'hard-to-use' oligomeric proteins to monomers limits the available function space open to the protein designer and engineer. Taking inspiration from nature and its concept of building up complex supramolecular structures from simple components (protein monomers), we can now envisage assembling the complex protein–based structure. Approaches available to expedite such 'bottom-up' assembly are expanding all the time taking in classical protein engineering to *in silico* design to the use of new chemistry types. Such potential is already being shown with the simple peptide-based system (reviewed by Beesley and Woolfson [85]). We should also consider incorporating different molecular types, even including abiological materials such as the exciting nano-carbon system [86]. So sometimes, it is better for proteins to be together.

Perspectives

- Oligomeric proteins complexes are the most common structural form of proteins found in nature.

- Taking inspiration from nature new artificial protein oligomeric systems are being constructed to sample new functional and structural features not accessible in monomeric proteins

- Major steps have been taken in constructing artificial protein oligomers but more is to come in terms of function, complexity, applications and even inclusion of non-biological components

Abbreviation

ddFP, dimerisation-dependent FP; FRET, Fluorescence resonance energy transfer; ncAAs, non-canonical amino acids; PPIs, protein–protein interactions; sfGFP, super-folder green fluorescent protein; SNR, signal-to-noise ratio; SPAAC, strain promoted azide-alkyne cycloaddition

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1   Marianayagam, N.J., Sunde, M. and Matthews, J.M. (2004) The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **29**, 618–625 https://doi.org/10.1016/j.tibs.2004.09.006

2   Mei, G., Di Venere, A., Rosato, N. and Finazzi-Agro, A. (2005) The importance of being dimeric. *FEBS J.* **272**, 16–27 https://doi.org/10.1111/j.1432-1033.2004.04407.x

3   Goodsell, D.S. and Olson, A.J. (2000) Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 https://doi.org/10.1146/annurev.biophys.29.1.105

4   Larsen, T.A., Olson, A.J. and Goodsell, D.S. (1998) Morphology of protein–protein interfaces. *Struct. Fold. Des.* **6**, 421–427 https://doi.org/10.1016/S0969-2126(98)00044-6

5   Oohora, K. and Hayashi, T. (2014) Hemoprotein-based supramolecular assembling systems. *Curr. Opin. Chem. Biol.* **19**, 154–161 https://doi.org/10.1016/j.cbpa.2014.02.014

6   Adams, M.J., Blundell, T.L., Dodson, E.J., Dodson, G.G., Vijayan, M., Baker, E.N. et al. (1969) Structure of rhombohedral 2 zinc insulin crystals. *Nature* **224**, 491–495 https://doi.org/10.1038/224491a0

7   Leibly, D.J., Arbing, M.A., Pashkov, I., DeVore, N., Waldo, G.S., Terwilliger, T.C. et al. (2015) A suite of engineered GFP molecules for oligomeric scaffolding. *Structure* **23**, 1754–1768 https://doi.org/10.1016/j.str.2015.07.008

8   Lewis, M.S. and Youle, R.J. (1986) Ricin subunit association – thermodynamics and the role of the disulfide bond in toxicity. *J. Biol. Chem.* **261**, 1571–1577 PMID: 3745156

9   Ali, M.H. and Imperiali, B. (2005) Protein oligomerization: how and why. *Bioorg. Med. Chem.* **13**, 5013–5020 https://doi.org/10.1016/j.bmc.2005.05.037

10   Ahnert, S.E., Marsh, J.A., Hernandez, H., Robinson, C.V. and Teichmann, S.A. (2015) Principles of assembly reveal a periodic table of protein complexes. *Science* **350**, aaa2245 https://doi.org/10.1126/science.aaa2245

11   Klotz, I.M. (1967) Protein subunits – a table. *Science* **155**, 697–698 https://doi.org/10.1126/science.155.3763.697

12   Monod, J., Wyman, J. and Changeux, J.P. (1965) On Nature of allosteric transitions – a plausable model. *J. Mol. Biol.* **12**, 88–118 https://doi.org/10.1016/S0022-2836(65)80285-6

13   Labeit, S. and Kolmerer, B. (1995) Titins – giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270**, 293–296 https://doi.org/10.1126/science.270.5234.293

14   Pauling, L. (1953) Protein interactions. aggregation of globular proteins. *Discuss. Faraday Soc.* **13**, 170–176 https://doi.org/10.1039/df9531300170

15   Crick, F.H.C. and Watson, J.D. (1956) Structure of small viruses. *Nature* **177**, 473–475 https://doi.org/10.1038/177473a0

16   Ljubetic, A., Gradisar, H. and Jerala, R. (2017) Advances in design of protein folds and assemblies. *Curr. Opin. Chem. Biol.* **40**, 65–71 https://doi.org/10.1016/j.cbpa.2017.06.020

17   Norn, C.H. and Andre, I. (2016) Computational design of protein self-assembly. *Curr. Opin. Struct. Biol.* **39**, 39–45 https://doi.org/10.1016/j.sbi.2016.04.002

18   Kobayashi, N. and Arai, R. (2017) Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Curr. Opin. Biotechnol.* **46**, 57–65 https://doi.org/10.1016/j.copbio.2017.01.001

19   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**, 2217–2221 https://doi.org/10.1073/pnas.041614998

20   Kobayashi, N., Yanase, K., Sato, T., Unzai, S., Hecht, M.H. and Arai, R. (2015) Self-assembling nano-architectures created from a protein nano-building block using an intermolecularly folded dimeric de novo protein. *J. Am. Chem. Soc.* **137**, 11285–11293 https://doi.org/10.1021/jacs.5b03593

21   Kobayashi, N., Inano, K., Sasahara, K., Sato, T., Miyazawa, K., Fukuma, T. et al. (2018) Self-assembling supramolecular nanostructures constructed from de novo extender protein nanobuilding blocks. *ACS Synth. Biol.* **7**, 1381–1394 https://doi.org/10.1021/acssynbio.8b00007

22   Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T. et al. (2012) Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl Acad. Sci. U.S.A.* **109**, E690–E697 https://doi.org/10.1073/pnas.1115485109

23   Leng, W., Pang, X., Xia, H., Li, M., Chen, L., Tang, Q. et al. (2013) Novel split-luciferase-based genetically encoded biosensors for noninvasive visualization of rho GTPases. *PLos One* **8**, e62230 https://doi.org/10.1371/journal.pone.0062230

24   Huang, P.-S., Oberdorfer, G., Xu, C., Pei, X.Y., Nannenga, B.L., Rogers, J.M. et al. (2014) High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 https://doi.org/10.1126/science.1257481

25   Mou, Y., Huang, P.-S., Hsu, F.-C., Huang, S.J. and Mayo, S.L. (2015) Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl Acad. Sci. U.S.A.* **112**, 10714–10719 https://doi.org/10.1073/pnas.1505072112

26   Thomson, A.R., Wood, C.W., Burton, A.J., Bartlett, G.J., Sessions, R.B., Brady, R.L. et al. (2014) Computational design of water-soluble alpha-helical barrels. *Science* **346**, 485–488 https://doi.org/10.1126/science.1257452

27   Song, W.J. and Tezcan, F.A. (2014) A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **346**, 1525–1528 https://doi.org/10.1126/science.1259680

28   Brodin, J.D., Ambroggio, X.I., Tang, C., Parent, K.N., Baker, T.S. and Tezcan, F.A. (2012) Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–382 https://doi.org/10.1038/nchem.1290

29 Mills, J.H., Sheffler, W., Ener, M.E., Almhjell, P.J., Oberdorfer, G., Pereira, J.H. et al. (2016) Computational design of a homotrimeric metalloprotein with a trisbipyridyl core. *Proc. Natl Acad. Sci. U.S.A.* **113**, 15012–15017 https://doi.org/10.1073/pnas.1600188113

30 Jones, D.D. and Barker, P.D. (2005) Controlling self-assembly by linking protein folding, DNA binding, and the redox chemistry of heme. *Angew. Chem. Int. Ed.* **44**, 6337–6341 https://doi.org/10.1002/anie.200463035

31 Fallas, J.A., Ueda, G., Sheffler, W., Nguyen, V., McNamara, D.E., Sankaran, B. et al. (2017) Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 https://doi.org/10.1038/nchem.2673

32 Gonen, S., DiMaio, F., Gonen, T. and Baker, D. (2015) Design of ordered two-dimensional arrays mediated by noncovalent protein–protein interfaces. *Science* **348**, 1365–1368 https://doi.org/10.1126/science.aaa9897

33 Sahasrabuddhe, A., Hsia, Y., Busch, F., Sheffler, W., King, N.P., Baker, D. et al. (2018) Confirmation of intersubunit connectivity and topology of designed protein complexes by native MS. *Proc. Natl Acad. Sci. U.S.A.* **115**, 1268–1273 https://doi.org/10.1073/pnas.1713646115

34 Chevalier, A., Silva, D.A., Rocklin, G.J., Hicks, D.R., Vergara, R., Murapa, P. et al. (2017) Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 https://doi.org/10.1038/nature23912

35 Aupic, J., Lapenta, F., Strmsek, Z. and Jerala, R. (2016) Towards designing new nano-scale protein architectures. *Synth. Biol.* **60**, 315–324 https://doi.org/10.1042/ebc20160018

36 Bale, J.B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C. et al. (2016) Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 https://doi.org/10.1126/science.aaf8818

37 Pessino, V., Citron, Y.R., Feng, S.Y. and Huang, B. (2017) Covalent protein labeling by SpyTag–SpyCatcher in fixed cells for super-resolution microscopy. *Chembiochem* **18**, 1492–1495 https://doi.org/10.1002/cbic.201700177

38 Bailey, J.B., Subramanian, R.H., Churchfield, L.A. and Tezcan, F.A. (2016) Metal-directed design of supramolecular protein assemblies. *Pept. Protein Enzyme Des.* **580**, 223–250 https://doi.org/10.1016/bs.mie.2016.05.009

39 Brodin, J.D., Smith, S.J., Carr, J.R. and Tezcan, F.A. (2015) Designed, helical protein nanotubes with variable diameters from a single building block. *J. Am. Chem. Soc.* **137**, 10468–10471 https://doi.org/10.1021/jacs.5b05755

40 Zhu, S., Riou, M., Yao, C.A., Carvalho, S., Rodriguez, P.C., Bensaude, O. et al. (2014) Genetically encoding a light switch in an ionotropic glutamate receptor reveals subunit-specific interfaces. *Proc. Natl Acad. Sci. U.S.A.* **111**, 6081–6086 https://doi.org/10.1073/pnas.1318808111

41 Yeates, T.O., Liu, Y.X. and Laniado, J. (2016) The design of symmetric protein nanomaterials comes of age in theory and practice. *Curr. Opin. Struct. Biol.* **39**, 134–143 https://doi.org/10.1016/j.sbi.2016.07.003

42 van Amsterdam, I.M.C., Ubbink, M., Einsle, O., Messerschmidt, A., Merli, A., Cavazzini, D. et al. (2002) Dramatic modulation of electron transfer in protein complexes by crosslinking. *Nat. Struct. Biol.* **9**, 48–52 https://doi.org/10.1038/nsb736

43 van Vught, R., Pieters, R.J. and Breukink, E. (2014) Site-specific functionalization of proteins and their applications to therapeutic antibodies. *Comput. Struct. Biotechnol. J.* **9**, e201402001 https://doi.org/10.5936/csbj.201402001

44 Hyun Bae, J., Rubini, M., Jung, G., Wiegand, G., Seifert, M.H.J., Azim, M.K. et al. (2003) Expansion of the genetic code enables design of a novel "gold'' class of green fluorescent proteins. *J. Mol. Biol.* **328**, 1071–1081 https://doi.org/10.1016/S0022-2836(03)00364-4

45 Baskin, J.M., Prescher, J.A., Laughlin, S.T., Agard, N.J., Chang, P.V., Miller, I.A. et al. (2007) Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl Acad. Sci. U.S.A.* **104**, 16793–16797 https://doi.org/10.1073/pnas.0707090104

46 Liu, C.C. and Schultz, P.G. (2010) Adding new chemistries to the genetic code. In *Annual Review of Biochemistry*, vol. *79*, pp. 413–444

47 Reddington, S., Watson, P., Rizkallah, P., Tippmann, E. and Jones, D.D. (2013) Genetically encoding phenyl azide chemistry: new uses and ideas for classical biochemistry. *Biochem. Soc. Trans.* **41**, 1177–1182 https://doi.org/10.1042/BST20130094

48 Dommerholt, J., Rutjes, F. and van Delft, F.L. (2016) Strain-promoted 1,3-dipolar cycloaddition of cycloalkynes and organic azides. *Top. Curr. Chem.* **374**, 16 https://doi.org/10.1007/s41061-016-0016-4

49 Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z. and Schultz, P.G. (2003) An expanded eukaryotic genetic code. *Science* **301**, 964–967 https://doi.org/10.1126/science.1084772

50 Hartley, A.M., Zaki, A.J., McGarrity, A.R., Robert-Ansart, C., Moskalenko, A.V., Jones, G.F. et al. (2015) Functional modulation and directed assembly of an enzyme through designed non-natural post-translation modification. *Chem. Sci.* **6**, 3712–3717 https://doi.org/10.1039/C4SC03900A

51 Hartley, A.M., Worthy, H.L., Reddington, S.C., Rizkallah, P.J. and Jones, D.D. (2016) Molecular basis for functional switching of GFP by two disparate non-native post-translational modifications of a phenyl azide reaction handle. *Chem. Sci.* **7**, 6484–6491 https://doi.org/10.1039/C6SC00944A

52 Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S. et al. (2010) Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* **38**, 8188–8195 https://doi.org/10.1093/nar/gkq707

53 Noren, C.J., Anthonycahill, S.J., Griffith, M.C. and Schultz, P.G. (1989) A general method for site-specific incorporation of unnatural amino-acids into proteins. *Science* **244**, 182–188 https://doi.org/10.1126/science.2649980

54 Reddington, S.C., Rizkallah, P.J., Watson, P.D., Pearson, R., Tippmann, E.M. and Jones, D.D. (2013) Different photochemical events of a genetically encoded phenyl azide define and modulate GFP fluorescence. *Angew. Chem. Int. Ed.* **52**, 5974–5977 https://doi.org/10.1002/anie.201301490

55 Reddington, S.C., Driezis, S., Hartley, A.M., Watson, P.D., Rizkallah, P.J. and Jones, D.D. (2015) Genetically encoded phenyl azide photochemistry drives positive and negative functional modulation of a red fluorescent protein. *RSC Adv.* **5**, 77734–77738 https://doi.org/10.1039/C5RA13552D

56 Kim, C.H., Axup, J.Y., Dubrovska, A., Kazane, S.A., Hutchins, B.A., Wold, E.D. et al. (2012) Synthesis of bispecific antibodies using genetically encoded unnatural amino acids. *J. Am. Chem. Soc.* **134**, 9918–9921 https://doi.org/10.1021/ja303904e

57 Eger, S., Scheffner, M., Marx, A. and Rubini, M. (2010) Synthesis of defined ubiquitin dimers. *J Am. Chem. Soc.* **132**, 16337–16339 https://doi.org/10.1021/ja1072838

58 Hatzakis, N.S., Engelkamp, H., Velonia, K., Hofkens, J., Christianen, P.C.M., Svendsen, A. et al. (2006) Synthesis and single enzyme activity of a clicked lipase-BSA hetero-dimer. *Chem. Commun.*, 2012–2014 https://doi.org/10.1039/B516551B

59 Schoffelen, S., Beekwilder, J., Debets, M.F., Bosch, D. and van Hest, J.C.M. (2013) Construction of a multifunctional enzyme complex via the strain-promoted azide-alkyne cycloaddition. *Bioconjug. Chem.* **24**, 987–996 https://doi.org/10.1021/bc400021j

60 Jang, S., Sachin, K., Lee, H.J., Kim, D.W. and Lee, H.S. (2012) Development of a simple method for protein conjugation by copper-free click reaction and its application to antibody-free western blot analysis. *Bioconjug. Chem.* **23**, 2256–2261 https://doi.org/10.1021/bc300364z

61   Zimmerman, E.S., Heibeck, T.H., Gill, A., Li, X., Murray, C.J., Madlansacay, M.R. et al. (2014) Production of site-specific antibody-drug conjugates using optimized non-natural amino acids in a cell-free expression system. *Bioconjug. Chem.* **25**, 351–361 https://doi.org/10.1021/bc400490z

62   Worthy, H.L., Auhim, H.S., Jamieson, W.D., Pope, J.R., Wall, A., Batchelor, R. et al. (2019) Positive functional synergy of structurally integrated artificial protein dimers assembled by Click chemistry. *Commun. Chem.* **2**, 99 https://doi.org/10.1038/s42004-019-0203-7

63   Brejc, K., Sixma, T.K., Kitts, P.A., Kain, S.R., Tsien, R.Y., Ormo, M. et al. (1997) Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl Acad. Sci. U.S.A.* **94**, 2306–2311 https://doi.org/10.1073/pnas.94.6.2306

64   Orm, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. and Remington, S.J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 https://doi.org/10.1126/science.273.5280.1392

65   Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 https://doi.org/10.1038/nbt1172

66   Tsien, R.Y. (1998) The green fluorescent protein. *Ann. Rev. Biochem.* **67**, 509–544 https://doi.org/10.1146/annurev.biochem.67.1.509

67   Tainaka, K., Sakaguchi, R., Hayashi, H., Nakano, S., Liew, F.F. and Morii, T. (2010) Design strategies of fluorescent biosensors based on biological macromolecular receptors. *Sensors* **10**, 1355–1376 https://doi.org/10.3390/s100201355

68   Giepmans, B.N.G., Adams, S.R., Ellisman, M.H. and Tsien, R.Y. (2006) Review – the fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–224 https://doi.org/10.1126/science.1124618

69   Wang, H.X., Nakata, E. and Hamachi, I. (2009) Recent progress in strategies for the creation of protein-based fluorescent biosensors. *Chembiochem* **10**, 2560–2577 https://doi.org/10.1002/cbic.200900249

70   Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M. et al. (1997) Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* **388**, 882–887 https://doi.org/10.1038/42264

71   Xu X, Gerard AL, Huang BC, Anderson DC, Payan DG, Luo Y (1998) Detection of programmed cell death using fluorescence energy transfer. *Nucleic Acids Res.* **26**, 2034–2035 https://doi.org/10.1093/nar/26.8.2034

72   Sapsford, K.E., Berti, L. and Medintz, I.L. (2006) Materials for fluorescence resonance energy transfer analysis: beyond traditional donor-acceptor combinations. *Angew. Chem. Int. Ed.* **45**, 4562–4589 https://doi.org/10.1002/anie.200503873

73   Leavesley, S.J. and Rich, T.C. (2016) Overcoming limitations of FRET measurements. *Cytometry A*, 325–327 https://doi.org/10.1002/cyto.a.22851

74   Ghosh, I., Hamilton, A.D. and Regan, L. (2000) Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* **122**, 5658–5659 https://doi.org/10.1021/ja994421w

75   Hu, C.D., Chinenov, Y. and Kerppola, T.K. (2002) Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **9**, 789–798 https://doi.org/10.1016/S1097-2765(02)00496-3

76   Romei, M.G. and Boxer, S.G. (2019) Split green fluorescent proteins: scope, limitations, and outlook. *Ann. Rev. Biophys.* **48**, 19–44 https://doi.org/10.1146/annurev-biophys-051013-022846

77   Nagai, T., Sawano, A., Park, E.S. and Miyawaki, A. (2001) Circularly permuted green fluorescent proteins engineered to sense Ca^{2+}. *Proc. Natl Acad. Sci. U.S.A.* **98**, 3197–3202 https://doi.org/10.1073/pnas.051636098

78   Nakai, J., Ohkura, M. and Imoto, K. (2001) A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nat. Biotechnol.* **19**, 137–141 https://doi.org/10.1038/84397

79   Arpino, J.A.J., Czapinska, H., Piasecka, A., Edwards, W.R., Barker, P., Gajda, M.J. et al. (2012) Structural basis for efficient chromophore communication and energy transfer in a constructed didomain protein scaffold. *J. Am. Chem. Soc.* **134**, 13632–13640 https://doi.org/10.1021/ja301987h

80   Hanna, D.A., Harvey, R.M., Martinez-Guzman, O., Yuan, X., Chandrasekharan, B., Raju, G. et al. (2016) Heme dynamics and trafficking factors revealed by genetically encoded fluorescent heme sensors. *Proc. Natl Acad. Sci. U.S.A.* **113**, 7539–7544 https://doi.org/10.1073/pnas.1523802113

81   Carpenter, A.C., Paulsen, I.T. and Williams, T.C. (2018) Blueprints for biosensors: design, limitations, and applications. *Genes* **9**, E375 https://doi.org/10.3390/genes9080375

82   Alford, S.C., Abdelfattah, A.S., Ding, Y.D. and Campbell, R.E. (2012) A fluorogenic red fluorescent protein heterodimer. *Chem. Biol.* **19**, 353–360 https://doi.org/10.1016/j.chembiol.2012.01.006

83   Alford, S.C., Ding, Y.D., Simmen, T. and Campbell, R.E. (2012) Dimerization-dependent green and yellow fluorescent proteins. *ACS Synth. Biol.* **1**, 569–575 https://doi.org/10.1021/sb300050j

84   Ding, Y., Li, J., Enterina, J.R., Shen, Y., Zhang, I., Tewson, P.H. et al. (2015) Ratiometric biosensors based on dimerization-dependent fluorescent protein exchange. *Nat. Methods* **12**, 195–198 https://doi.org/10.1038/nmeth.3261

85   Beesley, J.L. and Woolfson, D.N. (2019) The de novo design of alpha-helical peptides for supramolecular self-assembly. *Curr. Opin. Biotechnol.* **58**, 175–182 https://doi.org/10.1016/j.copbio.2019.03.017

86   Freeley, M., Worthy, H.L., Ahmed, R., Bowen, B., Watkins, D., Macdonald, J.E. et al. (2017) Site-specific one-to-one click coupling of single proteins to individual carbon nanotubes: a single-molecule approach. *J. Am. Chem. Soc.* **139**, 17834–17840 https://doi.org/10.1021/jacs.7b07362