Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Computational protein design — the next generation tool to expand synthetic biology applications

Pablo Gainza-Cirauqui[1,2] and Bruno Emanuel Correia[1,2]

One powerful approach to engineer synthetic biology pathways is the assembly of proteins sourced from one or more natural organisms. However, synthetic pathways often require custom functions or biophysical properties not displayed by natural proteins, limitations that could be overcome through modern protein engineering techniques. Structure-based computational protein design is a powerful tool to engineer new functional capabilities in proteins, and it is beginning to have a profound impact in synthetic biology. Here, we review efforts to increase the capabilities of synthetic biology using computational protein design. We focus primarily on computationally designed proteins not only validated *in vitro*, but also shown to modulate different activities in living cells. Efforts made to validate computational designs in cells can illustrate both the challenges and opportunities in the intersection of protein design and synthetic biology. We also highlight protein design approaches, which although not validated as conveyors of new cellular function *in situ*, may have rapid and innovative applications in synthetic biology. We foresee that in the near-future, computational protein design will vastly expand the functional capabilities of synthetic cells.

Addresses
[1] Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland
[2] Swiss Institute of Bioinformatics (SIB), Lausanne CH-1015, Switzerland

Corresponding author: Correia, Bruno Emanuel (bruno.correia@epfl.ch)

Current Opinion in Biotechnology 2018, **52**:145–152

This review comes from a themed issue on **Tissue, cell and pathway engineering**

Edited by **David Schaffer** and **Stanislav Y Shvartsman**

For a complete overview see the Issue and the Editorial

Available online 2nd May 2018

https://doi.org/10.1016/j.copbio.2018.04.001

Introduction

Synthetic biologists manipulate and engineer cellular pathways to forge new cellular functions. Important achievements in the field to date include novel biosynthetic pathways [1], cellular sensing with complex logic [2], and even creation of cells with therapeutic applications [3]. A way to engineer these pathways is by introducing exogenous protein *parts* [4,5] into an existing cellular pathway, knocking-out existing proteins in these pathways if necessary. Till now, most parts used in synthetic biology are sourced from natural organisms. Although *Nature* effectively constitutes a large catalog of protein-based parts, the catalog is intrinsically limited, and naturally sourced proteins with desired properties and activities may not be always readily available.

Protein engineering approaches hold the potential to create novel functional proteins, enabling the sourcing of custom, designer parts to build biological pathways. One fast growing protein engineering method is structure-based computational protein design (referred to as simply *protein design*), where structures and atomistic computational simulations guide the design of novel protein sequences, structures, and functions. Protein design has already delivered a number of hallmark achievements which illustrate the potential of this approach to design functional proteins with potential applications in synthetic biology, such as computationally designed enzymes [6–8], protein-based binders [9], vaccine-like immunogens [10], novel membrane transporters [11], and large macromolecular assemblies (*e.g.* [12]).

The use of protein design in synthetic biology, however, is only beginning. Although many computational protein designs are inspired by potential future applications in synthetic biology, the transition from *in vitro* assays to a synthetic *in situ* pathway may not always be straightforward. Studying efforts that have attempted to validate computational designs in cells can illustrate both the challenges and opportunities in the intersection of protein design and synthetic biology. Thus, in this review we cover primarily computationally designed proteins that have not only been validated *in vitro*, but have also been shown to control and manipulate different activities in living cells. These papers cover a broad range of synthetic biology applications, such as biofuel production, biosensors for toxic products, or chemical sensors for basic biology applications, among others. We have attempted to classify them into four broad categories, with corresponding subsections: enzyme engineering, protein specificity engineering, cellular pathway control, and high-order protein assemblies. Papers that meet our protein design and *in vivo* validation criterion are highlighted in Table 1, marked of special/outstanding interest, and the protein design methods used are described in the references' summary at the end of this article. We also highlight protein design efforts which, though not validated as conveyors of new cellular function *in situ*, were

Table 1

Main papers discussed in this review

Reference	Category	Design methodology	Test cellular system
[13•] Liu *et al.*	Enzyme engineering	Redesign of existing enzyme	Mammalian cells
[14•] Mak *et al.*	Enzyme engineering	Redesign of existing enzyme	*E coli*
[20] Kapp *et al.*	Protein specificity engineering	Redesign of existing PPI	Mammalian cells
[21•] De los Santos *et al.*	Protein specificity engineering	Redesign of a ligand binding site	*E coli*
[23•] Mandell *et al.*	Protein specificity engineering	Redesign of the core of essential enzymes	*E coli*
[31•] Dagliyan *et al.*	Cellular pathway control	Linker design/MD simulations	Mammalian cells/Zebrafish tissue
[32•] Dagliyan *et al.*	Cellular pathway control	Linker design	Mammalian cells
[33•] Tinberg *et al.* [34] Feng *et al.*	Cellular pathway control	Scaffold identification, redesign to completely new activity	Yeast, mammalian cells, *in vivo* plant system
[35•] Bick *et al.*	Cellular pathway control	Scaffold identification, redesign to completely new activity	Plant cells, *in vivo* plant cells
[36•] Rose *et al.*	Cellular pathway control	Linker design with flexible backbone	Multiple types of mammalian cells
Hsia *et al.* [38] [39•] Votteler *et al.*	High-order protein assembly	*De novo* design, starting from known trimers	Mammalian cells

inspired by and may have emerging applications in synthetic biology (Figure 1).

Engineering enzymes

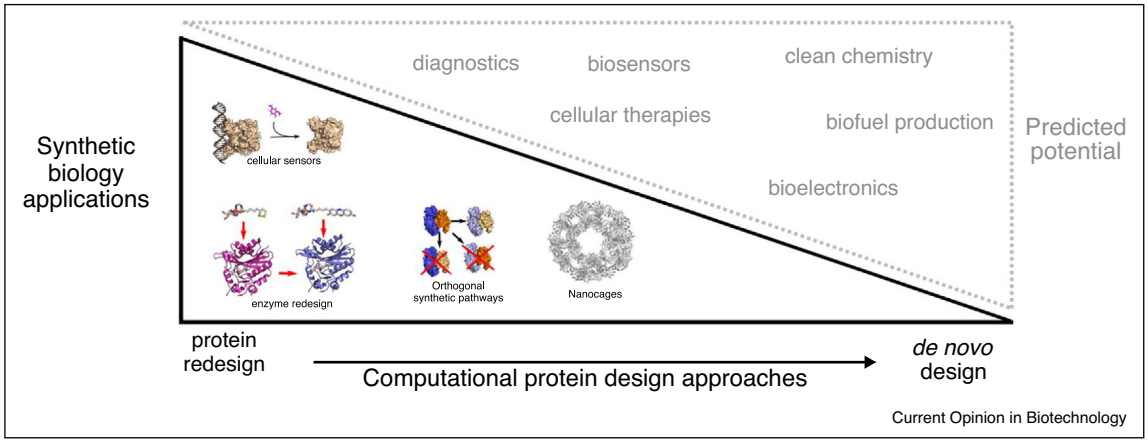
One of the most promising applications of synthetic biology is the development of new biocatalysts, such as cells that produce biofuel, manufacture expensive chemicals, or biodegrade toxic waste. Protein design can be a powerful tool to engineer enzymes for reactions that may not occur in nature or where the natural counterparts may not have the desired characteristic. Three landmark papers have shown that the current computational methods are powerful enough for *de novo* design of enzymes [6–8], though they have not yet been tested for cellular applications.

Instead, most efforts to design enzymes *in situ* have redesigned existing protein–ligand interactions for new substrate specificities, leveraging the existing catalytic machinery but changing the substrate recognized by the enzyme. Liu *et al.* [13•] repurposed a lipoic acid ligase (lpal) into a resorufin (a red fluorophore) ligase and tested its activity in living cells. The ultimate goal was to develop a method that would allow *in vivo* labelling of any protein containing a specific 13-amino acid tag. The authors computationally redesigned lpal, increasing the volume of the enzyme binding pocket to be able to recognize the resorufin substrates. The designed enzymes were validated *in vitro*, and then used to label tagged proteins *in vivo*, proving that the computationally designed enzyme could indeed perform the desired activity and enabling a number of microscopy approaches for protein visualization in living cells.

Mak *et al.* [14•] computationally redesigned the enzyme KIVD for specificity towards the substrate 2-ketooctanoate. The ultimate goal was to modify a synthetic alcohol-producing pathway to produce longer-chained alcohols, such as 1-octanol. Longer-chained alcohols are desirable because have a higher energy density, which may be important for biofuel production. The authors used RosettaDesign [15] to compute hydrophobic mutations in ten residues of the KIVD active site, leading to a library of 400 mutants. Single mutations that increased the specificity for 2-ketooctanoate were combined, resulting in a triple-point mutant with a 600-fold specificity switch. The designed enzyme was tested in *Escherichia coli* cells and showed an increase in the production of 1-heptanol and 1-hexanol. Although the synthetic cells increased the production of longer chained alcohols, they showed an overall lower rate of alcohol production compared with the native pathway. This decrease in alcohol production was attributed to the toxicity of longer-chained alcohols.

Other efforts in enzyme redesign are worth noting even if they were not directly tested *in situ* for a synthetic biology application. Siegel *et al.* [16] tackled the challenge of building enzyme products consisting of multiple carbons, from a one-carbon substrate. They computationally designed a novel enzyme, formalase, that transforms a molecule with a single carbon into one containing three. The enzyme was placed *in vitro* into an enzyme pathway where formate is converted into dihydroxyacetone phosphate. The authors also showed that the pathway could potentially function in cells by expressing it in *E. coli* and then testing cell lysates for activity. Chen et al. [17] used an ensemble-based scoring method in the OSPREY protein design program [18] to switch the specificity of an enzyme in the pathway of a nonribosomal synthetase that produces a natural antibiotic. After computationally designing distant stabilizing mutations the top enzyme showed ~20% of the wildtype enzyme's activity. Reeve et al. [19] also used OSPREY to mutate an enzyme towards resistance to an antibiotic, while maintaining catalytic activity. The computationally predicted mutants appeared in resistance selection experiments [19] with a

Figure 1



Landscape of computational design approaches and potential applications in synthetic biology. Some examples include enzyme redesign, enveloped protein nanocages, orthogonal synthetic pathways and cellular sensors for small molecule ligands. We predict that as computational protein design methods advance towards more reliable *de novo* methods, this will open up new possibilities in all areas of synthetic biology.

compensating mutation and minimal fitness loss, showing that computational methods can help in predict enzyme mutations that are fit *in situ*.

Designing a new catalytic pathway and placing it in a cellular context has two main challenges. The first challenge is designing a protein with the catalytic rate required by a specific application. This is particularly difficult in *de novo* enzyme design, where the catalytic machinery must be placed in a new scaffold protein, and must be stable in the optimal transition states of the enzyme. As the work of Mak *et al.* shows [14•], placing a designed enzyme in an endogenous pathway may result in cellular toxicity associated with the product of the reaction. Thus, the second challenge is actually placing the designed enzyme(s) in living cells, ensuring a correct interplay with the host cell environment and other enzymes in the pathway.

Engineering protein specificity

Redesigning naturally occurring proteins provides ample opportunities to manipulate cellular pathways. Typically this involves repurposing and modulating existing protein–protein interactions (PPI) or protein–ligand interactions in terms of affinity or specificity. Kapp *et al.* [20] were the first to use computational design methods to fine-tune a protein–protein interface relevant for a signaling pathway, and tested its effect *in situ*. Their work repurposed an endogenous activator-effector pair centered on the PPI between a GTPase (Cdc42) and its activator (Intersectin), and created an orthogonal (orthoCdc42-orthoIntersectin) pair that successfully interacted with itself and avoided interactions with the native counterparts. *In vitro* characterization of the designed proteins confirmed the *in silico* specificity switch prediction: the designed orthoCdc42-orthoIntersectin pair showed a 478 nM affinity, while no detectable affinity was measured for the natural counterparts. Structural characterization confirmed the presence of the hydrogen bond, even if some non-predicted structural distortions occurred in second-shell residues. NIH 3T3 cells were transfected with combinations of the native and designed components, and it was confirmed that the orthogonally designed components were functional *in situ* [20].

De los Santos *et al.* [21•] redesigned an existing protein–ligand interaction to develop a vanillin sensor. The ultimate aim was to design a feedback loop responsive to vanillin that could react according to changing conditions. Their starting point was the qacR protein, a transcription factor that can be inhibited allosterically by multiple ligands. The design objective was to alter the specificity of the effector and essentially make a vanillin-responsive qacR. Twenty-seven different computationally designed proteins were tested *in vitro* in a cell-free system (TX-TL) [22] where the activity of the designed qacR was tested and showed vanillin-dependent activity. Two qacR variants were also tested *in situ* in *E. coli* cells, where one of the variants was responsive to vanillin. However, this *E. coli* variant showed a diminished doubling time, which the authors hypothesized may be related to qacR toxicity.

Mandell *et al.* [23•] used protein design to select residues in essential enzymes from bacteria that could be redesigned to contain the non-natural amino acid (nnAA) L-4,4′-biphenylalanine (bipA) as part of their amino acid sequence. This effort had the final aim of creating biocontained auxotrophic bacteria dependent on external supplementation of bipA, which could be used to prevent genetically modified organisms from escaping to nature. The design strategy was to identify essential *E. coli*

enzymes where core residues could be mutated to bipA, while mutating the surrounding residues to accommodate bipA. Six essential genes were designed to contain bipA, two of which yielded bipA dependent bacterial strains. The designed enzymes replaced the wildtype counterparts in bacterial cells, and the authors showed that the biocontainment system was resilient against both escape mutations and horizontal gene transfer.

In addition to the aforementioned work, other recent methods in the design and manipulation of specificity have shown promise for synthetic biology, even if they were not directly tested in a cellular setting. Boyken *et al.* [24] designed protein homo-oligomers using hydrogen-bond networks in PPIs, and their method holds potential for the design of protein-binding specificity in engineered cells. Mills *et al.* computationally designed a metalloprotein by integrating a non-natural metal-binding amino acid into the core of a scaffold protein [25]. In a later work, these authors designed homotrimers nucleated by the same non-natural metal-binding amino acid [26].

Controlling cellular pathways

The ability to control protein activity using external stimuli (*e.g.* light or small molecules) has many important applications in both fundamental and synthetic biology, due to the need to control the precise timing and/or localization of particular protein activities in cells. Designed proteins can act as controllers or switches by conjugating them to other proteins that carry out a specific activity.

Protein engineering techniques have been successfully used to create light-controlled protein switches [27–30], but structure-based computational methods have had minor roles. In contrast, in small molecule-controlled activities, protein design has been shown to hold a greater potential. One strategy to create small molecule-controlled proteins fuses small molecule-responsive domains to the target effector proteins. In one such example, Dagliyan *et al.* [31•] used computational tools to perform structure-guided design of a small molecule-responsive protein, uniRapR, a synthetic regulatory domain responsive to rapamycin. UnirapR was fused to several kinases and was shown *in vitro* to yield constructs with rapamycin-dependent activity. Src kinase fused to uniRapR was then tested in HeLa cells, where the presence of rapamycin triggered a polarized spreading of the cells following Src activation. After demonstrating that the uniRapR was active *in situ*, the authors sought to test it *in vivo*, showing that in the presence of rapamycin, Src-uniRapR induced drastic alterations in the morphology of epidermal cells in zebra-fish embryos.

In later work, Dagliyan *et al.* [32•] devised a structure-based strategy to engineer allosteric control into signaling proteins. The approach was tested with three different signaling proteins: kinases, guanosine triphosphatases, and guanine exchange factors. Remarkably, the methodology was used to fuse both light-sensitive (LOV2) and drug-responsive domains (uniRapR) into the signaling proteins. The experimentally tested proteins were active and responsive to the external stimuli both *in vitro* and *in vivo*, showing major effects in cell motility.

Tinberg *et al.* [33•] computationally designed protein binders of the small molecule digoxigenin. Seventeen designed proteins were experimentally characterized and two were shown to bind. The top design, DIG.10, was further optimized by library selection using yeast surface display, with the designed proteins achieving sub-nanomolar affinity. In a follow-up study, Feng *et al.* [34] used these designs as proofs-of-concept in a general strategy to design biosensors from computationally designed ligand binders. Their strategy consisted of redesigning the ligand-binding proteins (DIG.10, and a related progesterone-binding protein) for instability in the apo state, and simultaneously, for stability in the ligand-bound state. These conditionally stable proteins should be targeted for degradation in cellular systems in the apo state. Feng *et al.* conjugated the protein sequence to transcription factors and fluorescent proteins, and validated this method in yeast cells, mammalian cells and a plant-based sensor eukaryotic system. A similar study was later performed by Bick *et al.* [35•] where the authors developed a new method to bind non-polar small molecules based on shape complementarity to scaffold proteins. Their method was used to design a plant sensor for the highly potent opioid fentanyl [35•].

Rose *et al.* [36•] used the BCL-XL/BIMBH3 protein–protein interaction and its existing inhibitors to create an intramolecularly regulated protein sensor. Specifically, they created a chemically induced activator of the RAS protein (CIAR). The core of this small molecule-regulated sensor includes the domain architecture adopted for its construction and the ability to inhibit the effector domain by steric occlusion of its active site. The BimBH3 peptide motif was fused onto the N-terminus of the effector domain and the BCL-XL motif onto the C-terminus. In a first validation *in vitro*, the CIARs showed a dose-dependent response to the small molecule. In a next stage, a construct with optimized linker lengths was tested in cells, and the authors then focused on using these molecules to interrogate key aspects of the RAS signaling pathway.

To enable the control of a wider range of protein activities using more diverse small molecules, it is critical to accurately design novel protein-small molecule binding pairs. To this end, several important contributions have been made during the last few years in the design of ligand-binding proteins (*e.g.* [33•,34]). Furthermore, as protein design techniques advance, it may become possible to

design light-sensitive proteins *de novo*, enabling the external control of cells through distinct light spectra, potentially avoiding the toxicity of the blue light intensities required in some of today's systems [37]. We can envision that in the future, synthetic biologists will be able to control cellular activities with designer-made proteins that are switched ON or OFF through safe amounts of light or well-tolerated small molecule drugs.

High-order protein assemblies

High-order protein assemblies are one of the most exciting areas in protein design with many potential applications in synthetic biology. This is particularly relevant in light of the many important roles that macromolecular machines play in central biological processes. Hsia *et al.* [38] designed a 60-subunit icosahedron particle. Their computational approach used known protein trimers as building blocks, which were then assembled into an icosahedron and the interfaces designed using Rosetta. Experimental characterization demonstrated that the particles were extremely stable, and structures determined through cryo-electron microscopy showed that the computational models very closely resembled the designed molecules. In a subsequent study, Votteler *et al.* [39•] engineered the icosahedron particles into nanocages capable of inducing the formation and releasing of extracellular vesicles. The authors refer to these biomaterials as enveloped protein nanocages (EPNs). The nanocages were designed and validated *in situ* to accomplish three distinct functionalities: membrane binding, self-assembly, and recruitment of the endosomal sorting complexes to promote release from the cell.

Using proteins as building blocks for larger assemblies has recently resulted in a number of important achievements relevant to protein design and with potential applications for synthetic biology. Tezcan and co-workers [40] rationally designed a tetrameric protein complex from monomeric redox enzymes. The tetramer displayed β-lactamase activity and was shown *in situ* to confer antibiotic resistance. Thomson *et al.* [41] developed and validated *in vitro* a structure-based method to design channel-containing soluble coiled-coils. King *et al.* developed a computational method to design nanocage-like particles with homo-oligomer subunits [42] and hetero-oligomer subunits consisting of two different proteins [12]. This methodology was later extended to megadalton 120-subunit proteins [43]. Fallas *et al.* [44] developed a method to design homo-oligomers with cyclic symmetry and validated it using dimers, trimers, tetramers and pentamers.

Outlook for synergies between protein design and synthetic biology

One of the major hurdles between protein engineers and synthetic biologists lies on the timespan that it takes to design an optimal functional protein to perform a specific purpose in the cell. Technical improvements in DNA synthesis and sequencing, high throughput screening, computational power, and computational protein design methodologies will expand our capability to quickly design and characterize novel protein sequences [45,46]. These advances will in turn significantly diminish the time requirement necessary to develop new, custom-built parts for synthetic pathway construction.

On the other hand, in a synthetic biology setting the complexity of the design objective is often incompletely defined. While we can clearly define the main function that a protein should have (*e.g.* binding, catalysis, assembly, *etc.*), generally we will not be able to anticipate all the possible events that may occur when a designed protein is placed in a cellular context, such as toxicity [14•] or interaction with endogenous proteins. These factors tremendously complicate the evaluation of the design outcome; if failure occurs, it may be difficult to uncover the exact reason for a poor outcome. However, technical advances such as high-end microscopy [47] proteomics and genomics can help us to profile cellular phenotypes at high resolution and thus illuminate potential pitfalls of the designed proteins.

There are a number of problems where structure-based design is likely to make an impact in synthetic biology applications. Beyond those highlighted in this review, there are other topics worth mentioning, for instance: the capability to design improved and/or controllable genome editing tools; optogenetics and the control of different proteins with light; and the design of membrane proteins for the generation of signaling pathways, which although in its infancy, recently saw development of the first *de novo* membrane channel [11].

Finally, technical advances in protein design, specifically flexible backbone approaches and *de novo* design approaches, will open new avenues for the design of functional proteins which can be used as building blocks for synthetic biology componentry. Efforts in these research directions have started to emerge and we anticipate that in the near future, we will see many more advances using structure-based strategies with applications that will profoundly impact synthetic biology.

Conflict of interest statement

Nothing declared.

Acknowledgements

We thank Sabrina Vollers for critical reading. BE Correia is a grantee from the European Research Council (Starting grant — 716058), the Swiss National Science Foundation, Novartis Foundation for Medical-Biological Research and the Biltema Foundation. P Gainza-Cirauqui is sponsored by an EPFL-Fellows grant funded by an H2020 Marie Sklodowska-Curie action.

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. Smanski MJ, Zhou H, Claesen J, Shen B, Fischbach M, Voigt CA: **Synthetic biology to access and expand nature's chemical diversity**. *Nat Rev Microbiol* 2016, **14**:135-149.

2. Brenner M, Cho JH, Wong WW: **Synthetic biology: sensing with modular receptors**. *Nat Chem Biol* 2017, **13**:131-132.

3. Lim WA, June CH: **The principles of engineering immune cells to treat cancer**. *Cell* 2017, **168**:724-740.

4. Grünberg R, Serrano L: **Strategies for protein synthetic biology**. *Nucleic Acids Res* 2010. gkq139.

5. Khalil AS, Collins JJ: **Synthetic biology: applications come of age**. *Nat Rev Genet* 2010, **11**:367-379.

6. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D: **Kemp elimination catalysts by computational enzyme design**. *Nature* 2008, **453**:190-195.

7. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, Hilvert D, Houk KN, Stoddard BL, Baker D: ***De novo* computational design of retro-aldol enzymes**. *Science* 2008, **319**:1387-1391.

8. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St. Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D: **Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction**. *Science* 2010, **329**:309-313.

9. Schreiber G, Fleishman SJ: **Computational design of protein–protein interactions**. *Curr Opin Struct Biol* 2013, **23**:903-910.

10. Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, Rupert P, Correnti C, Kalyuzhniy O, Vittal V, Connell MJ, Stevens E, Schroeter A, Chen M, MacPherson S, Serra AM, Adachi Y, Holmes MA, Li Y, Klevit RE, Graham BS, Wyatt RT, Baker D, Strong RK, Crowe JE, Johnson PR, Schief WR: **Proof of principle for epitope-focused vaccine design**. *Nature* 2014, **507**:201-206.

11. Joh NH, Wang T, Bhate MP, Acharya R, Wu Y, Grabe M, Hong M, Grigoryan G, DeGrado WF: ***De novo* design of a transmembrane Zn2+-transporting four-helix bundle**. *Science* 2014, **346**:1520-1524.

12. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D: **Accurate design of coassembling multi-component protein nanomaterials**. *Nature* 2014, **510**:103-108.

13. • Liu DS, Nivón LG, Richter F, Goldman PJ, Deerinck TJ, Yao JZ, Richardson D, Phipps WS, Ye AZ, Ellisman MH, Drennan CL, Baker D, Ting AY: **Computational design of a red fluorophore ligase for site-specific protein labeling in living cells**. *Proc Natl Acad Sci U S A* 2014, **111**:E4551-E4559.

Liu *et al.* designed a red flourophore ligase for *in situ* labelling of proteins. Their starting point was the enzyme lipoic acid ligase (lpal), which ligates lipoic acid to a specific 13 amino-acid peptide tag. Their computational strategy was to mutate residues in (lpal) close to the lipoyl binding pocket but away from the catalytic site in order to accommodate the much larger substrate resorufin-AMP, the red fluorophore. A rotamer library for resorufin-AMP was positioned in the enzyme active site pockets by superimposing the AMP moiety. Residues within 6–8 Å of the substrate were designed using Rosetta fixed backbone simulations. Top designs were selected using a composite scoring system based on: overall score, structural packing, satisfied polar residues and the ligand-protein interaction energy.

14. • Mak WS, Tran S, Marcheschi R, Bertolani S, Thompson J, Baker D, Liao JC, Siegel JB: **Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway**. *Nat Commun* 2015, **6**.

Mak *et al.* used two methods to discover a ketoacid decarboxylase specific for longer-chain ketoacids, yet one is primarily a genomic mining method, outside the focus of this review. Their structure-based protein design strategy consisted of mutating the ketoacid decarboxylase in the wildtype pathway (KIVD) for specificity towards longer alcohols. Ten residues in the active site were allowed to mutate to eleven hydrophobic amino acids. At each position a mutation was systematically introduced to a different hydrophobic while residues at the remaining positions were allowed to be mutated to other amino acids. The top 50 sequences were selected and were used as a base to build a library of 400 sequences for experimental characterization.

15. 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.

16. Siegel JB, Smith AL, Poust S, Wargacki AJ, Bar-Even A, Louw C, Shen BW, Eiben CB, Tran HM, Noor E, Gallaher JL, Bale J, Yoshikuni Y, Gelb MH, Keasling JD, Stoddard BL, Lidstrom ME, Baker D: **Computational protein design enables a novel one-carbon assimilation pathway**. *Proc Natl Acad Sci U S A* 2015, **112**:3704-3709.

17. Chen C-Y, Georgiev I, Anderson AC, Donald BR: **Computational structure-based redesign of enzyme activity**. *Proc Natl Acad Sci U S A* 2009, **106**:3764-3769.

18. Gainza P, Roberts KE, Georgiev I, Lilien RH, Keedy DA, Chen C-Y, Reza F, Anderson AC, Richardson DC, Richardson JS, Donald BR: **OSPREY: protein design with ensembles, flexibility, and provable algorithms**. *Methods Enzymol* 2013, **523**:87-107.

19. Reeve SM, Gainza P, Frey KM, Georgiev I, Donald BR, Anderson AC: **Protein design algorithms predict viable resistance to an experimental antifolate**. *Proc Natl Acad Sci U S A* 2015, **112**:749-754.

20. Kapp GT, Liu S, Stein A, Wong DT, Reményi A, Yeh BJ, Fraser JS, Taunton J, Lim WA, Kortemme T: **Control of protein signaling using a computationally designed GTPase/GEF orthogonal pair**. *Proc Natl Acad Sci U S A* 2012, **109**:5277-5282.

Kapp *et al.* used a flexible backbone approach (RosettaBackrub) to design a two-sided interface. Compensatory mutations in one binding partner, the intersectin interface, were predicted such that they could recover binding to its designed partner, an orthogonal cdc42. The enhanced backbone conformational sampling enabled design of an interfacial hydrogen bond (cdc42 F56R, intersectin S1373E), which was predicted to reach a two-state objective by stabilizing the designed cognate pair and destabilizing the non-cognate pair.

21. • de los Santos ELC, Meyerowitz JT, Mayo SL, Murray RM: **Engineering transcriptional regulator effector specificity using computational design and in vitro rapid prototyping: developing a vanillin sensor**. *ACS Synth Biol* 2016, **5**:287-295.

De los Santos *et al.* designed a vanillin sensor protein. The authors used Phoenix Match [48] to find potential placements for the vanillin ligand. Specifically, placements that would allow for pi-stacking and hydrogen bond interactions between protein and ligand were favored. To generate the new sequences, multiple strategies were employed: single state design — to optimize vanillin binding to the ligand-bound conformation; and multistate-design — to optimize vanillin binding to bound state but also to ensure that the mutations were compatible with the unbound state.

22. Shin J, Noireaux V: **An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells**. *ACS Synth Biol* 2012, **1**:29-41.

23. • Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, Gregg CJ, Stoddard BL, Church GM: **Biocontainment of genetically modified organisms by synthetic protein design**. *Nature* 2015, **518**:55-60.

Mandell *et al.* redesigned essential enzymes in bacteria to contain non-natural amino acids (nnAA) as part of their sequence. Rosetta fixed backbone simulations on the available crystal structures were used to scan core residues where the nnAA could be inserted with compensatory mutations in the neighboring residues. Through using a combination of energetic scores, the ultimate goal was to select sites in proteins where the nnAA was energetically favorable and the naturally occurring amino acid would become unfavorable due to all mutations performed to accommodate the nnAA.

24. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH, Sankaran B, Seelig G, Zwart PH, Baker D: ***De novo* design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity**. *Science* 2016, **352**:680-687.

25. Mills JH, Khare SD, Bolduc JM, Forouhar F, Mulligan VK, Lew S, Seetharaman J, Tong L, Stoddard BL, Baker D: **Computational design of an unnatural amino acid dependent metalloprotein**

with atomic level accuracy. *J Am Chem Soc* 2013, **135**:13393-13399.

26. Mills JH, Sheffler W, Ener ME, Almhjell PJ, Oberdorfer G, Pereira JH, Parmeggiani F, Sankaran B, Zwart PH, Baker D: **Computational design of a homotrimeric metalloprotein with a trisbipyridyl core**. *Proc Natl Acad Sci U S A* 2016, **113**:15012-15017.

27. Richter F, Fonfara I, Bouazza B, Schumacher CH, Bratovič M, Charpentier E, Möglich A: **Engineering of temperature- and light-switchable Cas9 variants**. *Nucleic Acids Res* 2016, **44**:10003-10014.

28. Wang H, Vilela M, Winkler A, Tarnawski M, Schlichting I, Yumerefendi H, Kuhlman B, Liu R, Danuser G, Hahn KM: **LOVTRAP: an optogenetic system for photoinduced protein dissociation**. *Nat Methods* 2016, **13**:755-758.

29. Yumerefendi H, Lerner AM, Zimmerman SP, Hahn K, Bear JE, Strahl BD, Kuhlman B: **Light-induced nuclear export reveals rapid dynamics of epigenetic modifications**. *Nat Chem Biol* 2016, **12**:399-401.

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

31. • Dagliyan O, Shirvanyants D, Karginov AV, Ding F, Fee L, Chandrasekaran SN, Freisinger CM, Smolen GA, Huttenlocher A, Hahn KM, Dokholyan NV: **Rational design of a ligand-controlled protein conformational switch**. *Proc Natl Acad Sci U S A* 2013, **110**:6800-6804.

Dagliyan *et al.* designed UniRapR, a small molecule-controlled protein conformational switch. UniRapR was designed through a succession of steps, which included circular permutations and linker design between the two domains that constitute UniRapR (12-kDa FK506-binding protein (FKBP12) and FKBP12-rapamycin binding protein (FRB)). After modeling the UniRapR protein, molecular dynamics simulations were used to assess the motion dynamics between the subdomains. One structural feature that arose from the MD trajectories is that a distinct change in the distance between the N and C termini of FKBP12 and FRB was modulated by rapamycin. UnirapR was fused to Src kinase domain and MD trajectories were performed, showing that in the absence of rapamycin, the ATP-binding site had higher conformational dynamics, consistent with its decreased activity observed *in vitro* and *in situ*. This finding strongly suggested that the UniRapR could be fused to the ATP-binding site of kinase catalytic domains and provide rapamycin-dependent allosteric control.

32. • Dagliyan O, Tarnawski M, Chu P-H, Shirvanyants D, Schlichting I, Dokholyan NV, Hahn KM: **Engineering extrinsic disorder to control protein activity in living cells**. *Science* 2016, **354**:1441-1444.

Dagliyan *et al.* proposed a method based on structure and sequence analysis to find allosteric sites that are amenable to a domain insertion and allosterically coupled to the active site. To address both requirements, a multistage method was developed to select the appropriate insertion sites: first, detection of solvent-exposed loops permissive to domain fusions; second, selection of loops with low sequence conservation within the protein family, which are likely less critical for function; third, coupling analysis to identify loops that are coupled to the active sites of proteins according to their three-dimensional contacts; and fourth, selection of tight loops to favor the 'switchability' of the fused constructs. Following these stages, domains were inserted using either no linkers or frequently used linkers (*e.g.* G, GSG or GPG) and the constructs were tested experimentally.

33. • Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D: **Computational design of ligand-binding proteins with high affinity and selectivity**. *Nature* 2013, **501**:212-216.

Tinberg et al. [33•] developed a computational method to design protein binders to small molecule ligands. Their method is based on disembodied side chains that optimize hydrogen bonds between ligand and receptor. A matching program in Rosetta was then used to find scaffolds that can be designed to display the disembodied side chains in an optimized orientation. Any designs that had poor shape complementarity compared to native ligand binders or that had poor organization of the active site in terms of conformational entropy were discarded.

34. Feng J, Jester BW, Tinberg CE, Mandell DJ, Antunes MS, Chari R, Morey KJ, Rios X, Medford JI, Church GM, Fields S, Baker D: **A general strategy to construct small molecule biosensors in eukaryotes**. *eLife* 2015, **4**.

35. • Bick MJ, Greisen PJ, Morey KJ, Antunes MS, La D, Sankaran B, Reymond L, Johnsson K, Medford JI, Baker D: **Computational design of environmental sensors for the potent opioid fentanyl**. *eLife* 2017, **6**:e28909.

Bick *et al.* proposed a method to design protein binders of hydrophobic ligands. Their method consisted of generating ligand conformers, including solvated ligand conformers, and then using a docking program to find scaffolds with shape complementarity with the ligand. Top docking conformations for each scaffold were selected and designed for shape complementarity as well as polar interaction with the ligand.

36. • Rose JC, Huang P-S, Camp ND, Ye J, Leidal AM, Goreshnik I, Trevillian BM, Dickinson MS, Cunningham-Bryant D, Debnath J, Baker D, Wolf-Yadlin A, Maly DJ: **A computationally engineered RAS rheostat reveals RAS-ERK signaling dynamics**. *Nat Chem Biol* 2017, **13**:119-126.

Rose *et al.* designed a protein-based switch that, upon administration of a small molecule, allows an enzyme to activate the RAS protein. Computational design was used to optimize the inter-domain linker lengths that favor a spatial location of the BCL-XL/PPI complex to occlude the binding site of the effector domain. To link the three domains, the authors used flexible backbone modeling for loop building with additional sampling of the inter-domain rigid body orientation. The linker sequence was also restricted to Glycine-Serine repeats.

37. Hallett RA, Zimmerman SP, Yumerefendi H, Bear JE, Kuhlman B: **Correlating in vitro and in vivo activities of light-inducible dimers: a cellular optogenetics guide**. *ACS Synth Biol* 2016, **5**:53-64.

38. Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang P-S, Ravichandran R, Yi S, Davis TN, Gonen T, King NP, Baker D: **Design of a hyperstable 60-subunit protein icosahedron**. *Nature* 2016, **535**:136-139.

39. • Votteler J, Ogohara C, Yi S, Hsia Y, Nattermann U, Belnap DM, King NP, Sundquist WI: **Designed proteins induce the formation of nanocage-containing extracellular vesicles**. *Nature* 2016, **540**:292-295.

Votteler *et al.* designed protein nanocages that direct their own exit from human cells. The I3-01 dodecahedron particle was designed using a computational method which performs enumerative sampling/docking in relation to the symmetry axes that are used to define the supra-molecular nano-structure. For the design of I3-01, trimeric building blocks were used to assemble the dodecahedral symmetry and docked by sampling different arrangements according to rotations and translations around the three-fold symmetry axis. The generated candidate assemblies were ranked according to the density of C_β atoms at the interface of the building blocks. Symmetry RosettaDesign was used to optimize the subunit interfaces and the designs for experimental characterization were filtered based on shape complementarity, interface surface area, buried unsatisfied hydrogen bonds and binding energy.

40. Song WJ, Tezcan FA: **A designed supramolecular protein assembly with in vivo enzymatic activity**. *Science* 2014, **346**:1525-1528.

41. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN: **Computational design of water-soluble α-helical barrels**. *Science* 2014, **346**:485-488.

42. King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, Gonen T, Yeates TO, Baker D: **Computational design of self-assembling protein nanomaterials with atomic level accuracy**. *Science* 2012, **336**:1171-1174.

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

44. Fallas JA, Ueda G, Sheffler W, Nguyen V, McNamara DE, Sankaran B, Pereira JH, Parmeggiani F, Brunette TJ, Cascio D, Yeates TR, Zwart P, Baker D: **Computational design of self-assembling cyclic protein homo-oligomers**. *Nat Chem* 2017, **9**:353-360.

45. Chevalier A, Silva D-A, Rocklin GJ, Hicks DR, Vergara R, Murapa P, Bernard SM, Zhang L, Lam K-H, Yao G, Bahl CD, Miyashita S-I, Goreshnik I, Fuller JT, Koday MT, Jenkins CM, Colvin T, Carter L, Bohn A, Bryan CM, Fernández-Velasco DA, Stewart L, Dong M, Huang X, Jin R, Wilson IA, Fuller DH, Baker D: **Massively parallel**

***de novo* protein design for targeted therapeutics**. *Nature* 2017, **550**:74-79.

46. Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houliston S, Lemak A, Carter L, Ravichandran R, Mulligan VK, Chevalier A, Arrowsmith CH, Baker D: **Global analysis of protein folding using massively parallel design, synthesis, and testing**. *Science* 2017, **357**:168-175.

47. Lukinavičius G, Umezawa K, Olivier N, Honigmann A, Yang G, Plass T, Mueller V, Reymond L, Corrêa IR Jr, Luo Z-G, Schultz C, Lemke EA, Heppenstall P, Eggeling C, Manley S, Johnsson K: **A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins**. *Nat Chem* 2013, **5**:132-139.

48. Lassila JK, Privett HK, Allen BD, Mayo SL: **Combinatorial methods for small-molecule placement in computational enzyme design**. *Proc Natl Acad Sci U S A* 2006, **103**:16710-16715.