Biochemical Society Transactions (2019) **47** 701–711
https://doi.org/10.1042/BST20180614

Review Article

Protein engineering: the potential of remote mutations

Matthew Wilding[1,2], Nansook Hong[1], Matthew Spence[1], Ashley M. Buckle[3] and Colin J. Jackson[1]

[1]Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory, Australia; [2]CSIRO Synthetic Biology Future Science Platform, Canberra, Australia; [3]Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

Correspondence: Matthew Wilding (matthew.wilding@anu.edu.au) or Colin Jackson (colin.jackson@anu.edu.au)

Engineered proteins, especially enzymes, are now commonly used in many industries owing to their catalytic power, specific binding of ligands, and properties as materials and food additives. As the number of potential uses for engineered proteins has increased, the interest in engineering or designing proteins to have greater stability, activity and specificity has increased in turn. With any rational engineering or design pursuit, the success of these endeavours relies on our fundamental understanding of the systems themselves; in the case of proteins, their structure–dynamics–function relationships. Proteins are most commonly rationally engineered by targeting the residues that we understand to be functionally important, such as enzyme active sites or ligand-binding sites. This means that the majority of the protein, i.e. regions remote from the active- or ligand-binding site, is often ignored. However, there is a growing body of literature that reports on, and rationalises, the successful engineering of proteins at remote sites. This minireview will discuss the current state of the art in protein engineering, with a particular focus on engineering regions that are remote from active- or ligand-binding sites. As the use of protein technologies expands, exploiting the potential improvements made possible through modifying remote regions will become vital if we are to realise the full potential of protein engineering and design.

Introduction

Enzymes have been used as biocatalysts for thousands of years, but our ability to understand how proteins function, and to utilise this knowledge to engineer and design new proteins for our own applications is a relatively recent development. Through advances in enzymology, structural biology and biophysics, bioinformatics, and computational simulations, we now have a relatively sophisticated molecular understanding of how proteins fold and function. This knowledge has dramatically improved our ability to engineer, design, and evolve proteins, which has transformed biochemistry and biotechnology and facilitated advances in neighbouring fields, such as metabolic engineering and molecular evolution. At the same time, protein engineering has become one of the best methods available with which to test our hypotheses regarding protein structure, function, and dynamics and has in turn deepened our fundamental understanding of these molecules.

Despite recent advances in our understanding of proteins, it is clear that there is still much to learn when it comes to understanding the structure–dynamics–function paradigm of proteins. For even the very best characterised proteins (a tiny fraction of the natural proteome), our understanding of these relationships is often limited to well-defined areas, typically substrate-binding sites. The contribution of the amino acids comprising the rest of the protein, remote from the active/binding site, is generally poorly understood. Part of the reason for this is the increasing complexity of the problem: while an active site might consist of <10 amino acids, the second shell could include ∼50, the third shell ∼200, and so on (Figure 1). Simplistically, it could be viewed that most of the molecule has evolved to

Received: 18 January 2019
Revised: 18 January 2019
Accepted: 18 February 2019

Version of Record published:
22 March 2019

Biochemical Society Transactions (2019) **47** 701–711
https://doi.org/10.1042/BST20180614

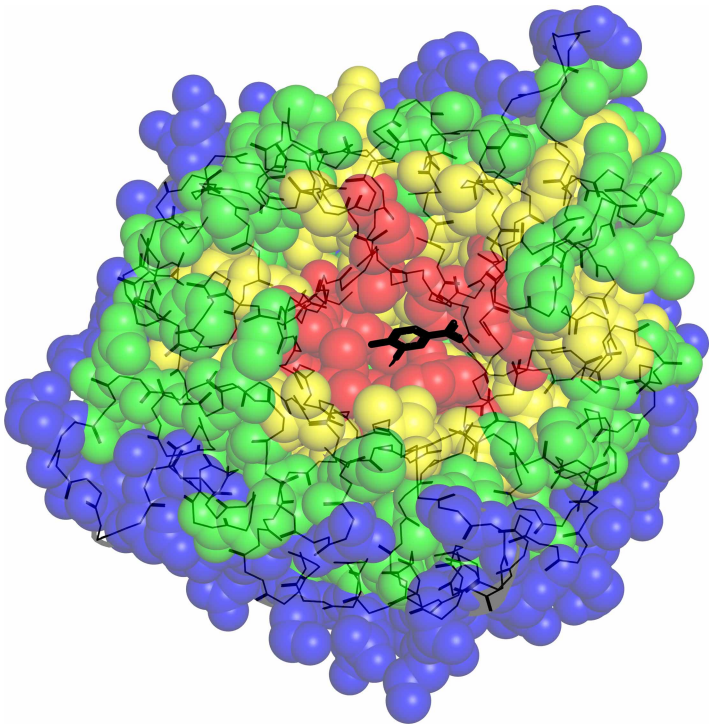

Figure 1. A protein can be simplistically viewed as an 'onion' with an inner shell (*red*; often the active or binding site) being surrounded by increasingly larger second (*yellow*), third (*green*), fourth (*blue*), and so on, shells.
In this example, the computationally designed Kemp Eliminase KE07 is illustrated with product bound at the active site (PDB ID: 5D2W).

simply stabilise the protein active site or binding site, but reports in the literature demonstrate that remote residues can contribute to changes in many properties including protein function, dynamics, expression, oligomerisation, and stability (see Table 1).

As the field of protein engineering and design continues to develop, and the demand for increasingly complex protein-based tools and catalysts grows, the importance of understanding proteins more holistically, i.e. attaining a more quantitative understanding of the role of remote amino acids, will become paramount. Random mutagenesis will remain a powerful tool for identifying remote mutations that can affect activity, even as its utility in protein engineering becomes replaced by rational/computational approaches. This minireview will discuss some of the approaches to protein engineering of remote regions that have been developed in recent years and highlight their importance in the context of a more comprehensive approach to protein design and engineering.

Directed evolution

In nature, genes and the proteins they encode are always evolving and, although often silent, a single mutation can significantly affect function. This is perhaps most obvious when the mutation results in a significant competitive advantage (e.g. antibiotic [1] or insecticide resistance [2]) or loss of function (e.g. hereditary diseases that result from deleterious mutations [3]). In either case, the role of the mutation is often relatively facile to characterise if the mutation occurs within the active site of the protein [2,4], but there are numerous examples in the literature where remote point mutations have resulted in significant functional changes [5–8]. Likewise, many engineering studies have noted that beneficial mutations can occur in remote regions of the protein [9–14]. Indeed, a recent study by Wrenbeck et al. [15] showed that beneficial single mutations can occur in many regions of a protein other than active/binding sites. These examples support the notion that diverse regions within a protein have the potential to influence function.

Table 1 Summary of investigations reporting remote mutations, the remote mutation positions within the protein and the effect(s) of those mutations

Part 1 of 2

Enzyme	Organism	Property	Remote mutation(s)	Reference
DNA Polymerase β	*Homo sapiens*	Activity, affinity	S229L, G231D	[66]
HIV protease variant PR-S17	*Homo sapiens*	Activity, affinity	A71V, L90M, I93L	[7]
Valosin containing protein p97	*Homo sapiens*	Activity, Protein–protein Interface	R95G, T262A, R155C/H/P, N387H, A232E, R191Q, L198W	[8]
Troponin C	*Homo sapiens*	Protein–protein Interface	F20L, N12D, R85H, N96S, N203S	[67]
Simvastatin synthase, LovD	*Aspergillus terreus*	Activity, stability, binding	Catalysis (N191S, N191G, L192I, L174F, A178L, S172N) stability (L361M, V370I, A383V, I35L) binding and PPI interfere (A247S, A9V, K26E, H404K, I4N, R28S) thermostability (Q241M, A261V, A261H) reduced aggregation (N43R, D96R, H404K)	[68]
Dialkylglycine decarboxylase	*Pseudomomas cepacia*	Activity, stability	S306F	[18]
New Delhi metallo-beta-lactamase 1	*Klebsiella pneumonia*	Activity, affinity	A233V, L49P, M154V, V88M, Q151R, D96A, N103K, N166T	[69]
Second carbapenem-hydrolysing metalloenzyme VIM-2	*Pseudomonas aeruginosa*	Activity, affinity, solubility, assembly	G27R, V41A, V46D, T64A, V72A, E150K, V195I, S202R, T263S, N264D	[69]
N-acyl homoserine lactonase	*Bacillus thuringiensis*	Activity, affinity	V69G, K139T, I230M, F64C, L33V	[13]
Ancestor node 1 for the methyl-parathion hydrolase AncDHCH1	Predicted ancestors	Activity, affinity, selectivity	Δ193S	[70]
Kemp eliminase KE07	*Thermotoga maritima* (scaffold)	Activity	V12L/M, K146T, F77I, F229S, I102F, H84Y, M207T	[56] [54]
Kemp eliminase KE15	*Thermotoga maritima* (scaffold)	Activity	D130K, I168M, G199A	[71]
Kemp eliminase KE59	*Sulfolobus solfataricus* (scaffold)	Activity, stability	K9E, L14R, F21V, N33K, S69A, Y75G, T94D, Y151L, N160H, L16Q, I48M, A76V, V80A, I104V, S179T, K190N, A208V, R222Y, L247Q	[72]
Diels-alderase DE20/CE20	*Loligo vulgaris* (scaffold)	Activity, affinity, selectivity	DE20 (R50H, V96I, T197R, E288D, L309S, D232V, H274L, R56S) CE20 (P48L, K53E, R56S, S55R, G57D)	[73] [55]
Retro-aldolase RA95.5-8F	*Sulfolobus solfataricus* (scaffold)	Activity	R75P, N90D, N135E, S151G, V178T, K210L, I213F, S214F, R216P, L231M	[53] [74] [52]
Photosensitiser protein conjugated to terpyridine PSP2T	*Pyrococcus abyssi* (scaffold)	Activity, quantum yield	E95C-terpyridine, 93Y97Y	[75] [76]
D-amino acid oxidase	*Homo sapiens*	Specificity, activity, solvent accessibility	Y55A, L56T	[77]
Adenylate kinase	*Escherichia coli*	Activity, affinity	Activity (A37G, A55G), Affinity (V135G, V142G)	[78]
Monoacylglycerol lipase	*Homo sapiens*	Activity	W289L, L232G	[79]
Phytochrome-based near-infrared fluorescent protein iRFP	*Rhodopseudomonas palustris*	Quantum yield	Not reported	[80,81]
Dormancy survival proteins DosS	*Mycobacterium tuberculosis*	Activity	E87A/G/D, H89A, R204A	[82]
Chromosomal zinc-regulated repressor CzrA	*Staphylococcus aureus*	DNA binding	V66A, V66A/L68V, V66A/L68A	[83]
Cytochrome P450	*Bacillus megaterium*	Activity, selectivity	F393H/A	[84]

Continued

Biochemical Society Transactions (2019) **47** 701–711
https://doi.org/10.1042/BST20180614

Table 1 Summary of investigations reporting remote mutations, the remote mutation positions within the protein and the effect(s) of those mutations Part 2 of 2

Enzyme	Organism	Property	Remote mutation(s)	Reference
DNA Polymerase	*Homo sapiens*	Activity, affinity, stability	K224A/K231A	[85]
RNA-guided DNA endonuclease enzyme Cas9	*Streptococcus pyogenes*	Activity, affinity	K974A	[86]
Cytochromes P450	*Rattus norvegicus*	Activity, affinity, product release	F240A	[87]
Interleukin 2	*Homo sapiens*	Activity, affinity	Q74H, L80F, R81D, L85V, I86V, I92F	[88,89]
Ancestor node 2 for cyclohexadienyl dehydratase	Predicted Ancestors	Activity	F25L, G99S, P102L, A155I	[28]
Transaminase	*Pseudomonas* sp.	Activity	E178D, G179R, Q142N	[34]
Phosphotriesterases	*Pseudomonas diminuta*	Activity	R22 (D233E, F306I, I274S, T172I, S269T, M138I, T199I, I272M, A80V, S111R, A204G, I130V, L271F, A49V, K77E, I140M, I313F, S137T, Q180H, T45A, E144V, M314T, I341, S102T, V176M), Rev12 (D233E, F306M, I274S, S269T, M138I, T199I, I272M, A80V, S111R, A204G, I130M, K77E, I140M, I313F, S137T, T45A, E144V, I341T, S102T, V176M, S308C, P135S, A203E, M293K, G194D, S258N, Y156H)	[14]
Ancestor node 1 of the CYP3 family	Predicted Ancestors	Stability	I93A, I160A, I196A, I207A, I225A, A252S, I360A, I398A, I400A	[30]
Superoxide dismutase	*Porphyromonas gingivalis*	Activity	G155T	[5]
HIV-1 protease	*Homo sapiens*	Activity	L76V, L90M, V32I, L33F	[6]

The most common approach to circumvent our limited understanding of protein structure–function–dynamics relationships is directed evolution. Directed evolution recreates the natural evolutionary process, mutating genes at random, but with the selection of variants under the control of the researcher/engineer [16]. This engineering approach is arguably the simplest and most holistic, as it requires little to no prior knowledge about the complex structure–function relationship within the protein being evolved and allows mutations to arise throughout the protein structure. Although this may appear to be a universal solution for the problem of engineering remote sites, directed evolution is not without drawbacks. Most notably, depending on the size of the protein, thousands of variants need to be screened, which (depending on the function of the protein being evolved) often limits such experiments to model systems or requires exhaustive work to screen the activity of the variants [17]. Directed evolution is therefore often relatively inefficient when time and resources are considered.

Despite these drawbacks, directed evolution has yielded numerous examples where remote mutations have been shown to be beneficial and, in doing so, identified regions of proteins that were previously thought to have no effect on function. Several examples of remote mutations improving thermostability [11,18] and protein expression [19] have been reported, but other studies that have focused on protein conformational landscapes have demonstrated that remote mutations can improve catalytic efficiency by virtue of conformational enrichment and active site remodelling [11,14,18,20–22]. For example, in a recent study by Otten et al. [20], the catalytic efficiency of a proline isomerase variant was rescued by second-shell mutations obtained through directed evolution. Analysis of the variants revealed that the remote mutations had restored catalytic activity through accelerated interconversion of conformational substates. Directed evolution can also provide insight into natural evolutionary processes. A series of directed evolution studies with the bacterial phosphotriesterases has shown that remote mutations can modulate the dynamics of surface loops, which in turn affects substrate diffusion and activity, that second-shell residues can control the conformational sampling of active site residues, and that variants with substantial differences in terms of their primary sequence can display very similar catalytic behaviour provided the remote mutations produce similar structures and conformational sampling [14,21,22].

Computational tools are becoming increasingly powerful when combined with directed evolution, displaying an ability to conduct multivariate analysis and predict the effects of different mutations in parallel [23,24]. A framework outlining a general procedure and some of the available tools for *in silico* screening of stabilising mutations was described by Wijma et al. [25]. They demonstrated that the directed evolution process, when initially conducted *in silico*, could be performed more efficiently than solely at the bench. They specifically excluded residues in and around the active site as part of the process then experimentally validated their predictions. Ultimately, the variants produced demonstrated significant improvements in thermostability, and superior catalytic activity at elevated temperatures, albeit at the expense of reduced efficiency at lower temperatures. Similarly, a semi-automated computational framework, CADEE (Computer-Aided Directed Evolution of Enzymes), has been demonstrated to expedite directed evolution *in silico* [26]. The study reported the mutation and analysis of the effects of 128 substitutions on the stability and function of triosephosphate isomerase.

In summary, when coupled with a suitable screening method, directed evolution is a powerful approach for protein engineering, especially because of its ability to incorporate remote beneficial mutations that we currently are unable to predict rationally. It has also been a valuable learning tool, with the rationalisation of the effects of remote mutations shedding new light on protein structure, function, and especially dynamics and the functional roles of remote regions.

Ancestral protein reconstruction and consensus design

In contrast with directed evolution, ancestral sequence reconstruction (ASR) is a retrospective evolutionary engineering strategy. ASR uses statistical models to infer hypothetical ancestral protein sequences based on the evolutionary relationships of related extant proteins. Consensus design is a similar approach, in which each residue of the new sequence represents the most conserved (consensus) amino acid at that position in a multiple sequence alignment of the protein family [27]. Both of these approaches are particularly relevant in the context of remote mutations because they focus on the entire sequence; as such, most of the sequence differences between extant proteins and variants generated by ASR or consensus design are in remote regions. As gene synthesis has become more cost effective, the study of complete ancestral proteins has become increasingly popular and studies involving these proteins have expanded beyond examination of evolutionary trajectories [28] or paleobiochemistry [29]. There are some advantages to using ASR or consensus design over directed evolution. Firstly, these approaches usually produce a relatively small number of variants rather than vast libraries of tens to millions of variants. Secondly, sequence changes are incorporated based on existing sequence data, rather than at random, and are thus less likely to significantly disrupt protein function or stability because the mutations are known to be tolerated by related proteins. In this sense, the sampling of sequence space is focused primarily on tolerable mutations, rather than directed evolution in which a large proportion of the mutations are deleterious.

ASR has been used to design protein variants with modified substrate specificity, thermostability, pH, and solvent tolerance [30–32]. For example, a recent study of ancestral p450 enzymes demonstrated that, in addition to their enhanced stability (≈30°C increase in melting temperature and >100-fold increases in incubation time) and solvent tolerance, the ancestral proteins were more active than their extant counterparts even at ambient temperatures [30]. This improvement was enhanced further when operating at elevated temperatures. Another recent study to exemplify engineering by ASR compared a highly promiscuous extant transaminase from *Pseudomonas* sp. strain AAC [31,33] with its ancestral counterparts. A comparison of the protein structures showed that despite strictly conserved active sites, the proteins exhibited differences in catalytic efficiency and substrate preference [34]. Likewise, consensus design has achieved significant success in improving stability [27,35] mainly because the likelihood of a positive contribution to stability at a given position is higher from the consensus amino acid than that from non-conserved amino acids. Consensus design has been used for at least two decades, but its popularity and robustness have benefitted from the recent explosion in genomic sequence data. For example, consensus design of a fibronectin domain using thousands more sequences than previous attempts led to an exceptional window of thermostability to tolerate deleterious mutations, providing downstream opportunities for functional design using highly stable scaffolds [36,37].

Consensus design is not only useful for increasing the stability of the native state of single domains, but has also been applied to relatively complex folds, for example serine protease inhibitors (serpins), in which complicated folding transitions are utilised for inhibitory function [38]. Serpins must exist on a knife-edge of stability to function, and it is remarkable that consensus design was successful, but suggests that this may have been achieved by remodelling of the folding landscape. This raises the exciting possibility that manipulation of the

Biochemical Society Transactions (2019) **47** 701–711
https://doi.org/10.1042/BST20180614

folding landscape in preference to stabilisation of the native state may be a more effective strategy for optimising the folding behaviour of proteins (for example the removal of aggregation-prone intermediates).

Much like directed evolution, the large numbers of mutations introduced in ASR/consensus design (often in a single step) makes analysis complex and structural or computational analysis is often essential to gain any insights. Even so, there are several examples of studies that have deepened our understanding of the effects of remote mutations. Two recent examples demonstrated the use of ASR in the functionalisation of a non-catalytic binding proteins [28,39]. Both articles demonstrated that multiple mutational pathways could yield catalytically active proteins from non-catalytic starting points, with early rounds of mutagenesis installing the catalytically active residues and subsequent rounds optimising residues progressively further from the active site (Figure 1). Over the course of evolution, remote residues refined the active site configuration, conformational sampling and substrate complementarity.

In summary, the decreased costs of gene synthesis and availability of bioinformatics tools have made ASR and consensus design increasingly popular engineering strategies. The ability to incorporate multiple mutations across the whole protein has led to new insights into protein structure–function relationships as well as protein evolution more broadly.

Rational and computational protein design

In situations where it is not practical or possible to utilise high-throughput screening to identify beneficial mutations among the thousands of neutral or deleterious mutations generated through directed evolution, rational design is the most commonly used alternative approach. Rational design involves the creation of a new protein variant incorporating specific, predicted changes. Our ability to successfully utilise rational design has increased in parallel with advances in computational power and our expanding knowledge of protein structure, function, and dynamics. In contrast with the evolutionary strategies described above, when mutations are *designed* there is an underlying rationale based on an understanding of the protein structure/function. This paradigm extends to remote mutations. Accordingly, rational design has historically focused on regions that are well understood, such as active/binding sites. However, many recent developments in the literature describing the influence of remote mutations have utilised rational design strategies.

There are a variety of computational methods that have been developed to aid protein engineering and design. Among these, the Rosetta suite provides perhaps the most extensive toolkit for predicting structural perturbations caused by mutations. Rosetta uses a scoring function for protein stability, derived from empirical data and physical models, which allows it to calculate the lowest energy protein conformations and the effects of mutations. In the context of remote mutations, Rosetta is particularly useful as it efficiently explores conformational space using a Metropolis Monte Carlo (MMC) search algorithm that samples and scores backbone dihedrals and side-chain conformations from empirical peptide and rotamer libraries [40,41]. Rosetta-based protein design applications explore sequence space for mutations that satisfy the MMC conformational search criteria. This has allowed many protocols to be developed, such as Rosetta supercharge, which guides the selection of remote surface mutations to charge protein pI values [42,43]. Recent adaptations to standard Rosetta design protocols have seen evolutionary and structural information being leveraged to restrict the sequence space that Rosetta can explore. For example, the Protein Repair One Stop Shop (PROSS) and FuncLib algorithms use phylogenetic profiles derived from homologues of the query protein (in this sense, they can be viewed as being inspired by ASR/consensus design) with the Rosetta suite to predict stabilising and activity-enhancing mutations, respectively [44,45]. Khersonsky et al. [45] used PROSS and FuncLib in conjunction to design a focussed library of 29 *Salmonella enterica* Acetyl-CoA synthetase (ACS) mutants that both improved expression in *E. coli* and increased catalytic activity by as much as 47-fold for isobutyrate ligation. Each design comprised 47 stabilising mutations that were distant from the active site and combinations of 3–6 mutations in the first- and second-shells that re-structured the ACS-binding pocket, emphasising the importance of mutations beyond immediate active site residues in enzyme design. The sequence space available to Rosetta can also be restrained according to deep-learning neural networks trained on residue-specific structural features from the protein databank (PDB) [46]. Protocols based on this principal remain in their infancy, but could provide next-generation computational protein design tools in the future. Other software packages that use empirical force fields, such as FoldX [47], can also allow non-expert users to model the effects of mutations on protein structure and stability with ease and accuracy, and bioinformatics-based approaches such as those used by SDM [48], HotSpot Wizard 3.0 [49], and Caver analyst [50] provide valuable resources for designing enzyme variants.

An extreme case of rational design is protein design. Advances in bioinformatics, quantum chemistry, structural biology, and enzymology now allow us to redesign entire regions (active/binding sites) or even entire proteins *de novo*. Most protein design experiments have aimed to design a new structure with improved function, with the key point being the design is focused on a single state. For example, a series of ground-breaking enzyme design papers have been published within over the past ∼10 years in which catalytic active sites were designed *in silico*, often with the use of quantum chemical calculations, and then transplanted into naturally occurring scaffolds [51]. This has successfully generated several new enzymes [52–55], many of which catalyse reactions not otherwise seen in nature. However, most of these initial designs could be substantially improved upon by multiple rounds of directed evolution, in which remote mutations were regularly observed to impart dramatic improvements. This was exemplified by studies of the Kemp eliminase, where analysis of the designed and evolved variants demonstrated that remote mutations affected the conformational sampling of the enzyme and eventually stabilised a new configuration of the designed active site residues [56].

Past enzyme design and directed evolution studies have highlighted the need for more than one conformational sub-state to be taken into account during protein design, and that conformational sampling is largely controlled by regions of the protein remote from the active site [14,57]. The ability to not only understand, but also control, the dynamic nature of proteins has great potential, and in recent years there have been advances in this area. For example, tools to design proteins based on multiple stable states (multistate design) have now been developed and utilised to design novel proteins that can conformationally interconvert, opening up exciting areas for exploration in protein design and synthetic biology [57,58].

Finally, full protein design, the creation of completely new proteins from first principles, is a powerful approach and tests our understanding of structure-dynamics-function relationships. At present, the complexity of a complete 'bottom-up' *de novo* design limits this approach to small proteins, but this exciting field is producing the first holistically designed biocatalysts and structural proteins [59–63]. The proteins typically comprise ∼50 amino acids, and can be synthesised chemically or biologically. Until recently, full protein design has been largely concerned with the principles underpinning structural elements, but recently those elements have been functionalised to generate small catalysts [64]. Another exciting area of application is in therapeutic proteins; computationally designed cytokine mimics have recently been shown to bind with higher affinity, reduced toxicity, and no apparent immunogenicity, demonstrating the potential to create designer protein therapeutics [65]. However, most of these designs are single-state proteins; the *de novo* design of large and complex structures that sample multiple defined conformational substates is still very challenging.

Future directions

Biotechnology and the applications of proteins have evolved beyond chemical production or fermentation. Proteins are now designed for use as sensors, materials and therapeutics and the demand for efficient and increasing complex protein engineering strategies continues to grow. It has become clear than many of the most functionally impressive proteins (highly efficient enzymes, multi-specific binding proteins, etc.) are capable of sampling multiple conformational sub-states, each with defined functions in a catalytic cycle or interaction network. Likewise, the abundance of remote mutations identified through directed evolution has made it obvious that regions remote from active/binding sites are of extreme functional importance. The future of protein engineering is therefore likely to involve an increased focus on improving activity through remote mutations. To do this, all varieties of engineering approaches will have a place, from directed evolution, to bioinformatics-based approaches like ASR, to rational design and *de novo* computational design (Figure 2). Our success in this field will also rely on our continued study of protein structure–function–dynamics to provide the theoretical basis for the increasingly powerful computational algorithms that will hopefully, eventually make designed proteins a ubiquitous part of life.

Perspectives

- Protein engineering has transformed industries and is a fundamental tool in biotechnology, metabolic engineering, and synthetic biology. As these fields continue to diversify, the need to extend the capabilities and efficiency of protein engineering and design will increase.

Biochemical Society Transactions (2019) **47** 701–711
https://doi.org/10.1042/BST20180614

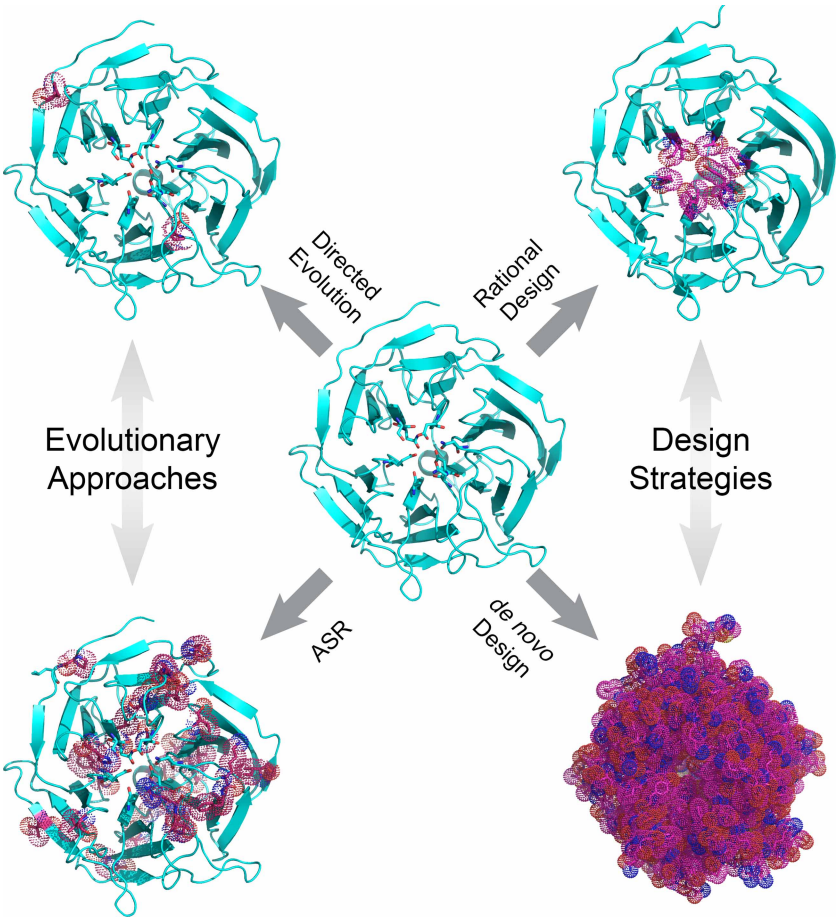

Figure 2. Approaches to protein engineering illustrated on the X-ray structure of diisopropylfluorophosphatase (DFPase; PDB ID: 1E1A).

The unmodified (cyan) structure is shown centrally with active site residues shown as sticks. Mutation sites are coloured magenta and highlighted in dot representation. Although rational design approaches have traditionally targeted residues in and around the active site (top right), directed evolution (top left), and ASR (bottom left) are more likely to identify remote mutations which effect protein properties, albeit ASR typically results in more extensive modifications. Finally, *de novo* methods (bottom right) look to design the entire protein.

- Our solid understanding of protein active/binding sites, and less-well developed understanding of the functional roles of remote regions of proteins have meant that protein engineering has predominantly focused on active/binding sites, yet there is ample evidence from directed evolution that large gains in efficiency can be made through remote mutations.
- Recent developments in our understanding of protein function, especially in the area of protein dynamics and conformational sampling, coupled with rapid advances in computational protein design now make engineering of protein remote sites a practical, albeit challenging, means of improving protein function.

Abbreviations

ACS, acetyl-CoA synthetase; ASR, ancestral sequence reconstruction; MMC, Metropolis Monte Carlo; PDB, protein databank; PROSS, protein repair one stop shop.

Competing Interests

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

References

1 Long, H., Miller, S.F., Strauss, C., Zhao, C., Cheng, L., Ye, Z. et al. (2016) Antibiotic treatment enhances the genome-wide mutation rate of target cells. *Proc. Natl Acad. Sci. U.S.A.* **113**, E2498 https://doi.org/10.1073/pnas.1601208113
2 Mabbitt, P.D., Correy, G.J., Meirelles, T., Fraser, N.J., Coote, M.L. and Jackson, C.J. (2016) Conformational disorganization within the active site of a recently evolved organophosphate hydrolase limits its catalytic efficiency. *Biochemistry* **55**, 1408–1417 https://doi.org/10.1021/acs.biochem.5b01322
3 Taipale, M. (2018) Disruption of protein function by pathogenic mutations: common and uncommon mechanisms. *Biochem. Cell Biol.* **97**, 46–57 https://doi.org/10.1139/bcb-2018-0007
4 Gable, K., Gupta, S.D., Han, G., Niranjanakumari, S., Harmon, J.M. and Dunn, T.M. (2010) A disease-causing mutation in the active site of serine palmitoyltransferase causes catalytic promiscuity. *J. Biol. Chem.* **285**, 22846–22852 https://doi.org/10.1074/jbc.M110.122259
5 Yamakura, F., Sugio, S., Hiraoka, B.Y., Ohmori, D. and Yokota, T. (2003) Pronounced conversion of the metal-specific activity of superoxide dismutase from porphyromonas gingivalis by the mutation of a single amino acid (Gly155Thr) located apart from the active site. *Biochemistry* **42**, 10790–9 https://doi.org/10.1021/bi0349625
6 Ragland, D.A., Nalivaika, E.A., Nalam, M.N.L., Prachanronarong, K.L., Cao, H., Bandaranayake, R.M. et al. (2014) Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. *J. Am. Chem. Soc.* **136**, 11956–11963 https://doi.org/10.1021/ja504096m
7 Agniswamy, J., Louis, J.M., Roche, J., Harrison, R.W. and Weber, I.T. (2016) Structural studies of a rationally selected multi-drug resistant HIV-1 protease reveal synergistic effect of distal mutations on flap dynamics. *PLoS ONE* **11**, e0168616 https://doi.org/10.1371/journal.pone.0168616
8 Schuetz, A.K. and Kay, L.E. (2016) A dynamic molecular basis for malfunction in disease mutants of p97/VCP. *eLife* **5**, e20143 https://doi.org/10.7554/eLife.20143
9 Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C. and Tawfik, D.S. (2004) The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73 https://doi.org/10.1038/ng1482
10 Lee, J. and Goodey, N.M. (2011) Catalytic contributions from remote regions of enzyme structure. *Chem. Rev.* **111**, 7595–7624 https://doi.org/10.1021/cr100042n
11 Morley, K.L. and Kazlauskas, R.J. (2005) Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **23**, 231–237 https://doi.org/10.1016/j.tibtech.2005.03.005
12 Schmidt, M., Hasenpusch, D., Kähler, M., Kirchner, U., Wiggenhorn, K., Langel, W. et al. (2006) Directed evolution of an esterase from pseudomonas fluorescens yields a mutant with excellent enantioselectivity and activity for the kinetic resolution of a chiral building block. *ChemBioChem* **7**, 805–809 https://doi.org/10.1002/cbic.200500546
13 Yang, G., Hong, N., Baier, F., Jackson, C.J. and Tokuriki, N. (2016) Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects. *Biochemistry* **55**, 4583–4593 https://doi.org/10.1021/acs.biochem.6b00561
14 Campbell, E., Kaltenbach, M., Correy, G.J., Carr, P.D., Porebski, B.T., Livingstone, E.K. et al. (2016) The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **12**, 944 https://doi.org/10.1038/nchembio.2175
15 Wrenbeck, E.E., Azouz, L.R. and Whitehead, T.A. (2017) Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* **8**, 15695 https://doi.org/10.1038/ncomms15695
16 Renata, H., Wang, Z.J. and Arnold, F.H. (2015) Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. *Angew. Chem. Int. Ed.* **54**, 3351–3367 https://doi.org/10.1002/anie.201409470
17 Zeymer, C. and Hilvert, D. (2018) Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **87**, 131–157 https://doi.org/10.1146/annurev-biochem-062917-012034
18 Taylor, J.L., Price, J.E. and Toney, M.D. (2015) Directed evolution of the substrate specificity of dialkylglycine decarboxylase. *Biochim. Biophys. Acta* **1854**, 146–155 https://doi.org/10.1016/j.bbapap.2014.12.003
19 Khersonsky, O., Röthlisberger, D., Dym, O., Albeck, S., Jackson, C.J., Baker, D. et al. (2010) Evolutionary optimization of computationally designed enzymes: kemp eliminases of the KE07 series. *J. Mol. Biol.* **396**, 1025–1042 https://doi.org/10.1016/j.jmb.2009.12.031
20 Otten, R., Liu, L., Kenner, L.R., Clarkson, M.W., Mavor, D., Tawfik, D.S. et al. (2018) Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.* **9**, 1314 https://doi.org/10.1038/s41467-018-03562-9
21 Jackson, C.J., Foo, J.L., Tokuriki, N., Afriat, L., Carr, P.D., Kim, H.K. et al. (2009) Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. *Proc. Natl Acad. Sci. U.S.A.* **106**, 21631 https://doi.org/10.1073/pnas.0907548106
22 Kaltenbach, M., Jackson, C.J., Campbell, E.C., Hollfelder, F. and Tokuriki, N. (2015) Reverse evolution leads to genotypic incompatibility despite functional and active site convergence. *eLife* **4**, e06492 https://doi.org/10.7554/eLife.06492
23 Barak, Y., Nov, Y., Ackerley, D.F. and Matin, A. (2007) Enzyme improvement in the absence of structural knowledge: a novel statistical approach. *ISME J.* **2**, 171 https://doi.org/10.1038/ismej.2007.100
24 Fox, R.J., Davis, S.C., Mundorff, E.C., Newman, L.M., Gavrilovic, V., Ma, S.K. et al. (2007) Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **25**, 338 https://doi.org/10.1038/nbt1286
25 Wijma, H.J., Floor, R.J., Jekel, P.A., Baker, D., Marrink, S.J. and Janssen, D.B. (2014) Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* **27**, 49–58 https://doi.org/10.1093/protein/gzt061
26 Amrein, B.A., Steffen-Munsberg, F., Szeler, I., Purg, M., Kulkarni, Y. and Kamerlin, S.C.L. (2017) CADEE: computer-aided directed evolution of enzymes. *IUCrJ* **4**, 50–64 https://doi.org/10.1107/S2052252516018017
27 Porebski, B.T. and Buckle, A.M. (2016) Consensus protein design. *Protein Eng. Des. Sel.* **29**, 245–251 https://doi.org/10.1093/protein/gzw015
28 Clifton, B.E., Kaczmarski, J.A., Carr, P.D., Gerth, M.L., Tokuriki, N. and Jackson, C.J. (2018) Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat. Chem. Biol.* **14**, 542–547 https://doi.org/10.1038/s41589-018-0043-2

Biochemical Society Transactions (2019) **47** 701–711
https://doi.org/10.1042/BST20180614

29 Gaucher, E.A., Thomson, J.M., Burgan, M.F. and Benner, S.A. (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425**, 285–288 https://doi.org/10.1038/nature01977

30 Gumulya, Y., Baek, J.-M., Wun, S.-J., Thomson, R.E.S., Harris, K.L., Hunter, D.J.B. et al. (2018) Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888 https://doi.org/10.1038/s41929-018-0159-5

31 Wilding, M., Peat, T.S., Kalyaanamoorthy, S., Newman, J., Scott, C. and Jermiin, L.S. (2017) Reverse engineering: transaminase biocatalyst development using ancestral sequence reconstruction. *Green Chem.* **19**, 5375–5380 https://doi.org/10.1039/C7GC02343J

32 Whitfield, J.H., Zhang, W.H., Herde, M.K., Clifton, B.E., Radziejewski, J., Janovjak, H. et al. (2015) Construction of a robust and sensitive arginine biosensor through ancestral protein reconstruction. *Protein Sci.* **24**, 1412–1422 https://doi.org/10.1002/pro.2721

33 Wilding, M., Walsh, E.F.A., Dorrian, S.J. and Scott, C. (2015) Identification of novel transaminases from a 12-aminododecanoic acid-metabolizing *Pseudomonas* strain. *Microb. Biotechnol.* **8**, 665–672 https://doi.org/10.1111/1751-7915.12278

34 Wilding, M., Scott, C. and Warden, A.C. (2018) Computer-guided surface engineering for enzyme improvement. *Sci. Rep.* **8**, 11998 https://doi.org/10.1038/s41598-018-30434-5

35 Sternke, M., Tripp, K.W. and Barrick, D. (2018) Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. *bioRxiv*, 466391 https://doi.org/10.1101/466391

36 Porebski, B.T., Nickson, A.A., Hoke, D.E., Hunter, M.R., Zhu, L., McGowan, S. et al. (2015) Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain. *Protein Eng. Des. Sel.* **28**, 67–78 https://doi.org/10.1093/protein/gzv002

37 Porebski, B.T., Conroy, P.J., Drinkwater, N., Schofield, P., Vazquez-Lombardi, R., Hunter, M.R. et al. (2016) Circumventing the stability-function trade-off in an engineered FN3 domain. *Protein Eng. Des. Sel.* **29**, 541–550 https://doi.org/10.1093/protein/gzw046

38 Porebski, B.T., Keleher, S., Hollins, J.J., Nickson, A.A., Marijanovic, E.M., Borg, N.A. et al. (2016) Smoothing a rugged protein folding landscape by sequence-based redesign. *Sci. Rep.* **6**, 33958 https://doi.org/10.1038/srep33958

39 Kaltenbach, M., Burke, J.R., Dindo, M., Pabis, A., Munsberg, F.S., Rabin, A. et al. (2018) Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* **14**, 548–555 https://doi.org/10.1038/s41589-018-0042-3

40 Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R. et al. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 https://doi.org/10.1016/B978-0-12-381270-4.00019-6

41 Bender, B.J., Cisneros, III, A., Duran, A.M., Finn, J.A., Fu, D., Lokits, A.D. et al. (2016) Protocols for molecular modeling with Rosetta3 and RosettaScripts. *Biochemistry* **55**, 4748–4763 https://doi.org/10.1021/acs.biochem.6b00444

42 Campbell, E.C., Grant, J., Wang, Y., Sandhu, M., Williams, R.J., Nisbet, D.R. et al. (2018) Hydrogel-immobilized supercharged proteins. *Adv. Biosyst.* **2**, 1700240 https://doi.org/10.1002/adbi.201700240

43 Miklos Aleksandr, E., Kluwe, C., Der Bryan, S., Pai, S., Sircar, A., Hughes Randall, A. et al. (2012) Structure-based design of supercharged, highly thermoresistant antibodies. *Chem. Biol.* **19**, 449–455 https://doi.org/10.1016/j.chembiol.2012.01.018

44 Goldenzweig, A., Goldsmith, M., Hill, S.E., Gertman, O., Laurino, P., Ashani, Y. et al. (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell* **63**, 337–346 https://doi.org/10.1016/j.molcel.2016.06.012

45 Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H. et al. (2018) Automated design of efficient and functionally diverse enzyme repertoires. *Mol. Cell* **72**, 178–86.e5 https://doi.org/10.1016/j.molcel.2018.08.033

46 Wang, J., Cao, H., Zhang, J.Z.H. and Qi, Y. (2018) Computational protein design with deep learning neural networks. *Sci. Rep.* **8**, 6349 https://doi.org/10.1038/s41598-018-24760-x

47 Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388 https://doi.org/10.1093/nar/gki387

48 Pandurangan, A.P., Ochoa-Montaño, B., Ascher, D.B. and Blundell, T.L. (2017) SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* **45**, W229–W235 https://doi.org/10.1093/nar/gkx439

49 Sumbalova, L., Stourac, J., Martinek, T., Bednar, D. and Damborsky, J. (2018) Hotspot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Res.* **46**, W356–W362 https://doi.org/10.1093/nar/gky417

50 Jurcik, A., Bednar, D., Byska, J., Marques, S.M., Furmanova, K., Daniel, L. et al. (2018) CAVER analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* **34**, 3586–3588 https://doi.org/10.1093/bioinformatics/bty386

51 Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T. et al. (2012) Principles for designing ideal protein structures. *Nature* **491**, 222 https://doi.org/10.1038/nature11600

52 Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Röthlisberger, D., Zanghellini, A. et al. (2008) De Novo computational design of retro-aldol enzymes. *Science* **319**, 1387 https://doi.org/10.1126/science.1152692

53 Obexer, R., Godina, A., Garrabou, X., Mittl, P.R.E., Baker, D., Griffiths, A.D. et al. (2017) Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 https://doi.org/10.1038/nchem.2596

54 Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J. et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190 https://doi.org/10.1038/nature06879

55 Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., Lambert, A.R., St.Clair, J.L. et al. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular DIELS-ALDER reaction. *Science* **329**, 309 https://doi.org/10.1126/science.1190239

56 Hong, N.-S., Petrović, D., Lee, R., Gryn'ova, G., Purg, M., Saunders, J. et al. (2018) The evolution of multiple active site configurations in a designed enzyme. *Nat. Commun.* **9**, 3900 https://doi.org/10.1038/s41467-018-06305-y

57 Davey, J.A., Damry, A.M., Euler, C.K., Goto, N.K. and Chica, R.A. (2015) Prediction of stable globular proteins using negative design with non-native backbone ensembles. *Structure* **23**, 2011–2021 https://doi.org/10.1016/j.str.2015.07.021

58 Davey, J.A., Damry, A.M., Goto, N.K. and Chica, R.A. (2017) Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.* **13**, 1280 https://doi.org/10.1038/nchembio.2503

59 Discher, B.M., Noy, D., Strzalka, J., Ye, S., Moser, C.C., Lear, J.D. et al. (2005) Design of amphiphilic protein maquettes: controlling assembly, membrane insertion, and cofactor interactions. *Biochemistry* **44**, 12329–12343 https://doi.org/10.1021/bi050695m

60 Joh, N.H., Wang, T., Bhate, M.P., Acharya, R., Wu, Y., Grabe, M. et al. (2014) De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle. *Science* **346**, 1520 https://doi.org/10.1126/science.1261172

61 Korendovych, I.V., Senes, A., Kim, Y.H., Lear, J.D., Fry, H.C., Therien, M.J. et al. (2010) De novo design and molecular assembly of a transmembrane diporphyrin-binding protein complex. *J. Am. Chem. Soc.* **132**, 15516–8 https://doi.org/10.1021/ja107487b
62 Marcos, E., Basanta, B., Chidyausiku, T.M., Tang, Y., Oberdorfer, G., Liu, G. et al. (2017) Principles for designing proteins with cavities formed by curved β sheets. *Science* **355**, 201 https://doi.org/10.1126/science.aah7389
63 Shen, H., Fallas, J.A., Lynch, E., Sheffler, W., Parry, B., Jannetty, N. et al. (2018) De novo design of self-assembling helical protein filaments. *Science* **362**, 705 https://doi.org/10.1126/science.aau3775
64 Lalaurie, C.J., Dufour, V., Meletiou, A., Ratcliffe, S., Harland, A., Wilson, O. et al. (2018) The de novo design of a biocompatible and functional integral membrane protein using minimal sequence complexity. *Sci. Rep.* **8**, 14564 https://doi.org/10.1038/s41598-018-31964-8
65 Silva, D.-A., Yu, S., Ulge, U.Y., Spangler, J.B., Jude, K.M., Labão-Almeida, C. et al. (2019) De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 https://doi.org/10.1038/s41586-018-0830-7
66 Eckenroth, B.E., Towle-Weicksel, J.B., Nemec, A.A., Murphy, D.L., Sweasy, J.B. and Doublié, S. (2017) Remote mutations induce functional changes in active site residues of human DNA polymerase β. *Biochemistry* **56**, 2363–2371 https://doi.org/10.1021/acs.biochem.6b01287
67 Jubb, H.C., Pandurangan, A.P., Turner, M.A., Ochoa-Montaño, B., Blundell, T.L. and Ascher, D.B. (2017) Mutations at protein-protein interfaces: small changes over big surfaces have large impacts on human health. *Prog. Biophys. Mol. Biol.* **128**, 3–13 https://doi.org/10.1016/j.pbiomolbio.2016.10.002
68 Jiménez-Osés, G., Osuna, S., Gao, X., Sawaya, M.R., Gilson, L., Collier, S.J. et al. (2014) The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat. Chem. Biol.* **10**, 431 https://doi.org/10.1038/nchembio.1503
69 Baier, F., Hong, N., Yang, G., Pabis, A., Miton, C.M., Barrozo, A. et al. (2019) Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. *eLife* **8**, e40789 https://doi.org/10.7554/eLife.40789.001
70 Yang, G., Anderson, D.W., Baier, F., Dohmen, E., Hong, N., Carr, P.D. et al. (2018) Higher-order epistatic networks underlie the evolutionary fitness landscape of a xenobiotic- degrading enzyme. *bioRxiv*, 504811 https://doi.org/10.1101/504811
71 Vaissier, V., Sharma, S.C., Schaettle, K., Zhang, T. and Head-Gordon, T. (2018) Computational optimization of electric fields for improving catalysis of a designed kemp eliminase. *ACS Catal.* **8**, 219–227 https://doi.org/10.1021/acscatal.7b03151
72 Khersonsky, O., Kiss, G., Röthlisberger, D., Dym, O., Albeck, S., Houk, K.N. et al. (2012) Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl Acad. Sci. U.S.A.* **109**, 10358 https://doi.org/10.1073/pnas.1121063109
73 Preiswerk, N., Beck, T., Schulz, J.D., Milovník, P., Mayer, C., Siegel, J.B. et al. (2014) Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc. Natl Acad. Sci. U.S.A.* **111**, 8013 https://doi.org/10.1073/pnas.1401073111
74 Romero-Rivera, A., Garcia-Borràs, M. and Osuna, S. (2017) Role of conformational dynamics in the evolution of retro-aldolase activity. *ACS Catal.* **7**, 8524–8532 https://doi.org/10.1021/acscatal.7b02954
75 Liu, X., Kang, F., Hu, C., Wang, L., Xu, Z., Zheng, D. et al. (2018) A genetically encoded photosensitizer protein facilitates the rational design of a miniature photocatalytic CO_2-reducing enzyme. *Nat. Chem.* **10**, 1201–1206 https://doi.org/10.1038/s41557-018-0150-4
76 Pearson, A.D., Mills, J.H., Song, Y., Nasertorabi, F., Han, G.W., Baker, D. et al. (2015) Trapping a transition state in a computationally designed protein bottle. *Science* **347**, 863 https://doi.org/10.1126/science.aaa2424
77 Subramanian, K., Góra, A., Spruijt, R., Mitusińska, K., Suarez-Diez, M., Martins dos Santos, V. et al. (2018) Modulating D-amino acid oxidase (DAAO) substrate specificity through facilitated solvent access. *PLoS ONE* **13**, e0198990 https://doi.org/10.1371/journal.pone.0198990
78 Saavedra, H.G., Wrabl, J.O., Anderson, J.A., Li, J. and Hilser, V.J. (2018) Dynamic allostery can drive cold adaptation in enzymes. *Nature* **558**, 324–328 https://doi.org/10.1038/s41586-018-0183-2
79 Tyukhtenko, S., Rajarshi, G., Karageorgos, I., Zvonok, N., Gallagher, E.S., Huang, H. et al. (2018) Effects of distal mutations on the structure, dynamics and catalysis of human monoacylglycerol lipase. *Sci. Rep.* **8**, 1719 https://doi.org/10.1038/s41598-017-19135-7
80 Buhrke, D., Velazquez Escobar, F., Sauthof, L., Wilkening, S., Herder, N., Tavraz, N.N. et al. (2016) The role of local and remote amino acid substitutions for optimizing fluorescence in bacteriophytochromes: a case study on iRFP. *Sci. Rep.* **6**, 28444 https://doi.org/10.1038/srep28444
81 Filonov, G.S., Piatkevich, K.D., Ting, L.-M., Zhang, J., Kim, K. and Verkhusha, V.V. (2011) Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat. Biotechnol.* **29**, 757 https://doi.org/10.1038/nbt.1918
82 Basudhar, D., Madrona, Y., Yukl, E.T., Sivaramakrishnan, S., Nishida, C.R., Moënne-Loccoz, P. et al. (2016) Distal hydrogen-bonding interactions in ligand sensing and signaling by *Mycobacterium tuberculosis* DosS. *J. Biol. Chem.* **291**, 16100–16111 https://doi.org/10.1074/jbc.M116.724815
83 Capdevila, D.A., Braymer, J.J., Edmonds, K.A., Wu, H. and Giedroc, D.P. (2017) Entropy redistribution controls allostery in a metalloregulatory protein. *Proc. Natl Acad. Sci. U.S.A.* **114**, 4424 https://doi.org/10.1073/pnas.1620665114
84 Gober, J.G., Rydeen, A.E., Schwochert, T.D., Gibson-O'Grady, E.J. and Brustad, E.M. (2018) Enhancing cytochrome P450-mediated non-natural cyclopropanation by mutation of a conserved second-shell residue. *Biotechnol. Bioeng.* **115**, 1416–1426 https://doi.org/10.1002/bit.26571
85 Genna, V., Colombo, M., De Vivo, M. and Marcia, M. (2018) Second-shell basic residues expand the two-metal-ion architecture of DNA and RNA processing enzymes. *Structure* **26**, 40–50.e2 https://doi.org/10.1016/j.str.2017.11.008
86 Zhang, F., Gao, L., Zetsche, B. and Slaymaker, I. (2015) Crispr enzyme mutations reducing off-target effects patent WO2016205613A1
87 Navrátilová, V., Paloncýová, M., Berka, K., Mise, S., Haga, Y., Matsumura, C. et al. (2017) Molecular insights into the role of a distal F240A mutation that alters CYP1A1 activity towards persistent organic pollutants. *Biochim. Biophys. Acta Gen. Subj.* **1861**, 2852–2860 https://doi.org/10.1016/j.bbagen.2017.08.002
88 Mei, L., Zhou, Y., Zhu, L., Liu, C., Wu, Z., Wang, F. et al. (2018) Site-mutation of hydrophobic core residues synchronically poise super interleukin 2 for signaling: identifying distant structural effects through affordable computations. *Int. J. Mol. Sci.* **19**, E916 https://doi.org/10.3390/ijms19030916
89 Levin, A.M., Bates, D.L., Ring, A.M., Krieg, C., Lin, J.T., Su, L. et al. (2012) Exploiting a natural conformational switch to engineer an interleukin-2 'superkine'. *Nature* **484**, 529 https://doi.org/10.1038/nature10975