Review Article

The road to fully programmable protein catalysis

https://doi.org/10.1038/s41586-022-04456-z

Received: 8 July 2021

Accepted: 21 January 2022

Published online: 1 June 2022

Check for updates

Sarah L. Lovelock[1], Rebecca Crawshaw[1], Sophie Basler[2], Colin Levy[1], David Baker[3,4,5]✉, Donald Hilvert[2]✉ & Anthony P. Green[1]✉

The ability to design efficient enzymes from scratch would have a profound effect on chemistry, biotechnology and medicine. Rapid progress in protein engineering over the past decade makes us optimistic that this ambition is within reach. The development of artificial enzymes containing metal cofactors and noncanonical organocatalytic groups shows how protein structure can be optimized to harness the reactivity of nonproteinogenic elements. In parallel, computational methods have been used to design protein catalysts for diverse reactions on the basis of fundamental principles of transition state stabilization. Although the activities of designed catalysts have been quite low, extensive laboratory evolution has been used to generate efficient enzymes. Structural analysis of these systems has revealed the high degree of precision that will be needed to design catalysts with greater activity. To this end, emerging protein design methods, including deep learning, hold particular promise for improving model accuracy. Here we take stock of key developments in the field and highlight new opportunities for innovation that should allow us to transition beyond the current state of the art and enable the robust design of biocatalysts to address societal needs.

Enzymes are exceptionally powerful catalysts that use sophisticated active sites to process chemical transformations. The enormous rate accelerations and unrivalled selectivities achievable with enzymes make them attractive catalysts for use in sustainable manufacturing processes[1–3]. The field of biocatalysis has advanced to the stage where it is now viewed as a key enabling technology for the development of a greener and more efficient chemical industry[4,5]. Rapid progress has been underpinned by several major methodological innovations, including: the availability of rapid, accurate and low cost DNA synthesis and sequencing services; the development of advanced bioinformatics tools and computational modelling methods; and increasingly sophisticated experimental workflows for high-throughput structural and biochemical enzyme characterization[6]. These advances have led to the availability of an increasingly diverse portfolio of natural enzymes with interesting catalytic functions, which can in principle be exploited by synthetic chemists when devising routes to target molecules. However, natural enzymes are seldom suitable for direct use in chemical processes, and protein engineering is typically required to optimize their properties for practical applications. Recent years have seen the emergence of high-throughput protein engineering strategies, most notably directed evolution, for developing biocatalysts that process non-native substrates with high efficiency and selectivity, and that operate effectively under commercially viable process conditions[5,7–9] (Fig. 1a).

Although powerful, directed evolution is costly and time consuming and this restricts the potential impact of biocatalysis on many industrial processes. Furthermore, for many desirable chemical transformations there are no known enzymes that can serve as starting templates for evolutionary optimization. To address these limitations, ultrahigh-throughput screening methods[10–13] and continuous evolution platforms have been developed, which offer exciting avenues to accelerate protein engineering[14–16], and the discovery of mechanistically promiscuous enzymes provides a gateway to catalytic functions not found naturally[17–20]. However, these strategies are only amenable to a handful of chemical transformations and so do not offer general solutions.

In light of these limitations it is timely to consider how we as a field will deliver a step change to allow rapid, reliable and cost-effective development of biocatalysts for a broad range of applications. Although top-down engineering of natural enzymes undoubtedly remains the gold standard for biocatalyst development, this approach can only take us so far. In our view, bottom-up or de novo enzyme design, in which entirely new catalytic centres are created within protein hosts, could offer a general solution to both the speed and scope of biocatalyst delivery in the future[21–23] (Fig. 1b). Here we will review key developments to illustrate recent progress in this nascent field, including the development of artificial metalloenzymes, enzymes with noncanonical organocatalytic groups and the computational design of de novo enzymes from first principles. This analysis serves as a platform to discuss the limitations of current design approaches, and how these challenges may be addressed moving forward.

[1]Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, Manchester, UK. [2]Laboratory of Organic Chemistry, ETH Zürich, Zürich, Switzerland. [3]Department of Biochemistry, University of Washington, Seattle, WA, USA. [4]Institute for Protein Design, University of Washington, Seattle, WA, USA. [5]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. ✉e-mail: dabaker@uw.edu; donald.hilvert@org.chem.ethz.ch; anthony.green@manchester.ac.uk

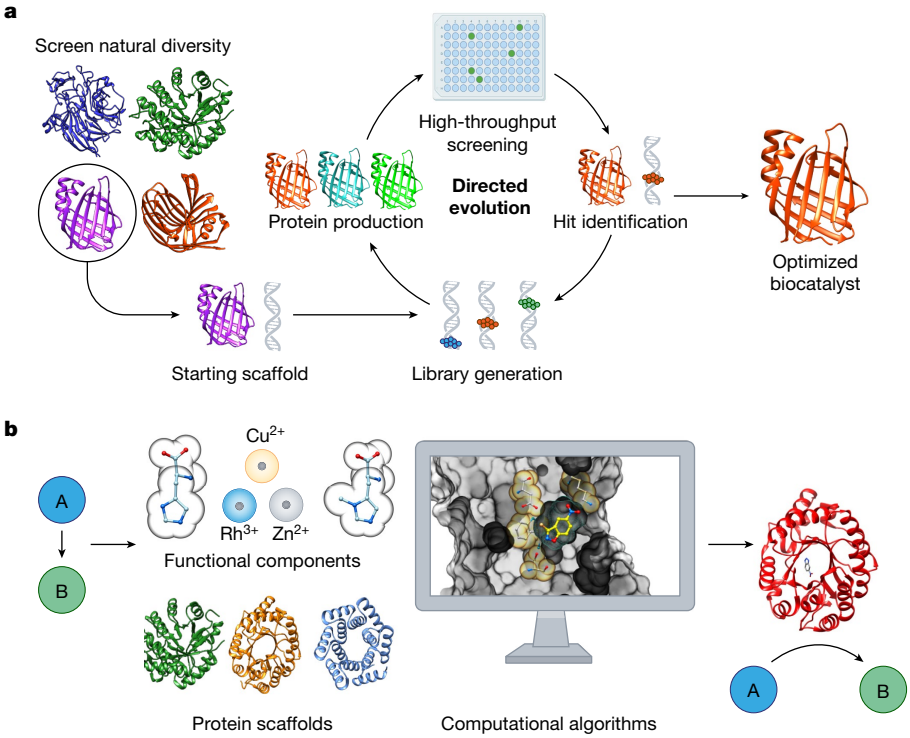

Fig. 1 | Top-down enzyme engineering versus bottom-up design. **a**, Workflow for the development of practically useful biocatalysts. Natural enzymes with desired catalytic activities are identified and their properties optimized via directed evolution. **b**, Ambition of de novo enzyme design. Following selection of a target transformation, computational methods can be used to predict protein sequences with desired catalytic function. Both natural and de novo proteins can serve as host scaffolds for new catalytic sites. Canonical and noncanonical amino acid side chains and metal ion cofactors can serve as key functional components in de novo active sites.

Artificial metalloenzymes

The functional capabilities of natural enzymes are greatly expanded by the recruitment of metal ion cofactors that facilitate redox chemistry, radical processes and challenging functional group conversions. These metalloenzymes benefit from the synergistic action of the metal cofactor and protein scaffold to accelerate some of the most challenging naturally occurring transformations. The enviable catalytic properties of these systems have inspired the development of complementary approaches to design artificial metalloenzymes. A strategy that has proven particularly versatile involves anchoring pre-assembled transition metal complexes into selected protein scaffolds. This approach has given rise to active catalysts for a wide range of nonbiological transformations, including alkene metathesis and transfer hydrogenations[24–26]. However, in general designing productive interactions between protein, substrate and transition metal complex has proven challenging, and consequently the catalytic efficiencies achieved by these hybrid systems are often lower than the isolated small-molecule complex. A notable exception involved the design of an enantioselective benzannulase comprising a biotinylated rhodium(III) complex bound to a streptavidin scaffold[27] (Fig. 2a). This artificial metalloenzyme accelerates the coupling of benzamides and alkenes by approximately 100-fold over the isolated complex to generate dihydroisoquinolone products with enantiomeric ratios as high as 93:7. This rate acceleration can be attributed to a designed aspartate or glutamate which serves as a catalytic base and works in tandem with the rhodium cofactor to promote the key C–H activation/orthometallation process.

An alternative approach to metalloenzyme design that has proven particularly effective involves reengineering natural metalloproteins to install new functional elements that work in concert with the native cofactor. Key examples include the development of functional mimics of haem copper oxidases and nitric oxide reductases by engineering copper (Cu_B) and non-haem iron (Fe_B) binding sites, respectively, into the distal pocket of the haem protein myoglobin[28,29]. In contrast to the native enzymes, the heteronuclear centres of these de novo metalloenzymes are built into a small and robust protein scaffold that can be easily produced, engineered and crystallized for high-resolution structural characterization. Such systems provide an ideal basis for elucidating key structural and mechanistic features giving rise to the high efficiencies and selectivities achieved by natural metalloenzymes. The power of this approach was recently exemplified through the design of an artificial enzyme that catalyses sulfite reduction, a transformation that has thus far eluded synthetic catalysts[30] (Fig. 2b). Rosetta Matcher and Enzyme Design algorithms were used to design an iron sulfur cluster into the proximal pocket of cytochrome *c* peroxidase, along with a bridging cysteine ligand that coordinates the native haem cofactor and the designed [4Fe–4S] cluster. The sulfite reductase activity of the initial design improved more than 60-fold through the targeted introduction of positively charged Arg and Lys residues in the substrate binding pocket and a Cys235 residue close to the [4Fe–4S] cluster. The activity of this optimized variant is only about 5-fold lower than the native sulfite reductase from *Mycobacterium tuberculosis*, with approximately 10% of the products formed arising from complete (six electron and seven proton) reduction to hydrogen sulfide.

Creating metalloenzymes 'from scratch', where new protein scaffolds are designed to bind metal cofactors and modulate catalysis, offers the prospect of complete control over metalloprotein sequence, structure and function. Most research in this area has focused on introducing binding sites for metal ions and metalloporphyrin cofactors into designed α-helical bundles, giving rise to protein catalysts for hydrolytic reactions, redox processes and carbene transfers[31–36]. The introduction of complex dinuclear cofactors such as carboxylate-bridged diiron centres has led to a family of Due Ferri proteins with various O_2-dependent activities, including the substrate-gated four-electron

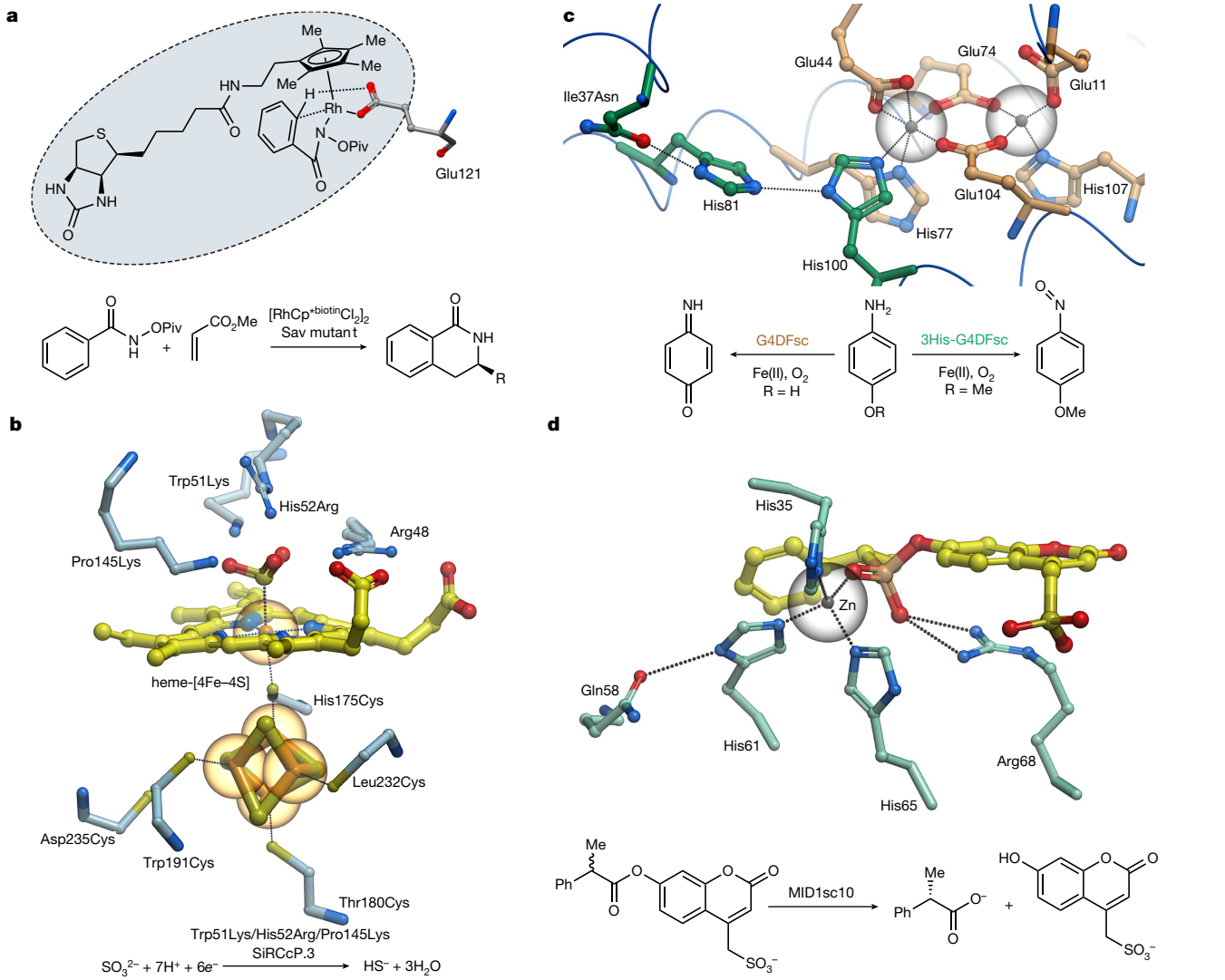

Fig. 2 | Approaches to de novo metalloenzymes. a, Supramolecular anchoring of pre-assembled transition metal complexes into host scaffolds. This approach was used for the development of an enantioselective benzannulase comprising a biotinylated rhodium(III) complex bound to streptavidin. A designed aspartate or glutamate serves as a catalytic base and works in tandem with the rhodium cofactor to promote the key C–H activation/orthometallation process. **b**, Introduction of new functional components into existing metalloenzymes can lead to new functions. For example, the design of a [4Fe–4S] cluster into an engineered myoglobin gave rise to an artificial sulfite reductase. **c**, New protein scaffolds have been designed from scratch to bind metal cofactors and modulate catalysis. The Due Ferri protein G4DFsc (gold) is a hydroquinone oxidase that employs a carboxylate bridged diiron cofactor as a catalytic centre. Introduction of a third histidine ligand (green) into the metal binding cavity transforms G4DFsc from a hydroquinone oxidase to an arylamine *N*-hydroxylase. **d**, A designed homodimeric peptide (MID1) containing two interfacial zinc binding sites served as a starting point for evolutionary optimization to afford an efficient and enantioselective zinc hydrolase MID1sc10.

reduction of oxygen to water[33,37–39]. Interestingly, the catalytic function of these Due Ferri proteins can be altered through rational reprogramming of the metal coordination environment (Fig. 2c). Specifically, G4DFsc was successfully transformed from a hydroquinone oxidase to an arylamine *N*-hydroxylase by introducing a third His ligand to the metal binding cavity[39].

Metal binding sites have also been designed at the interface of polypeptides or protein subunits to direct the assembly of higher-order structures[40]. In one instance, a designed homodimeric peptide containing two interfacial zinc binding sites was found to display serendipitous activity for ester bond hydrolysis resulting from a vacant metal coordination site adjacent to a hydrophobic pocket[41,42]. Fusion of the N and C termini of the dimer subunits and removal of one of the two zinc binding sites afforded a single chain variant, which was subjected to extensive laboratory evolution to deliver a highly efficient ($k_{cat}/K_M \approx 10^6$ M^{-1} s^{-1}) and enantioselective zinc hydrolase[43] (Fig. 2d). A high-resolution crystal structure of the evolved enzyme complexed with a transition state analogue sheds light on the catalytic mechanism, revealing that the catalytic zinc ion is coordinated by three histidine ligands and activates the nucleophilic water as a metal hydroxide, while an active-site Arg64 stabilizes the anionic transition states through bidentate hydrogen bonding. This simple helical bundle scaffold was subsequently transformed into a proficient catalyst for a bimolecular hetero-Diels–Alder reaction[44]. Evolution afforded the chemo- and stereoselective metalloenzyme DA7, which uses Lewis acid catalysis and a strategically positioned hydrogen bond network for effective transition state stabilization.

Adding noncanonical amino acids

Enzyme design strategies are typically reliant on twenty canonical amino acids, which contain a narrow set of functional groups. This limited

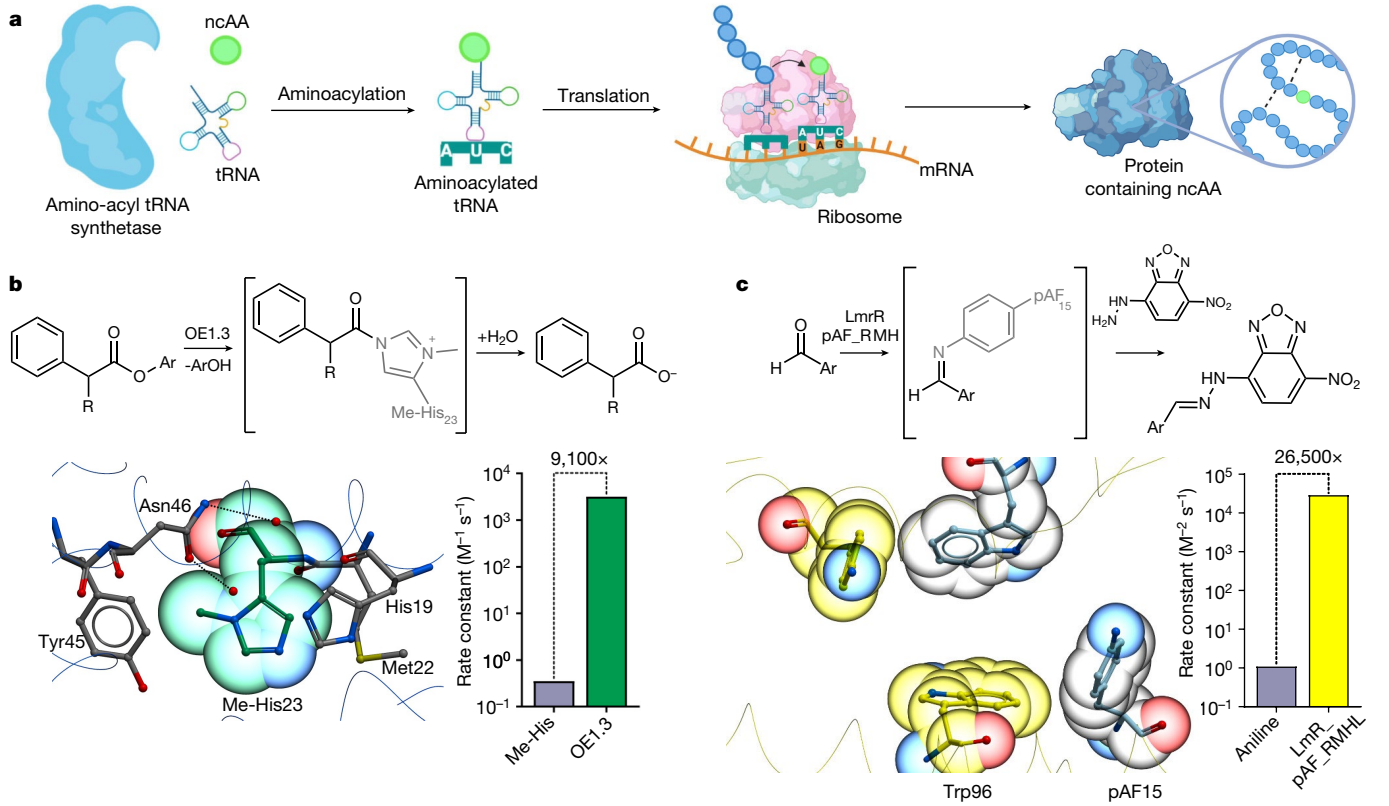

Fig. 3 | Enzymes with an expanded amino acid alphabet. a, An orthogonal aminoacyl-tRNA synthetase aminoacylates its cognate tRNA with a noncanonical amino acid (ncAA). The aminoacylated tRNA is decoded on the ribosome in response to a UAG codon in the mRNA during translational elongation, leading to the addition of an ncAA to the growing polymer. **b**, Genetically encoded N_δ-methyl histidine (Me-His) residues can serve as noncanonical catalytic nucleophiles to promote enantioselective ester hydrolysis in de novo active sites. Histidine methylation is essential for catalytic function as it prevents the formation of unreactive acyl-enzyme intermediates derived from canonical histidine nucleophiles. The bar chart shows the approximately 9,000-fold rate acceleration achieved by OE1.3 compared with free N_δ-methyl histidine in solution. **c**, Post-translational chemical reduction of a *p*-azidophenylalanine (pAzF) residue installed into an engineered LmrR unmasks a *p*-aminophenylalanine (pAF) catalytic nucleophile designed to promote hydrazone formation.

functionality restricts the catalytic mechanisms that can be installed into de novo active sites. However, a wider range of functional components can now be accessed using genetic code expansion methods that allow the selective incorporation of structurally diverse noncanonical amino acids (ncAAs) into proteins[45,46]. These methods typically employ an orthogonal aminoacyl-tRNA synthetase–tRNA pair to direct the incorporation of an ncAA in response to an unassigned codon (most commonly UAG) introduced into the gene of interest (Fig. 3a). Genetically encoded ncAAs provide new avenues to explore how enzymes operate at the molecular level[47–51], and have been used to improve biocatalyst activity and stability[52,53]. The availability of an expanded set of amino acid building blocks also provides exciting opportunities to design enzymes with new catalytic mechanisms that do not occur naturally[54].

This approach was recently showcased through the design of a de novo hydrolase (OE1) that employs N_δ-methyl histidine (Me-His) as a noncanonical catalytic nucleophile with a similar mode of reactivity to the widely employed nucleophilic catalyst DMAP[55] (4-dimethylaminopyridine; Fig. 3b). Histidine methylation was essential for catalytic function, as it prevented the accumulation of unreactive acyl-enzyme intermediates that compromised the catalytic activity of earlier de novo hydrolases equipped with canonical nucleophiles[56–60]. Optimization of OE1 was achieved over iterative rounds of evolution using workflows adapted to an expanded genetic code to afford a variant OE1.3, which is four orders of magnitude more efficient than equivalent small molecule catalysts in promoting ester hydrolysis, and OE1.4, which is able to promote enantioselective transformations.

A related approach was used to generate de novo enzymes to promote the synthesis of oximes and hydrazones (Fig. 3c). A *p*-azidophenylalanine residue introduced into the transcriptional regulator protein LmrR was reduced with tris(2-carboxyethyl)phosphine to unmask a *p*-aminophenylalanine residue containing a reactive aniline side chain[61]. Although the initial LmrR_pAF variant gave modest activity improvements over the parent protein LmrR, subsequent optimization through targeted rounds of directed evolution afforded a quadruple mutant with a 55-fold improvement in k_{cat} and a 26,000-fold increased efficiency over aniline in solution[62]. Combined, these studies suggest that the introduction of 'organocatalytic' motifs into evolvable protein scaffolds can offer a general strategy to deliver enzymes that are orders of magnitude more efficient than small organic catalysts used in isolation.

Stabilizing key transition states

If we are to capitalize on our ability to install new functional components into proteins, we must learn how to reliably design enzymes on the basis of the fundamental principles of transition state stabilization that underpin catalysis. To this end, early efforts exploited the mammalian immune system to raise antibodies towards stable transition state analogues[63–68]. This approach has delivered catalytic antibodies for a wide range of chemical transformations. However, in general antibodies fail to achieve the efficiencies of natural enzymes and many energetically demanding reactions have proved intractable to this approach.

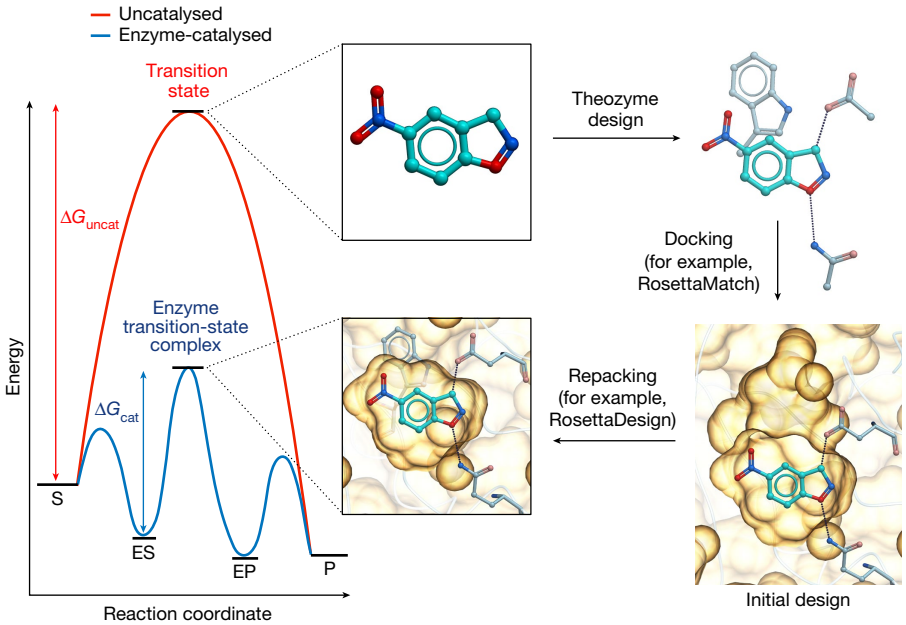

Fig. 4 | Computational design of enzymes. Reaction profiles of an uncatalysed (red) and enzyme-catalysed (blue) single-step reaction are shown, where S is the substrate, P is the product, ES is the enzyme–substrate complex, and EP is the enzyme–product complex. Computational methods are used to design new active sites in proteins on the basis of their ability to stabilize rate-limiting transitions states. A typical procedure begins with a quantum mechanically calculated transition state of a target transformation, termed the theozyme. Key functional groups, such as surrogates of amino acid chains required to stabilize the transition state, are explicitly included in the calculation. The resulting ensemble, which represents an idealized model of a minimal active site, is then docked into structurally characterized protein scaffolds using programmes such as RosettaMatch. The selected binding pockets are then computationally repacked, for example with RosettaDesign, to optimize interactions between the substrate and the transition state. The designs are subsequently ranked and tested experimentally. Promising designs can be optimized experimentally using directed evolution.

More recently, computational enzyme design has emerged as a powerful and more flexible approach that does not depend on the availability of imperfect transition state analogues and is not restricted to the antibody fold[21–23] (Fig. 4). The design process involves the following general steps. (1) Design and generation of a 'theozyme'—an idealized active-site model comprising a quantum mechanically calculated transition state and key functional groups from amino acid side chains required for transition state stabilization. (2) Docking of the theozyme into structurally characterized proteins to identify sterically complementary scaffolds that can accommodate the key catalytic groups as amino acid side chains linked to the protein backbone. (3) Redesign of residues in and around the active site to optimize packing of the theozyme. So far, this process has enabled the design of protein catalysts for a handful of model transformations[56,69–72]. Although the activities of the starting designs are typically low, they can be optimized through laboratory evolution[73–76]. In favourable cases, this combination of computational design and directed evolution has afforded biocatalysts with efficiencies comparable to natural enzymes. A comprehensive understanding of the molecular changes giving rise to improved activity can then be used to inform the development of improved design protocols.

The Kemp elimination, involving the conversion of benzisoxazoles into salicylonitriles, has been a popular target transformation for enzyme designers and provides a valuable model system for studying proton transfers from carbon[69,70] (Fig. 5a). The evolved enzyme HG3.17 is the most efficient Kemp eliminase reported so far and catalyses the deprotonation of 5-nitrobenzisoxazole with a $k_{cat} \approx 700\ s^{-1}$, which represents a 1,000-fold improvement over the parent design (HG3) and is in the range of proton transfer rates observed in many naturally occurring enzymes[73]. This extraordinary activity can be attributed to highly effective bifunctional catalysis, involving the designed catalytic base Asp127 and Gln50, which emerged during evolution to stabilize developing negative charge on the phenoxide leaving group, within an active site that is perfectly tuned to accommodate the substrate in a productive pose for catalysis. Subsequent characterization of HG3, HG3.17 and the intermediate variant HG3.7 using a combination of cryo- and high-temperature crystallography, and nuclear magnetic resonance (NMR) spectroscopy reveals that the conformational ensemble of the protein backbone is altered during evolution, to minimize unproductive conformers observed in the HG3 design in favour of a highly active conformational substate[77,78]. Interestingly, this active conformer is not present in the original xylanase scaffold used as the template for design. This analysis suggests that more efficient catalyst sequences could be predicted using improved design protocols that explicitly sample energetically accessible backbone conformers.

Beyond simple proton transfer reactions, the interplay of computational design and directed evolution has afforded highly efficient catalysts for bimolecular aldol[71,75,79,80] and Diels–Alder reactions[72,74,81]. The most active aldolase so far, RA95.5-8F, was generated following extensive evolution of the RA95 design (Fig. 5b). It cleaves the fluorogenic substrate (*R*)-methodol with a k_{cat} of $10.5\ s^{-1}$, which is in the range of natural type-I aldolases[75]. Here the discovery of highly efficient enzymes was facilitated by ultrahigh-throughput screening of variant libraries by fluorescence-activated droplet sorting, enabling the evaluation of approximately 2,000 sequences per second. Interestingly, the originally designed catalytic Lys210, which operates via the formation of Schiff-base intermediates along the reaction coordinate, was abandoned during the course of evolution in favour of Lys83, which is a more reactive catalytic nucleophile. Structural and biochemical characterization of RA95.5-8F reveals that Lys83 forms part of a sophisticated catalytic centre comprising a Lys-Tyr-Asn-Tyr tetrad that emerged adjacent to a designed hydrophobic pocket during laboratory evolution.

In contrast to extensive active-site changes observed during optimization of RA95 and HG3, the structure of the most highly evolved Diels–Alderase CE20, which promotes selective cycloaddition of 4-carboxybenzyl-*trans*-1,3-butadiene-1-carbamate and *N*,*N*-dimethylacrylamide, shows good overall agreement to DA20_00,

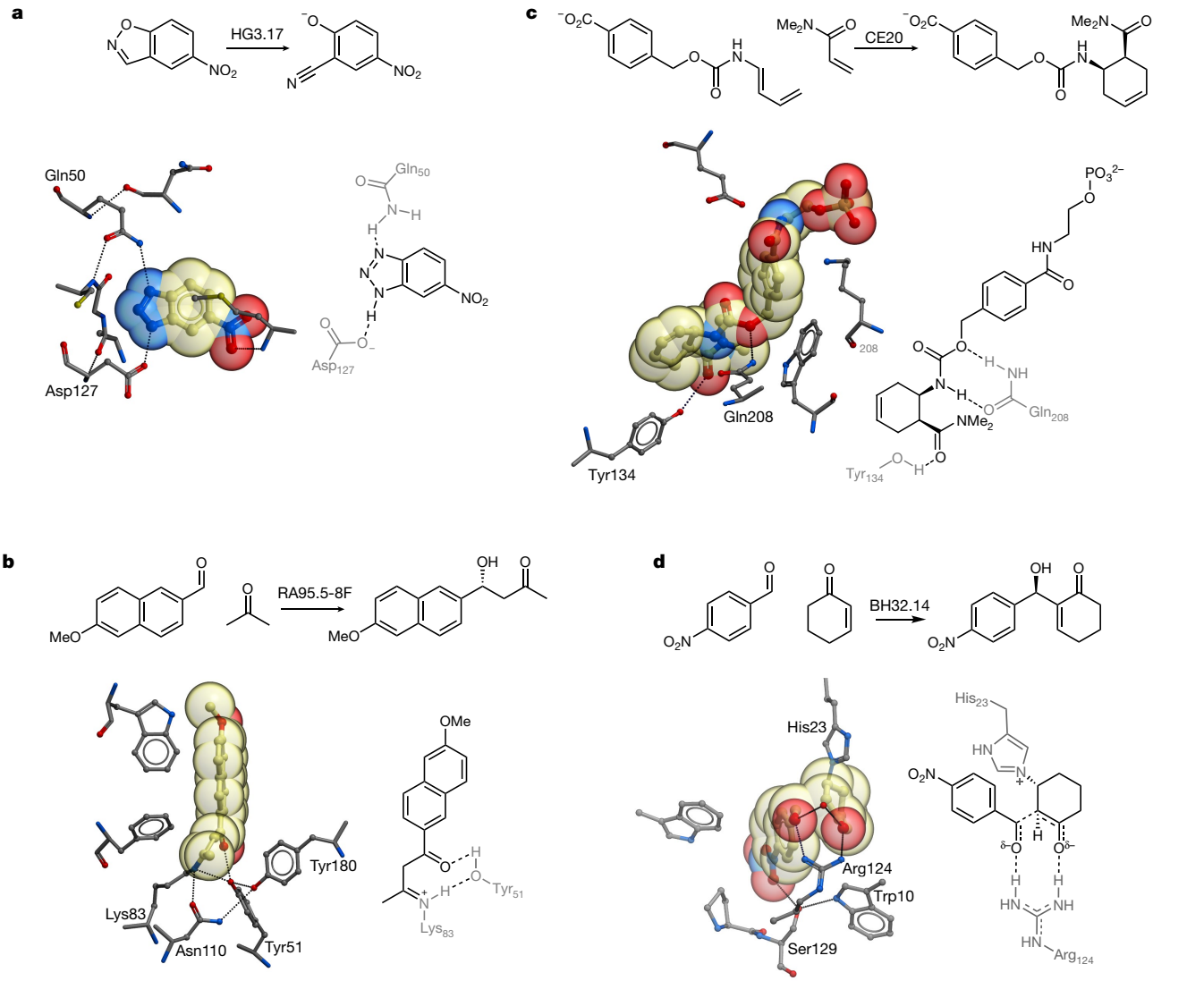

Fig. 5 | De novo enzymes through computational design and directed evolution. a, The Kemp eliminase HG3.17 employs a designed Asp127 catalytic base and Gln50 as an oxyanion stabilizer to catalyse the deprotonation of 5-nitrobenzisoxazole. **b**, The evolved retro-aldolase RA95.5-8F utilizes a Lys-Tyr-Asn-Tyr catalytic tetrad to cleave the fluorogenic substrate (*R*)-methodol. **c**, The Diels–Alder reaction of 4-carboxybenzyl-*trans*-1, 3-butadiene-1-carbamate and *N*,*N*-dimethylacrylamide is accelerated by CE20. Catalytic side chains Gln208 and Tyr134 form hydrogen bonding interactions with the diene and dienophile to reduce the HOMO–LUMO energy gap. **d**, The enantioselective Morita–Baylis–Hillmanase BH32.14 employs a designed His23 nucleophile and catalytic Arg124 to promote the coupling of 4-nitrobenzaldehyde and 2-cyclohexen-1-one. Arg124 shuttles between conformational states to stabilize multiple oxyanion intermediates and transition states along the complex reaction coordinate.

the original design model[74] (Fig. 5c). In particular, the orientation of the bound product and the conformations of the catalytic side chains of Gln208 and Tyr134—which form hydrogen bonding interactions with the diene and dienophile to reduce the energy gap between the highest-occupied molecular orbital (HOMO) of the former and the lowest-occupied molecular orbital (LUMO) of the latter—closely match the design model and changed minimally over the entire evolutionary trajectory. Instead, molecular changes introduced during enzyme optimization gradually reshaped the active-site pocket to achieve more effective preorganization of the reactants into productive conformations for the bimolecular reaction. Notably, substantial activity gains were achieved through the introduction of a 24-residue helix–turn–helix motif, which was designed in a crowdsourcing experiment using the problem solving skills of online game players and serves as a lid element to close off the solvent-exposed active site[81].

Given the conceptual similarities between computational enzyme design and catalytic antibody technology, early computational design efforts targeted chemical transformations that were previously achieved with antibodies. If we are to establish design as a useful source of biocatalysts for practical applications, we must move beyond the functional capabilities of antibodies and develop enzymes for more complex chemical processes for which no effective protein catalysts are known. To this end, an efficient and enantioselective enzyme (BH32.14) for bimolecular Morita–Baylis–Hillman (MBH) reactions was recently developed following extensive evolutionary optimization of an initial computational design[76,82] (Fig. 5d). Crystallographic, biochemical and computational studies reveal that selective catalysis by BH32.14 is achieved through a sophisticated active-site arrangement comprising a designed His23 paired with a judiciously positioned Arg124 that emerged during evolution. This catalytic arginine shuttles between conformational states to provide a highly economical means of stabilizing multiple oxyanion intermediates and transition states along the complex reaction coordinate. Arg124 serves as a genetically encoded surrogate of bidentate

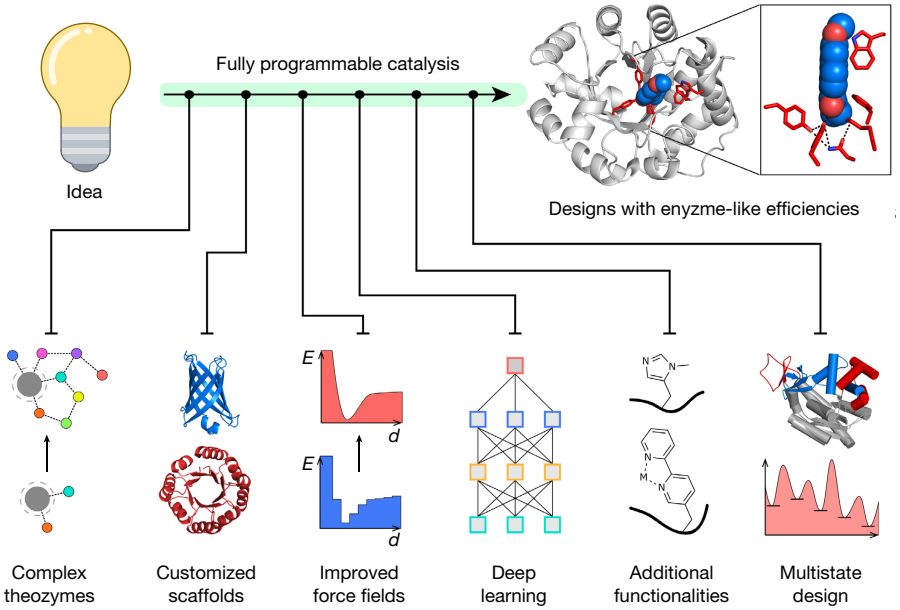

Fig. 6 | A roadmap to better designer enzymes. An illustration of anticipated methodological innovations that will facilitate the design of protein catalysts with enzyme-like efficiencies for a broad range of chemical transformations.

hydrogen bonding catalysts commonly used in organic synthesis to promote a wealth of chemical transformations, including the MBH reaction.

A roadmap to designer enzymes

The examples presented in this Review illustrate great progress made in the field of enzyme design and engineering over the past decade and offer a glimpse of the exciting opportunities that lie ahead (Fig. 6). If design is to achieve, or even surpass, the level of practical utility achieved by more established top-down approaches to biocatalyst development there are two central challenges that the community must now address.

First, we must learn how to design highly active enzymes with efficiencies more akin to natural systems. At present, even for relatively simple transformations many designs must be produced and experimentally tested to identify a few that display desired activity, and extensive evolutionary optimization is required to bridge the efficiency gap to natural enzymes. The development of ultrahigh-throughput enzyme design and screening protocols would facilitate the search for more potent catalysts and could offer a practical route to highly active designs in the medium term. However, if we are to overcome our reliance on high-throughput experimentation, we must consider the factors that make enzyme design so challenging. Efficient protein catalysis requires an extremely high degree of precision to achieve effective discrimination of the transition state from the ground state, and even angstrom level inaccuracies in side-chain positioning can have a catastrophic impact on catalysis. The design challenge is amplified when targeting multistep reactions, where carefully orchestrated conformational adjustments are needed for precise recognition of multiple chemical states. Striking the balance between active-site preorganization and conformational dynamics will be critical to the future success of enzyme design.

The low success rates and modest activities achieved thus far can, in part, be attributed to limitations in conformational sampling methods and energy functions used by existing design algorithms. Although these design methods allow for rapid exploration of protein sequence space, this increased speed inevitably comes at the expense of accuracy. Structural characterization of designed enzymes reveals that key catalytic elements are often not positioned as intended[57,71,73,82]. Accurate placement of polar side chains and the generation of hydrogen bonding networks have proven especially challenging. To address these limitations, more sophisticated molecular force fields, which allow accurate treatment of electrostatics and interactions with solvent, are needed to increase model accuracy. Likewise, more intensive calculations, including hybrid quantum mechanics/molecular mechanics (QM/MM) methods and molecular dynamics simulations with explicit solvent, can have an important role in evaluating and refining computational designs to enable more effective discrimination of the transition state relative to the ground state[83–85]. Although too slow for routine screening of hundreds of designs, as computational power increases we can expect these methods to be more widely integrated into enzyme design processes in the future.

More efficient designs could also be generated through the use of more sophisticated theozyme arrangements. So far, designs have been generated based on simple theozymes containing a small number of functional side chains. In all cases, evolutionary optimization of these designs led to complex arrangements of secondary and tertiary interactions to orientate and fine-tune the reactivity of key catalytic residues, underscoring the importance of extended networks of active-site residues[73–76,86]. Although the transition to more complex theozymes could provide a gateway to more active designs, it will be more challenging to identify protein scaffolds with suitable backbone geometries to accommodate the increased number of functional components. New ensemble-based design methods that account for backbone flexibility along with methods for sculpting protein backbones could facilitate this search[87,88]. More ambitious is the design of new protein folds with backbone geometries specifically tailored to accommodate complex theozyme arrangements[89]. Although yet to be applied to catalysis, the recent design of a fluorescence-activating beta-barrel with a backbone custom built to bind a small fluorescent cofactor[90], and the design of proteins that undergo conformational exchanges reminiscent of those observed in natural enzymes[87,91], hints at the future potential of this approach.

The emergence of deep learning algorithms that allow accurate prediction of protein structure directly from primary sequence provides

exciting new opportunities to design such customized scaffolds[92–96]. Beyond the prediction of protein structure, machine learning has been used to more intelligently navigate sequence space during directed evolution of protein function[97–100], and to produce proteins from scratch that satisfy sets of constraints associated with binding interfaces[101]. A data-driven strategy using evolutionary sequence information was also recently employed to generate active mutases that convert chorismate to prephenate[102]. Interestingly, this transformation has thus far proved intractable to energy-based computational approaches. A crucial next step is to understand how these 'big data' approaches can be adapted to tackle new chemical transformations where extensive evolutionary sequence information is lacking. To this end, deep learning methods are now being extended to design proteins containing active sites defined by a collection of residues and their relative geometries. Given a description of such a site and a structure prediction method such as RoseTTAFold[94] or AlphaFold[92,95], sequences that fold to proteins harbouring the site can be generated by explicitly optimizing a loss function assessing the extent to which the active site is recapitulated, or by filling in the missing sequence and structure information in a forward pass through versions of the networks optimized to recover both sequence and structure information[103]. Moving forward, we anticipate that hybrid design strategies that combine beneficial aspects of deep learning and fundamental biophysical understanding will prove a fruitful avenue for exploration. Irrespective of the particular design method employed, directed evolution will likely continue to have a central role in refining the catalytic sites of de novo enzymes for the foreseeable future.

Second, we must expand the range of chemistries achievable with de novo enzymes, and develop catalysts for valuable chemical processes that can be implemented at scale. To maximize synthetic utility, particular emphasis should be placed on nonbiological classes of reaction for which no natural enzymes are known. In these instances, mechanistic strategies employed in small molecule catalysis can be used to inspire theozyme design. Here, broader collaboration between organic chemists and protein designers will prove especially valuable to identify suitable active-site arrangements for new target transformations. The range of accessible chemistries can be greatly extended by engineering cellular translation to introduce new functional amino acids into proteins, that can be used to modulate catalysis by metal ion cofactors or that serve as genetically encoded surrogates of small molecule organic catalysts[50,55,61,62,104,105]. In the field of organocatalysis, large numbers of chemical transformations can be accelerated using a handful of generic activation modes[106–110]. Consequently, the addition of a few key amino acids to the genetic code could lead to an explosion of new activities in designed active sites. In the future, enzyme designers and engineers will continue to push the boundaries of the field by developing catalysts for increasingly complex transformations. Here it is probable that existing design protocols that use a single transition structure to approximate all species along the reaction coordinate and generate a static model of the protein–transition state complex will become progressively less effective. Instead, more comprehensive design methods that model multiple chemical states along the reaction coordinate will be needed to tackle complex reactions involving multiple high-energy transition states[111,112].

In summary, although there remain considerable challenges to overcome, we are cautiously optimistic that fully programmable catalysis—in which new protein sequences can be predicted from scratch to deliver efficient biocatalysts with desired functions—will become a reality in the future. In our view, this ambitious goal can only be achieved through a collaborative and multidisciplinary effort drawing on expertise in computational chemistry and biology, organic chemistry, enzymology, structural biology, protein design, directed evolution and beyond.

1. Savile, C. K. et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).
2. Huffman, M. A. et al. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* **366**, 1255–1259 (2019). erratum **368**, eabc1954 (2020).
3. Schober, M. et al. Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nat. Catal.* **2**, 909–915 (2019).
4. Devine, P. N. et al. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2**, 409–421 (2018).
5. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567–573 (2009).
6. Bornscheuer, U. T. et al. Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
7. Zeymer, C. & Hilvert, D. Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).
8. Arnold, F. H. Directed evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed.* **57**, 4143–4148 (2018).
9. Qu, G., Li, A., Acevedo-Rocha, C. G., Sun, Z. & Reetz, M. T. The crucial role of methodology development in directed evolution of selective enzymes. *Angew. Chem. Int. Ed.* **59**, 13204–13231 (2020).
10. Fernandez-Gacio, A., Uguen, M. & Fastrez, J. Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.* **21**, 408–414 (2003).
11. Becker, S., Schmoldt, H. U., Adams, T. M., Wilhelm, S. & Kolmar, H. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Curr. Opin. Biotechnol.* **15**, 323–329 (2004).
12. Agresti, J. J. et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl Acad. Sci. USA* **107**, 4004–4009 (2010).
13. Debon, A. et al. Ultrahigh-throughput screening enables efficient single-round oxidase remodelling. *Nat. Catal.* **2**, 740–747 (2019).
14. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
15. Bryson, D. I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
16. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* **175**, 1946–1957 (2018).
17. Zhang, R. K. et al. Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp^3 C–H functionalization. *Nature* **565**, 67–72 (2018).
18. Chen, K., Huang, X., Kan, S. B. J., Zhang, R. K. & Arnold, F. H. Enzymatic construction of highly strained carbocycles. *Science* **360**, 71–75 (2018).
19. Biegasiewicz, K. F. et al. Photoexcitation of flavoenzymes enables a stereoselective radical cyclization. *Science* **364**, 1166–1169 (2019).
20. Ji, P., Park, J., Gu, Y., Clark, D. S. & Hartwig, J. F. Abiotic reduction of ketones with silanes catalysed by carbonic anhydrase through an enzymatic zinc hydride. *Nat. Chem.* **13**, 312–318 (2021).
21. Kiss, G., Celebi-Olcum, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed.* **52**, 5700–5725 (2013).
22. Hilvert, D. Design of protein catalysts. *Annu. Rev. Biochem.* **82**, 447–470 (2013).
23. Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817–1819 (2010).
24. Jeschek, M. et al. Directed evolution of artificial metalloenzymes for in vivo metathesis. *Nature* **537**, 661–665 (2016).
25. Zhao, J. et al. Genetic engineering of an artificial metalloenzyme for transfer hydrogenation of a self-immolative substrate in *Escherichia coli*'s periplasm. *J. Am. Chem. Soc.* **140**, 13171–13175 (2018).
26. Rebelein, J. G. & Ward, T. R. In vivo catalyzed new-to-nature reactions. *Curr. Opin. Biotechnol.* **53**, 106–114 (2018).
27. Hyster, T. K., Knorr, L., Ward, T. R. & Rovis, T. Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C–H activation. *Science* **338**, 500–503 (2012). **Demonstration that transition metal complexes embedded in protein hosts can work in synergy with amino acid side chains to accelerate a challenging C–H activation process**.
28. Bhagi-Damodaran, A. et al. Why copper is preferred over iron for oxygen activation and reduction in haem-copper oxidases. *Nat. Chem.* **9**, 257–263 (2017).
29. Yeung, N. et al. Rational design of a structural and functional nitric oxide reductase. *Nature* **462**, 1079–1082 (2009).
30. Mirts, E. N., Petrik, I. D., Hosseinzadeh, P., Nilges, M. J. & Lu, Y. A designed heme-[4Fe–4S] metalloenzyme catalyzes sulfite reduction like the native enzyme. *Science* **361**, 1098–1101 (2018). **This study shows how the introduction of new functional elements into metalloproteins can generate artificial enzymes for challenging chemical tranformations that have thus far eluded synthetic catalysts**.
31. Hill, R. B., Raleigh, D. P., Lombardi, A. & DeGrado, W. F. De novo design of helical bundles as models for understanding protein folding and function. *Acc. Chem. Res.* **33**, 745–754 (2000).
32. Koder, R. L. & Dutton, P. L. Intelligent design: the de novo engineering of proteins with specified functions. *Dalton Trans.* **25**, 3045–3051 (2006).
33. Faiella, M. et al. An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **5**, 882–884 (2009).
34. Smith, B. A. & Hecht, M. H. Novel proteins: from fold to function. *Curr. Opin. Chem. Biol.* **15**, 421–426 (2011).
35. Zastrow, M. L., Peacock, A. F., Stuckey, J. A. & Pecoraro, V. L. Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **4**, 118–123 (2011).
36. Stenner, R., Steventon, J. W., Seddon, A. & Anderson, J. L. R. A de novo peroxidase is also a promiscuous yet stereoselective carbene transferase. *Proc. Natl Acad. Sci. USA* **117**, 1419–1428 (2020).
37. Chino, M. et al. A de novo heterodimeric Due Ferri protein minimizes the release of reactive intermediates in dioxygen-dependent oxidation. *Angew. Chem. Int. Ed.* **56**, 15580–15583 (2017).

38. Lombardi, A., Pirro, F., Maglio, O., Chino, M. & DeGrado, W. F. De novo design of four-helix bundle metalloproteins: one scaffold, diverse reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019).
39. Reig, J. A. et al. Alteration of the oxygen-dependent reactivity of de novo Due Ferri proteins. *Nat. Chem.* **4**, 900–906 (2012).
Demonstration that the catalytic function of de novo Due Ferri proteins can be altered through rational reprogramming of the metal coordination environment.
40. Salgado, E. N., Faraone-Mennella, J. & Tezcan, F. A. Controlling protein–protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.* **129**, 13374–13375 (2007).
41. Der, B. S. et al. Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* **134**, 375–385 (2012).
42. Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* **51**, 3933–3940 (2012).
43. Studer, S. et al. Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, 1285–1288 (2018).
This study uses a combination of design and evolution to transform a designed zinc-binding peptide into a globular metalloenzyme that accelerates ester hydrolysis with high efficiency.
44. Basler, S. et al. Efficient Lewis acid catalysis of an abiological reaction in a de novo protein scaffold. *Nat. Chem.* **13**, 231–235 (2021).
In this study, a de novo metalloenzyme is engineered to accelerate an abiological hetero-Diels–Alder reaction with high specificity and a catalytic proficiency that exceeds all previously characterized Diels–Alderases.
45. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
46. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
47. Seyedsayamdost, M. R., Xie, J., Chan, C. T. Y., Schultz, P. G. & Stubbe, J. Site-specific insertion of 3-aminotyrosine into subunit α2 of *E. coli* ribonucleotide reductase: direct evidence for involvement of Y_{730} and Y_{731} in radical propagation. *J. Am. Chem. Soc.* **129**, 15060–15071 (2007).
48. Faraldos, J. A. et al. Probing eudesmane cation–π interactions in catalysis by aristolochene synthase with non-canonical amino acids. *J. Am. Chem. Soc.* **133**, 13906–13909 (2011).
49. Wu, Y. & Boxer, S. G. A critical test of the electrostatic contribution to catalysis with noncanonical amino acids in ketosteroid isomerase. *J. Am. Chem. Soc.* **138**, 11890–11895 (2016).
50. Ortmayer, M. et al. Rewiring the 'push–pull' catalytic machinery of a heme enzyme using an expanded genetic code. *ACS Catal.* **10**, 2735–2746 (2020).
51. Ortmayer, M. et al. A noncanonical tryptophan analogue reveals an active site hydrogen bond controlling ferryl reactivity in a heme peroxidase. *JACS Au* **1**, 913–918 (2021).
52. Li, J. C., Liu, T., Wang, Y., Mehta, A. P. & Schultz, P. G. Enhancing protein stability with genetically encoded noncanonical amino acids. *J. Am. Chem. Soc.* **140**, 15997–16000 (2018).
53. Green, A. P., Hayashi, T., Mittl, P. R. & Hilvert, D. A chemically programmed proximal ligand enhances the catalytic properties of a heme enzyme. *J. Am. Chem. Soc.* **138**, 11344–11352 (2016).
54. Zhao, J., Burke, A. J. & Green, A. P. Enzymes with noncanonical amino acids. *Curr. Opin. Chem. Biol.* **55**, 136–144 (2020).
55. Burke, A. J. et al. Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **570**, 219–223 (2019).
Demonstrates how introducing noncanonical amino acids can expand the chemical reactivity and catalytic mechanisms accessible with designed enzymes.
56. Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **98**, 14274–14279 (2001).
57. Richter, F. et al. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).
58. Rajagopalan, S. et al. Design of activated serine-containing catalytic triads with atomic-level accuracy. *Nat. Chem. Biol.* **10**, 386–391 (2014).
59. Moroz, Y. S. et al. New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. *J. Am. Chem. Soc.* **137**, 14905–14911 (2015).
60. Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L. & Woolfson, D. N. Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **8**, 837–844 (2016).
61. Drienovska, I., Mayer, C., Dulson, C. & Roelfes, G. A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue. *Nat. Chem.* **10**, 946–952 (2018).
Demonstrates that the introduction of noncanonical amino acids can open up new modes of reactivity within proteins.
62. Mayer, C., Dulson, C., Reddem, E., Thunnissen, A. W. H. & Roelfes, G. Directed evolution of a designer enzyme featuring an unnatural catalytic amino acid. *Angew. Chem. Int. Ed.* **58**, 2083–2087 (2019).
63. Tramontano, A., Janda, K. D. & Lerner, R. A. Catalytic antibodies. *Science* **234**, 1566–1570 (1986).
64. Wagner, J., Lerner, R. A. & Barbas, C. F. III. Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. *Science* **270**, 1797–1800 (1995).
65. Gouverneur, V. E. et al. Control of the exo and endo pathways of the Diels–Alder reaction by antibody catalysis. *Science* **262**, 204–208 (1993).
66. Wentworth, P. Jr. et al. Antibody catalysis of the oxidation of water. *Science* **293**, 1806–1811 (2001).
67. Hsieh, L. C., Yonkovich, S., Kochersperger, L. & Schultz, P. G. Controlling chemical reactivity with antibodies. *Science* **260**, 337–339 (1993).
68. Hilvert, D. Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **69**, 751–793 (2000).
69. Rothlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
70. Privett, H. K. et al. Iterative approach to computational enzyme design. *Proc. Natl Acad. Sci. USA* **109**, 3790–3795 (2012).
71. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).
72. Siegel, J. B. et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels–Alder reaction. *Science* **329**, 309–313 (2010).
Computational design and experimental characterization of enzymes catalysing a bimolecular Diels–Alder reaction, an important carbon–carbon bond-forming process.
73. Blomberg, R. et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418–421 (2013).
Demonstrates that artificial enzymes can be evolved to accelerate elementary chemical reactions with efficiencies comparable to natural enzymes.
74. Preiswerk, N. et al. Impact of scaffold rigidity on the design and evolution of an artificial Diels–Alderase. *Proc. Natl Acad. Sci. USA* **111**, 8013–8018 (2014).
75. Obexer, R. et al. Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).
Ultrahigh-throughput screening facilitates the development of artificial enzymes with efficiencies comparable to natural systems.
76. Crawshaw, R. et al. Engineering an efficient and enantioselective enzyme for the Morita–Baylis–Hillman reaction. *Nat. Chem.* **14**, 313–320 (2022).
A demonstration that laboratory evolution of designed enzymes can deliver sophisticated active sites to accelerate complex nonbiological transformations.
77. Otten, R. et al. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **370**, 1442–1446 (2020).
78. Broom, A. et al. Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nat. Commun.* **11**, 4808 (2020).
79. Althoff, E. A. et al. Robust design and optimization of retroaldol enzymes. *Protein Sci.* **21**, 717–726 (2012).
80. Giger, L. et al. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494–498 (2013).
81. Eiben, C. B. et al. Increased Diels–Alderase activity through backbone remodeling guided by Foldit players. *Nat. Biotechnol.* **30**, 190–192 (2012).
82. Bjelic, S. et al. Computational design of enone-binding proteins with catalytic activity for the Morita–Baylis–Hillman reaction. *ACS Chem. Biol.* **8**, 749–757 (2013).
83. Kiss, G., Rothlisberger, D., Baker, D. & Houk, K. N. Evaluation and ranking of enzyme designs. *Protein Sci.* **19**, 1760–1773 (2010).
84. Frushicheva, M. P., Cao, J., Chu, Z. T. & Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial Kemp eliminase. *Proc. Natl Acad. Sci. USA* **107**, 16869–16874 (2010).
85. Bunzel, H. A. et al. Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nat. Chem.* **13**, 1017–1022 (2021).
86. Weitzner, B. D., Kipnis, Y., Daniel, A. G., Hilvert, D. & Baker, D. A computational method for design of connected catalytic networks in proteins. *Protein Sci.* **28**, 2036–2041 (2019).
87. Davey, J. A., Damry, A. M., Goto, N. K. & Chica, R. A. Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.* **13**, 1280–1285 (2017).
88. Pan, X. et al. Expanding the space of protein geometries by computational design of de novo fold families. *Science* **369**, 1132–1136 (2021).
89. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
90. Dou, J. et al. De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).
91. Wei, K. Y. et al. Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. *Proc. Natl Acad. Sci. USA* **117**, 7208–7215 (2020).
92. Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
Development of AlphaFold, a deep learning algorithm for accurate prediction of protein structure from primary sequence.
93. Hiranuma, N. et al. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nat. Commun.* **12**, 1340 (2021).
94. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
Development of RoseTTAFold, a freely available deep learning programme for fast and accurate prediction of protein structure.
95. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
96. Anishchenko, I. et al. De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
97. Mazurenko, S., Prokop, Z. & Damborsky, J. Machine learning in enzyme engineering. *ACS Catal.* **10**, 1210–1223 (2020).
98. Ma, E. J. et al. Machine-directed evolution of an imine reductase for activity and stereoselectivity. *ACS Catal.* **11**, 12433–12445 (2021).
99. Bedbrook, C. N., Yang, K. K., Rice, A. J., Gradinaru, V. & Arnold, F. A. Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization. *PLoS Comput. Biol.* **13**, e1005786 (2017).
100. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl Acad. Sci. USA* **116**, 8852–8858 (2019).
101. Tischer, D. et al. Design of proteins presenting discontinuous functional sites using deep learning. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.29.402743.
102. Russ, W. P. et al. An evolution-based model for designing chorismate mutase enzymes. *Science* **369**, 440–445 (2020).
103. Wang, J. et al. Deep learning methods for designing proteins scaffolding functional sites. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.10.468128.
104. Hayashi, T. et al. Capture and characterization of a reactive haem–carbenoid complex in an artificial metalloenzyme. *Nat. Catal.* **1**, 578–584 (2018).

105. Carminati, D. M. & Fasan, R. Stereoselective cyclopropanation of electron-deficient olefins with a cofactor redesigned carbene transferase featuring radical reactivity. *ACS Catal.* **9**, 9683–9687 (2019).
106. Erkkila, A., Majander, I. & Pihko, P. M. Iminium catalysis. *Chem. Rev.* **107**, 5416–5470 (2007).
107. Mukherjee, S., Yang, J. W., Hoffmann, S. & List, B. Asymmetric enamine catalysis. *Chem. Rev.* **107**, 5471–5569 (2007).
108. Doyle, A. G. & Jacobsen, E. N. Small-molecule H-bond donors in asymmetric catalysis. *Chem. Rev.* **107**, 5713–5743 (2007).
109. Wurz, R. P. Chiral dialkylaminopyridine catalysts in asymmetric synthesis. *Chem. Rev.* **107**, 5570–5595 (2007).
110. Beeson, T. D., Mastracchio, A., Hong, J. B., Ashton, K. & Macmillan, D. W. Enantioselective organocatalysis using SOMO activation. *Science* **316**, 582–585 (2007).
111. St-Jacques, A. D., Eyahpaise, M.-È. C. & Chica, R. A. Computational design of multisubstrate enzyme specificity. *ACS Catal.* **9**, 5480–5485 (2019).
112. Davey, J. A. & Chica, R. A. Multistate approaches in computational protein design. *Protein Sci.* **21**, 1241–1252 (2012).

Acknowledgements We thank the European Research Council (ERC Starter Grant, no. 757991 to A.P.G.), the Biotechnology and Biological Sciences Research Council (David Phillips Fellowship BB/M027023/1 to A.P.G.), UK Research and Innovation (Future Leader Fellowship MR/T041722/1 to S.L.L.), the Swiss National Science Foundation (D.H.), ETH Zürich (D.H.) and the Howard Hughes Medical Institute (D.B.) for generous support.

Author contributions All authors discussed the content of the manuscript, including selection of key studies highlighted and opportunities for future innovations. R.C., S.B. and C.L. prepared the figures. S.L.L., D.B., D.H. and A.P.G. wrote the manuscript text.

Competing interests The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to David Baker, Donald Hilvert or Anthony P. Green.
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.