Expanding the enzyme universe with genetically encoded unnatural amino acids

Drienovska, Ivana; Roelfes, Gerard

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

[Link to publication in University of Groningen/UMCG research database](#)

Expanding the enzyme universe with genetically encoded unnatural amino acids

Ivana Drienovská[a] and Gerard Roelfes[b]

[a]Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria

[b]Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[*]*email:* i.drienovska@tugraz.at; j.g.roelfes@rug.nl

Abstract

The emergence of robust methods to expand the genetic code allows incorporation of non-canonical amino acids into the polypeptide chain of proteins, thus making it possible to introduce unnatural chemical functionalities in enzymes. In this perspective, we show how this powerful methodology is used to create enzymes with improved and novel, even new-to-nature, catalytic activities. We provide an overview of the current state of the start and discuss the potential benefits of developing and using enzymes with genetically encoded non-canonical amino acids compared to enzymes containing canonical amino acids only.

Enzymes are immensely powerful catalysts that have evolved over millions of years to catalyse the reactions that sustain life on earth. Mankind has harnessed the power of these biological catalysts for its own applications and nowadays biocatalysis is considered a key element of the transition to a more sustainable chemical synthesis.[1] Enzymes can readily be adapted, through site-directed mutagenesis and directed evolution, to accept non-natural substrates and, even, to

catalyse reactions that do not occur in nature.[2,3] Yet, the structural and functional diversity of the 20 canonical amino acids (cAAs) of which enzymes are made is relatively small and this supposedly limits the catalytic repertoire of enzymes, at least compared to the vast number of chemical reactions available to the synthetic chemist. Nature itself has addressed this by employing cofactors and/or post-translational modifications of canonical amino acids to introduce novel functionalities and, with this, novel reactivities.[4] These approaches are a source of inspiration for scientists looking to create novel enzymes.

In the past two decades, we have witnessed the emergence of robust expanded genetic code methods, that allow reprograming of the genetic code to enable incorporation of unnatural or non-canonical amino acids (ncAAs) into proteins.[5–7] This powerful methodology is now also increasingly applied to biocatalysis.[8] The resulting xenobiotic enzymes containing abiological functionalities in their peptide side chains are envisioned to give access to new biocatalysts for the catalysis of, in principle, any reaction, including those that are new-to-nature.

As with any revolutionary technique that holds such great promise, it is good to take a step back and critically evaluate the benefits of using this methodology in the field of biocatalysis. The size of natural protein sequence space is already far beyond our comprehension and even with the emergence of powerful directed evolution methods, only minuscule parts of it have been explored.[9,10] Hence, is it at all necessary to incorporate ncAAs into proteins to achieve

improved and new enzymatic catalysis? Or should we just invest more time and resources into deeper explorations of the existing sequence space.[11–13]

In this perspective, we show how genetically encoded non-canonical amino acids have been used to improve the catalytic activity and selectivity of enzymes and can, possibly, give rise to the catalysis of reactions that have no equivalents in nature. Finally, the benefits of using enzymes with genetically encoded non-canonical amino acids compared to using canonical enzymes will be discussed.

Expanded genetic code methods.

With the rare exception of pyrrolysine[14] and selenocysteine[15], the genetic code of all organisms encodes the same 20 amino acids.[16] The amino acid sequence in polypeptides is specified by the nucleotide sequence of mRNA. The four nucleotides can form 64 three-letter codons and 61 of those code for the 20 cAAs and the three remaining codons, termed termination codons, signal a stop in translation. Thus, in order to introduce novel ncAAs *in-vivo*, additional codons or reprogramming of existing ones has to be achieved. A variety of methods have been developed, with Selective Pressure Incorporation (SPI) and Stop Codon Suppression (SCS), being the most frequently applied (**Fig. 1**).[17,18] SPI relies on the promiscuity of the natural translation machinery as some transfer RNAs (tRNAs) can be loaded with ncAAs structurally similar to their cognate cAAs, albeit at lower efficiency.[19] If the organism itself cannot produce the cAA (it is an auxotroph), it relies on exogenously supplied ones. In the event that these are not present, the organism may use the analogue. The advantage of this methodology is that it is technologically rather simple, provided that the auxotrophic bacterial strain is available. On the other hand, the number of ncAAs that can be introduced using this technology is limited to close structural analogues of canonical amino acids. Additionally, a consequence of this approach is that only global replacement is

achievable, *i.e.* it is all or nothing (**Fig. 1a**). This often leads to substantial loss of activity and/or stability due to perturbations in the folded structure. A powerful alternative is the SCS method[20] which uses one of the three stop codons (UAG, UAA, UGA) and reprograms it to incorporate the desired ncAAs. Since the amber codon (UAG) is used the least in *E. coli* (7-8%), it has been the preferred choice for recoding. SCS uses an orthogonal translation system, that is, a translation system that has no cross-talk with the endogeneous translation system, comprising an orthogonal aminoacyl-tRNA synthetase (aaRS) in combination with an orthogonal tRNA specifically engineered for a given ncAA (**Fig. 1b**). For example, the tyrosyl-tRNA synthetase TyrRS/tRNA$_{CUA}$ pair from *Methanocaldococcus jannaschi* and the pyrrolysyl-tRNA synthetase PylRS/tRNA$_{CUA}$ pairs from *Methanosarcina mazei* and *Methanosarcina barkeri* are widely used.[5,14,20–22] Up to date, more than 200 ncAAs have been incorporated into proteins using this method.[18]

Applications of genetically encoded ncAAs in biocatalysis

Below, an overview of the current applications of genetically encoded ncAAs in biocatalysis is given, grouped by the effect that these ncAAs have on catalytic activity and selectivity. Non-canonical amino acids are also used to improve chemical or thermal stability of proteins and to introduce bio-orthogonal handles for conjugation of transition metal catalysts. However, this will not be discussed here and we refer the reader to other recent reports and reviews.[8,23–27]

Improving the activity of natural enzymes. Complementarity of the active site structure to the transition state of the catalysed reaction is key to the large rate accelerations achieved by enzymes.[28] Hence, as has been shown often in directed evolution studies,[29] any changes in the global protein structure, even remote from the active site, can have an effect on catalysis if this causes even small changes in the structure of the active site. With this in mind, ncAAs

can be used to fine-tune the steric and electronic interactions in the active site in ways that would not be readily achievable with cAAs.

Indeed, a number of groups have shown that global replacement of cAAs with ncAAs using SPI can improve enzyme activity and selectivity. As early as 2001, all four phenylalanine residues of the restriction endonuclease PvuII were related with *o*-, *m*-, and *p*-fluorophenylalanines. The *o*- and *p*-fluorophenylalanine variants exhibited the same or lower specific activity, while the *m*-fluorophenylalanine variant exhibited a twofold higher activity compared to the wildtype (WT) enzyme. Notably, none of the phenylalanine residues belonged to the catalytic or DNA-binding site.[30]

Global replacement of tyrosine for *m*-fluorotyrosine within a ω-transaminase enzyme led to enhanced stability and an increased activity in the synthesis of (*S*)-1-phenylethylamine.[31] Replacement of methionine with its analogue norleucine, has been used in cytochrome P450 peroxygenase[32] from *Bacillus megaterium* and lipase from *Thermoanaerobacter thermohydrosulfiricus*,[33] leading to variants that showed modest two to ten-fold increases in activity.

With SCS more surgical changes can be made, which in principle makes it possible to introduce specific interactions in the active site not available from cAAs. For example, a phenylalanine residue within the active site of a nitroreductase was substituted by several cAAs and ncAAs. *p*-Nitrophenylalanine showed a more than twofold increase of catalytic efficiency (k_{cat}/K_M) compared to other substitutions.[34]

Enhanced enantioselectivity was reported for a diketoreductase (DKR), an industrially relevant enzyme that is capable of stereoselective reduction of ketones to chiral alcohols.[35] A gate-keeper tryptophan residue at position 222 of the enzyme that determines the orientation of a substrate was identified. Incorporation of ncAAs with bulkier side-chains than tryptophan, which is the bulkiest AA in the canonical pool, such as *O-tert*-butyltyrosine and

p-phenylphenylalanine at this position enhanced the enantioselectivity. Interestingly, smaller side-chains caused the inversion of the enantioselectivity towards the *S* enantiomer.

Fasan and colleagues reported that the replacement of the active site Leu75 with *p*-aminophenylalanine in a cytochrome P450 resulted in a significant increase in the oxidation of (*S*)-ibuprofen.[36] Furthermore, placing ncAAs in the position of Ala82 led to dramatically increased yield in the oxidation of (+)-nootkatone. Finally, with *p*-acetylphenylalanine at position Ala78, the completely stereoselective oxidation of (+)-nootkatone to (9*R*)-hydroxynootkatone was observed, which was not possible with the native enzyme.

The squalene hopene cyclase (SHC) from *Alicyclobacillus acidocaldarius* catalyses the polycyclization of linear molecules of squalene to polycyclic triterpene. Active site residues Phe365 and Phe605 have been proposed to stabilize the cationic intermediates formed during the polycyclization cascade and the activity of the enzyme relates to the strength of the cation–π interactions in its active site. *o*-methyltyrosine and mono-, di-, and trifluorophenylalanines were incorporated at these two positions.[37] Fluorophenylalanines reduce cation−π binding interactions and, as a consequence, the reaction became slower. In contrast, *o*-methyltyrosine enhances these interactions and, at temperatures below 40 °C, a more active enzyme was obtained.

Ugwumba *et al.* showed that replacement of Tyr309 with L-(7-hydroxycoumarin-4-yl)ethylglycine (Hco) in a bacterial phosphotriesterase (arPTE) increased its turnover rate towards the hydrolysis of the agrochemical paraoxon 8-fold. This was shown to be related to increases in the rate of the rate-determining product release step as a result of increased charge repulsion of Hco, which has a lower p*K*a of the side chain than Tyr, with the products. Previous attempts to optimize the activity of this enzyme using canonical amino acids had not resulted in improved variants.[38]

While in the previous examples the ncAAs were introduced at predetermined positions, ncAAs have also been employed in directed evolution enzymes. Xiao *et al.* reported a library of mutant enzymes with single substitutions, introducing 10 structurally distinct ncAAs randomly throughout TEM-1 β-lactamase.[39] A growth screen on the β-lactam antibiotic cephalexin led to discovery of a unique *p*-acrylamidophenylalanine mutation at valine-216 that showed an eightfold increase in catalytic efficiency. In comparison, none of the natural amino acids at this position afforded similar enhancement.

Ancillary functions. Enzymes often contain amino acids that have ancillary functions supporting catalysis, including mediation of electron transfer and electron and proton shuttling. Tyrosine (Tyr), for example, has a unique ability to donate both an electron and a proton in enzymatic reactions. ncAAs can be used to modulate these properties and have proven to be powerful biophysical probes for deciphering the role of such AAs in the catalytic mechanism. In landmark studies, Stubbe and co-workers used a variety of non-canonical tyrosine analogues, such as 3,4-dihydroxyphenylalanine, *m*-aminotyrosine, *m*-nitrotyrosine or fluorotyrosines, spanning a range of oxidation potentials, to elucidate the electron transfer pathway in ribonucleotide reductases.[40–46]

To investigate the ancillary role of Tyr in a cytochrome *c* oxidase, a functional model of this enzyme was prepared by introducing two histidines (His29 and His43) and one Tyr (Tyr33) into the distal pocket of sperm whale myoglobin (Mb). Next, different variants of engineered myoglobin (Mb) were created by replacing the introduced Tyr33 with three halogenated tyrosine analogues: 3-chlorotyrosine (ClTyr), 3,5-difluorotyrosine (F2Tyr), and 2,3,5-difluorotyrosine (F3Tyr), which all have a lower pKa value than tyrosine. 3-Methoxytyrosine (MeOTyr) has a similar pKa to Tyr, but a significantly lower reduction potential. The results demonstrated that the catalytic activity of the model oxidase is strongly dependent on the pKa value of Tyr33 as well as its reduction potential; thus confirming the important role played by

Tyr in the oxidase reaction.[47,48] In addition to the aforementioned, 3-methyltyrosine (MtTyr) and MeOTyr significantly contributed to the elucidation of Tyr417's function in a nonheme iron enzyme (OvoA) that catalyses a four electron oxidative coupling between L-His and L-Cys.[49,50]

Nature itself also modulates the properties of tyrosine by posttranslational modifications, to optimize its properties for catalysis. The textbook example is the posttranslational formation of a tyrosine-cysteine cross-link in galactose oxidase, which is key to its activity in the oxidation of a broad range of alcohols to their corresponding aldehydes, concomitant with reduction of O_2 to H_2O_2. It is, generally, rather difficult to generate these modifications in other enzymes. However, ncAAs can be utilized to mimic this type of modification to study the mechanism or exploit the potential of these modifications in catalysis (**Fig. 2a**). Liu *et al.*, for example, used SCS methodology to incorporate 2-amino-3-(4-hydroxy-3-(1H-imidazol-1-yl)phenyl)propanoic acid (ImiTyr) into sperm whale myoglobin, which mimicked the Tyr-His, a cross-link essential for the function of heme-copper oxidase (HCO).[51] The variant Mb_Phe33imiTyr, a functional model of HCO, exhibited an oxygen consumption rate of 2.2 min^{-1} and showed a three-fold faster reduction of O_2 compared to Mb_Phe33Tyr, which had tyrosine at the same position but lacked the Tyr-His link. Furthermore, this improved variant also released fewer reactive oxygen species as side products. As a mimic of tyrosine-cysteine cross-link (which is present throughout diverse metalloenzymes)[52], Wang and co-workers introduced 3-methylthiotyrosine (MtSTyr) at the position of 33 of myoglobin, which resulted in an enzyme that catalysed selective hydroxylamine reduction to ammonia at a rate of 800 min^{-1}.[53]

In addition to mimicking natural post-translational modifications, ncAA-containing groups with ancillary functions that have no equivalent in nature can be introduced. Wang and co-workers exchanged a Tyr66 in superfolder yellow fluorescent protein with a ncAA with

benzophenone in its side-chain (BpA).[54]   Benzophenone is a commonly used photosensitizer in organic photocatalysis.[55] In the wild-type protein, Tyr66 participates in the generation of a highly fluorescent p-hydroxybenzylidene-5-imidizolinone species through the autocatalytic conversion of the tripeptide Gly65-Tyr66-Gly67. It was envisioned, that BpA66 can undergo the same conversion, resulting in (E)−4-(4-benzoylbenzylidene)-1,2-dimethyl-1H-imidazole-5(4H)-one, a chromophore with a significantly increased lifetime of triplet excited state, which, upon its reaction with sacrificial reductant, turned into a super-reducing radical with extremely low redox potential of -1.5 V. By combining this artificial fluorescent protein with organic nickel-terpyridine cofactors, light energy was harvested to reduce CO_2 with more than 100 total turnover numbers and 2.6% quantum yield[54], which is higher than most reported CO_2 photoreduction catalysts (**Fig. 2b**).

The SCS methodology also allows the incorporation of abiological moieties that can be used to control the catalytic activity. This allows for non-invasive spatial and temporal control of protein activity in complex cellular systems. One of the most commonly used strategies relies on site-specific incorporation of a caged ncAA, followed by the light-induced removal of the protecting groups, which activates the enzyme.[56] Photocaged cysteines, lysines, tyrosines or serines have been widely used in chemical biology to control different cell or enzyme functions (**Fig. 2c**).[57–61]

Uncaging is an irreversible process. Reversible control over enzyme activity might be achieved by incorporation of photoisomerizable ncAAs, for example containing azobenzene side chains.[62,63] While azobenzene has been shown to allow for control over protein function,[64,65] no control over catalysis has been demonstrated yet with genetically encoded azobenzenes.

Novel metal-binding sites. One approach followed by nature to achieve catalytic activity that is not realizable through the 20 canonical amino acids is the recruitment of cofactors such as metal ions for catalysis. The catalytic activity of metal ions is determined to a large extent by its ligands; in those cases where the ligation by natural AAs does result in the desired catalytic properties, metal complexes such as, for example, heme, are recruited. Hence, metal-binding ncAAs provide an attractive approach to bind and modulate the properties of catalytic metal ions, resulting in artificial metalloenzymes (**Fig. 3a**).[66,67]

In an early example, the Schultz group installed (2,2′-bipyridin-5-yl)alanine (BpyA,) into the *E. coli* catabolite activator protein, which upon binding of copper or iron cations displayed nuclease activity.[68]

Drienovská *et al.* incorporated BpyA, at various positions into the structure of the Lactococcal multidrug resistance Regulator (LmrR). This homodimeric protein with no natural catalytic function has a promiscuous hydrophobic pocket that can bind a wide range of hydrophobic organic substrates, which makes it an attractive protein scaffold for enzyme design.[69] The resulting artificial metalloenzymes catalysed the Friedel–Crafts alkylation of indoles with α,β-unsaturated 2-acylimidazoles, with the best results obtained by incorporation at position 89 (LmrR_M89BpyA): up to 80% *ee* was obtained. This could be improved to 83% *ee* and 94% conversion by introducing additional active site mutations (**Fig. 3a**).[70] Other multidrug resistance regulator proteins, namely QacR, CgmR and RamR, were used as scaffolds as well.[71] Similar to LmrR, these proteins contain large hydrophobic pockets and show promiscuous binding of hydrophobic substrates. Four positions in the hydrophobic pocket of each protein were selected for introduction of BpyA. Most of the variants were able to achieve low to moderate enantioselectivity. However, excellent results were attained with the variant QacR_Y123BpyA, giving rise to good conversion (82%) and 94% *ee*. Interestingly, the opposite enantiomer compared to the LmrR variants was preferred.

Therefore, both enantiomers of the Friedel-Crafts product could be accessed by judicious choice of the biomolecular scaffold (**Fig. 3b**).

Next, the more challenging reaction of enantioselective hydration of prochiral enones was achieved using the LmrR-based metalloenzyme. The variant LmrRM89_BpyA was further redesigned *in silico* using a combination of cluster model calculations (QM), protein–ligand docking and molecular dynamics simulations.[72] The introduction of a glutamate residue was proposed within the required distance of the β position of the enone to function as a general base and form a pre-reactive conformation with a water nucleophile. The selected mutant V15E gave rise to 3-fold increase in catalytic efficiency and an increased enantioselectivity of the β-hydroxyketone product. Finally, LmrR_M89BpyA containing a variety of different divalent first row transition metal ions was found capable of binding and stabilizing the radical semiquinone, albeit that no catalysis was observed.[73]

The Baker group applied Rosetta computational protein design to find/design a suitable protein scaffold for the incorporation of ncAAs with metal-binding moieties.[74] A variant containing a buried active site with an octahedral coordination geometry supported by two water molecules and other scaffold residues was found to bind various divalent metal cations, such a Co^{2+}, Zn^{2+}, Ni^{2+} or Fe^{2+} with micro to picomolar affinities. However, no catalytic activity was reported. Another reported example of a computational Rosetta design of a threefold symmetric, self-assembling protein homotrimer contained a highly stable BpyA-mediated metal complex.[75]

Hilvert and co-workers reported a mutated ascorbate peroxidase enzyme, where heme-coordinating His163 was specifically replaced by the ncAA N_δ-methylhistidine (NMH). This new variant demonstrated a higher catalytic efficiency and longer overall catalytic stability compared to the wildtype.[76]

Similarly, a myoglobin with the proximal histidine ligand of heme replaced with NMH, showed a modest 3.7-fold increase in the catalytic efficiency during the guaiacol oxidation (**Fig. 3c**). A comparison of the variant's crystal structure to that of the wildtype protein revealed significant structural changes, including a rotation of the imidazole plane of the proximal ligand and alterations to hydrogen bonding between the serine residue and heme. Further optimization through a combination of rational design and several rounds of directed evolution led to an improved variant with 1140-fold improvement in catalytic efficiency compared to the wildtype, which is one order of magnitude higher than a natural peroxidase enzyme (**Fig. 3c**).[77] Additionally, an engineered variant was shown to allow cyclopropanation of styrene with ethyl diazoacetate. The reaction proceeds in the absence of a reductant under aerobic conditions. The X-ray crystal structure of an unusual bridging metallocarbene species, Fe(III)–C(carbene)–N(pyrrole), was reported. This species itself is inert towards further reaction with styrene, but is in rapid equilibrium with the end-on carbene species that is responsible for the cyclopropanation of the substrate.[78]

In addition, Fasan and co-workers substituted the heme-coordinating histidine residue with 3-(3′-pyridyl)-alanine, *p*-aminophenylalanine (pAF), and β-(3-thienyl)-alanine and investigated the impact on carbene and nitrene transfer reactions (**Fig. 3c**).[79] All three ncAAs were well tolerated in myoglobin, albeit that a slight decrease in selectivity and conversion in the cyclopropanation of styrene with ethyl α-diazoacetate was observed. No significant beneficial effects were observed on the C-H amination reaction with 2,4,6-triisopropylbenzenesulfonyl azide, albeit that the variant with pAF was shown to be the most efficient biocatalyst under aerobic conditions.

NcAAs as catalytic residues. An alternative approach used by nature to expand the catalytic scope of enzymes involves the posttranslational modification of amino acid side chains to

introduce non-canonical residues with novel functionalities and activities.[4] Well-known examples include formyl glycine (sulfatases)[80] and 4-methylideneimidazole-5-one (MIO) (lyases).[81] Introducing these non-canonical residues selectively at the desired position is difficult to emulate, as this generally requires a defined protein sequence, limiting the possibility to adapt this. However, ncAAs can be used to introduce different unnatural side chain reactive groups into proteins by SCS (**Fig. 4a**).

The first example of a new-to-nature reaction catalysed by a designer enzyme containing a genetically encoded ncAA as a catalytic residue was reported by Roelfes and co-workers, who introduced *p*-aminophenylalanine (pAF) to LmrR. The catalytic activity of the aniline side chain in these designer enzymes was demonstrated in abiological hydrazine and oxime formation reactions. The best initial design with pAF at position 15 of LmrR, LmrR_V15pAF, boosted activity by a factor of approximately 560 in a model hydrazone formation in comparison to an aniline-catalysed reaction and by 11-fold compared to the isostructural LmrR_V15Y variant.[82] The catalysis was attributed to a combination of effective molarity effects resulting from binding of the substrates by LmrR and iminium ion activation of the benzaldehyde. The enzyme was further optimized by directed evolution using several rounds of site-saturation mutagenesis. This resulted in an enzyme variant containing a total of four additional mutations (LmrR_V15pAF_A11L_N19M_A92R_F93H) that displayed approximately 100-times higher turnover frequency (k_{cat}) and improved apparent catalytic efficiencies in comparison to the parent designer enzyme. Furthermore, this variant outperformed aniline in these reactions by more than four orders of magnitude (**Fig. 4b**).[83]

Green and co-workers reported a novel hydrolase that uses N_δ-methylhistidine as a catalytic nucleophile.[84] The NMH was incorporated into a designed protein scaffold BH32. This protein was previously designed using Rosetta for catalysing Morita–Baylis–Hillman

reactions and contains a histidine nucleophile (His23) built into the cap domain of haloacid dehalogenase from *Pyrococcus horikoshii*. This designer enzyme was evaluated in the hydrolysis of fluresceinyl esters. However, it was found that BH32 exhibited formation of the acyl-enzyme intermediate but was stable to further hydrolysis. The incorporation of NMH results in a more reactive acyl-imidazolium intermediate, which allows for subsequent ester hydrolysis. The modest enzymatic activity was optimized using combinatorial active-site-saturation mutagenesis and error-prone PCR, to identify 21 residues which were then individually randomized *via* saturation mutagenesis. DNA shuffling of beneficial mutations led to the identification of a variant containing six mutations that gave a 15-fold improvement in activity over the parent template and more than 9000-fold increased efficiency over free N_δ-methylhistidine in solution (**Fig. 4c**). The hydrolase was evaluated also in the hydrolysis of the chiral fluorescein 2-phenylpropanoate substrate. While both the catalytic activity and enantiodiscrimination were low, another directed evolution study was performed. This resulted in the identification of more active variant which showed an 8-fold higher activity towards the (*R*)-enantiomer compared to the (*S*)-enantiomer.

Summary and Outlook

The presented examples show that genetically encoded ncAAs already have a considerable impact on the field of biocatalysis. However, while ncAAs increase the possibilities and versatility of enzyme catalysis, they also further increase the number of variables in the already complex problem of enzyme design. Hence, just because we can, does not necessarily mean we should use ncAAs. Clearly, the decision to use ncAAs should be based on the expected benefit compared to using cAAs only. For example, when used in existing enzymes, ncAAs have proven their value as mechanistic probe because they allow making subtle changes that can give information about active site structure and interactions. But the value of

using ncAAs to improve activity and selectivity of existing enzymes is more ambiguous: most likely, the improvements achieved by incorporating ncAAs will have to be dramatic to justify the additional effort and, most likely, costs of using ncAAs.

In our opinion, the greatest benefit of using genetically encoded ncAAs in enzymes is that this can give rapid access to novel enzymes with new catalytic activities that have currently no equivalent in nature, which can then be optimized further using well-established biotechnological methods such as site-directed mutagenesis and/or directed evolution to achieve optimal catalysis and selectivity. Of course, it is not inconceivable that, yet undiscovered, canonical enzymes capable of catalysing similar novel reactions do exist. However, finding these enzymes in the vast sequence space is a formidable challenge, even with powerful methods such as directed evolution. It is not feasible to screen novel parts of sequence at random looking for new activity because of the numbers problem: an enzyme consisting of n canonical amino acids has 20^n possible sequences. Therefore, to have any chance of success, directed evolution requires a good starting point, that is, an enzyme that already has a basic level of the desired, or closely related, promiscuous catalytic activity.[3] ncAA containing enzymes can provide such a starting point, if the ncAA already has, or supports, the new catalytic activity and the protein in which it is incorporated is evolvable, *i.e.* it possesses the structural flexibility that allows it to be subjected to directed evolution.

To achieve these initial activities, we can take advantage of decades of research in homogeneous transition metal catalysis and organocatalysis, especially in water or aqueous conditions,[85–87] which provides plenty of inspiration for the selection of catalytic groups that could be the basis for novel enzymes and enzymatic catalysis. This illustrated rapid developments in the field of artificial metalloenzymes.[66] The challenge will be to translate the

structure of these catalysts, be it a ligand for a transition metal ion or an organocatalyst, into an ncAA that can be incorporated into a protein using SCS.

An additional attractive feature of genetically encoded ncAAs is that they are incorporated into the protein during biosynthesis, which increases the potential of using ncAA-containing enzymes in whole cell biocatalysis. This is attractive for many applications, including directed evolution and integration into existing biosynthetic pathways to create novel hybrid metabolic pathways.[88]

An important question is of course: is the expanded genetic code methodology, and in particular SCS, capable of delivering on this promise? Over the years, SCS has become more and more robust and is now readily accessible for non-experts. The number of ncAAs that can be incorporated using SCS is already large and steadily increasing, but there is clearly a demand for more orthogonal translation systems (OTS) that can incorporate ncAAs containing ligands for metal ions or organocatalytic side chains. A potential issue here is the moderate structural diversity of ncAAs that can be incorporated, which relates to the fact that the most popular orthogonal translation systems are derived from those of only two amino acids: tyrosine and pyrrolysine. Many of the ncAAs that are attractive for enzyme design would require significant expansion of the structural versatility of OTS systems. Also, the fact that many ncAAs are not commercially available can be inhibitory, as their synthesis is often not trivial and quite costly. In this regard, the development of novel biosynthetic pathways towards ncAAs, that can be integrated into the metabolism of host cells represents a promising approach as it would allow ncAAs *in-vivo* production as cheaply and readily as cAAs.[89,90]

Other important points to consider, especially with regard to directed evolution studies, are the protein expression yields that can be achieved: depending on the protein, the position of

ncAA incorporation and the OTS (suppression efficiency), the yield can sometimes be significantly decreased. Similarly, the fidelity of ncAA incorporation is important. While some OTS are highly specific for a particular ncAA, others are more promiscuous[44,91–93]. This can be advantageous for expanding the scope of ncAAs, but obviously a significant degree of misincorporation of, for example, Phe or Tyr will influence screening and/or selection.

Yet, while there are still many challenges to address, progress in this field is fast and it is realistic to expect that many of these issues will be solved in the near future and that the number of robust and selective OTS for incorporation of novel ncAAs will increase further. Other promising developments are the special release factor 1 knockout bacterial strains[94–98], which reduce the problem with unwanted termination during expression, and the methods for incorporation of multiple diverse ncAAs,[6] which raises the prospect of synergistic catalysis[99] of multiple abiological catalytic groups.

In conclusion, we firmly believe that enzymes containing genetically encoded unnatural amino acids will be an integral part of the future of biocatalysis, especially for reactions that have no equivalent in biology. In this regard, we consider it at least on par with other popular approaches to biocatalyst development, including the use of promiscuous natural enzymes[3], artificial enzymes with abiological cofactors[66] and computationally designed enzymes[100]. Most likely, successful design of novel enzymes will entail a synergistic combination of these approaches, as is illustrated by some early examples described above: for example, the use of ncAAs to create artificial metalloenzymes and the application of computational-design algorithms in combination with directed evolution efforts to improve the catalytic activities of non-canonical enzymes. With so many exciting developments, achieving the important goal of creating biocatalysts for any desired natural or new-to-nature transformation seems tantalizingly close.
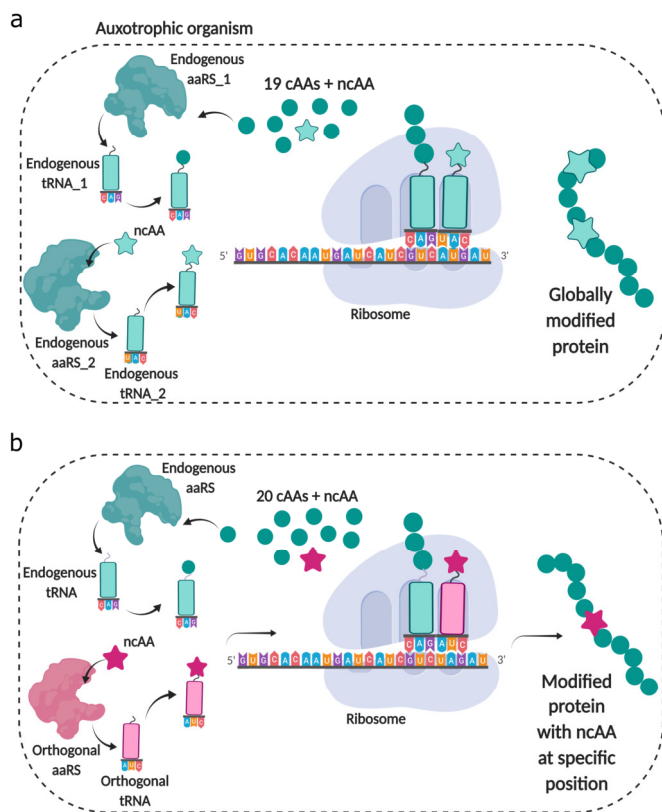
Fig. 1. Schematic representation of the two main strategies for *in-vivo* incorporation of non-canonical amino acids (ncAAs). (a) Selective Pressure Incorporation method, which uses the endogenous translation machinery of an auxotrophic organism. **(b)** Stop Codon Suppression methodology for the site-specific incorporation of ncAAs using an orthogonal aminoacyl-tRNA synthetase (aaRS)/transfer RNA (tRNA) pair. The orthogonal aaRS/tRNA pair is used to recode UAG codons in the mRNA of a gene of interest. Figure was created with Biorender.com
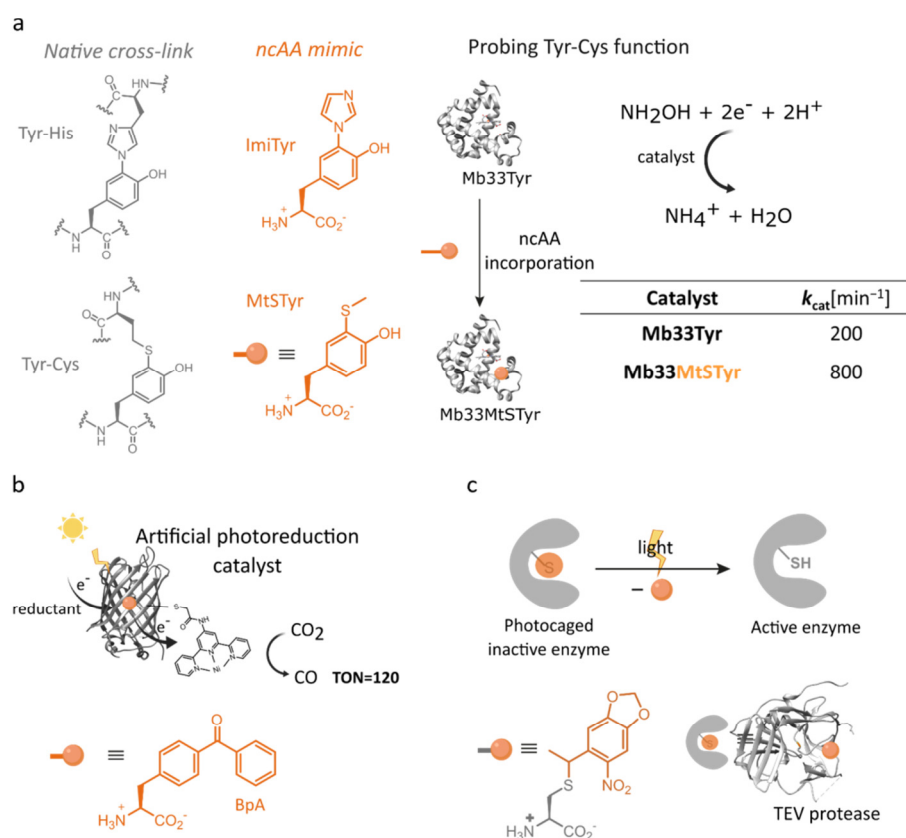
Fig. 2. Ancillary functions of non-canonical amino acids (ncAAs). (**a**) 2-amino-3-(4-hydroxy-3-(1H-imidazol-1-yl)phenyl) propanoic acid (ImiTyr) and 3-methylthiotyrosine (MtSTyr) function as mimics of Tyr-His and Tyr-Cys cross-link, respectively. When MtSTyr incorporated into engineered sperm whale myoglobin (Mb, PDB: 4FWX), Mb33MtSTyr can catalyse hydroxylamine reduction to ammonia with high efficiency.[51,53] (**b**) Artificial fluorescent protein, created by incorporation of benzophenone-alanine (BpA) into the superfolder yellow fluorescent protein (PDB: 3ED8), combined with organic nickel-terpyridine cofactor leads to an efficient CO_2 photoreduction catalyst. The turnover number (TON) of 120 reported for the best variant PSP2_95C_93Y97Y.[54] (**c**) ncAAs can control enzyme activity *via* incorporation of photocaged or photoswitchable amino acids. The most common photocaging strategy relies on *o*-nitrobenzyl group or derivatives to be installed on catalytic residues, which are cleaved off upon irradiation with (UV) light. Here, caging the catalytic cysteine of TEV protease (C151) leads to an inactive enzyme, which is activated by removing the caging group with light.[61]

20
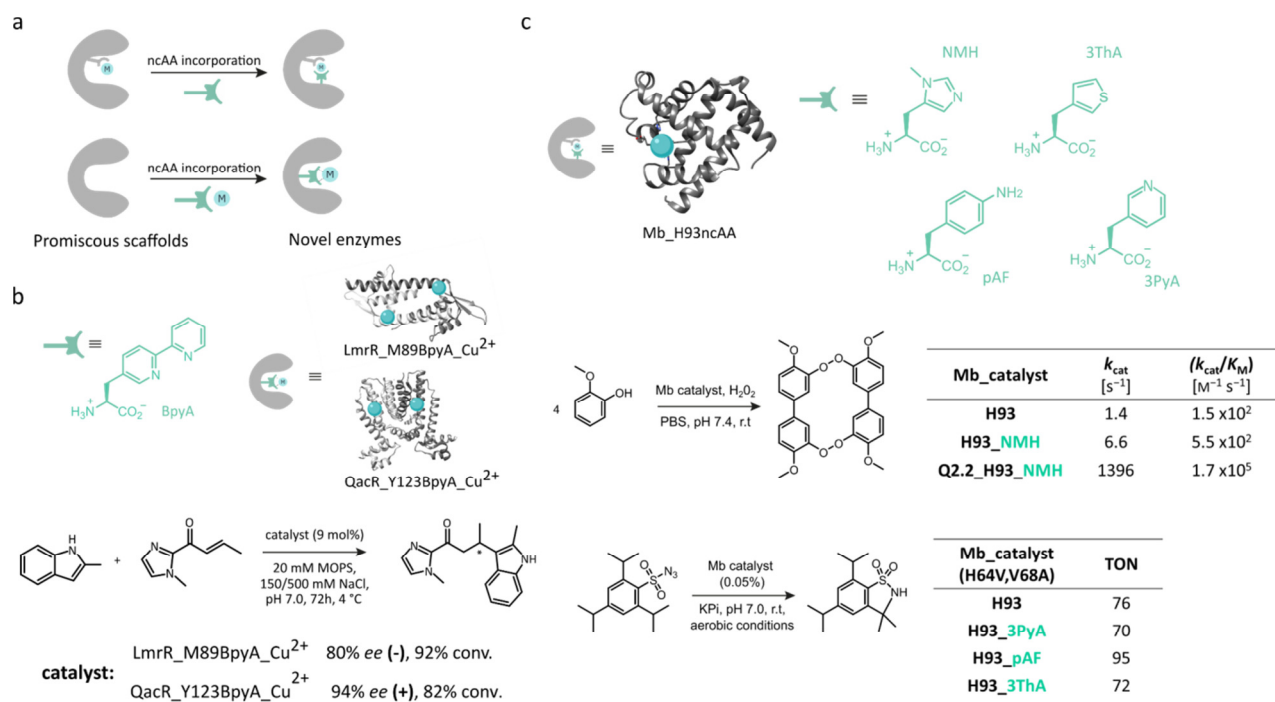
Fig. 3. Non-canonical amino acids (ncAAs) as metal-binding and metal-coordinating residues. **(a)** Schematic representation of genetic incorporation of a non-canonical amino acid containing a metal binding moiety as additional ligand in a metalloenzyme (top) or creation of a novel metal binding site by introduction of a metal-binding ncAA (bottom). **(b)** Incorporation of (2,2-bipyridin-5yl)alanine (BpyA) into Lactococcal multidrug resistance Regulator (LmrR, PDB: 3F8F), and Quarternary ammonium compound Regulator (QacR) (PDB: 1JTY). These novel metalloenzymes catalyse Friedel-Crafts reaction of 2-methylindole with 1-(1-methyl-1*H*-imidazol-2-yl)but-2-en-1-one, resulting in opposite enantiomers.[70,71] **(c)** Incorporation of different ncAAs, namely N$_\delta$-methylhistidine (NMH),), β-(3-thienyl)-alanine (3ThA), *p*-aminophenylalanine (pAF) and 3-(3′-pyridyl)-alanine (3PyA), with metal-coordination properties into myoglobin (Mb, PDB: 1A6K). Resulting enzymes tested in guaiacol oxidation or C-H amination with 2,4,6-triisopropylbenzenesulfonyl azide. Q2.2_H93_NMH corresponds to evolved variant containing N$_\delta$-methylhistidine and 11 mutations (T39I, R45D, F46L, I107F, V21A, I28T, D45G, K63E, T67A, T95A, K140T).[77,79]
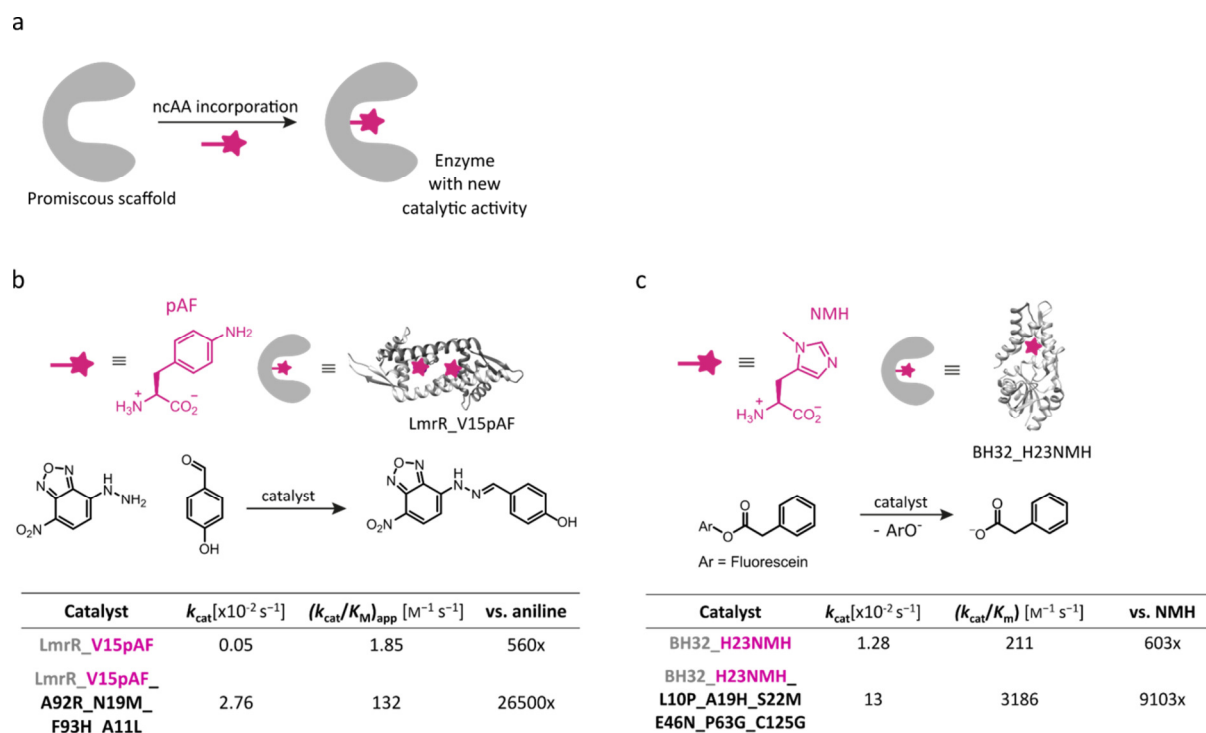
Fig. 4. Non-canonical amino acids (ncAAs) as catalytic residues. (a) Schematic representation of genetic incorporation of a non-canonical amino acid containing a reactive functional group into promiscuous scaffold, which creates enzymes with novel activities. **(b)** Artificial enzyme, Lactococcal multidrug resistance Regulator (LmrR) containing ncAA *p*-aminophenylalanine (pAF) (PDB: 6I8N) catalysing chromogenic hydrazone formation between 4-hydrazino-7-nitro-2,1,3-benzoxadiazole and 4-hydroxybenzaldehyde. The progress of the reaction was monitored by following product formation at 472 nm.[82,83] **(c)** Designed BH32 protein containing N_δ-methylhistidine (NMH) (PDB: 6Q7O) catalysing the hydrolysis of fluorescein 2-phenylacetate. The progress of the reaction was monitored by an increase in absorbance at 500 nm due to the formation of fluorescein.[84] Tables summarizing catalytic parameters obtained for the initial non-optimized variant and the best variant after a mutagenesis study.

Catalyst	k_{cat} [×10⁻² s⁻¹]	$(k_{cat}/K_M)_{app}$ [M⁻¹ s⁻¹]	vs. aniline
LmrR_V15pAF	0.05	1.85	560x
LmrR_V15pAF_ A92R_N19M_ F93H_A11L	2.76	132	26500x

Catalyst	k_{cat} [×10⁻² s⁻¹]	(k_{cat}/K_m) [M⁻¹ s⁻¹]	vs. NMH
BH32_H23NMH	1.28	211	603x
BH32_H23NMH_ L10P_A19H_S22M E46N_P63G_C125G	13	3186	9103x

Acknowledgment

Author contributions

I.D. researched data for the article. I.D. and G.R. wrote the article and both authors contributed to the discussion, reviewing and editing of the manuscript before submission.

Competing interests

The authors declare no competing interests.

References

1. Wohlgemuth, R. Biocatalysis-key to sustainable industrial chemistry. *Curr. Opin. Biotechnol.* **21**, 713–724 (2010).

2. Reetz, M. T. Biocatalysis in organic chemistry and biotechnology: past, present, and future. *J. Am. Chem. Soc.* **135**, 12480–12496 (2013).

3. Arnold, F. H. Directed evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed.* **57**, 4143–4148 (2018).

4. Okeley, N. M. & van der Donk, W. A. Novel cofactors via post-translational modifications of enzyme active sites. *Chem. Biol.* **7**, 159–171 (2000).

5. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).

6. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).

7. Young, D. D. & Schultz, P. G. Playing with the molecules of life. *ACS Chem. Biol.* **13**, 854–870 (2018).

8. Agostini, F. *et al.* Biocatalysis with unnatural amino acids: enzymology meets xenobiology. *Angew. Chem. Int. Ed.* **56**, 9680–9703 (2017).

9. Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *ChemBioChem* **9**, 1797–1804 (2008).

10. Currin, A., Swainston, N., Day, P. J. & Kell, D. B. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **44**, 1172–1239 (2015).

11. Li, G., Dong, Y. & Reetz, M. T. Can machine learning revolutionize directed evolution of selective enzymes? *Adv. Synth. Catal.* **361**, 2377–2386 (2019).

12. 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).

13. Currin, A. *et al.* GeneORator: an effective strategy for navigating protein sequence space more efficiently through boolean OR-type DNA libraries. *ACS Synth. Biol.* **8**, 1371–1378 (2019).

14. Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462 (2002).

15. Böck, A. *et al.* Selenocysteine: the 21st amino acid. *Mol. Microbiol.* **5**, 515–520 (1991).

16. Crick, F. H., Griffith, J. S. & Orgel, L. E. Codes without commas. *Proc. Natl. Acad. Sci. USA.* **43**, 416–421 (1957).

17. Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14**, 774–780 (2010).

18. Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment-expanding the chemistry in biology. *Chem. Sci.* **6**, 50–69 (2015).

19. Munier, R. & Cohen, G. N. Incorporation of structural analogues of amino acids into bacterial proteins during their synthesis in vivo. *Biochim. Biophys. Acta* **31**, 378–391 (1959).

20. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli. Science* **292**, 498–500 (2001).

21. Blight, S. K. *et al.* Direct charging of tRNA_{CUA} with pyrrolysine in vitro and in vivo. *Nature* **431**, 333–335 (2004).

22. Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **1844**, 1059–1070 (2014).

23. 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).

24. Yu, Y., Hu, C., Xia, L. & Wang, J. Artificial metalloenzyme design with unnatural amino acids and non-native cofactors. *ACS Catal.* **8**, 1851–1863 (2018).

25. Yang, H., Srivastava, P., Zhang, C. & Lewis, J. C. A general method for artificial metalloenzyme formation through strain-promoted azide-alkyne cycloaddition. *ChemBioChem* **15**, 223–227 (2014).

26. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat. Commun.* **6**, 7789 (2015).

24

27. Yang, H. *et al.* Evolving artificial metalloenzymes via random mutagenesis. *Nat. Chem.* **10**, 318-324 (2018).

28. Pauling, L. Nature of forces between large molecules of biological interest. *Nature* **161**, 707–709 (1948).

29. Lee, J. & Goodey, N. M. Catalytic contributions from remote regions of enzyme structure. *Chem. Rev.* **111**, 7595–7624 (2011).

30. Dominguez, M. A., Thornton, K. C., Melendez, M. G. & Dupureur, C. M. Differential effects of isomeric incorporation of fluorophenylalanines into PvuII endonuclease. *Proteins Struct. Funct. Genet.* **45**, 55–61 (2001).

31. Deepankumar, K. *et al.* Enhancing thermostability and organic solvent tolerance of ω-transaminase through global incorporation of fluorotyrosine. *Adv. Synth. Catal.* **356**, 993–998 (2014).

32. Cirino, P. C., Tang, Y., Takahashi, K., Tirrell, D. A. & Arnold, F. H. Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. *Biotechnol. Bioeng.* **83**, 729–734 (2003).

33. Hoesl, M. G. *et al.* Lipase congeners designed by genetic code engineering. *ChemCatChem* **3**, 213–221 (2011).

34. Jackson, J. C., Duffy, S. P., Hess, K. R. & Mehl, R. A. Improving nature's enzyme active site with genetically encoded unnatural amino acids. *J. Am. Chem. Soc.* **128**, 11124–11127 (2006).

35. Ma, H., Yang, X., Lu, Z., Liu, N. & Chen, Y. The 'gate keeper' role of Trp222 determines the enantiopreference of diketoreductase toward 2-chloro-1-phenylethanone. *PLoS One* **9**, e103792 (2014).

36. Kolev, J. N., Zaengle, J. M., Ravikumar, R. & Fasan, R. Enhancing the efficiency and regioselectivity of P450 oxidation catalysts by unnatural amino acid mutagenesis. *ChemBioChem* **15**, 1001–1010 (2014).

37. Morikubo, N. *et al.* Cation-π interaction in the polyolefin cyclization cascade uncovered by incorporating unnatural amino acids into the catalytic sites of squalene cyclase. *J. Am. Chem. Soc.* **128**, 13184–13194 (2006).

38. Ugwumba, I. N. *et al.* Improving a natural enzyme activity through incorporation of unnatural amino acids. *J. Am. Chem. Soc.* **133**, 326–333 (2011).

39. Xiao, H. *et al.* Exploring the potential impact of an expanded genetic code on protein function. *Proc. Natl. Acad. Sci. USA.* **112**, 6961–6966 (2015).

40. Stubbe, J. A., Nocera, D. G., Yee, C. S. & Chang, M. C. Y. Radical initiation in the class I ribonucleotide reductase: long-range proton-coupled electron transfer? *Chem. Rev.* **103**, 2167–2201 (2003).

41. Yee, C. S., Chang, M. C. Y., Ge, J., Nocera, D. G. & Stubbe, J. 2,3-Difluorotyrosine at position 356 of ribonucleotide reductase R2: a probe of long-range proton-coupled electron

transfer. *J. Am. Chem. Soc.* **125**, 10506–10507 (2003).

42. Seyedsayamdost, M. R., Reece, S. Y., Nocera, D. G. & Stubbe, J. A. Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis. *J. Am. Chem. Soc.* **128**, 1569–1579 (2006).

43. Yokoyama, K., Uhlin, U. & Stubbe, J. Site-specific incorporation of 3-nitrotyrosine as a probe of pka perturbation of redox-active tyrosines in ribonucleotide reductase. *J. Am. Chem. Soc.* **132**, 8385–8397 (2010).

44. Minnihan, E. C., Young, D. D., Schultz, P. G. & Stubbe, J. Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J. Am. Chem. Soc.* **133**, 15942–15945 (2011).

45. Minnihan, E. C., Nocera, D. G. & Stubbe, J. Reversible, long-range radical transfer in *E. coli* class Ia ribonucleotide reductase. *Acc. Chem. Res.* **46**, 2524–2535 (2013).

46. Oyala, P. H. *et al.* Biophysical characterization of fluorotyrosine probes site-specifically incorporated into enzymes: *E. coli* ribonucleotide reductase as an example. *J. Am. Chem. Soc.* **138**, 7951–7964 (2016).

47. Yu, Y. *et al.* Defining the role of tyrosine and rational tuning of oxidase activity by genetic incorporation of unnatural tyrosine analogs. *J. Am. Chem. Soc.* **137**, 4594–4597 (2015).

48. Yu, Y. *et al.* Significant improvement of oxidase activity through the genetic incorporation of a redox-active unnatural amino acid. *Chem. Sci.* **6**, 3881–3885 (2015).

49. Chen, L. *et al.* Use of a tyrosine analogue to modulate the two activities of a nonheme iron enzyme OvoA in ovothiol biosynthesis, cysteine oxidation versus oxidative C–S bond formation. *J. Am. Chem. Soc.* **140**, 4604–4612 (2018).

50. Chen, L. *et al.* Mechanistic studies of a nonheme iron enzyme OvoA in ovothiol biosynthesis using a tyrosine analogue, 2-Amino-3-(4-hydroxy-3-(methoxyl) phenyl) propanoic acid (MeOTyr). *ACS Catal.* **9**, 253–258 (2019).

51. Liu, X. *et al.* Significant increase of oxidase activity through the genetic incorporation of a tyrosine-histidine cross-link in a myoglobin model of heme-copper oxidase. *Angew. Chem. Int. Ed.* **51**, 4312–4316 (2012).

52. Jones, L. H., Narayanan, A. & Hett, E. C. Understanding and applying tyrosine biochemical diversity. *Mol. Biosyst.* **10**, 952–969 (2014).

53. Zhou, Q. *et al.* Probing the function of the Tyr-Cys cross-link in metalloenzymes by the genetic incorporation of 3-methylthiotyrosine. *Angew. Chem. Int. Ed.* **52**, 1203–1207 (2013).

54. Liu, X. *et al.* A genetically encoded photosensitizer protein facilitates the rational design of a miniature photocatalytic CO_2-reducing enzyme. *Nat. Chem.* **10**, 1201–1206 (2018).

55. Romero, N. A. & Nicewicz, D. A. Organic photoredox catalysis. *Chem. Rev.* **116**, 10075–10166 (2016).

56. Givens, R. S., Weber, J. F. W., Jung, A. H. & Park, C.-H. New photoprotecting groups: desyl

and *p*-hydroxyphenacyl phosphate and carboxylate esters. *Methods Enzymol.* **291**, 1–29 (1998).

57. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew. Chem. Int. Ed.* **45**, 2728–2731 (2006).

58. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 14306–14307 (2004).

59. Luo, J., Torres-Kolbus, J., Liu, J. & Deiters, A. Genetic encoding of photocaged tyrosines with improved light-activation properties for the optical control of protease function. *ChemBioChem* **18**, 1442–1447 (2017).

60. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **3**, 769–772 (2007).

61. Nguyen, D. P. *et al.* Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* **136**, 2240–2243 (2014).

62. Beharry, A. A. & Woolley, G. A. Azobenzene photoswitches for biomolecules. *Chem. Soc. Rev.* **40**, 4422–4437 (2011).

63. Hoppmann, C. *et al.* Genetically encoding photoswitchable click amino acids in *Escherichia coli* and mammalian cells. *Angew. Chem. Int. Ed.* **53**, 3932–3936 (2014).

64. Tsai, Y.-H., Essig, S., James, J. R., Lang, K. & Chin, J. W. Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nat. Chem.* **7**, 554–561 (2015).

65. Luo, J., Samanta, S., Convertino, M., Dokholyan, N. V & Deiters, A. Reversible and tunable photoswitching of protein function through genetic encoding of azobenzene amino acids in mammalian cells. *ChemBioChem* **19**, 2178–2185 (2018).

66. Schwizer, F. *et al.* Artificial metalloenzymes: reaction scope and optimization strategies. *Chem. Rev.* **118**, 142–231 (2018).

67. Hayashi, T., Hilvert, D. & Green, A. P. Engineered metalloenzymes with non-canonical coordination environments. *Chem. Eur. J.* **24**, 11821–11830 (2018).

68. Lee, H. S. & Schultz, P. G. Biosynthesis of a site-specific DNA cleaving protein. *J. Am. Chem. Soc.* **130**, 13194–13195 (2008).

69. Roelfes, G. LmrR: a privileged scaffold for artificial metalloenzymes. *Acc. Chem. Res.* **52**, 545–556 (2019).

70. Drienovská, I., Rioz-Martínez, A., Draksharapu, A. & Roelfes, G. Novel artificial metalloenzymes by in vivo incorporation of metal-binding unnatural amino acids. *Chem. Sci.* **6**, 770–776 (2015).

71. Bersellini, M. & Roelfes, G. Multidrug resistance regulators (MDRs) as scaffolds for the design of artificial metalloenzymes. *Org. Biomol. Chem.* **15**, 3069–3073 (2017).

72. Drienovská, I. *et al.* Design of an enantioselective artificial metallo-hydratase enzyme containing an unnatural metal-binding amino acid. *Chem. Sci.* **8**, 7228–7235 (2017).

73. Ségaud, N., Drienovská, I., Chen, J., Browne, W. R. & Roelfes, G. Artificial metalloproteins for binding and stabilization of a semiquinone radical. *Inorg. Chem.* **56**, 13293–13299 (2017).

74. Mills, J. H. *et al.* Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *J. Am. Chem. Soc.* **135**, 13393–13399 (2013).

75. Mills, J. H. *et al.* Computational design of a homotrimeric metalloprotein with a trisbipyridyl core. *Proc. Natl. Acad. Sci. USA.* **113**, 15012–15017 (2016).

76. Green, A. P., Hayashi, T., Mittl, P. R. E. & Hilvert, D. A chemically programmed proximal ligand enhances the catalytic properties of a heme enzyme. *J. Am. Chem. Soc.* **138**, 11344–11352 (2016).

77. Pott, M. *et al.* A noncanonical proximal heme ligand affords an efficient peroxidase in a globin fold. *J. Am. Chem. Soc.* **140**, 1535–1543 (2018).

78. Hayashi, T. *et al.* Capture and characterization of a reactive haem–carbenoid complex in an artificial metalloenzyme. *Nat. Catal.* **1**, 578–584 (2018).

79. Moore, E. J. & Fasan, R. Effect of proximal ligand substitutions on the carbene and nitrene transferase activity of myoglobin. *Tetrahedron* **75**, 2357–2363 (2019).

80. Schmidt, B., Selmer, T., Ingendoh, A. & Figurat, K. von. A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* **82**, 271–278 (1995).

81. Schwede, T. F., Rétey, J. & Schulz, G. E. Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry* **38**, 5355–5361 (1999).

82. Drienovská, 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).

83. Mayer, C., Dulson, C., Reddem, E., Thunnissen, A.-M. W. H. & Roelfes, G. Directed evolution of a designer enzyme featuring an unnatural catalytic amino acid. *Angew. Chem. Int. Ed.* **58**, 2083–2087 (2019).

84. Burke, A. J. *et al.* Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **570**, 219–223 (2019).

85. Lindström, U. M. Stereoselective organic reactions in water. *Chem. Rev.* **102**, 2751-2772 (2002).

86. Lipshutz, B. H., Ghorai, S. & Cortes-Clerget, M. The hydrophobic effect applied to organic synthesis: recent synthetic chemistry "in water". *Chem. Eur. J.* **24**, 6672–6695 (2018).

87. van der Helm, M. P., Klemm, B. & Eelkema, R. Organocatalysis in aqueous media. *Nat. Rev. Chem.* **3**, 491–508 (2019).

88. Jeschek, M., Panke, S. & Ward, T. R. Artificial metalloenzymes on the verge of new-to-nature

metabolism. *Trends Biotechnol.* **36**, 60–72 (2018).

89.   Almhjell, P. J., Boville, C. E. & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* **47**, 8980–8997 (2018).

90.   Mehl, R. A *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).

91.   Brustad, E., Bushey, M. L., Brock, A., Chittuluru, J. & Schultz, P. G. A promiscuous aminoacyl-tRNA synthetase that incorporates cysteine, methionine, and alanine homologs into proteins. *Bioorg. Med. Chem. Lett.* **18**, 6004–6006 (2008).

92.   Young, D. D. *et al.* An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894–1900 (2011).

93.   Wang, Y.-S., Fang, X., Wallace, A. L., Wu, B. & Liu, W. R. A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J. Am. Chem. Soc.* **134**, 2950–2953 (2012).

94.   Mukai, T. *et al.* Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* **38**, 8188–8195 (2010).

95.   Johnson, D. B. F. *et al.* RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **7**, 779–786 (2011).

96.   Johnson, D. B. F. *et al.* Release factor one is nonessential in *Escherichia coli*. *ACS Chem. Biol.* **7**, 1337–1344 (2012).

97.   Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).

98.   Mukai, T. *et al.* Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci. Rep.* **5**, 9699 (2015).

99.   Allen, A. E. & MacMillan, D. W. C. Synergistic catalysis: a powerful synthetic strategy for new reaction development. *Chem. Sci.* **3**, 633-658 (2012).

100.   Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed.* **52**, 5700–5725 (2013).