doi:10.2533/chimia.2023.395 Chimia 77 (2023) 395–402

Advances in Noncanonical Amino Acid Incorporation for Enzyme Engineering Applications

Sandro Giger[ab] and Rebecca Buller*[a]

Abstract: Incorporation of noncanonical amino acids (ncAAs) *via* genetic code expansion (GCE) opens up new possibilities for chemical biology. The technology has led to the development of novel xenobiotic enzymes with tailored properties which can serve as entry points into a multitude of applications, including protein conjugation, immobilization, or labeling. In this review, we discuss recent progress in the use of GCE to create biocatalysts possessing reaction repertoires that lie beyond what is achievable with canonical amino acids (cAAs). Furthermore, we highlight how GCE enables to gain mechanistic insights into protein function by the incorporation of judiciously selected ncAAs. As the amino acid alphabet continues to grow and improved tools for ncAA incorporation are being developed, we anticipate the creation of additional powerful biological catalysts for synthetic applications merging the chemical versatility of anthropogenic building blocks with the exquisite selectivities of enzymes.

Keywords: Chemical synthesis · Enzyme engineering · Enzyme mechanism · Genetic code expansion · New-to-nature enzymes · Noncanonical amino acid

Sandro Giger completed his apprenticeship as a chemical laboratory technician at Novartis before studying chemistry at the Zurich University of Applied Sciences (ZHAW). After his Master thesis under the supervision of Prof. Dr. Rebecca Buller, Sandro Giger joined the biocatalysis group at the ZHAW for his PhD in collaboration with the Vienna University of Technology. His research focuses on enzyme engineering for synthetic applications.

Rebecca Buller is a biological chemist and Professor for Biotechnological Methods, Systems and Processes at the Zurich University of Applied Sciences. Rebecca Buller studied chemistry at the Westfälische – Wilhelms Universität Münster (D) and the University of California Santa Barbara (US). After completing her PhD with a focus on enzyme engineering at ETH Zurich (CH), Rebecca Buller accepted a position as laboratory head at the flavour and fragrance company Firmenich (CH). In 2015, she relocated to the Zurich University of Applied Sciences where she founded the Competence Center for Biocatalysis (CCBIO). Research in Rebecca Buller's laboratory focusses on the expansion of the biocatalytic toolbox by sourcing and engineering enzymes for synthetic applications.

1. Introduction

Nature combinatorically builds its powerful catalysts – enzymes – from a defined set of building blocks, the twenty canonical amino acids (cAAs). If considering a sequence length of 300 amino acids, enzyme engineers can thus explore 20^{300} enzyme variants, a wealth of possible sequences effectively dwarfing the number of atoms in the universe.[1] This enzyme space is explored in enzyme engineering by methods such as directed evolution[2] and/or (semi)-rational design[3] and enables researchers to create enzymes with improved functions, including catalytic, biophysical, and molecular recognition properties. Optimized enzymes are applied in medicinal chemistry,[4,5] the synthesis of commodity chemicals[6,7] or for the manufacture of pharmaceutical intermediates.[8,9] It should be noted, however, that the number of exploitable enzyme variants is greatly reduced compared to the number of theoretical enzyme variants. The reduction of sequence space is caused, among others, by the necessity to have folded protein structures which require the formation of certain basic secondary protein structure elements, such as α-helices and β-sheets. Additionally, the twenty canonical amino acids used in ribosomal protein biosynthesis possess a limited functional and structural diversity and are therefore restricting the accessible catalytic repertoire of enzymes. While nature addresses these limitations by the utilization of co-factors[10] and posttranslational modifications,[11] an additional avenue towards a broader synthetic utility of enzymes is to explore the use of noncanonical amino acids (ncAAs) for the construction of biocatalysts.[12]

The successful incorporation of ncAAs into proteins is an entry point to many applications, including conjugation,[13] immobilization,[14] labeling,[15] or stability improvement[16] of enzymes. The technique is not limited to *E. coli* strains but has also been utilized in different prokaryotic or mammalian cells[17–19] to facilitate the production of otherwise difficult-to-express proteins. In addition, certain biological processes need to be studied in mammalian cells to give meaningful information.

This review will highlight how the construction of ncAA-containing enzymes in *E. coli* can afford mechanistic knowledge and/or can be profitably employed to broaden the synthetic scope of the targeted enzymes. It will focus on advances made in the last

*Correspondence: Prof. R. M. Buller[a], E-mail: rebecca.buller@zhaw.ch
[a]Zurich University of Applied Sciences, School of Life Sciences and Facility Management, Institute of Chemistry and Biotechnology, Einsiedlerstrasse 31, CH-8820 Wädenswil, Switzerland; [b]Vienna University of Technology, Faculty of Technical Chemistry, Institute of Applied Synthetic Chemistry, Getreidemarkt 9, 1060 Vienna, Austria

three years and delineate how this knowledge can be employed in future enzyme engineering campaigns with the aim to develop powerful new-to-nature catalysts in chemical synthesis. For a more comprehensive overview of additional ncAA applications the reader is referred elsewhere.[20–22]

1.1 *Noncanonical Amino Acid Incorporation*

Today, powerful methods of genetic code expansion (GCE) allow the incorporation of ncAAs at predefined positions into proteins enabling the creation of xenobiotic enzymes. Typically, *in vivo* incorporation of ncAAs into proteins is performed *via* one of the following two methods: Selective pressure incorporation (SPI) or stop codon suppression (SCS).

The SPI method utilizes the translational machinery of a cell to incorporate a ncAA that is structurally related to a cAA.[23] For this method, an auxotrophic *E. coli* strain unable to produce a specific target cAA is employed as the host strain for protein expression. Before inducing the translation of a target protein, an excess amount of the ncAA is added to the culture. Without the presence of the cAA, the desired ncAA is utilized as a substrate for the cognate aminoacyl-tRNA synthetase (aaRS) and consequently, an enzyme that incorporates the ncAA at every native position of the absent cAA is produced. Notably, since the SPI method comprehensively replaces the selected cAA with the desired ncAA in the entire protein, it is impossible to generate a xenobiotic enzyme that incorporates the ncAA only at specific positions within the enzyme sequence.

In contrast, the SCS method constitutes an avenue for a more targeted incorporation of ncAAs: Using stop codon suppression, it is possible to incorporate a ncAA site-specifically into an enzyme structure while the rest of the enzyme sequence remains untouched. Developed in the group of Peter Schultz,[24] SCS makes use of an orthogonal pair of aaRS and tRNA that can recognize the desired ncAA and incorporates it in response to the appropriate DNA codon. Since the amber codon (UAG) is the least frequently used stop codon in *E. coli*, it is usually reassigned to encode the desired ncAA (Fig. 1). Furthermore, to enable successful incorporation of ncAAs using orthogonal aaRSs, it is critical that these aaRSs exclusively recognize the associated orthogonal tRNA and target ncAA as a substrate, without any recognition of the endogenous tRNA or cAAs of the microbial host strains. Of course, the orthogonal tRNA may equally not act as a substrate for any endogenous aaRS. Although different strategies have been explored for the generation of orthogonal aaRS/tRNA pairs,[25,26] the most straightforward and efficient method entails the introduction of a heterologous aaRS/tRNA pair derived from a distinct domain of life.[27] Several aaRS/tRNA pairs have been used to successfully incorporate ncAAs into enzymes in *E. coli*, such as the TyrRS/$tRNA_{CUA}$ from *Methanococcus jannaschii* (*Mj*),[24] PylRS/$tRNA_{CUA}$ from the methanogen *Methanosarcina barkeri (Mb),*[28] or the LeuRS/$tRNA_{CUA}$ pair for which the aaRS is derived from *Methanobacterium thermoautotrophicum* and the tRNA from *Halobacterium sp. NRC-1*.[29] These aaRS/tRNA pairs are usually evolved towards incorporation of the desired ncAA in positive and negative screening rounds of aaRS variant libraries. The positive round ensures incorporation of any amino acid in answer to a UAG codon located on a viability associated gene. The following negative screening in absence of the ncAA has a UAG codon located on a lethal gene, guaranteeing that none of the cAAs are incorporated by the aaRS/tRNA pair.[26] Altogether, more than 200 ncAAs have been reported to be incorporated by SCS (selected examples of ncAAs in Fig. 2),[20] providing a versatile toolbox to specifically modify enzymes for targeted applications.

Importantly, ncAA incorporation in *E. coli* through reassignment of the UAG stop codon is in direct competition with the action of release factor 1 (RF1), which terminates translation in response to UAG and UAA codons. To address this issue, Lajoie *et al*. replaced all known UAG stop codons in *E. coli* MG1655 with UAA codons which permitted the deletion of RF1.[30] The optimized strain, named C321.ΔA.exp, was successfully used to express enzymes with ncAAs[31] but the genetic modification se-

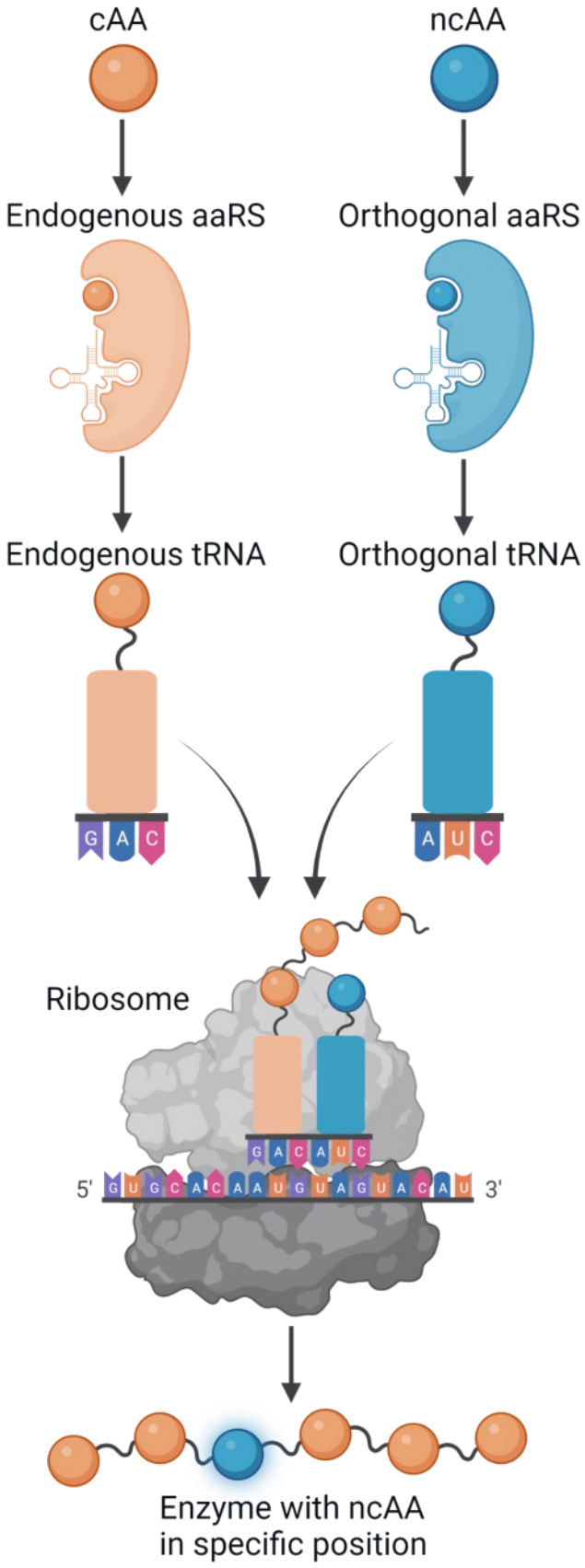

Fig. 1. Illustration of genetic code expansion by the SCS method utilizing the amber stop codon. The orthogonal aaRS aminoacylates the ncAA to the orthogonal tRNA, which recognizes the UAG codon.

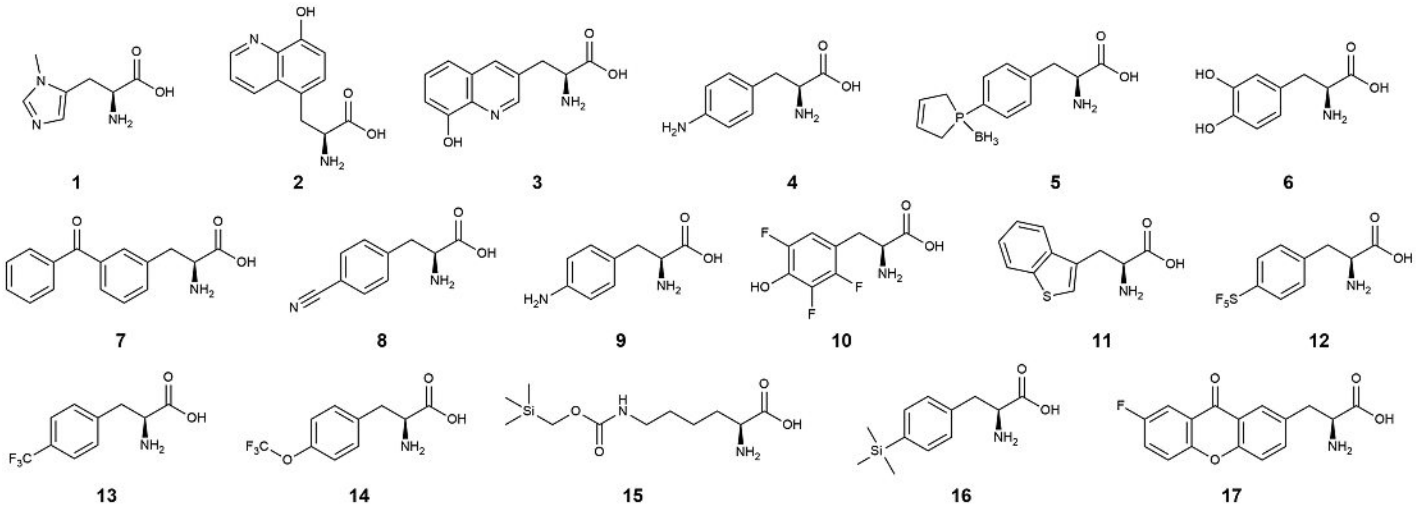

Fig. 2. Overview of ncAAs described in the examples of this review.

verely affected the strain's growth rate and increased its doubling time by 60%. In response to this observation, Sakamoto and coworkers opted to replace only 95 out of a total of 178 selected UAG codons. The resulting strain, B95.ΔA, retained its reproductive strength[32] and has been successfully employed in the expression of ncAA-containing enzymes.[33–38]

2. Enzyme Engineering Using Non-canonical Amino Acids

The ability to introduce ncAAs into enzymes at precise locations has resulted in the creation of enzymes with intriguing catalytic capabilities. Examples of such enzymes include a computationally designed hydrolase containing N_δ-methylhistidine (**1**, NMH)[39] and a lactococcal multidrug-resistance regulator (LmrR) that was transformed into a designer enzyme for hydrazone and oxime formation.[40] In this section, we provide a detailed discussion of the latest additions to these xenobiotic enzymes.

Artificial metalloenzymes (ArMs), which are created by incorporating an abiotic transition metal cofactor into a protein scaffold, are versatile tools in bioengineering. The new-to-nature enzymes can achieve transition metal reactivity with substrate specificity and regio- and enantioselectivity reminiscent of natural enzymes.[41] For example, HaloTag, a self-labeling protein, was used as a scaffold for the generation of an ArM that catalyzes olefin metathesis.[42] Building on this work, Ward and coworkers further tailored the HaloTag ArM through the incorporation of ncAAs. In order to assemble an artificial deallylase, the metal-chelating ncAA 2-amino-3-(8-hydroxyquinolin-5-yl)propanoic acid (**2**, HQ-Ala-1) was introduced into three different positions (F144, A145, and M175) that lie in proximity of the binding-cleft of HaloTag.[43] HQ-Ala-1 (**2**) was selected as the preferred ncAA candidate to position $[(\eta^5\text{-}C_5H_5)Ru(MeCN)_3]^+$ as prior experiments had shown that *in vivo* CpRu-catalyzed deallylation is more efficient in the presence of hydroxyquinoline-based bidentate ligands.[44] Additionally, the synthesis of HQ-Ala-1 (**2**) was found to be considerably less complex than other hydroxyquinoline-based ncAAs. To assess the catalytic performance, the conveniently quantifiable deallylation of *O*-allyl carbamate-protected coumarin to 7-aminocoumarin was performed. Notably, the turnover number (TON) of all three variants was increased 1.3- to 1.5-fold compared to wildtype HaloTag. The authors hypothesized that the improved catalytic performance is due to the increased ruthenium content of ncAA-containing HaloTag variants compared to wildtype HaloTag and conclude that this study might serve as a starting point for the creation of unique HaloTag-based ArMs through the incorporation of ncAAs.[43]

Using a similar concept, Roelfes and coworkers incorporated 2-amino-3-(8-hydroxyquinolin-3-yl)propanoic acid (**3**, HQ-Ala-2), a structural isomer of HQ-Ala-1 (**2**), into the homodimeric LmrR at positions V15 and M89 and complexed the resulting LmrR variants with a variety of transition metal ions such as Cu^{II} and Zn^{II}. Following assembly, the resulting enzyme variants were examined for their ability to catalyze selected reactions: the Zn^{II}-containing variants were found to be capable of hydrolysing the amide bond between a model tripeptide and *p*-nitrophenylalanine, whereas Cu^{II}-containing variants catalyzed a vinylogous Friedel-Crafts alkylation reaction (Fig. 3a).[45] In further experiments, valine at position 15 of LmrR was replaced by *para*-aminophenylalanine (**4**, pAF) to create LmrR_pAF, which was shown to facilitate a Friedel-Crafts alkylation between α,β-unsaturated aldehydes and indoles in a metal-free fashion without the need for a chelating group (Fig. 3b). Using LmrR_pAF as a starting point for further enzyme engineering, several key residues around the enzyme's active site were selected for randomization utilizing the NDT codon encoding 12 chemical diverse cAAs. The experiments identified a pAF-containing (**4**) triple mutant (L18R, S95G, M89N) characterized by a >2-fold improved yield of the alkylation product (87%), compared to LmrR_pAF (42%). In substrate scope experiments, the evolved variant additionally outperformed the parent in various substrate combinations and was shown to transform α-substituted acroleins, which are sterically particularly demanding substrates (Fig. 3b).[46,47] In a further study, wildtype LmrR served as a scaffold for the introduction of a phosphine ligand with the aim to bridge the gap between non-biological phosphine ligands and protein chemistry. (*S*)-2-amino-3-(4-(2,5-dihydro-1H-phosphol-1-borane-1-yl)phenyl)propanoic acid (**5**, P3BF), a borane-protected phosphine derivative of phenylalanine, was inserted at position V15 of LmrR using the SCS technology and subsequently chemically transformed into the metal-binding phosphine ligand P3F *via* the addition of a palladacycle dimer under aerobic conditions (Fig. 3c). This straightforward approach provided an efficient one-pot deboronation and palladium coordination of the artificial metalloenzyme in aqueous solution under aerobic conditions and demonstrates the first example of a phosphine ligand containing enzyme created *in vitro*, laying the groundwork for the development of numerous novel ArMs.[48]

A further example of how enzymatic properties can be modulated by introducing ncAAs is the incorporation of L-3,4-dihydroxyphenylalanine (**6**, L-DOPA) into alcohol dehydrogenase II (ADHII) from *Zymomonas mobilis* at position H277. The resulting xenobiotic protein exhibited an almost 3-fold higher bind-

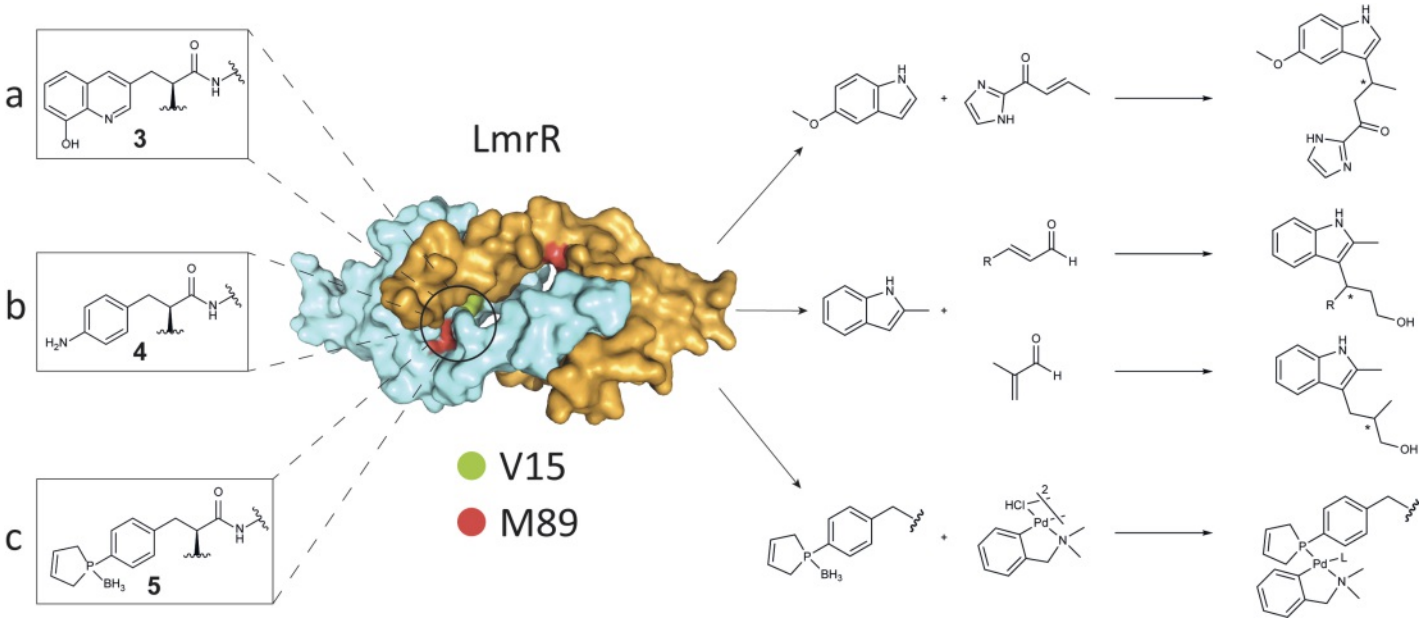

Fig. 3. Example reactions of xenobiotic LmrR variants, homodimeric proteins. a) Incorporation of HQ-Ala-2 (**3**) at V15 and M89 enables vinylogous Friedel-Crafts alkylation.[44] b) Incorporation of pAF (**4**) at V15 enables an analogous Friedel-Crafts alkylation between α,β-unsaturated aldehydes and indoles in a metal-free fashion.[45] Further directed evolution of the ncAA containing enzyme increased substrate scope towards acroleins with substituents in the α-position.[46] c) Incorporation of P3BF (**5**) provides a tool for an efficient one-pot deboronation and palladium coordination with LmrR in aqueous solution under aerobic conditions.[47]

ing affinity towards zinc compared to wildtype ADHII, while Fe binding was abolished. As a result, the L-DOPA (**6**) containing variant was rendered tolerant to oxygen and showed activity both for acetaldehyde reduction and ethanol oxidation reactions under aerobic conditions.[49]

Going beyond the incorporation of metal-coordinating residues to induce new reactivities into enzymes, ncAAs can also be used to improve wildtype enzymatic activities or improve protein stability. In this context, a *(R)*-amine transaminase, which had previously been engineered for the conversion of the aromatic amine substrate (*R*)-phenylethylamine and the amine acceptor pyruvate to yield acetophenone and D-alanine, respectively,[50] was used as a scaffold for ncAA introduction. Informed by the original analysis, which had highlighted that three of the six beneficial mutations (Y31F, H86F and Y88F) heavily contributed towards the hydrophobicity of the active site, replacement of F88 residue with *para*-benzoylphenylalanine (**7**, pBpA) was targeted. As expected, the pBpA-containing variant further increased the hydrophobicity of the active site and exhibited increased relative activity for a range of amine donors, such as 1-(4-fluorophenyl)ethan-1-amine and 1-phenylpropan-1-amine, compared to the parental enzyme. While F88pBpA displayed thermostability identical to that of the parent enzyme, both enzyme variants only retained 45% of their activity at 55 °C. In contrast, a double variant, F86A/F88pBpA, demonstrated improved thermostability and kept 85% of its activity when exposed to the identical temperature challenge. In summary, the results of this study provide evidence for the potential to tune the activity of enzymes through the incorporation of ncAAs without detrimental effects on protein stability.[51]

Another example of improving enzymatic catalytic efficiency and stability through the incorporation of ncAAs can be found in an engineered *E. coli* transketolase created by Dalby and coworkers. The *E. coli* transketolase was previously evolved towards acceptance of aromatic aldehydes which were not converted by the wildtype enzyme.[52] In the improved variant (S385Y/D469T/R520Q), the introduction of tyrosine at position 385 was identified as a critical factor for the binding of aromatic substrates. The triple mutant was further fine-tuned by incorporation of pAF (**4**), *para*-cyanophenylalanine (**8**, pCNF), and *para*-nitrophenylalanine (**9**, pNTF) in place of Y385. This strategy enabled the introduction of side chains characterized by aromatic ring electron densities distinct from those available within the repertoire of the twenty cAAs. Compared to the previous best variant, the pCNF (**8**) variant exhibited a 43-fold increase in specific activity for the conversion of 50 mM 3-hydroxybenzaldehyde and a doubled k_{cat} (k_{cat} (S385Y/D469T/R520Q) = 18.7 ± 0.5 min^{-1}; k_{cat} (S385pCNF/D469T/R520Q) = 36.7 ± 0.6 min^{-1}), while the K_m of the pNTF (**9**) variant was improved by approximately 3-fold (K_m (S385Y/D469T/R520Q) = 44.8 ± 16.3 mM; K_m (S385pNTF/D469T/R520Q) = 14.8 ± 4.2 mM). Finally, the pAF (**4**) variant exhibited a >3-fold increase in k_{cat}/K_M (k_{cat}/K_M (S385Y/D469T/R520Q) = 6.9 ± 1.1 $s^{-1}M^{-1}$; k_{cat}/K_M (S385pAF/D469T/R520Q) = 23.8 ± 0.8 $s^{-1}M^{-1}$) and decreased substrate inhibition in addition to causing a 7.8 °C increase in the thermal transition mid-point T_m. Notably, the pAF (**4**) variant thus constitutes the first example for which a simultaneous enhancement of catalytic activity and stability through site-specific ncAA incorporation was achieved.[53]

3. Structural and Mechanistic Studies with ncAAs

Installation of ncAAs characterized by properties which are distinct from the native cAA at key positions within enzymes offers the possibility to study enzymatic interactions, dynamics, or mechanisms in more detail. Common examples are the replacement of crucial tyrosine, tryptophane or histidine residues, such as the tyrosine residues in *E. coli* ribonucleotide reductases.[54,55]

The installation of tyrosine analogs has been found to be a valuable tool for mechanistic investigations of enzymes following a radical reaction scheme in which such residues often play a key role. In this context, ribonucleotide reductases (RNR) are involved in the production of deoxyribonucleotide pools needed for DNA biosynthesis and repair. For catalysis in class I RNRs, the creation of a thiyl radical in the α2β2 complex is a key step. It occurs on the RNR α2 subunit at C439 and involves a long-range proton-coupled electron transfer (PCET) from Y122 located in the β2-subunit bridging around 35 Å *via* a series of conserved aromatic amino acids. By substituting the critical tyrosine residue with (2,3,5)-trifluorotyrosine (**10**, F_3Y) in the context of the double mutant E52Q/Y122F_3Y, the Drennan group could trap the

active state of the α2β2 RNR complex and obtain its structure with an overall resolution of 3.6 Å by cryo-electron microscopy (Fig. 4a). The reduced dissociation rate of the α2β2 complex from the second (wildtype) to the minute time scale (E52Q/Y122F_3Y variant) enabled the imaging of a 32 Å-long pathway for the PCET for the first time.[56]

Similarly looking into tyrosine replacement, several analogs of this aromatic amino acid were employed as probes to study the mechanism of endoperoxide installation by the verruculogen synthase (FtmOx1), an Fe(II)/α-ketoglutarate-dependent dioxygenase. During the reaction, a tyrosine residue donates a hydrogen atom to the C26• radical of the substrate. Replacing either Y224 or Y68 of FtmOx1 with phenylalanine revealed that the Y224F variant produced wildtype-like yields, whereas the Y68F variant produced unidentified alternative products. To further probe the role of these two residues, ring-halogenated tyrosines and pAF (**4**) were introduced in the place of either Y224 or Y68 to allow for mechanistic studies with UV-Vis and electron paramagnetic resonance (EPR). The experiments established that Y68 is the hydrogen donor to C26• of the substrate verruculogen giving crucial mechanistic insights into the installation of the endoperoxide bridge of the mycotoxin (Fig. 4b).[57]

Moving beyond replacing tyrosine residues to evaluate their roles in radical mechanisms, targeting metal coordinating residues has also provided important insights into enzymatic mechanisms. Amino acid coordination of metal centers is a prevalent motif in metalloenzymes, such as cytochrome *c*, myoglobin, or matrix metalloproteases. Harnessing the modulating nature of the coordinating amino acid residues on the properties of the metal, Green and coworkers created a cytochrome *c* peroxidase (C*c*P) variant with NMH (**1**) instead of histidine as proximal ligand. The rationale behind the choice of the less electron donating NMH (**1**) was the anticipated impact of the ncAA on the catalytic cycle of C*c*P while having only a minor effect on enzyme structure. During the C*c*P reaction, compound I (CpdI), containing an iron (IV)-oxo species – often referred to as the 'ferryl' heme species – and a tryptophan radical cation, is initially formed. This enzymatic intermediate is reduced to compound II (CpdII) by a single electron transfer *via* C*c*P's biological redox partner ferrous cytochrome *c* before further reduction events return the enzyme to the resting state. While the introduction of the proximal NMH-ligand had little impact on CpdI reactivity, it was found to result in a 10-fold slower reduction of CpdII, providing a direct link between axial ligand electron donation and CpdII reactivity. In their analysis of the results, the authors suggest that the pK_a of Fe(IV)=O is an important factor determining compound II reactivity in C*c*P and other heme peroxidases, with reduced electron donation of the NMH (**1**) ligand affording an electron-deficient ferryl oxygen with decreased proton affinity compared to the wild-type enzyme (Fig. 4d). Interestingly, the observed deleterious effect of NMH (**1**) on C*c*P catalysis could be fully compensated by a W51F mutation, located on the opposite site of the coordinated iron. By removing

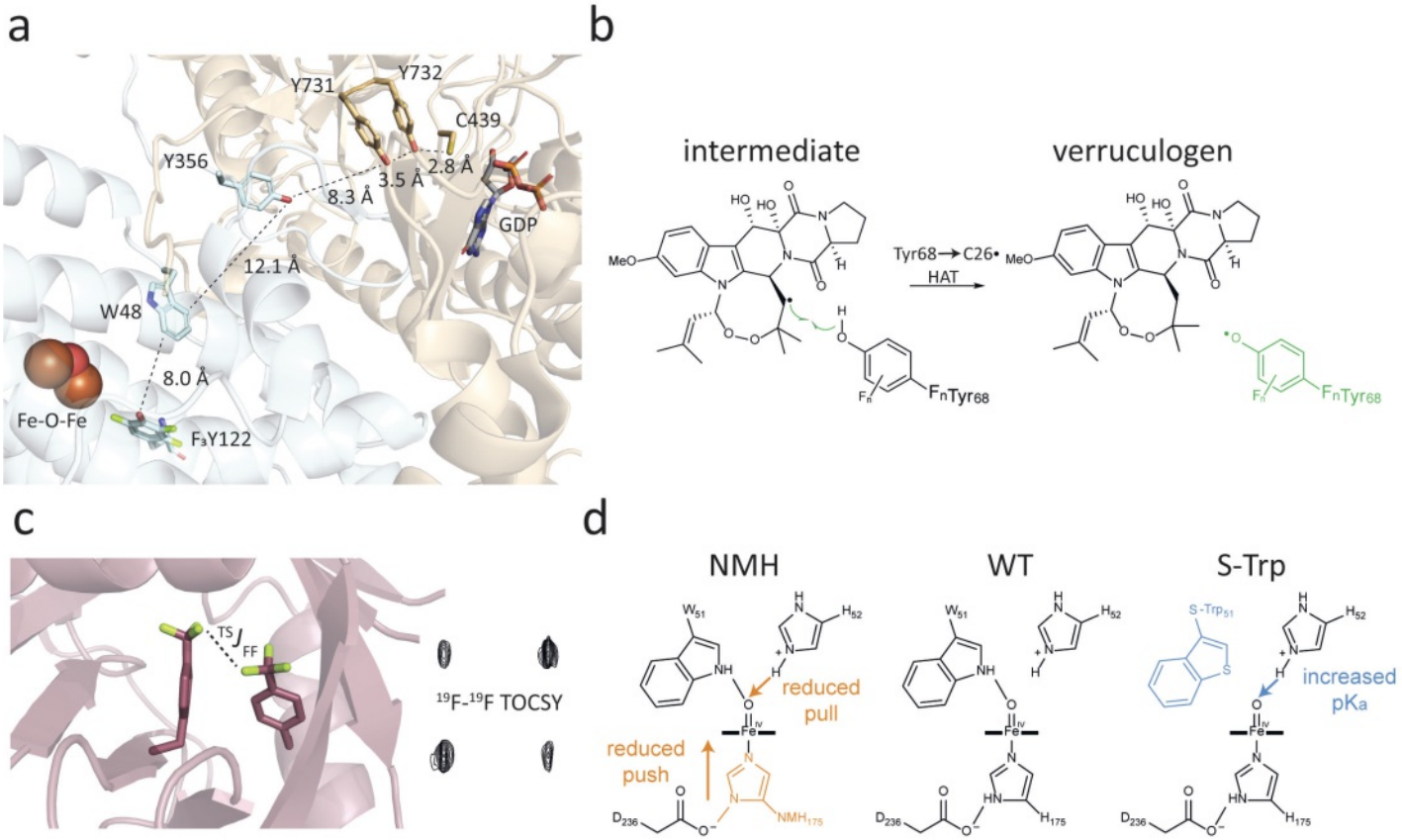

Fig. 4. Examples of enzymes which were mechanistically or structurally studied with GCE. a: Replacement of Y122 with F3Y (**10**) (shown with green sticks) in the enzyme RNR enabled the entrapment and analysis of the active state of the α2β2 complex by cryo-electron microscopy. Amino acids involved in the 32 Å-long PCET pathway connecting F3Y122 to C349 are highlighted as sticks. Relevant distances are shown with dashed lines. Figure adapted from Kang *et al*.[56] b: Replacement of two tyrosine residues in FtmOx1 with ring-halogenated tyrosines and the subsequent analysis of the resulting enzyme variants with UV-Vis and EPR established that Y68 is the hydrogen donor to the C26• radical of the substrate verruculogen. Figure adapted from Lin *et al*.[57] c: Incorporation of CF3-Phe (**13**) into PpiB at positions F27 and F98 allowed to detect through-space (TS) ^{19}F–^{19}F couplings which confirmed the spatial proximity of the two enzyme residues. Figure adapted from Orton *et al.*[33] d: Replacement of the proximal histidine ligand of CcP with the less electron-donating NMH (**1**) resulted in a 10-fold slower reduction of CpdII caused by a reduced proton affinity of the ferryl oxygen compared to the wild-type enzyme. Conversely, replacement of W51 with S-Trp (**11**) resulted in an increased lifetime of CpdI, while the lifetime of CpdII was decreased. Figure adapted from Ortmayer *et al*.[58,59]

a hydrogen bond that is formed in the wildtype between W51 and the ferryl oxygen, the authors hypothesize that the W51F mutation increased the 'electron pull' of the ferryl oxygen.[58] To investigate this effect further in the context of the wildtype enzyme, Green and co-workers replaced W51 with 3-benzothienyl-L-alanine (**11**, S-Trp), a sulfur-containing tryptophan analog unable to serve as a hydrogen bond donor. The S-Trp (**11**) substitution led to an increased lifetime of CpdI, while decreasing the lifetime of CpdII. Supplementing the experimental findings with density-functional theory (DFT) calculations, the authors could confirm their suggestion that tryptophan at position 51 suppresses reactivity and proton affinity of CpdII through hydrogen bonding to the ferryl oxygen (Fig. 4d).[59] In conclusion, these two studies highlight how ncAAs can serve as valuable probes of enzyme mechanism and provide insights into heme biochemistry. Looking forward, the delineated ncAA strategy can be employed in similar attempts to deconstruct bioinorganic mechanisms.

Beyond harnessing their influence on the reaction mechanism of enzymes, ncAAs can be also used to introduce unique labels for spectrometric analysis. Since fluorine is only rarely found in biological systems, the ^{19}F NMR spectra of proteins are devoid of unwanted background resonances rendering it a useful tool for studying proteins. Employing this methodology with the goal to confirm the spatial proximity between two enzyme residues, the incorporation of *para*-pentafluorosulfanyl-phenylalanine (**12**, SF_5Phe) into *E. coli* peptidyl-prolyl *cis/trans*-isomerase B (PpiB) at F27 and F98 enabled Huber and coworkers to carry out ^{19}F–^{19}F nuclear Overhauser effect spectroscopy (NOESY) experiments. The study allowed the direct experimental confirmation of short contacts between the investigated residues and was the first example of observation of intramolecular ^{19}F–^{19}F nuclear Overhauser effects in a protein.[37] In a next step, *para*-trifluoromethyl-phenylalanine (**13**, CF_3-Phe) and *O*-trifluoromethyl-tyrosine (**14**, CF_3-Tyr) were introduced at the same positions in PpiB. The use of less bulky noncanonical amino acids resulted in narrower line widths and thus improved spectral resolution and sensitivity for the detection of through-space ^{19}F–^{19}F couplings (Fig. 4c). Additionally, the reduced space demand of CF_3-Phe (**13**) and CF_3-Tyr (**14**) compared to SF_5Phe minimized their impact on the natural protein structure.[33]

Apart from the utilization of fluorine labels, the introduction of trimethylsilyl groups can also be a valuable strategy to facilitate the NMR studies of proteins as these entities generate strong and narrow singlet signals in the ^{1}H NMR spectrum around 0 ppm, a region that is relatively devoid of other proton resonances. In this spirit, Otting and coworkers incorporated N^6-(((trimethylsilyl)-methoxy)carbonyl)-L-lysine (**15**, TMSK) into PbiB, *Bacillus stearothermophilus* DnaB, and a *Methanocaldococcus jannaschii p*-cyano-L-phenylalanyl-tRNA synthetase. Incorporation of TMSK (**15**) enabled the assignment of the TMS signal from the ^{1}H NMR resonance at 0 ppm and facilitated measurement of the dissociation constant K_d of tightly bound ligands and the determination of the exchange rate between open and closed enzyme conformations.[35] Following a similar experimental setup, 4-(trimethylsilyl)phenylalanine (**16**, TMSiPhe) was incorporated into arrestin, a key signal transducer downstream of most G-protein-coupled receptors (GPCRs). To investigate the protein mechanism, TMSiPhe (**16**) was genetically introduced at seven different positions of arrestin allowing to track conformational changes of the membrane protein complex at residue resolution with only little perturbation to the target protein. Interestingly, the ncAA technology can thus be used to improve our understanding of how different ligands induce conformational changes in arrestin and lead to the development of more specific and effective drugs that target arrestin-mediated signaling pathways.[60]

Going beyond NMR, the incorporation of ncAAs can facilitate the use of powerful analysis techniques that were prevously less well accessible. In this context, the phototrigger (*S*)-2-amino-3-(7-fluoro-9-oxo-9*H*-xanthen-2-yl)propanoic acid (**17**, FXO) was introduced into a rationally designed human liver fatty-acid binding protein mutant (XOM) at position 63. Upon light absorption, FXO (**17**) created a new C–C bond to the methylthio-group of a close-by methionine (M71). The intermediates of this reaction were captured *via* time-resolved serial femtosecond crystallography (TR-SFX). Through this approach, the excited intermediate state, a ketone triplet, responsible for precise C-H bond activation was identified. As this intermediate is functionally equivalent to high-valent metal-oxo species which are key in many metalloenzymes, the authors hypothesize that the gained structural insights will be helpful in future enzyme design experiments. In addition, they postulate that *via* FXO (**17**) incorporation, additional enzymes can be examined using TR-SFX to uncover reaction intermediates important for their efficiency and selectivity.[61]

4. Challenges and Future Directions

Incorporation of ncAAs *via* the method of SCS has enabled the creation of xenobiotic enzymes with unique properties and broad application potential. Tailored ncAA containing enzymes serve as tools for organic synthesis or can be used as valuable mechanistic probes.[62] Yet, to evolve the technology further, several challenges need to be addressed. While there are many aaRS/tRNA pairs readily available for the incorporation of ncAAs, there is a certain bias towards analogs of tyrosine and pyrrolysine since over two-thirds of aaRS/tRNA pairs are derived from TyrRS/$tRNA^{Tyr}$ or PylRS/$tRNA^{Pyl}$ couples.[63] Consequently, to facilitate the incorporation of a wider range of artificial amino acids, additional orthogonal translation systems must be evolved.

Furthermore, the cost and availability of ncAAs can be problematic. Synthetic routes toward ncAAs are often inefficient requiring multiple steps characterized by moderate yields. Additional synthetic issues are often caused by the absence of enantio- and stereoselective methods to prepare the amino acid analogues along with the requirement to protect the reactive amine and carboxylic groups. Looking forward, enzymatic synthesis of ncAAs could allow to address this issue: cascades of engineered enzymes could deliver the desired amino acid analogs selectively generating a more economically viable supply of the building blocks. One example of this approach is the production of various polyfluorinated tyrosines through the use of a tyrosine phenol lyase.[55,64] Further expanding on the idea of ncAA synthesis by enzyme cascades, limited ncAA availability could also be overcome by the engineering of autonomous microbial strains that utilize the desired ncAA as a 21st amino acid building block and produce the targeted xenobiotic enzyme.[65] In illustration of this concept, Xiao and coworkers created an *E. coli* strain that biosynthesizes L-DOPA (**6**) and shuttles the ncAA into DOPA-containing proteins. The engineered *E. coli* strains produced the designer proteins with higher yields than obtainable through exogenously fed cells.[66]

Finally, while the powerful method of directed evolution has already been applied to improve enzymes harboring ncAAs,[39,46,67] recent advances in rational design through the availability of highly accurate enzyme models[68] and machine learning approaches in enzyme evolution[69] have yet to be employed to further fine-tune the xenobiotic proteins. Particularly, the optimization of ncAA containing enzymes could profit from *in silico* screenings, as any reduction of experiments will save the required quantity of the expensive ncAA and, of course, additional resources.[4]

Acknowledgements

This work was created as part of NCCR Catalysis, a National Centre of Competence in Research funded by the Swiss National Science Foundation (Grant number 180544).

Received: May 5, 2023

[1] N. J. Turner, *Nat. Chem. Biol.* **2009**, *5*, 567, https://doi.org/10.1038/nchembio.203.
[2] F. H. Arnold, *Angew. Chem.* **2018**, *57*, 4143, https://doi.org/10.1002/ange.201708408.
[3] K. Steiner, H. Schwab, *Comput. Struct. Biotechnol. J.* **2012**, *2*, https://doi.org/10.5936/csbj.201209010.
[4] J. Büchler, S. H. Malca, D. Patsch, M. Voss, N. J. Turner, U. T. Bornscheuer, O. Allemann, C. Le Chapelain, A. Lumbroso, O. Loiseleur, R. Buller, *Nat. Commun.* **2022**, *13*, 1, https://doi.org/10.1038/s41467-022-27999-1.
[5] M. Voss, S. Hüppi, D. Schaub, T. Hayashi, M. Ligibel, E. Sager, K. Schroer, R. Snajdrova, R. M. U. Buller, *ChemCatChem* **2022**, *14*, https://doi.org/10.1002/cctc.202201115.
[6] F. Meyer, R. Frey, M. Ligibel, E. Sager, K. Schroer, R. Snajdrova, R. Buller, *ACS Catal.* **2021**, *11*, 6261, https://doi.org/10.1021/acscatal.1c00678.
[7] A. Papadopoulou, J. Meierhofer, F. Meyer, T. Hayashi, S. Schneider, E. Sager, R. Buller, *ChemCatChem* **2021**, *13*, 3914, https://doi.org/10.1002/cctc.202100591.
[8] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185, https://doi.org/10.1038/nature11117.
[9] T. Hayashi, M. Ligibel, E. Sager, M. Voss, J. Hunziker, K. Schroer, R. Snajdrova, R. Buller, *Angew. Chem. Int. Ed.* **2019**, *58*, 18535, https://doi.org/10.1002/anie.201907245.
[10] J. D. Fischer, G. L. Holliday, J. M. Thornton, *Bioinformatics* **2010**, *26*, 2496, https://doi.org/10.1093/bioinformatics/btq442.
[11] M. M. Müller, *Biochemistry* **2018**, *57*, 177, https://doi.org/10.1021/acs.biochem.7b00861.
[12] T. S. Young, P. G. Schultz, *JBC* **2010**, *285*, 11039, https://doi.org/10.1074/jbc.R109.091306.
[13] S. I. Lim, J. Cho, I. Kwon, *Chem. Commun.* **2015**, *51*, 13607, https://doi.org/10.1039/c5cc04611d.
[14] N. Ayyadurai, N. S. Prabhu, K. Deepankumar, Y. J. Jang, N. Chitrapriya, E. Song, N. Lee, S. K. Kim, B. G. Kim, N. Soundrarajan, S. Lee, H. J. Cha, N. Budisa, H. Yun, *Bioconjug. Chem.* **2011**, *22*, 551, https://doi.org/10.1021/bc2000066.
[15] J. Wang, J. Xie, P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 8738, https://doi.org/10.1021/ja062666k.
[16] J. C. Li, T. Liu, Y. Wang, A. P. Mehta, P. G. Schultz, *J. Am. Chem. Soc.* **2018**, *140*, 15997, https://doi.org/10.1021/jacs.8b07157.
[17] C. Xu, Q. Zou, J. Tian, M. Li, B. Xing, J. Gong, J. Wang, Y. X. Huo, S. Guo, *ACS Synth. Biol.* **2023**, *12*, 583, https://doi.org/10.1021/acssynbio.2c00604.
[18] H. S. Jang, X. Gu, R. B. Cooley, J. J. Porter, R. L. Henson, T. Willi, J. A. Didonato, S. L. Hazen, R. A. Mehl, *ACS. Chem. Biol.* **2020**, *15*, 562, https://doi.org/10.1021/acschembio.9b01026.
[19] M. D. Bartoschek, E. Ugur, T. A. Nguyen, G. Rodschinka, M. Wierer, K. Lang, S. Bultmann, *Nucleic Acids Res.* **2021**, *49*, E62, https://doi.org/10.1093/nar/gkab132.
[20] A. D. Pagar, M. D. Patil, D. T. Flood, T. H. Yoo, P. E. Dawson, H. Yun, *Chem. Rev.* **2021**, *121*, 6173, https://doi.org/10.1021/acs.chemrev.0c01201.
[21] D. D. Young, P. G. Schultz, *ACS Chem. Biol.* **2018**, *13*, 854, https://doi.org/10.1021/acschembio.7b00974.
[22] Z. Birch-Price, C. J. Taylor, M. Ortmayer, A. P. Green, *Protein Engineering, Design and Selection* **2023**, *36*, https://doi.org/10.1093/protein/gzac013.
[23] R. Munier, G. N. Cohen, *Biochim. Biophys. Acta* **1959**, *31*, 378, https://doi.org/10.1016/0006-3002(59)90011-3.
[24] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498, https://doi.org/10.1126/science.1060077.
[25] D. R. Liu, P. G. Schultz, *Proc Natl. Acad. Sci. USA* **1999**, *96*, 4780, https://doi.org/10.1073/pnas.96.9.4780
[26] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883, https://doi.org/10.1016/S1074-5521(01)00063-1.
[27] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413, https://doi.org/10.1146/annurev.biochem.052308.105824.
[28] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720, https://doi.org/10.1021/ja900553w.
[29] J. C. Anderson, P. G. Schultz, *Biochemistry* **2003**, *42*, 9598, https://doi.org/10.1021/bi034550w.
[30] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, *Science* **2013**, *342*, 357, https://doi.org/10.1126/science.1241459.
[31] Z. Fan, J. Song, T. Guan, X. Lv, J. Wei, *ACS Synth. Biol.* **2018**, *7*, 249, https://doi.org/10.1021/acssynbio.7b00290.
[32] T. Mukai, H. Hoshi, K. Ohtake, M. Takahashi, A. Yamaguchi, A. Hayashi, S. Yokoyama, K. Sakamoto, *Sci. Rep.* **2015**, *5*, https://doi.org/10.1038/srep09699.
[33] H. W. Orton, H. Qianzhu, E. H. Abdelkader, E. I. Habel, Y. J. Tan, R. L. Frkic, C. J. Jackson, T. Huber, G. Otting, *J. Am. Chem. Soc.* **2021**, *143*, 19587, https://doi.org/10.1021/jacs.1c10104.
[34] H. Qianzhu, E. H. Abdelkader, I. D. Herath, G. Otting, T. Huber, *ACS Sens.* **2022**, *7*, 44, https://doi.org/10.1021/acssensors.1c02467.
[35] E. H. Abdelkader, H. Qianzhu, Y. J. Tan, L. A. Adams, T. Huber, G. Otting, *J. Am. Chem. Soc.* **2021**, *143*, 1133, https://doi.org/10.1021/jacs.0c11971.
[36] M. Hauf, F. Richter, T. Schneider, T. Faidt, B. M. Martins, T. Baumann, P. Durkin, H. Dobbek, K. Jacobs, A. Möglich, N. Budisa, *ChemBioChem* **2017**, *18*, 1819, https://doi.org/10.1002/cbic.201700327.
[37] H. Qianzhu, A. P. Welegedara, H. Williamson, A. E. McGrath, M. C. Mahawaththa, N. E. Dixon, G. Otting, T. Huber, *J. Am. Chem. Soc.* **2020**, *142*, 17277, https://doi.org/10.1021/jacs.0c07976.
[38] J. N. Beyer, P. Hosseinzadeh, I. Gottfried-Lee, E. M. Van Fossen, P. Zhu, R. M. Bednar, P. A. Karplus, R. A. Mehl, R. B. Cooley, *J. Mol. Biol.* **2020**, *432*, 4690, https://doi.org/10.1016/j.jmb.2020.06.014.
[39] A. J. Burke, S. L. Lovelock, A. Frese, R. Crawshaw, M. Ortmayer, M. Dunstan, C. Levy, A. P. Green, *Nature* **2019**, *570*, 219, https://doi.org/10.1038/s41586-019-1262-8.
[40] I. Drienovská, C. Mayer, C. Dulson, G. Roelfes, *Nat. Chem.* **2018**, *10*, 946, https://doi.org/10.1038/s41557-018-0082-z.
[41] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, *118*, 142, https://doi.org/10.1021/acs.chemrev.7b00014.
[42] S. Fischer, T. R. Ward, A. D. Liang, *ACS Catal.* **2021**, *11*, 6343, https://doi.org/10.1021/acscatal.1c01470.
[43] A. Stein, A. D. Liang, R. Sahin, T. R. Ward, *J. Organomet. Chem.* **2022**, *962*, https://doi.org/10.1016/j.jorganchem.2022.122272.
[44] T. Völker, E. Meggers, *ChemBioChem* **2017**, *18*, 1083, https://doi.org/10.1002/cbic.201700168.
[45] I. Drienovská, R. A. Scheele, C. Gutiérrez de Souza, G. Roelfes, *ChemBioChem* **2020**, *21*, 3077, https://doi.org/10.1002/cbic.202000306.
[46] R. B. Leveson-Gower, Z. Zhou, I. Drienovská, G. Roelfes, *ACS Catal.* **2021**, *11*, 6763, https://doi.org/10.1021/acscatal.1c00996.
[47] R. B. Leveson-Gower, R. M. de Boer, G. Roelfes, *ChemCatChem* **2022**, *14*, https://doi.org/10.1002/cctc.202101875.
[48] H. Z. Duan, C. Hu, Y. L. Li, S. H. Wang, Y. Xia, X. Liu, J. Wang, Y. X. Chen, *J. Am. Chem. Soc.* **2022**, *144*, 22831, https://doi.org/10.1021/jacs.2c09683.
[49] A. K. Bhagat, H. Buium, G. Shmul, L. Alfonta, *ACS Catal.* **2020**, *10*, 3094, https://doi.org/10.1021/acscatal.9b03739.
[50] M. Voss, C. Xiang, J. Esque, A. Nobili, M. J. Menke, I. André, M. Höhne, U. T. Bornscheuer, *ACS Chem. Biol.* **2020**, *15*, 416, https://doi.org/10.1021/acschembio.9b00888.
[51] A. D. Pagar, H. Jeon, T. P. Khobragade, S. Sarak, P. Giri, S. Lim, T. H. Yoo, B. J. Ko, H. Yun, *Front. Chem.* **2022**, *10*, https://doi.org/10.3389/fchem.2022.839636.
[52] P. E. Affaticati, S. B. Dai, P. Payongsri, H. C. Hailes, K. Tittmann, P. A. Dalby, *Sci. Rep.* **2016**, *6*, https://doi.org/10.1038/srep35716.
[53] H. C. Wilkinson, P. A. Dalby, *FEBS J.* **2021**, *288*, 1935, https://doi.org/10.1111/febs.15560.
[54] E. C. Minnihan, M. R. Seyedsayamdost, U. Uhlin, J. Stubbe, *J. Am. Chem. Soc.* **2011**, *133*, 9430, https://doi.org/10.1021/ja201640n.
[55] E. C. Minnihan, D. D. Young, P. G. Schultz, J. Stubbe, *J. Am. Chem. Soc.* **2011**, *133*, 15942, https://doi.org/10.1021/ja207719f
[56] G. Kang, A. T. Taguchi, J. Stubbe, C. L. Drennan, *Science* **2020**, *368*, 424, https://doi.org/10.1126/ science.aba6794
[57] C. Y. Lin, A. L. Muñoz Hernández, T. N. Laremore, A. Silakov, C. Krebs, A. K. Boal, J. M. Bollinger, *ACS Catal.* **2022**, *12*, 6968, https://doi.org/10.1021/acscatal.2c01037.
[58] M. Ortmayer, K. Fisher, J. Basran, E. M. Wolde-Michael, D. J. Heyes, C. Levy, S. L. Lovelock, J. L. R. Anderson, E. L. Raven, S. Hay, S. E. J. Rigby, A. P. Green, *ACS Catal.* **2020**, *10*, 2735, https://doi.org/10.1021/acscatal.9b05129.
[59] M. Ortmayer, F. J. Hardy, M. G. Quesne, K. Fisher, C. Levy, D. J. Heyes, C. R. A. Catlow, S. P. De Visser, S. E. J. Rigby, S. Hay, A. P. Green, *J. Am. Chem. Soc.* **2021**, *1*, 913, https://doi.org/10.1021/jacsau.1c00145.
[60] Q. Liu, Q. He, X. Lyu, F. Yang, Z. Zhu, P. Xiao, Z. Yang, F. Zhang, Z. Yang, X. Wang, P. Sun, Q. wen Wang, C. Qu, Z. Gong, J. Lin, Z. Xu, S. Song, S. Huang, S. Guo, M. jie Han, K. kai Zhu, X. Chen, A. W. Kahsai, K.-H. Xiao, W. Kong, F. Li, K. Ruan, Z. Li, X. Yu, X. Niu, C. Jin, J. Wang, J. Sun, *Nat. Commun.* **2020**, *11*, https://doi.org/10.1038/s41467-020-18433-5.
[61] X. Liu, P. Liu, H. Li, Z. Xu, L. Jia, Y. Xia, M. Yu, W. Tang, X. Zhu, C. Chen, Y. Zhang, E. Nango, R. Tanaka, F. Luo, K. Kato, Y. Nakajima, S. Kishi, H. Yu, N. Matsubara, S. Owada, K. Tono, S. Iwata, L. J. Yu, J. R. Shen, J. Wang, *Nat. Chem.* **2022**, *14*, 1054, https://doi.org/10.1038/s41557-022-00992-3.
[62] J. Zhao, A. J. Burke, A. P. Green, *Curr. Opin. Chem. Biol.* **2020**, *55*, 136, https://doi.org/10.1016/j.cbpa.2020.01.006.
[63] Y. Li, P. A. Dalby, *Biosci. Rep.* **2022**, *42*, https://doi.org/10.1042/BSR20220168.
[64] M. R. Seyedsayamdost, S. Y. Reece, D. G. Nocera, J. A. Stubbe, *J. Am. Chem. Soc.* **2006**, *128*, 1569, https://doi.org/10.1021/ja055926r.
[65] R. A. Mehl, J. C. Anderson, S. W. Santoro, L. Wang, A. B. Martin, D. S. King, D. M. Horn, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 935, https://doi.org/10.1021/ja0284153.
[66] Y. Chen, A. Loredo, A. Chung, M. Zhang, R. Liu, H. Xiao, *J. Mol. Biol.* **2022**, *434*, https://doi.org/10.1016/j.jmb.2021.167412.

[67] J. S. Trimble, R. Crawshaw, F. J. Hardy, C. W. Levy, M. J. B. Brown, D. E. Fuerst, D. J. Heyes, R. Obexer, A. P. Green, *Nature* **2022**, *611*, 709, https://doi.org/10.1038/s41586-022-05335-3.
[68] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583, https://doi.org/10.1038/s41586-021-03819-2.
[69] D. Patsch, R. Buller, *CHIMIA* **2023**, *77*, https://doi.org/10.2533/chimia.2023.116.

License and Terms

The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2022.395