

Review Article

# Engineering of enzymes using non-natural amino acids

Yiwen Li and  Paul A. Dalby

Department of Biochemical Engineering, University College London, London, Gower Street WC1E 6BT, U.K.

Correspondence: Paul A. Dalby (p.dalby@ucl.ac.uk)



In enzyme engineering, the main targets for enhancing properties are enzyme activity, stereoselective specificity, stability, substrate range, and the development of unique functions. With the advent of genetic code extension technology, non-natural amino acids (nnAAs) are able to be incorporated into proteins in a site-specific or residue-specific manner, which breaks the limit of 20 natural amino acids for protein engineering. Benefitting from this approach, numerous enzymes have been engineered with nnAAs for improved properties or extended functionality. In the present review, we focus on applications and strategies for using nnAAs in enzyme engineering. Notably, approaches to computational modelling of enzymes with nnAAs are also addressed. Finally, we discuss the bottlenecks that currently need to be addressed in order to realise the broader prospects of this genetic code extension technique.

## Introduction

As biocatalysts, enzymes offer many advantages over traditional chemical catalysts. They are produced by living cells, offer higher catalytic efficiency, and also provide the capability of catalysing reactions in a highly stereoselective manner. Most importantly, as a biodegradable substance, enzymes can replace many polluting catalysts and reduce tons of environmentally unfriendly substances, in line with the concept of 'green chemistry' [1]. In recent decades, with continuous technological breakthroughs in enzyme engineering, their application in industry, agriculture, medicine and health, energy development, and environmental engineering have become increasingly widespread. According to a market analysis report, the global biocatalysts market will witness a CAGR of 6.90% during the forecast period of 2022–2029 [2]. All indications showed that there is a very promising market for biocatalysts, and with the development of related technologies, the engineering of enzyme functions, and properties including activity, stereoselective specificity, substrate range, thermal, and solvent stability, etc. has become a hot topic [3].

The main traditional strategies for enzyme engineering include direct evolution, rational design, and semirational design [4,5]. However, the limited ranges of side chains from the 20 natural amino acids limit the possibilities to develop more exotic functionalities. In living organisms, although cofactors or chemical modifications have expanded the range of amino acids to some extent, they still have certain drawbacks. With the development of genetic code expansion (GCE) technology, non-natural amino acids (nnAAs), also known as noncanonical amino acids (ncAAs) or nonstandard amino acids, can be efficiently and accurately translated into proteins *in vivo* and *in vitro* with the help of orthogonal incorporation systems [6–8]. When designing enzymes with artificial residues, the two key points are the choice of substitution sites and the type of nnAAs. Thus, nnAAs and enzyme backbones should first be considered separately. For example, the substitution sites within the scaffold can be selected by traditional protein design methods, considering location, potential impact on stability and function, and any known mutability of the site. The main considerations in selecting potential nnAAs are their chemistry (and potential impact on stability and function), current availability or ease of synthesis, and then the cost and efficiency of their corresponding incorporation systems. The combination of these options might then be considered by the

Received: 16 May 2022  
Revised: 05 July 2022  
Accepted: 20 July 2022

Accepted Manuscript online:  
20 July 2022  
Version of Record published:  
05 August 2022

following rules [9]: (1) The replacement nnAA should be structurally similar to the original one. (2) The nnAA is more likely to be successfully inserted into sites close to the N-terminal than the C-terminal of the protein. (3) The introduction of more than two nnAAs into a protein can significantly reduce the yield. In addition, due to the need to use live cells for expression, the inserted nnAAs need to be harmless to the cells and stable. Following these basic principles, numerous nnAAs have been successfully utilised in enzymes to improve their catalytic properties or to develop unique functions over the last two decades [9,10]. In the present review, we have focused on the applications and strategies for using nnAAs in enzyme engineering. Notably, approaches to computational modelling of enzymes with nnAAs are also addressed. Finally, we discuss the bottlenecks that currently need to be addressed in order to realise the broader prospects of using nnAAs in enzyme engineering.

## Orthologous aminoacyl-tRNA synthetase–tRNA pairs: key for the site-specific incorporation of nnAAs

To transfer the genetic information on mRNA into the encoded protein through the translation system, engineered tRNAs must only link the desired nnAA but not canonical amino acids to the reassigned codon [11]. New exogenous aminoacyl-tRNA synthetase (aaRS) is also required in this step to aminoacylate only the corresponding tRNA and no endogenous tRNAs. These exogenous elements, including orthogonal aaRS/tRNA pairs, reassigned codons, and nnAAs, form an orthogonal translation system, which means a system with little or no cross-over with the natural biological system (Figure 1). To date, more than two-thirds of nnAAs expressed in *Escherichia coli* (*E. coli*) and mammalian cells have been achieved by adopting *Methanocaldococcus jannaschii* (*Mj*) TyrRS-tRNA<sup>Tyr</sup> and *Methanosarcina mazei*/*Methanosarcina barkeri* PylRS-tRNA<sup>Pyl</sup> pairs or their evolved variants [11–13]. The former has been shown to introduce more than 100 unnatural side chains in different organisms, mainly phenylalanine and its derivatives, while the latter showed the ability to incorporate more than 50 tyrosine analogues [14,15]. In addition, a growing number of newly developed aaRS/tRNA pairs have been shown to introduce a wider variety of nnAAs, such as LeuRS/tRNA<sub>CUA</sub> [16], TrpRS/tRNA<sub>CUA</sub> [17], SepRS/tRNA<sub>CUA</sub> [18], and LysRS/tRNA<sub>CUA</sub> [19] pairs. However, they often suffer from small host range, a small variety of suitable nnAAs, and low efficiency; therefore, it is extremely important to evolve more efficient orthologous systems.

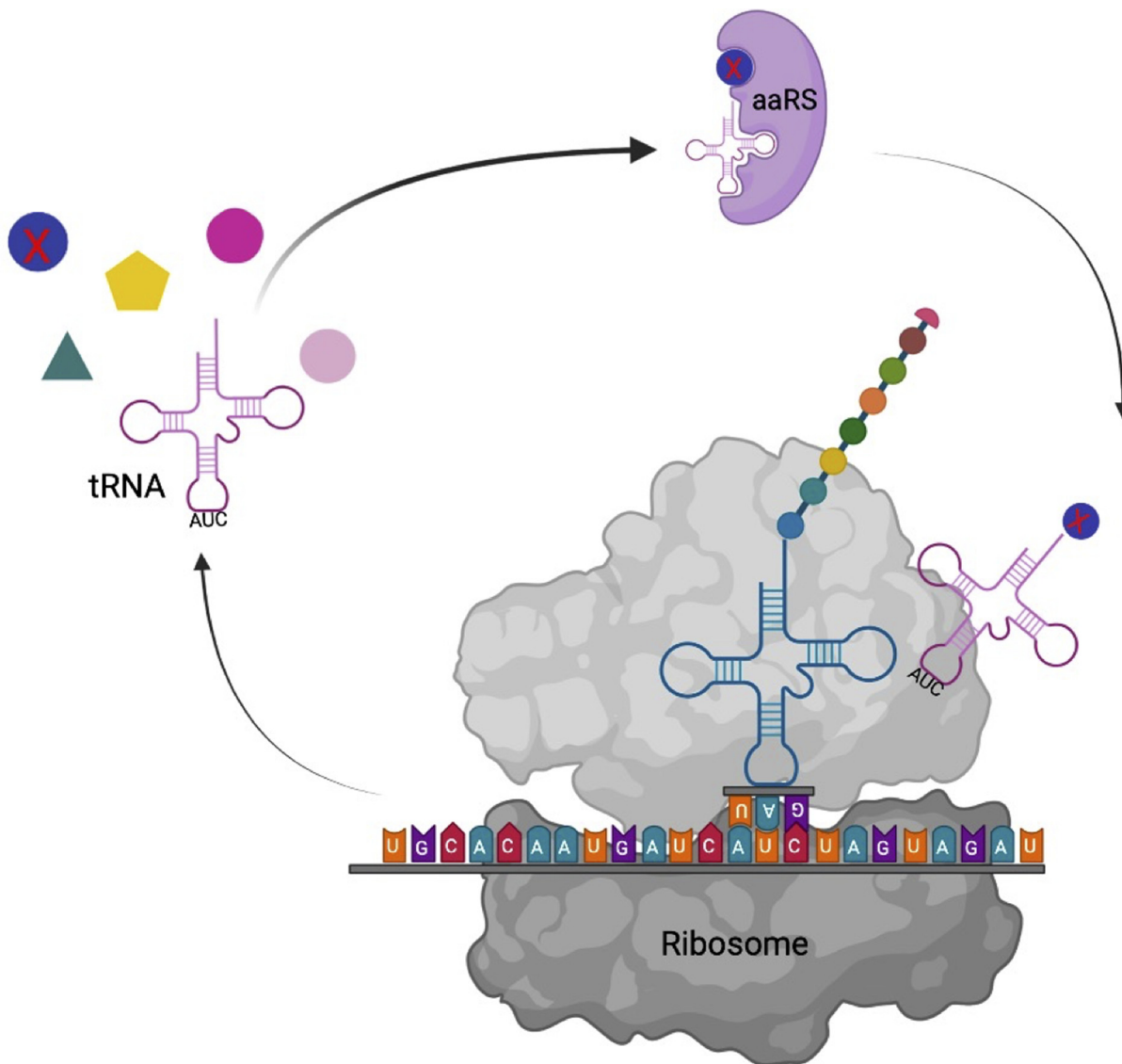
Directed evolution provides a powerful tool for screening orthologous aaRS–tRNA pairs with high specificity [20,21]. Target residues within the aaRS active site are mutated to generate libraries, which are then subjected to a two-step screening process over multiple cycles to select mutant aaRS/tRNAs that accept only the desired nnAAs. Based on this method, an increasing number of orthologous pairs have been developed, enabling more than 200 nnAAs to be successfully incorporated into proteins [22]. Using Pyrrolysyl-tRNA synthetase as an example, its high tolerance to substrate side chains and low selectivity towards  $\alpha$ -amine and tRNA anticodons allow it to be relatively easily evolved to accommodate more nnAAs or  $\alpha$ -hydroxy acids into proteins [23]. Comparing to native PylRS, these engineered variants exhibit a higher efficiency of nnAA incorporation, or can carry lysine and phenylalanine derivatives that the native one does not recognise. Among these variants, the N346A/C348A double-mutant constructed by Liu's group was shown to act on a wide range of substrates including a series of *p*-alkoxy-Phe, *m*-substituted Phe, and *o*-substituted Phe derivatives in *E. coli* and mammalian cells [24–27]. Notably, the vast majority of evolved aaRS–tRNA pairs only target the amber codon [13], which means that incorporating multiple nnAAs would require the use of alternative codons or new strategies. In general, to further increase the variety of side chains that can be genetically encoded, it is essential to evolve more incorporation systems with better properties.

## Applications

### Improve the activity of enzymes

The use of 'hotspots' to guide the engineering of enzymes has gradually grown as a viable and efficient strategy [28]. Changes to structure-based hotspots, including active-site residues, tunnel sites, flexible sites, distal sites coupled to the active centre, and subunit interfacial sites, are highly likely to alter the activity, stability, selective specificity, and other properties of the enzyme. With the widespread use of unnatural amino acid-based techniques, the replacement of these sites with unnatural amino acids has led to a significant increase in the activity of numerous enzymes.

Due to the high incorporation efficiency and ease of synthesis, derivatives of phenylalanine are widely used in enzyme engineering, with 25 commonly used phenylalanine derivatives listed in Figure 2. Transaminases (TAMs) are one of the most promising biocatalysts in organic synthesis for the preparation of chiral amino compounds. Replacement of the Phe88 residue in the active site of (*R*)-amine TAM with *p*-benzoylphenylalanine (*p*BzF, **4**) showed a pronounced improvement in the activity towards 1-phenylpropan-1-amine and benzaldehyde [29]. The introduction

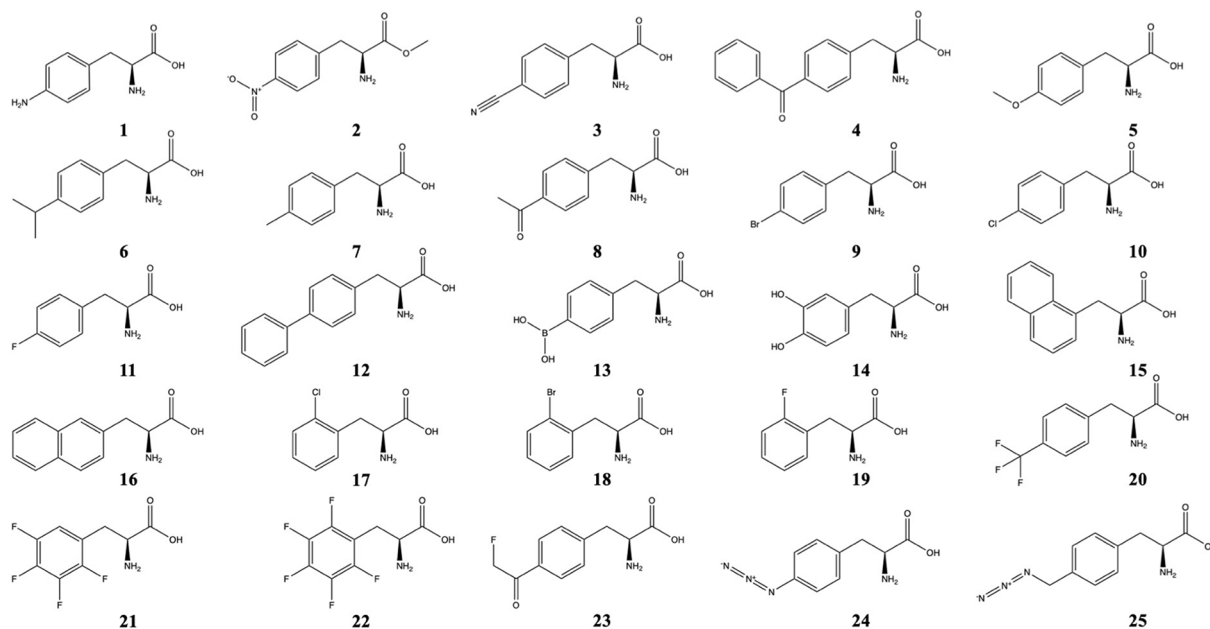


**Figure 1.** Schematic showing the process of incorporation of nnAAs into proteins via orthologous aaRS–tRNA pairs

of a single *pBzF* into the active residue reshaped the size of the active pocket while maintaining hydrophobicity, thus further enhancing the range of substrates.

Another example targeted the aromatic residue Phe385 in transketolase (TK), which has shown its critical role in new acceptor substrate binding. In 2020, this site was replaced with a series of derivatives of phenylalanine (1–3) on a previously created TK scaffold to reduce the aromatic ring electron density [30]. The results showed that the *p*-cyanophenylalanine (*p*CNF, 3) mutant simultaneously increased the activity and stability of TK towards 3-hydroxybenzaldehyde (3-HBA), most likely due to the conversion of this functional group from an electron donor to an acceptor, which formed a new hydrogen bond with G262 to stabilise a local helix-turn structure.

*E. coli* nitroreductase (NTR) is being used as a prodrug activator in cancer therapy [31]. Guided by structural information, Phe124 in the active site acts as the key to substrate binding. This site was replaced by eight nnAAs with the help of an evolved *Mj*TyrRS/tRNA<sup>Tyr</sup><sub>CUA</sub> pair. Variants incorporating *p*AMF 1, *p*Bpa 4, *p*MF 7, *p*tfmF 20, and *p*NF-NTR 2 showed better catalytic efficiency towards its substrates CB1954 or LH7. Of these, *p*AF-NTR showed a more than 30-fold increase in activity, probably due to the enhanced  $\pi$ -stacking with polarised aromatic substrates at the 124 site, which also facilitated hydride transfer [32].

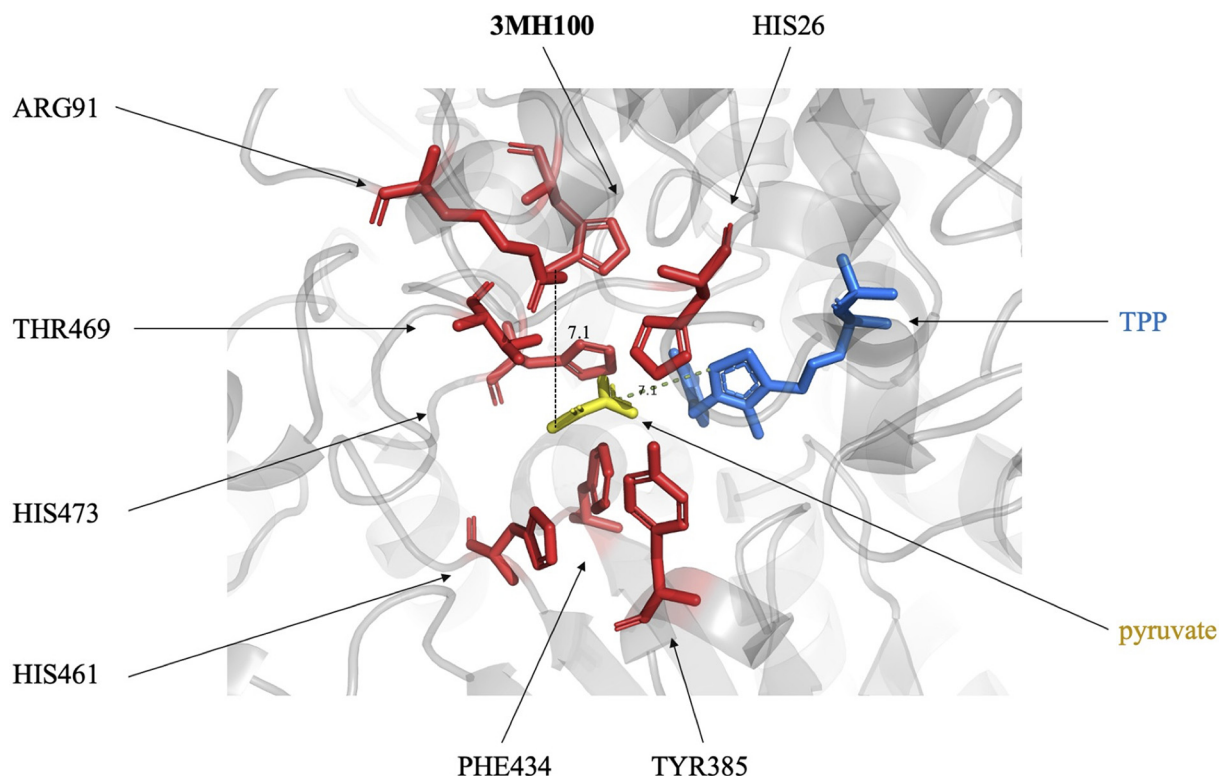


**Figure 2. Structures of 25 commonly used phenylalanine derivatives**

1, *p*-aminophenylalanine (pAMF); 2, *p*-nitrophenylalanine (pNTF); 3, *p*-cyanophenylalanine (pCNF); 4, *p*-benzoylphenylalanine (pBzF); 5, *p*-methoxyphenylalanine (pMOF); 6, *p*-isopropyl-phenylalanine; 7, *p*-methylphenylalanine (pMF); 8, *p*-acetylphenylalanine; 9, *p*-bromophenylalanine (pBrF); 10, *p*-chlorophenylalanine (pClF); 11, *p*-fluorophenylalanine (pFF); 12, *p*-biphenylalanine; 13, *p*-boronophenylalanine; 14, L-3,4-dihydroxyphenylalanine (DOPA); 15, 3-(1-naphthyl)-L-alanine (Na-pAla); 16, 3-(2-Naphthyl)-L-alanine; 17, *o*-chlorophenylalanine (oClF); 18, *o*-bromophenylalanine (oBrF); 19, *o*-fluorophenylalanine; 20, *p*-trifluoromethylphenylalanine (ptfmF); 21, 2,3,4,5-tetrafluorophenylalanine (F<sub>4</sub>F); 22, pentafluoro-phenylalanine (F<sub>5</sub>F); 23, *p*-fluoroacetylphenylalanine; 24, *p*-azidophenylalanine (pAzF); 25, *p*-azidomethyl-phenylalanine (pAzMeF).

Histidine is an essential amino acid, present in the active site of countless enzymes due to the ability of the imidazole side chain to act as a general acid and general base in catalysis, or to chelate metal ions such as in metalloproteins [33]. So far, about 20 analogues of histidine have been reported to be incorporated into proteins [34–37]. A mutant of pyrrolysyl-tRNA synthetase (PylRS – N346A/C348A) can recognise two of the most widely used histidine mimetics [37], 3-methylhistidine (3MH) and thiazole alanine (TA). Replacement of three active-site histidine residues in alanine racemase with 3MH and TA, led to a dramatic change in catalytic efficiency [37]. Similar results were obtained when a histidine in the substrate binding pocket of TK was replaced by 3MH. By analysing the crystal structures of TK complexed with donor substrates, the previous studies showed H100 of *E. coli* and yeast TK forms hydrogen bonds with the C1-hydroxy groups of its donor substrates including D-xylulose-5-phosphate (X5P) and D-fructose-6-phosphate (F6P) [38–40], and so switching the substrate to pyruvate would remove this hydrogen bond due to the lack of the hydroxyl group. Thus, we conjectured that by adding a methyl group to H100, its hydrophobic interaction with pyruvate may be enhanced, thus allowing the substrate to be anchored. Briefly, by re-assigning the sense codon at position H100 of *E. coli* TK to an amber codon, the original histidine was successfully replaced by 3MH with the help of an evolved *Mb*PylRS/tRNA<sup>Pyl</sup> pair (PylHRS; L270I, Y271F, L274G, C313F, and Y349F) [35]. However, the results showed a decrease in activity (unpublished data). The docking conformation of pyruvate with H100-3MH TK containing cofactor thiamine pyrophosphate (TPP) predicted by Autodock Vina [41] indicated that the methyl group on 3MH may have become too distant from pyruvate for hydrophobic interaction to occur. Combining the docking conformation and a molecular dynamics simulation in GROMACS [42], the addition of the methyl group also appeared to induce a conformational change in the surrounding amino acids, which ultimately shifted the pyruvate away from TPP, and so leading to a decrease in activity (Figure 3, unpublished data).

The metal-chelating ligands in metalloenzymes are extremely important and the introduction of nnAAs into key ligand sites can dramatically alter their catalytic properties [43,44]. It was reported that N<sub>δ</sub>-methyl histidine was incorporated into several metalloenzymes to create augmented properties and novel activities [45]. For instance, N<sub>δ</sub>-methyl histidine installed as the proximal heme ligand in the oxygen-binding protein, myoglobin (Mb), can cause



**Figure 3.** The docking conformation between pyruvate and the H100-3MH TK variant containing cofactor TPP predicted by Autodock Vina

The structure of the H100-3MH TK variant was built on a triple-mutant variant (PDB 5HHT) by Molecular Operating Environment (MOE) [48], energy minimisation was performed by GROMACS in amber99sb-ild forcefield [42]. Pyruvate was drawn in Chem3D-17.0 and energy minimised by MM2 calculations. The active site residues within 4 Å of the substrate are shown as red sticks, donor substrate pyruvate, and cofactor TPP are shown as yellow and blue sticks, respectively. The distance between methyl group on 3MH100 and pyruvate is shown as black-dashed line (7.1 Å), between pyruvate and cofactor TPP is shown as yellow-dashed line (7.1 Å).

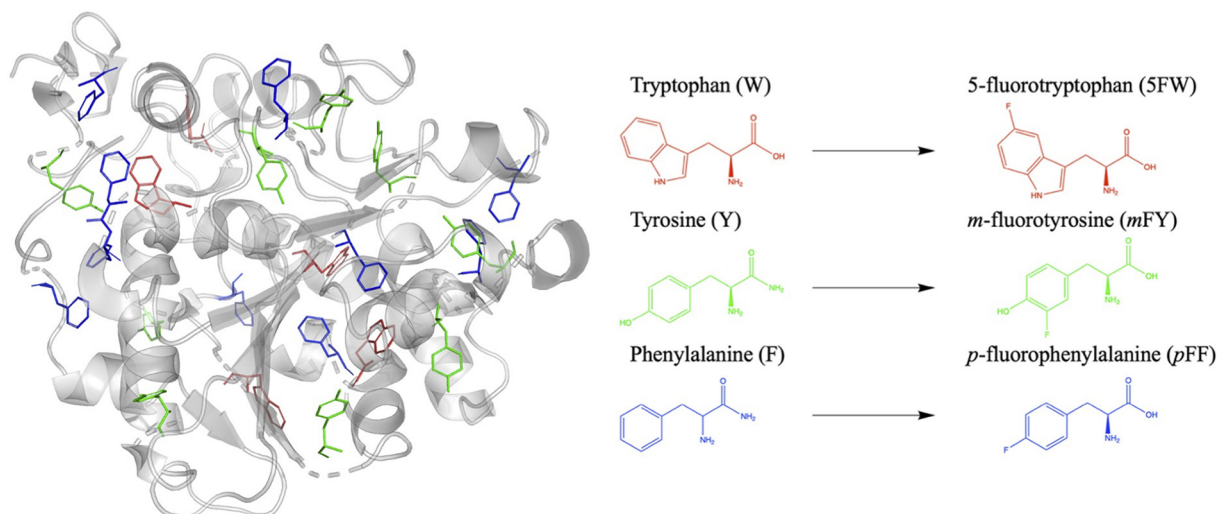
a dramatic increase in heme redox potential and promiscuous activity [46]. Similarly, the incorporation of the same nnAA into a modified ascorbate peroxidase resulted in improved robustness towards irreversible deactivation [47].

In addition to the examples above, enzymes engineered with analogues of tyrosine [11,49], proline [50], tryptophan [51], and lysine [52,53] have also been shown to improve catalytic properties. In a functional Mb-based oxidase, replacement of the conserved tyrosine with a series of Tyr analogues with reduced pKa in the phenolic ring demonstrated enhanced activity due to increased protonatability of Tyr [54]. In general, the substitution of amino acids with analogues at the active sites of enzymes has become an ideal method for fine-tuning enzyme activity. It is particularly advantageous that analogues can provide both a higher 'atom level' resolution and wider range of modifications to side chain chemistry in terms of sterics, electronics and noncovalent bonding, than the canonical 20 amino acids, as these are more likely to be compatible with the modified site, while also enabling large changes in catalytic properties.

### Improve the stability of enzymes

Stability is a fundamental property of enzymes and has significant impact on their structure, expression, solubility, and function [55,56]. The stability of proteins is due to a combination of factors such as the packing of the hydrophobic core, the level of burial in the hydrophobic surface region, the stability of secondary structure elements, the number of hydrogen bonds, and salt bridges on the protein surface, etc. [57]. The abundant range of side chains of nnAAs certainly provides more options for engineering these factors with the aim of improving the stability of enzymes.

For dimeric enzymes, the use of nnAAs can strengthen the interaction between the monomers and improve the stability of the enzyme. In the research conducted by Schultz group in 2018, the substitution of Phe 21 with



**Figure 4.** The fluorination of aromatic amino acid residues in *Candida antarctica* lipase B N74D

The structures of tryptophan, tyrosine, and phenylalanine in the original PDB structure model (PDB 1TCC) are shown in red, green, and blue sticks, respectively. They were fluorinated experimentally to 5-fluorotryptophan (5FW), *m*-fluorotyrosine (*m*FY), and *p*FF, respectively.

*p*-benzoylphenylalanine (pBzF **4**) in *E. coli* homoserine O-succinyltransferase (*metA*) increased the melting temperature by 21 °C [58]. This may result from the formation of a covalent adduct between the keto groups of Cys 90 and pBzF, which stabilises the dimeric structure of the enzyme. Extensively fluorinated (fluorous) amino acids at dimeric interfaces also seem particularly effective in increasing protein stability [59]. Substitution of a phenylalanine with *p*-fluorophenylalanine (*p*FF **11**) successfully increased the refoldability and thermoactivity of S5 phosphotriesterase (PTE) [60]. Similarly, the  $T_m$  increased by 19 °C in glutathione transferase when all four tryptophan residues in each monomer were replaced with 5-fluorotryptophan [61]. However, of the many members of the fluorinated amino acid family, is it true that the greater the fluorination the more effective it is in enhancing stability? In 2008, three phenylalanine-fluorinated analogues: pentafluoro-L-phenylalanine (F<sub>5</sub>F **22**), 2,3,4,5-tetrafluoro-L-phenylalanine (F<sub>4</sub>F **21**), and 2,3,5,6-tetrafluoro-L-phenylalanine (Zp) were incorporated into the villin headpiece subdomain (HP35) to investigate the relationship between the degree of fluorination and stability. The results showed that Phe10F<sub>4</sub>F had a melting temperature that was 14 °C higher than that of WT, while Phe10F<sub>5</sub>F was only 2 °C higher. This strongly suggested that the **highly but incompletely** fluorinated side chains could provide enhanced hydrophobicity while retaining the ArH... $\pi$  interactions, thus providing a stable protein fold [62].

Compared with site-specific substitution, residue-specific incorporation enables global incorporation of nnAAs, which is widely used as a strategy for obtaining mutants with desired properties [63]. In general, residue-specific incorporation of nnAAs can be achieved by multiple biochemical methods in addition to genetic code extension techniques, as the latter is currently inefficient on its own for incorporating nnAAs at multiple sites. A commonly used approach for global fluorination is to supply appropriate nnAAs in auxotrophic strains during expression of desired proteins. By applying this method, all tryptophan, tyrosine, and phenylalanine residues in a glycosylation-deficient mutant of *Candida antarctica* lipase B were substituted by 5-fluoro-L-tryptophan, *o*-fluoro-(DL)-tyrosine, and *p*-fluoro-L-phenylalanine **11** (Figure 4), respectively [64]. This global fluorination prolonged the shelf life of the lipase activity, which is an essential beneficial feature for storage. Following the same strategy, when all prolines were substituted with fluorinated analogues in GFP, the tendency of *cis-trans* isomerisation of the Pro residue changed, and therefore the stability of protein. In comparison, site-specific substitution through genetic encoding has minimal disturbance to the structure. Both methods have their advantages and limitations, thus they should be chosen appropriately according to the purpose.

In addition to activity and stability, the use of nnAAs to improve other enzymatic properties, including stereoselectivity, substrate specificity, reversibility, and immobilisation capacity have also been extensively reported [65,66], and details will not be discussed in the present review. Overall, the emergence of unnatural amino acids has given a new lease of life to enzyme engineering and holds great prospects for the future.

**Table 1 Summary for caged tyrosines and the responding aaRS**

nnAA	aaRS	Protein	Strain	Reference
ONBY	ONBTyrRS	Sperm-whale myoglobin	DH10B <i>E. coli</i>	[71]
		$\beta$ -galactosidase	DH10B <i>E. coli</i>	[71]
	BpaRS	<i>Taq</i> DNA polymerase	BL21 <i>E. coli</i>	[79]
		T7 RNA polymerase	BL21 <i>E. coli</i>	[81]
	MjYRSONBY	Zinc-Finger nuclease	BL21 <i>E. coli</i> , HEK 293T	[82]
		Imidazole glycerol phosphate synthase	BL21 <i>E. coli</i>	[85]
	ONBYRS	Superfold green fluorescent protein	DH10B <i>E. coli</i> , HEK 293T	[77]
	ONB-DOPARS-1	3,4-Dihydroxyphenyl-alanine (DOPA)-rich mussel adhesive proteins (MAPs)	B-95. $\Delta$ A <i>E. coli</i>	[80]
	ONBYRS-1	Superfold green fluorescent protein	BL21(DE3) <i>E. coli</i>	[88]
	NPYRS	Mutated firefly luciferase	HEK 293T	[78]
NPY	NPYRS	Mutated firefly luciferase	HEK 293T	[78]
		TEV protease	HEK 293T	[78]
	MjYRSONBY	Imidazole glycerol phosphate synthase	BL21 <i>E. coli</i>	[78]
		Mutated firefly luciferase	HEK 293T	[78]
MNPY	NPYRS	TEV protease	HEK 293T	[78]
		Mutated firefly luciferase	HEK 293T	[78]
NPPY	NPYRS	Mutated firefly luciferase	HEK 293T	[78]
		TEV protease	HEK 293T	[78]

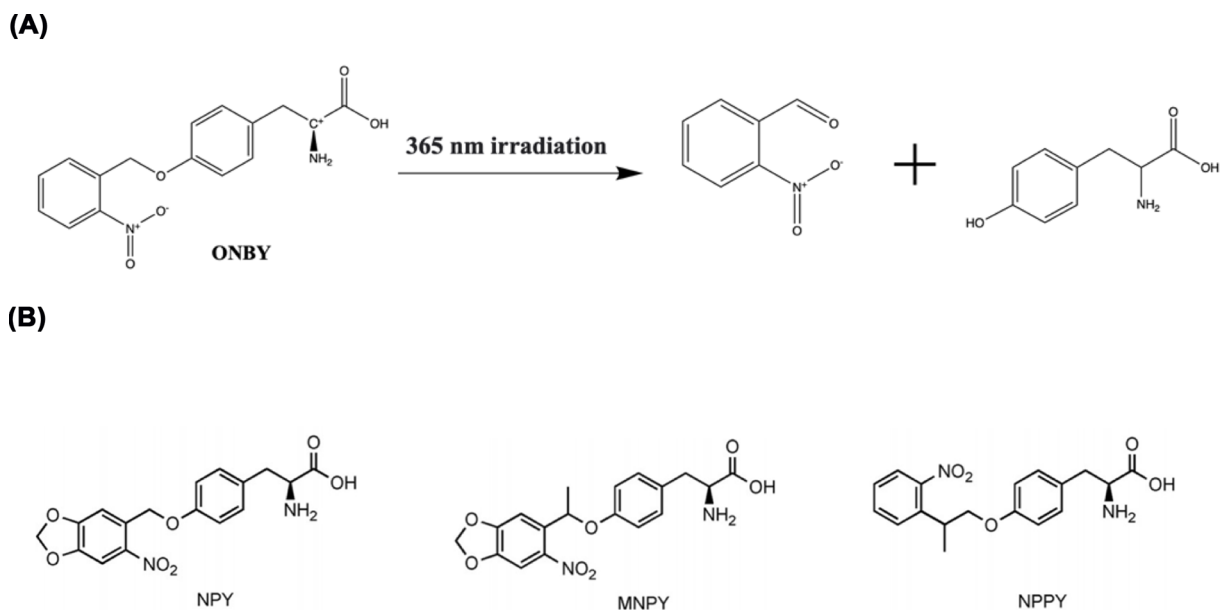
NPY, *o*-nitropiperonyltyrosine.

## Optical control of enzymes

The precise spatial and temporal control of enzymes has always been a hot topic. Compared with chemical activation methods, the use of light as an external trigger minimises adverse effects and is therefore of wide interest. The genetic encoding of nnAAs with light-sensitive groups enables the control of a wide range of biological processes *in vivo* and *in vitro* [67]. For instance, by replacing the sulfhydryl and hydroxyl groups of cysteine, serine, and tyrosine with nitrobenzyl groups, photomodulation of proteins can be achieved by the decaging of protecting groups under 365 nm light. To date, the main types of nnAAs that can be inserted into the enzyme to act as protectors are caged lysines [68,69], caged tyrosines [70,71], photoswitchable phenylalanine derivatives [72,73], caged cysteines [74], and selenocysteines [75,76].

Among all photocaging groups, nitrobenzyl groups and their derivatives are the most prestigious, and have been utilised in enzymes through a variety of evolved incorporation systems (Table 1). In the catalogue of tyrosine analogues, the most used is *o*-nitrobenzyltyrosine (ONBY) (Figure 5A). A number of studies have focused on designing enzymes with ONBY to achieve optical control of them [71,77–81]. One example is the creation of light-activated DNA and RNA polymerases. By rational design, ONBY was used to substitute Y61 in the *Thermus aquaticus* (*Taq*) DNA polymerase active site, where its *o*-nitrophenyl group impedes the extensive mechanical movement of the O-helix and finger-like structural domains, also interferes with the ability to accommodate dNTPs. Normal function of the *Taq* polymerase can be easily restored with the shedding of the protecting group by a 5-min exposure to 365-nm UV light [79]. Following the same strategy, a light-activatable bacteriophage T7 RNA polymerase (T7RNAP) was further generated. The light regulation of enzyme is achieved by removing the *o*-nitrophenyl group to allow the incoming nucleotides to reach the active site of the enzyme [81]. Moreover, the use of ONBY has had an impact in gene-editing. Zinc-finger nucleases (ZFN) have been developed as an important tool for gene editing due to their function in sequence-selective dsDNA cleavage. Photoactivated ZFN would allow conditional generation of gene editing with unprecedented accuracy while tackling toxicity issues [82].

Although ONBY has demonstrated its strength in generating photoactivated proteins, it has been noted that *o*-nitrobenzyl has an absorption maximum at 254 nm, resulting in a ‘blue shift’ from the commonly used 365-nm light source. In contrast, the three other caged tyrosine analogues, namely methyl *o*-nitropiperonyltyrosine (MNPY), *o*-nitropiperonyltyrosine (NPY) and *o*-nitrophenylpropyltyrosine (NPPY) (Figure 5B) showed better photochemical properties. Their incorporation into a protease demonstrated the ability of off-to-on switching [78]. However, due to



**Figure 5. Use of caged-tyrosine analogues**

**(A)** The deprotected process for ONBY. **(B)** The structure of o-nitropiperonyltyrosine (NPY), MNPY, and NPPY.

the complex synthetic routes, low incorporation efficiency and fidelity, these derivatives are far less applicable than ONBY.

Carrying the same nitrobenzyl-derivative groups, caged lysines have also been used to replace lysine residues in numerous enzymes including kinases [83], Cre recombinase [70,84], Cas9 nuclease [84], isocitrate dehydrogenase [85], imidazole glycerol phosphate synthase [86], and so forth to achieve optical control. Intriguingly, spontaneous decaging of nitrobenzene groups and their derivatives in *E. coli* in the absence of UV irradiation has also been reported (with the decaging ratio of 24–85%), possibly due to bacterial NTR-mediated elimination [87]. Furthermore, bulky side chains of caged nnAAs make the selection of insertion sites challenging. Thus, although nitrobenzyl groups have demonstrated their ability to optically control proteins *in vivo* and *in vitro*, its effectiveness in decaging varies widely in reports and remains as a challenge to be addressed.

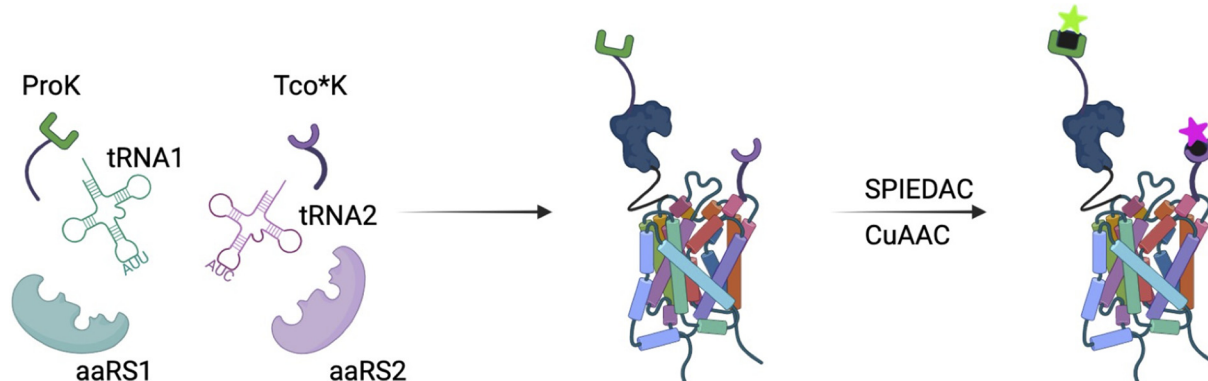
### nnAAs in enzymes as probes

In addition to the applications mentioned above, enzymatic engineering with the participation of nnAAs also demonstrates its benefits in an ever-increasing range of aspects as the variety of available sidechains increases. Some nnAAs with functional groups can be incorporated into enzymes as probes, which play important roles in the characterisation of protein structures, protein interactions, and the detection of protein dynamics. In general, nnAAs probes can be classified into the following categories: (1) infrared (IR) and NMR probes, (2) photocross-linkers, (3) probes with heavy atoms, (4) fluorescent probes, and (5) spin-label probes [89].

Fluorescence approaches have been widely used to elucidate the dynamics of proteins. However, global, non-specific fluorescence often alters the natural structure of proteins. The site-specific insertion of nnAAs-carrying fluorescent moieties is a promising way to overcome this challenge. For instance, since the incorporation system of L-3-(6-acetylnaphthalen-2-ylamino)-2-aminopropanoic acid (ANAP) was first evolved in 2003, it has been widely used as a fluorescent probe in eukaryotic and prokaryotic systems, benefiting from its small size and high sensitivity [90]. The dynamics of numerous proteins, including luciferase [91], GFP, and voltage-sensing phosphatases [92], have been revealed by the nature of ANAP's spectral redshift in response to environmental polarisation. It has also shown great ability in depicting the structure of multiple channels and gates, which includes but is not limited to KCNH channels [93], CNGA1 channels [94], ASIC channels [95], TRPV1 channels [96], and Nav1.5 inactivation gates [92].

The click reaction, copper(I)-catalysed alkyl-azide cyclisation (CuAAC, click reaction) is another classical method for labelling [97]. The introduction of nnAAs containing azide or alkynes allows labelling of an increasing number of proteins, including epidermal growth factor receptor (EGFR) [98], *E. coli* outer membrane protein C (Omp C) [99], mannitol dehydrogenase (MDH), and formate dehydrogenase (FDH) [100]. However, the cytotoxicity caused by the





**Figure 6.** Scheme illustrating dual incorporation of Tco\*K and ProK into a receptor and the subsequent dual bio-orthogonal labelling

Cu(I) metal catalyst limited its application in living cells. To enable a broader application in bioorthogonal labelling, numerous strategies including strain-promoted Cu(I)-free azide-alkyne cycloaddition (SPAAC), chelation-assisted CuAAC, and ligand-assisted CuAAC have been developed [101].

Furthermore, dual labelling has become feasible due to the emergence of multiple nnAA incorporation systems, which is especially useful in Förster resonance energy transfer (FRET) experiments. The reported coupled nnAAs include *p*-azidophenylalanine (pAzF)/N6-((2-propynoxy) carbonyl)-L-lysine [102], *p*-acetylphenylalanine/alkynyllysine [103], pAzF/Tet3.0 [104], 4-biphenyl-L-phenylalanine/L-(7-hydroxycoumarin-4-yl)-ethylglycine [105], and have been successfully employed to illustrate the dynamic changes in a wide range of proteins. An elegant example of this is the dual site-specific fluorescent labelling toolbox developed by Meineke *et al.* By combining *trans*-cyclooct-2-ene-L-lysine (TCO\*K) and *N*-Propargyl-L-lysine (ProK) incorporation with CuAAC/SPIEDAC, the labelling of Notch receptors and a G protein-coupled receptor (GPCR) has been achieved for the first time in living cells (Figure 6). These two bio-orthogonal reactions allow precise control and direction for probing and manipulation of proteins in space and time, and therefore hold promise for a wide range of applications [106].

## Designing enzymes with nnAAs *in silico*

Advances in computational technology have greatly facilitated the identification of hotspots for enzyme modification, including active-site residues, access tunnel sites, flexible sites, interface sites, etc. [28]. In particular, with the help of molecular modelling software such as GROMACS [42], CHARMM [107], AMBER [108], and FoldX [109], numerous studies have been carried out on enzymes to investigate their biochemical reactivity, thermodynamics, and interactions with substrates and solvents. However, most of them cannot be applied to simulations of proteins containing nnAAs as few contain parameterisations for unnatural side chains. To achieve simulations of proteins containing a wider variety of nnAAs, manually adding parameters for nonstandard residues obtained from multiple tools is an option, but this is a tedious and time-consuming process. More importantly, tools for autogenerated parameters are not compatible with all force fields and often lack good validation. Therefore, easier, faster, and experimentally validated methods for the *in-silico* introduction of nnAAs have become an urgent need. The good news is that extensive studies have recently provided parameters for many commonly used nnAAs in different force fields (Table 2). The Zagrovic group has developed over 250 post-translational modification parameters on the new GROMOS 54a7 force field, which has been verified to exhibit near-chemical accuracy in measuring hydration-free energies. This high-resolution computational tool greatly extends the range of artificial side chains that can be modelled [110]. Moreover, the Flodas lab released an online tool called Forcefield\_NCAA that provides charge parameters for a large library of 147 nnAAs compatible with AMBERff03. Similarly, developed parameters of 18 derivatives of phenylalanine and tyrosine were validated in Amber ff14SB [111]. Recently, more and more nnAAs with specific sidechains such as halogenated groups, azido and alkynyl groups [112] are being parametrised.

Rosetta, a powerful tool for the *ab initio* engineering of enzymes has been applied to design variants with diverse goals [113]. A number of enzymes involved in the Dies-Alder [114], Retro-aldol [115], and Kemp elimination [116] reactions have been designed using this approach and validated experimentally. However, while tools have been designed in Rosetta to handle artificial groups [117–119], the use of Rosetta to design nnAAs-containing enzymes has hardly been applied in practise.

**Table 2 Summary for forcefields modified with nnAAs parameters**

Ref. of modified forcefields	Compatible forcefield	Applicable nnAAs
[120]	CHARMM36	406 nnAAs
[121]	CHARMM19	Methylated lysines and arginines, acetylated lysines
[121]	CHARMM127	Methylated lysines and arginines, acetylated lysines
[111]	Amber ff14SB	18 phenylalanine and tyrosine derivatives
[122]	AMBERff03	147 nnAAs
[123]	AMBER ff15ipq	Fluorinated nnAAs
[124]	CHARMM22	13 nnAAs
[125]	CHARMM general force field (CGenFF)	nnAAs with Azido and Alkynyl R-Groups
[112]	COMPASS force field	nnAAs with Aliphatic Azide Chains
[110]	GROMOS 45a3 and 54a7	250 different types of Post-Translational Modification

## Challenges and prospects

The examples presented show that genetically encoded nnAAs have certainly become a hot spot in the field of biocatalyst engineering. However, it is of concern that its development is still limited by multiple factors. Firstly, a large number of nnAAs have very complex synthetic and downstream purification routes, which limit their mass production and lead to high costs [126]. This also directly contributes to the difficulty of producing protein products containing nnAAs on a large scale in industry. To overcome this challenge, it is of great importance to develop more cost-effective methods for synthesising nnAAs. Two general routes are chemical synthetic routes and biocatalytic routes. The former often suffers from a lack of enantioselectivity and stereoselectivity [127]. In contrast, biosynthetic routes are highly attractive due to their simple reaction conditions, low energy consumption, and high enantiomeric excess (*e.e.*). For instance, L-tert-leucine is an important precursor for the manufacture of protease inhibitors, which are of great value in clinical applications. The chemical synthesis of L-tert-leucine is limited by low stereoselectivity, low yields, and high pollution. In the development of its biosynthetic method, an evolved leucine dehydrogenase was screened out using a high-throughput spectrophotometric screening assay, resulting in a high yield of 1170 g/day [128].

Another issue of great importance is the incorporation efficiency/fidelity of nnAAs. To date, hundreds of nnAAs have been reported to be genetically encoded, but it is doubtful how many of these can be introduced by orthogonal translation systems at high efficiency and with high fidelity. In the research of our group, a methodology to quantify incorporation fidelity from deconvoluted mass spectra was introduced, and it was used to evaluate the incorporation performance of some of the most commonly used phenylalanine analogues in a specific site of TK [30]. Of these, *p*CNF had a misincorporation rate of around 22%, while it was as high as about 55% for *p*AMF. Although this depends on a range of factors, including protein structure, the specific site of insertion and other elements on the plasmid, it should still place a high priority on the most vital factor: the effectiveness of the orthologous tRNA/synthase pairs. As mentioned above, more than two-thirds of nnAA expression was achieved by adopting TyrRS-tRNA<sup>Tyr</sup> or PylRS-tRNA<sup>Pyl</sup> systems, which leads to the result that only analogues of tyrosine and pyrrolysine can be introduced in a relatively efficient manner [6]. To enable the use of a high diversity of artificial residues, more orthogonal translation systems need to be evolved.

In summary, nnAAs have already been extensively used in enzyme engineering. Their powerful functional side chains have conferred new features or improved the properties of numerous enzymes, and there is undoubtedly considerable potential to expand their applications. However, there are still many obstacles to overcome to make this more wide-ranging, efficient, and economical, though we are already on the way.

### Competing Interests

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

### Open Access

Open access for the present article was enabled by the participation of University College London in an all-inclusive *Read & Publish* agreement with Portland Press and the Biochemical Society under a transformative agreement with JISC.

### CRedit Author Contribution

**Yiwen Li:** Conceptualisation, Writing—original draft. **Paul A. Dalby:** Conceptualisation, Supervision, Writing—review & editing.

## Acknowledgements

We thank the Engineering and Physical Sciences Research Council (EPSRC) Future Targeted Healthcare Manufacturing Hub [grant no. EP/P006485/1] for financial support.

## Abbreviations

3-HBA, 3-hydroxybenzaldehyde; 3MH, 3-methyl-histidine; aaRS, aminoacyl-tRNA synthetase; ANAP, L-3-(6-acetylnaphthalen-2-ylamino)-2-aminopropanoic acid; CAGR, Compound Annual Growth Rate; CuAAC, copper(I)-catalysed alkyl-azide cyclisation; DOPA, 3,4-dihydroxyphenyl-alanine (DOPA; F4F, 2,3,4,5-tetrafluoro-L-phenylalanine; F5F, pentafluoro-L-phenylalanine; F6P, D-fructose-6-phosphate; FRET, Förster resonance energy transfer; GCE, genetic code expansion; GFP, green fluorescent protein; IR, infrared; Mb, myoglobin; Mj, *Methanocaldococcus jannaschii*; MNPY, methyl o-nitropiperonyltyrosine; MOE, molecular operating environment; ncAA, noncanonical amino acid; nnAA, non-natural amino acid; NMR, nuclear magnetic resonance; NPPY, o-nitrophenylpropyltyrosine; NPY, o-nitropiperonyltyrosine; NTR, nitroreductase; ONBY, o-nitrobenzyltyrosine; pAzF, p-azidophenylalanine; pBzF, p-benzoylphenylalanine; pCNF, p-cyanophenylalanine; PDB, protein databank; pFF, p-fluorophenylalanine; ProK, N-propargyl-L-lysine; PTE, phosphotriesterase; PylRS, pyrrolysyl-tRNA synthetase; TA, thiazole alanine; TAm, transaminase; Taq, *thermus aquaticus*; TCO\*K, trans-cyclooct-2-ene-L-lysine; TK, transketolase; TPP, thiamine pyrophosphate; UV, ultraviolet; X5P, D-xylulose-5-phosphate; ZFN, zinc-finger nuclease.

## References

- Sheldon, R.A. (2008) E factors, green chemistry and catalysis: An odyssey. *Chem. Commun.* 3352–3365, <https://doi.org/10.1039/b803584a>
- Tiorienteering Biocatalysis & biocatalysts market analysis, size, share, growth, trends, forecast to 2029. <https://tiorienteering.ch/biocatalysis-biocatalysts-market-analysis-size-share-growth-trends-forecast-to-2029>
- Sharma, A., Gupta, G., Ahmad, T., Mansoor, S. and Kaur, B. (2021) Enzyme engineering: current trends and future perspectives. *Food Rev. Int.* **37**, 121–154, <https://doi.org/10.1080/87559129.2019.1695835>
- Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D. and Arnold, F.H. (2005) Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15**, 447–452, <https://doi.org/10.1016/j.sbi.2005.06.004>
- Chen, R. (2001) Enzyme engineering: rational redesign versus directed evolution. *Trends Biotechnol.* **19**, 13–14, [https://doi.org/10.1016/S0167-7799\(00\)01522-5](https://doi.org/10.1016/S0167-7799(00)01522-5)
- Wang, L. and Schultz, P.G. (2004) Expanding the genetic code. *Angew. Chem. Int. Ed.* **44**, 34–66
- Liu, C.C. and Schultz, P.G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444, <https://doi.org/10.1146/annurev.biochem.052308.105824>
- Young, D.D. and Schultz, P.G. (2018) Playing with the molecules of life. *ACS Chem. Biol.* **13**, 854–870, <https://doi.org/10.1021/acscchembio.7b00974>
- Zhao, J., Burke, A.J. and Green, A.P. (2020) Enzymes with noncanonical amino acids. *Curr. Opin. Chem. Biol.* **55**, 136–144, <https://doi.org/10.1016/j.cbpa.2020.01.006>
- Jackson, J.C., Duffy, S.P., Hess, K.R. and Mehl, R.A. (2006) Improving nature's enzyme active site with genetically encoded unnatural amino acids. *J. Am. Chem. Soc.* **128**, 11124–11127, <https://doi.org/10.1021/ja061099y>
- Vargas-Rodriguez, O., Sevostyanova, A., Söll, D. and Crnković, A. (2018) Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Curr. Opin. Chem. Biol.* **46**, 115–122, <https://doi.org/10.1016/j.cbpa.2018.07.014>
- Hohsaka, T. and Sisido, M. (2002) Incorporation of non-natural amino acids into proteins. *Curr. Opin. Chem. Biol.* **6**, 809–815, [https://doi.org/10.1016/S1367-5931\(02\)00376-9](https://doi.org/10.1016/S1367-5931(02)00376-9)
- Wan, W., Tharp, J.M. and Liu, W.R. (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta (BBA) - Proteins and Proteomics* **1844**, 1059–1070, <https://doi.org/10.1016/j.bbapap.2014.03.002>
- Xie, J. and Schultz, P.G. (2006) A chemical toolkit for proteins — an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **7**, 775–782, <https://doi.org/10.1038/nrm2005>
- Krahn, N., Tharp, J.M., Crnković, A. and Söll, D. (2020) Engineering aminoacyl-tRNA synthetases for use in synthetic biology. *Enzymes* **48**, 351, <https://doi.org/10.1016/bs.enz.2020.06.004>
- Hyun, S.L., Guo, J., Lemke, E.A., Dimla, R.D. and Schultz, P.G. (2009) Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* **131**, 12921–12923, <https://doi.org/10.1021/ja904896s>
- Ikeda-Boku, A., Ohno, S., Hibino, Y., Yokogawa, T., Hayashi, N. and Nishikawa, K. (2013) A simple system for expression of proteins containing 3-azidotyrosine at a pre-determined site in *Escherichia coli*. *J. Biochem.* **153**, 317–326, <https://doi.org/10.1093/jb/mvs153>
- Park, H.S., Hohn, M.J., Umehara, T., Guo, L.T., Osborne, E.M., Benner, J. et al. (2011) Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* **333**, 1151–1154, <https://doi.org/10.1126/science.1207203>
- Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S. and Schultz, P.G. (2004) An expanded genetic code with a functional quadruplet codon. *PNAS* **101**, 7566–7571, <https://doi.org/10.1073/pnas.0401517101>
- O'Donoghue, P. and Luthy-Schulten, Z. (2003) On the evolution of structure in aminoacyl-tRNA Synthetases. *Microbiol. Mol. Biol. Rev.* **67**, 550–573, <https://doi.org/10.1128/MMBR.67.4.550-573.2003>
- Dumas, A., Lercher, L., Spicer, C.D. and Davis, B.G. (2014) Designing logical codon reassignment - expanding the chemistry in biology. *Chem. Sci.* **6**, 50–69, <https://doi.org/10.1039/C4SC01534G>

- 22 Adhikari, A., Bhattarai, B.R., Aryal, A., Thapa, N., Kc, P., Adhikari, A. et al. (2021) Reprogramming natural proteins using unnatural amino acids. *RSC Adv.* **11**, 38126–38145, <https://doi.org/10.1039/D1RA07028B>
- 23 Crnković, A., Suzuki, T., Söll, D. and Reynolds, N.M. (2016) Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion. *Croat. Chem. Acta* **89**, 163–174, <https://doi.org/10.5562/cca2825>
- 24 Wang, Y.S., Fang, X., Wallace, A.L., Wu, B. and Liu, W.R. (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant has a broad substrate spectrum. *J. Am. Chem. Soc.* **134**, 2950, <https://doi.org/10.1021/ja211972x>
- 25 Tuley, A., Wang, Y.S., Fang, X., Kurra, Y., Rezenom, Y.H. and Liu, W.R. (2014) The genetic incorporation of thirteen novel non-canonical amino acids. *Chem. Commun. (Camb.)* **50**, 2673, <https://doi.org/10.1039/C3CC49068H>
- 26 Tharp, J.M., Wang, Y.S., Lee, Y.J., Yang, Y. and Liu, W.R. (2014) Genetic incorporation of seven ortho-substituted phenylalanine derivatives. *ACS Chem. Biol.* **9**, 884–890, <https://doi.org/10.1021/cb400917a>
- 27 Vatansaver, E.C., Yang, K.S., Geng, Z.Z., Qiao, Y., Li, P., Xu, S. et al. (2022) A designed, highly efficient pyrrolysyl-tRNA synthetase mutant binds o-chlorophenylalanine using two halogen bonds. *J. Mol. Biol.* **434**, 167534, <https://doi.org/10.1016/j.jmb.2022.167534>
- 28 Yu, H., Ma, S., Li, Y. and Dalby, P.A. (2022) Hot spots-making directed evolution easier. *Biotechnol. Adv.* **56**, 107926, <https://doi.org/10.1016/j.biotechadv.2022.107926>
- 29 Pagar, A.D., Jeon, H., Khobragade, T.P., Sarak, S., Giri, P., Lim, S. et al. (2022) Non-canonical amino acid-based engineering of (R)-amine transaminase. *Front. Chem.* **10**, 839636, <https://doi.org/10.3389/fchem.2022.839636>
- 30 Wilkinson, H.C. and Dalby, P.A. (2021) Fine-tuning the activity and stability of an evolved enzyme active-site through noncanonical amino-acids. *FEBS J.* **288**, 1935–1955, <https://doi.org/10.1111/febs.15560>
- 31 Searle, P.F., Chen, M.J., Hu, L., Race, P.R., Lovering, A.L., Grove, J.I. et al. (2004) Nitroreductase: a prodrug-activating enzyme for cancer gene therapy. *Clin. Exp. Pharmacol. Physiol.* **31**, 811–816, <https://doi.org/10.1111/j.1440-1681.2004.04085.x>
- 32 Jackson, J.C., Duffy, S.P., Hess, K.R. and Mehl, R.A. (2006) Improving nature's enzyme active site with genetically encoded unnatural amino acids. *J. Am. Chem. Soc.* **128**, 11124–11127, <https://doi.org/10.1021/ja061099y>
- 33 Holeček, M. (2020) Histidine in health and disease: metabolism, physiological importance, and use as a supplement. *Nutrients* **12**, 848, <https://doi.org/10.3390/nu12030848>
- 34 Xiao, H., Peters, F.B., Yang, P.Y., Reed, S., Chittuluru, J.R. and Schultz, P.G. (2014) Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. *ACS Chem. Biol.* **9**, 1092–1096, <https://doi.org/10.1021/cb500032c>
- 35 Gan, F., Liu, R., Wang, F. and Schultz, P.G. (2018) Functional replacement of histidine in proteins to generate noncanonical amino acid dependent organisms. *J. Am. Chem. Soc.* **140**, 3829–3832, <https://doi.org/10.1021/jacs.7b13452>
- 36 Ikeda, Y., Kawahara, S.I., Taki, M., Kuno, A., Hasegawa, T. and Taira, K. (2003) Synthesis of a novel histidine analogue and its efficient incorporation into a protein in vivo. *Protein Eng.* **16**, 699–706, <https://doi.org/10.1093/protein/gzg084>
- 37 Sharma, V., Wang, Y.S. and Liu, W.R. (2016) Probing the catalytic charge-relay system in alanine racemase with genetically encoded histidine mimetics. *ACS Chem. Biol.* **11**, 3305–3309, <https://doi.org/10.1021/acschembio.6b00940>
- 38 Yu, H., Hernández López, R.I., Steadman, D., Méndez-Sánchez, D., Higson, S., Cázares-Körner, A. et al. (2020) Engineering transketolase to accept both unnatural donor and acceptor substrates and produce  $\alpha$ -hydroxyketones. *FEBS J.* **287**, 1758–1776, <https://doi.org/10.1111/febs.15108>
- 39 Asztalos, P., Parthier, C., Golbik, R., Kleinschmidt, M., Hübner, G., Weiss, M.S. et al. (2007) Strain and near attack conformers in enzymic thiamin catalysis: X-ray crystallographic snapshots of bacterial transketolase in covalent complex with donor ketoses xylulose 5-phosphate and fructose 6-phosphate, and in noncovalent complex with acceptor aldose ribose 5-phosphate. *Biochemistry* **46**, 12037–12052, <https://doi.org/10.1021/bi700844m>
- 40 Wikner, C., Nilsson, U., Meshalkina, L., Udekwi, C., Lindqvist, Y. and Schneider, G. (1997) Identification of catalytically important residues in yeast transketolase. *Biochemistry* **36**, 15643–15649, <https://doi.org/10.1021/bi971606b>
- 41 Eberhardt, J., Santos-Martins, D., Tillack, A.F. and Forli, S. (2021) AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* **61**, 3891–3898, <https://doi.org/10.1021/acs.jcim.1c00203>
- 42 Bauer, P., Hess, B. and Lindahl, E. (2022) GROMACS 2022.1 Manual. <https://manual.gromacs.org/2022.1/index.html>
- 43 Mirts, E.N., Bhagi-Damodaran, A. and Lu, Y. (2019) Understanding and modulating metalloenzymes with unnatural amino acids, non-native metal ions, and non-native metal cofactors. *Acc. Chem. Res.* **52**, 935, <https://doi.org/10.1021/acs.accounts.9b00011>
- 44 Jeschek, M., Panke, S. and Ward, T.R. (2018) Artificial metalloenzymes on the verge of new-to-nature metabolism. *Trends Biotechnol.* **36**, 60–72, <https://doi.org/10.1016/j.tibtech.2017.10.003>
- 45 Hayashi, T., Hilvert, D. and Green, A.P. (2018) Engineered metalloenzymes with non-canonical coordination environments. *Chem. Eur. J.* **24**, 11821–11830, <https://doi.org/10.1002/chem.201800975>
- 46 Pott, M., Hayashi, T., Mori, T., Mittl, P.R.E., Green, A.P. and Hilvert, D. (2018) A noncanonical proximal heme ligand affords an efficient peroxidase in a globin fold. *J. Am. Chem. Soc.* **140**, 1535–1543, <https://doi.org/10.1021/jacs.7b12621>
- 47 Green, A.P., Hayashi, T., Mittl, P.R.E. and Hilvert, D. (2016) A chemically programmed proximal ligand enhances the catalytic properties of a heme enzyme. *J. Am. Chem. Soc.* **138**, 11344–11352, <https://doi.org/10.1021/jacs.6b07029>
- 48 *Molecular Operating Environment (MOE)*, 2022.02 Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2022.
- 49 Seyedsayamdost, M.R., Xie, J., Chan, C.T.Y., Schultz, P.G. and Stubbe, J. (2007) Site-specific insertion of 3-aminotyrosine into subunit  $\alpha$ 2 of E. coli ribonucleotide reductase: direct evidence for involvement of Y730 and Y731 in radical propagation. *J. Am. Chem. Soc.* **129**, 15060–15071, <https://doi.org/10.1021/ja076043y>
- 50 Merkel, L., Schauer, M., Antranikian, G. and Budisa, N. (2010) Parallel incorporation of different fluorinated amino acids: on the way to “teflon” proteins. *Chem. Bio. Chem.* **11**, 1505–1507, <https://doi.org/10.1002/cbic.201000295>

- 51 Parsons, J.F., Xiao, G., Gilliland, G.L. and Armstrong, R.N. (1998) Enzymes harboring unnatural amino acids: mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan. *Biochemistry* **37**, 6286–6294, <https://doi.org/10.1021/bi980219e>
- 52 Ugwumba, I.N., Ozawa, K., Xu, Z.Q., Ely, F., Foo, J.L., Herlt, A.J. et al. (2011) Improving a natural enzyme activity through incorporation of unnatural amino acids. *J. Am. Chem. Soc.* **133**, 326–333, <https://doi.org/10.1021/ja106416g>
- 53 Ai, H.W., Lee, J.W. and Schultz, P.G. (2010) A method to site-specifically introduce methyllysine into proteins in *E. coli*. *Chem. Commun.* **46**, 5506–5508, <https://doi.org/10.1039/c0cc00108b>
- 54 Yu, Y., Lv, X., Li, J., Zhou, Q., Cui, C., Hosseinzadeh, P. et al. (2015) 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, <https://doi.org/10.1021/ja5109936>
- 55 Polizzi, K.M., Bommaris, A.S., Broering, J.M. and Chaparro-Riggers, J.F. (2007) Stability of biocatalysts. *Curr. Opin. Chem. Biol.* **11**, 220–225, <https://doi.org/10.1016/j.cbpa.2007.01.685>
- 56 Yeoman, C.J., Han, Y., Dodd, D., Schroeder, C.M., Mackie, R.I. and Cann, I.K.O. (2010) Thermostable enzymes as biocatalysts in the biofuel industry. *Adv. Appl. Microbiol.* **70**, 1–55, [https://doi.org/10.1016/S0065-2164\(10\)70001-0](https://doi.org/10.1016/S0065-2164(10)70001-0)
- 57 Goldenzweig, A. and Fleishman, S.J. (2018) Principles of protein stability and their application in computational design. **87**, 105–129, <https://doi.org/10.1146/annurev-biochem-062917-012102>
- 58 Li, J.C., Liu, T., Wang, Y., Mehta, A.P. and Schultz, P.G. (2018) Enhancing protein stability with genetically encoded noncanonical amino acids. *J. Am. Chem. Soc.* **140**, 15997–16000, <https://doi.org/10.1021/jacs.8b07157>
- 59 Buer, B.C. and Marsh, E.N.G. (2012) Fluorine: a new element in protein design. *Protein Sci.: A Publ. Protein Soc.* **21**, 453, <https://doi.org/10.1002/pro.2030>
- 60 Baker, P.J. and Montclare, J.K. (2011) Enhanced refoldability and thermoactivity of fluorinated phosphotriesterase. *Chem. Bio. Chem.* **12**, 1845–1848, <https://doi.org/10.1002/cbic.201100221>
- 61 Parsons, J.F., Xiao, G., Gilliland, G.L. and Armstrong, R.N. (1998) Enzymes harboring unnatural amino acids: mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan. *Biochemistry* **37**, 6286–6294, <https://doi.org/10.1021/bi980219e>
- 62 Zheng, H., Comeforo, K. and Gao, J. (2009) Expanding the fluorine arsenal: tetrafluorinated phenylalanines for protein design. *J. Am. Chem. Soc.* **131**, 18–19, <https://doi.org/10.1021/ja8062309>
- 63 Johnson, J.A., Lu, Y.Y., van Deventer, J.A. and Tirrell, D.A. (2010) Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14**, 774–780, <https://doi.org/10.1016/j.cbpa.2010.09.013>
- 64 Budisa, N., Wenger, W. and Wilschi, B. (2010) Residue-specific global fluorination of *Candida antarctica* lipase B in *Pichia pastoris*. *Mol. Biosyst.* **6**, 1630–1639, <https://doi.org/10.1039/c002256j>
- 65 Zhao, J., Burke, A.J. and Green, A.P. (2020) Enzymes with noncanonical amino acids. *Curr. Opin. Chem. Biol.* **55**, 136–144, <https://doi.org/10.1016/j.cbpa.2020.01.006>
- 66 Won, Y., Pagar, A.D., Patil, M.D., Dawson, P.E. and Yun, H. (2019) Recent advances in enzyme engineering through incorporation of unnatural amino acids. *Biotechnol. Bioprocess Eng.* **24**, 592–604, <https://doi.org/10.1007/s12257-019-0163-x>
- 67 Courtney, T. and Deiters, A. (2018) Recent advances in the optical control of protein function through genetic code expansion. *Curr. Opin. Chem. Biol.* **46**, 99–107, <https://doi.org/10.1016/j.cbpa.2018.07.011>
- 68 Luo, J., Uprety, R., Naro, Y., Chou, C., Nguyen, D.P., Chin, J.W. et al. (2014) Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation. *J. Am. Chem. Soc.* **136**, 15551–15558, <https://doi.org/10.1021/ja5055862>
- 69 Luo, J., Arbely, E., Zhang, J., Chou, C., Uprety, R., Chin, J.W. et al. (2016) Genetically encoded optical activation of DNA recombination in human cells. *Chem. Commun.* **52**, 8529–8532, <https://doi.org/10.1039/C6CC03934K>
- 70 Edwards, W.F., Young, D.D. and Deiters, A. (2009) Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS Chem. Biol.* **4**, 441–445, <https://doi.org/10.1021/cb900041s>
- 71 Deiters, A., Groff, D., Ryu, Y., Xie, J. and Schultz, P.G. (2006) A genetically encoded photocaged tyrosine. *Angew. Chem. Int. Ed.* **45**, 2728–2731, <https://doi.org/10.1002/anie.200600264>
- 72 Hoppmann, C., Lacey, V.K., Louie, G.v., Wei, J., Noel, J.P. and Wang, L. (2014) Genetically encoding photoswitchable click amino acids in *Escherichia coli* and mammalian cells. *Angew. Chem. Int. Ed.* **53**, 3932–3936, <https://doi.org/10.1002/anie.201400001>
- 73 Hoppmann, C., Maslennikov, I., Choe, S. and Wang, L. (2015) In situ formation of an azo bridge on proteins controllable by visible light. *J. Am. Chem. Soc.* **137**, 11218–11221, <https://doi.org/10.1021/jacs.5b06234>
- 74 Uprety, R., Luo, J., Liu, J., Naro, Y., Samanta, S. and Deiters, A. (2014) Genetic encoding of caged cysteine and caged homocysteine in bacterial and mammalian cells. *ChemBioChem* **15**, 1793–1799, <https://doi.org/10.1002/cbic.201400073>
- 75 Mousa, R., Notis Dardashti, R. and Metanis, N. (2017) Selenium and selenocysteine in protein chemistry. *Angew. Chem. Int. Ed.* **56**, 15818–15827, <https://doi.org/10.1002/anie.201706876>
- 76 Gupta, N., Demong, L.W., Banda, S. and Copeland, P.R. (2013) Reconstitution of selenocysteine incorporation reveals intrinsic regulation by SECIS elements. *J. Mol. Biol.* **425**, 2415–2422, <https://doi.org/10.1016/j.jmb.2013.04.016>
- 77 Arbely, E., Torres-Kolbus, J., Deiters, A. and Chin, J.W. (2012) Photocaging of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **134**, 11912–11915, <https://doi.org/10.1021/ja3046958>
- 78 Luo, J., Torres-Kolbus, J., Liu, J. and Deiters, A. (2017) Genetic encoding of photocaged tyrosines with improved light-activation properties for the optical control of protease function. *ChemBioChem* **18**, 1442–1447, <https://doi.org/10.1002/cbic.201700147>

- 79 Chou, C., Young, D.D. and Deiters, A. (2009) A light-activated DNA polymerase. *Angew. Chem. Int. Ed.* **48**, 5950–5953, <https://doi.org/10.1002/anie.200901115>
- 80 Hauf, M., Richter, F., Schneider, T., Faidt, T., Martins, B.M., Baumann, T. et al. (2017) Photoactivatable mussel-based underwater adhesive proteins by an expanded genetic code. *ChemBioChem* **18**, 1819–1823, <https://doi.org/10.1002/cbic.201700327>
- 81 Chou, C., Young, D.D. and Deiters, A. (2010) Photocaged T7 RNA polymerase for the light activation of transcription and gene function in pro- and eukaryotic cells. *Chembiochem: Eur. J. Chem. Biol.* **11**, 972, <https://doi.org/10.1002/cbic.201000041>
- 82 Chou, C. and Deiters, A. (2011) Light-activated gene editing with a photocaged zinc-finger nuclease. *Angew. Chem. Int. Ed. Engl.* **50**, 6839, <https://doi.org/10.1002/anie.201101157>
- 83 Gautier, A., Deiters, A. and Chin, J.W. (2011) Light-activated kinases enable temporal dissection of signaling networks in living cells. *J. Am. Chem. Soc.* **133**, 2124–2127, <https://doi.org/10.1021/ja1109979>
- 84 Hemphill, J., Borchardt, E.K., Brown, K., Asokan, A. and Deiters, A. (2015) Optical control of CRISPR/Cas9 gene editing. *J. Am. Chem. Soc.* **137**, 5642–5645, <https://doi.org/10.1021/ja512664v>
- 85 Walker, O.S., Elsässer, S.J., Mahesh, M., Bachman, M., Balasubramanian, S. and Chin, J.W. (2016) Photoactivation of mutant isocitrate dehydrogenase 2 reveals rapid cancer-associated metabolic and epigenetic changes. *J. Am. Chem. Soc.* **138**, 718–721, <https://doi.org/10.1021/jacs.5b07627>
- 86 Kneuttinger, A.C., Winter, M., Simeth, N.A., Heyn, K., Merkl, R., König, B. et al. (2018) Artificial light regulation of an allosteric bienzyme complex by a photosensitive ligand. *Chembiochem: Eur. J. Chem. Biol.* **19**, 1750–1757, <https://doi.org/10.1002/cbic.201800219>
- 87 Liu, L., Liu, Y., Zhang, G., Ge, Y., Fan, X., Lin, F. et al. (2018) Genetically encoded chemical decaging in living bacteria. *Biochemistry* **57**, 446–450, <https://doi.org/10.1021/acs.biochem.7b01017>
- 88 Baumann, T., Hauf, M., Richter, F., Albers, S., Möglich, A., Ignatova, Z. et al. (2019) Computational aminoacyl-tRNA synthetase library design for photocaged tyrosine. *Int. J. Mol. Sci.* **20**, 2343, <https://doi.org/10.3390/ijms20092343>
- 89 Lee, K.J., Kang, D. and Park, H.S. (2019) Site-specific labeling of proteins using unnatural amino acids. *Mol. Cells* **42**, 386
- 90 Xiang, Z. and Wang, L. (2011) Enantiospecific synthesis of genetically encodable fluorescent unnatural amino acid l-3-(6-acetylnaphthalen-2-ylamino)-2-aminopropanoic acid. *J. Org. Chem.* **76**, 6367–6371, <https://doi.org/10.1021/jo2007626>
- 91 Hsieh, T.Y., Nillegoda, N.B., Tyedmers, J., Bukau, B., Mogk, A. and Kramer, G. (2014) Monitoring protein misfolding by site-specific labeling of proteins in vivo. *PLoS ONE* **9**, e99395, <https://doi.org/10.1371/journal.pone.0099395>
- 92 Shandell, M.A., Quejada, J.R., Yazawa, M., Cornish, V.W. and Kass, R.S. (2019) Detection of Nav1.5 conformational change in mammalian cells using the noncanonical amino acid ANAP. *Biophys. J.* **117**, 1352–1363, <https://doi.org/10.1016/j.bpj.2019.08.028>
- 93 Dai, G., James, Z.M. and Zagotta, W.N. (2018) Dynamic rearrangement of the intrinsic ligand regulates KCNH potassium channels. *J. Gen. Physiol.* **150**, 625–635, <https://doi.org/10.1085/jgp.201711989>
- 94 Aman, T.K., Gordon, S.E. and Zagotta, W.N. (2016) Regulation of CNGA1 channel gating by interactions with the membrane. *J. Biol. Chem.* **291**, 9939–9947, <https://doi.org/10.1074/jbc.M116.723932>
- 95 Wen, M., Guo, X., Sun, P., Xiao, L., Li, J., Xiong, Y. et al. (2015) Site-specific fluorescence spectrum detection and characterization of hASIC1a channels upon toxin mambalgin-1 binding in live mammalian cells. *Chem. Commun.* **51**, 8153–8156, <https://doi.org/10.1039/C5CC01418B>
- 96 Zagotta, W.N., Gordon, M.T., Senning, E.N., Munari, M.A. and Gordon, S.E. (2016) Measuring distances between TRPV1 and the plasma membrane using a noncanonical amino acid and transition metal ion FRET. *J. Gen. Physiol.* **147**, 201–216, <https://doi.org/10.1085/jgp.201511531>
- 97 Hong, V., Steinmetz, N.F., Manchester, M. and Finn, M.G. (2010) Labeling live cells by copper-catalyzed alkyne-azide click chemistry. *Bioconjug. Chem.* **21**, 1912–1916, <https://doi.org/10.1021/bc100272z>
- 98 Ge, Y., Fan, X. and Chen, P.R. (2016) A genetically encoded multifunctional unnatural amino acid for versatile protein manipulations in living cells. *Chem. Sci.* **7**, 7055–7060, <https://doi.org/10.1039/C6SC02615J>
- 99 Link, A.J. and Tirrell, D.A. (2003) Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition. *J. Am. Chem. Soc.* **125**, 11164–11165, <https://doi.org/10.1021/ja036765z>
- 100 Lim, S.I., Mizuta, Y., Takasu, A., Kim, Y.H. and Kwon, I. (2014) Site-specific bioconjugation of a murine dihydrofolate reductase enzyme by copper(I)-catalyzed azide-alkyne cycloaddition with retained activity. *PLoS ONE* **9**, e98403, <https://doi.org/10.1371/journal.pone.0098403>
- 101 Li, L. and Zhang, Z. (2016) Development and applications of the copper-catalyzed azide-alkyne cycloaddition (CuAAC) as a bioorthogonal reaction. *Molecules* **21**, 1393, <https://doi.org/10.3390/molecules21101393>
- 102 Wang, K., Sachdeva, A., Cox, D.J., Wilf, N.W., Lang, K., Wallace, S. et al. (2014) Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* **6**, 393–403, <https://doi.org/10.1038/nchem.1919>
- 103 Kim, J., Seo, M.H., Lee, S., Cho, K., Yang, A., Woo, K. et al. (2013) Simple and efficient strategy for site-specific dual labeling of proteins for single-molecule fluorescence resonance energy transfer analysis. *Anal. Chem.* **85**, 1468–1474, <https://doi.org/10.1021/ac303089v>
- 104 Bednar, R.M., Jana, S., Kuppa, S., Franklin, R., Beckman, J., Antony, E. et al. (2021) Genetic incorporation of two mutually orthogonal bioorthogonal amino acids that enable efficient protein dual-labeling in cells. *ACS Chem. Biol.* **16**, 2612–2622, <https://doi.org/10.1021/acscchembio.1c00649>
- 105 Chen, S., Fahmi, N.E., Wang, L., Bhattacharya, C., Benkovic, S.J. and Hecht, S.M. (2013) Detection of dihydrofolate reductase conformational change by FRET using two fluorescent amino acids. *J. Am. Chem. Soc.* **135**, 12924–12927, <https://doi.org/10.1021/ja403007r>
- 106 Meineke, B., Heimgärtner, J., Eirich, J., Landreh, M. and Elsässer, S.J. (2020) Site-specific incorporation of two ncAAs for two-color bioorthogonal labeling and crosslinking of proteins on live mammalian cells. *Cell Reports* **31**, 107811, <https://doi.org/10.1016/j.celrep.2020.107811>
- 107 Brooks, B.R., Brooks, C.L., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B. et al. (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545, <https://doi.org/10.1002/jcc.21287>
- 108 Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M. et al. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688, <https://doi.org/10.1002/jcc.20290>

- 109 Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic. Acids. Res.* **33**, W382–388, <https://doi.org/10.1093/nar/gki387>
- 110 Petrov, D., Margreitter, C., Grandits, M., Oostenbrink, C. and Zagrovic, B. (2013) A systematic framework for molecular dynamics simulations of protein post-translational modifications. *PLoS Comput. Biol.* **9**, e1003154, <https://doi.org/10.1371/journal.pcbi.1003154>
- 111 Wang, X. and Li, W. (2020) Development and testing of force field parameters for phenylalanine and tyrosine derivatives. *Front. Mol. Biosci.* **7**, 350, <https://doi.org/10.3389/fmolb.2020.608931>
- 112 Smith, A.K., Wilkerson, J.W. and Knotts, T.A. (2020) Parameterization of unnatural amino acids with azido and alkynyl R-groups for use in molecular simulations. *J. Phys. Chem. A* **124**, 6246–6253, <https://doi.org/10.1021/acs.jpca.0c04605>
- 113 Leman, J.K., Weitzner, B.D., Lewis, S.M., Adolf-Bryfogle, J., Alam, N., Alford, R.F. et al. (2020) Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **17**, 665–680, <https://doi.org/10.1038/s41592-020-0848-2>
- 114 Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., Lambert, A.R., St. Clair, J.L. et al. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science (New York, N.Y.)* **329**, 309, <https://doi.org/10.1126/science.1190239>
- 115 Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Röthlisberger, D., Zanghellini, A. et al. (2008) De novo computational design of retro-aldol enzymes. *Science (New York, N.Y.)* **319**, 1387, <https://doi.org/10.1126/science.1152692>
- 116 Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J. et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195, <https://doi.org/10.1038/nature06879>
- 117 Renfrew, P.D., Choi, E.J., Bonneau, R. and Kuhlman, B. (2012) Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS ONE* **7**, e32637, <https://doi.org/10.1371/journal.pone.0032637>
- 118 Mills, J.H., Khare, S.D., Bolduc, J.M., Forouhar, F., Mulligan, V.K., Lew, S. et al. (2013) Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *J. Am. Chem. Soc.* **135**, 13393–13399, <https://doi.org/10.1021/ja403503m>
- 119 Watkins, A.M., Craven, T.W., Renfrew, P.D., Arora, P.S. and Bonneau, R. (2017) Rotamer libraries for the high-resolution design of  $\beta$ -amino acid foldamers. *Structure* **25**, 1771.e3–1780.e3, <https://doi.org/10.1016/j.str.2017.09.005>
- 120 Croitoru, A., Park, S.J., Kumar, A., Lee, J., Im, W., Mackerell, A.D. et al. (2021) Additive CHARMM36 force field for nonstandard amino acids. *J. Chem. Theory Comput.* **17**, 3554–3570, <https://doi.org/10.1021/acs.jctc.1c00254>
- 121 Grauffel, C., Stote, R.H. and DeJaegere, A. (2010) Force field parameters for the simulation of modified histone tails. *J. Comput. Chem.* **31**, 2434–2451, <https://doi.org/10.1002/jcc.21536>
- 122 Khoury, G.A., Smadbeck, J., Tamamis, P., Vandris, A.C., Kieslich, C.A. and Floudas, C.A. (2014) Forcefield-NCAA: Ab initio charge parameters to aid in the discovery and design of therapeutic proteins and peptides with unnatural amino acids and their application to complement inhibitors of the compstatin family. *ACS Synthetic Biol.* **3**, 855–869, <https://doi.org/10.1021/sb400168u>
- 123 Yang, D.T., Gronenborn, A.M. and Chong, L.T. (2022) Development and validation of fluorinated, aromatic amino acid parameters for use with the AMBER ff15ipq protein force field. *J. Physical Chem. A* **126**, 2286–2297, <https://doi.org/10.1021/acs.jpca.2c00255>
- 124 Olubiya, O.O. and Strodel, B. (2016) Topology and parameter data of thirteen non-natural amino acids for molecular simulations with CHARMM22. *Data in Brief* **9**, 642, <https://doi.org/10.1016/j.dib.2016.09.051>
- 125 Smith, A.K., Wilkerson, J.W. and Knotts, T.A. (2020) Parameterization of unnatural amino acids with azido and alkynyl R-groups for use in molecular simulations. *J. Phys. Chem. A* **124**, 6246–6253, <https://doi.org/10.1021/acs.jpca.0c04605>
- 126 Narancic, T., Almaboub, S.A. and O'Connor, K.E. (2019) Unnatural amino acids: production and biotechnological potential. *World. J. Microbiol. Biotechnol.* **35**, 1–11, <https://doi.org/10.1007/s11274-019-2642-9>
- 127 Bommarius, A.S., Schwarm, M., Stingl, K., Kottenhahn, M., Huthmacher, K. and Drauz, K. (1995) Synthesis and use of enantiomerically pure tert-leucine. *Tetrahedron Asymmetry* **6**, 2851–2888, [https://doi.org/10.1016/0957-4166\(95\)00377-0](https://doi.org/10.1016/0957-4166(95)00377-0)
- 128 Zhu, L., Wu, Z., Jin, J.M. and Tang, S.Y. (2016) Directed evolution of leucine dehydrogenase for improved efficiency of l-tert-leucine synthesis. *Appl. Microbiol. Biotechnol.* **100**, 5805–5813, <https://doi.org/10.1007/s00253-016-7371-5>