Chemical modifications of proteins and their applications in metalloenzyme studies

Nathchar Naowarojna, Ronghai Cheng, Juan Lopez, Christina Wong, Lu Qiao, Pinghua Liu

Department of Chemistry, Boston University, Boston, MA, 02215, United States

ARTICLE INFO

Keywords:
Metalloenzymes
Bioorthogonal reactions
Protein engineering
Non-canonical amino acids
Biocatalysis

ABSTRACT

Protein chemical modifications are important tools for elucidating chemical and biological functions of proteins. Several strategies have been developed to implement these modifications, including enzymatic tailoring reactions, unnatural amino acid incorporation using the expanded genetic codes, and recognition-driven transformations. These technologies have been applied in metalloenzyme studies, specifically in dissecting their mechanisms, improving their enzymatic activities, and creating artificial enzymes with non-natural activities. Herein, we summarize some of the recent efforts in these areas with an emphasis on a few metalloenzyme case studies.

Proteins have a remarkable range of catalytic, structural, and regulatory functions using the 20 canonical amino acids (cAAs) as their building blocks. The functions of proteins are further diversified to a level well beyond those possessed by the 20 cAAs through enzymatic-driven post-translational modifications (PTMs), such as methylations, sulfations, phosphorylations, and glycosylations. Moreover, chemical modifications of a protein of interest (POI) with special probes enable functional characterizations of the POI, including dynamics, localization, and protein-protein interactions [1,2]. In pharmaceutical settings, the modification of protein therapeutics with polyethylene glycol (PEG) enhances their stability and circulation half-life [3]. Due to these reasons, a wide variety of protein modification methodologies have been explored to acquire exquisite control of these macromolecules. In this review, we discuss some of the recent strategies for protein modifications as well as several metalloenzyme case studies focusing on mechanistic investigation and enzyme engineering.

1. Chemical modifications of proteins

Protein chemical modification approaches can be roughly classified into three categories: 1) modifications via the reactivities of cAAs; 2) ribosomal-mediated incorporation of noncanonical amino acids (ncAAs); 3) modifications via affinity-driven ligand-directed reactions. This section presents the recent progress in each of these approaches.

1.1. Chemical modifications of canonical amino acids (cAAs)

Chemical modifications of amino acids offer straightforward approaches to modify proteins. Both the reactivity and relative abundance of these residues play crucial roles in determining the selectivity of the chemical modifications. Reactions targeting highly abundant residues such as Lys often result in a mixture of modified proteins; nevertheless, significant progress has been made to obtain site-selective Lys modifications [1,2]. On the other hand, low abundance residues such as Cys and Tyr might offer an opportunity for site-selective protein modifications. Many of these reactions can be performed within living cells as part of routine experiments. Currently, methods applicable for studies in multicellular organisms are of great interest in this field.

1.1.1. Lysine modifications

Among the 20 cAAs, Lys is an attractive choice for modifications due to both its high natural abundance and the number of biocompatible chemical reactions with its nucleophilic primary amine. Lys can be modified using electrophiles, such as squaric acids, isothiocyanates, sulfonyl chlorides, and 2-amino-2-methoxyethyl (Fig. 1a–e) [1,4,5]. Because of its high natural abundance, Lys modifications are often useful in cases...
where site-selectivity is not crucial or where multiple modification sites are desired. For instance, the NHS-based ortho-nitrobenzyl esters and aldehydes are used to modify vitronectin, an extracellular matrix glycoprotein, in photopatternable hydrogels. This system provides simplified environments for studying cellular responses to physiological cues [6]. By reversibly patterning the presentation of vitronectin on the hydrogel, spatial and temporal control over osteogenic stem cell differentiation is achieved.

Due to its high natural abundance, site-specific modifications of Lys residues are challenging. In recent years, methods for achieving site-selectivity have emerged. In one case, by taking advantage of the different reactivities of Lys residues in various local microenvironments, Matos et al. computationally designed sulfonyl acrylates which afforded chemo- and regio-selective modifications [7]. Using the designed sulfonyl acrylates, a single Lys residue in several proteins, including the therapeutic antibody Trastuzumab, was modified without disturbing the structure and function of the POIs. With this method, an improved site-selectivity is achieved because the reaction favors Lys residues with the lowest $pK_a$ in the POIs (Lys33 of lysozyme with a $pK_a$ of 9.5, Lys300 of annexin V with a $pK_a$ of 10.3, and Lys100 of C2Am with a $pK_a$ of 10.1). However, this strategy might not be applicable in some POIs where the Lys with the lowest $pK_a$ is inaccessible. In another case, to prepare an antibody-drug conjugate (ADC), Rader and co-workers site-specifically conjugated $\beta$-lactam derivatives to a Lys residue ($pK_a \sim 6$) in the h38C2 antibody via aldol and retro-aldol reactions [8]. This approach results in homogenous ADC, which might be useful for further therapeutic development.

### 1.1.2. N-terminal protein modifications

In addition to Lys residues, the $N$-terminal primary amine of a protein also displays unique reactivity. This site-selective modification is commonly used in peptide synthesis and has been applied in native chemical ligation (NCL) reactions (Fig. 2a) [9]. In NCL reactions, the thiolate of the $N$-terminal Cys residue of one peptide reacts with the C-terminal thioester of another peptide. The resulting thioester intermediate undergoes an acyl shift to form a native peptide bond (Fig. 2a). An alternative method for $N$-terminal modifications is the expressed reactivity of Lys residues in various local microenvironments, Matos et al. computationally designed sulfonyl acrylates which afforded chemo- and regio-selective modifications [7]. Using the designed sulfonyl acrylates, a single Lys residue in several proteins, including the therapeutic antibody Trastuzumab, was modified without disturbing the structure and function of the POIs. With this method, an improved site-selectivity is achieved because the reaction favors Lys residues with the lowest $pK_a$ in the POIs (Lys33 of lysozyme with a $pK_a$ of 9.5, Lys300 of annexin V with a $pK_a$ of 10.3, and Lys100 of C2Am with a $pK_a$ of 10.1). However, this strategy might not be applicable in some POIs where the Lys with the lowest $pK_a$ is inaccessible. In another case, to prepare an antibody-drug conjugate (ADC), Rader and co-workers site-specifically conjugated $\beta$-lactam derivatives to a Lys residue ($pK_a \sim 6$) in the h38C2 antibody via aldol and retro-aldol reactions [8]. This approach results in homogenous ADC, which might be useful for further therapeutic development.
protein ligation (EPL) method, where a recombinant protein can be conjugated to a synthetic peptide containing the desired modification (Fig. 2b) [10-12]. This approach enables site-selective modifications when the desired site is close to the N- or C-terminus of a POI. For example, the N-terminal ultrafast split intein was genetically fused to the C-terminus of histone H2B, which then reacted with a synthetic C-terminal intein fragment bearing a pre-ubiquitinated Lys120. The reaction between the two intein fragments results in a ubiquitinated H2B (H2B-K120Ub) [13]. The semisynthetic H2B-K120Ub is operable and can induce chromatin signaling in isolated nuclei. Additionally, the N-terminus has been used to generate reactive ketones through a pyridoxal-5′-phosphate (PLP) catalyzed transamination reaction; however, this reaction is not compatible with some N-terminal residues, such as Lys and Gln due to side reactions [14-16]. The resulting ketone or aldehyde can be used for the oxime/hydrazine ligation. For instance, the N-terminal residue over other nucleophilic residues such as Lys and His can be targeted by selective modification of a Cys residue using one peptide with an α-halocarbonyl (iodoacetyl, bromoacetyl, or chloroacetyl) [21,22], maleimides [23,24], methanesulfonate, or phenylthiosulfonate (Fig. 3a) [25,26]. Maleimides have been widely used to synthesize ADCs, including the FDA-approved ADCs, e.g., brentuximab vedotin, trastuzumab emtansine, and certolizumab pegol [2]. However, the thioether of maleimide conjugates can undergo cleavage via the thiol-exchange or hydrolysis-driven reactions. The resulting products from thiol exchange lead to premature loss of the drug’s efficacy and increased toxicity in vivo [27]. These issues can be addressed by minimizing the thiol exchange reaction or by converting the maleimide-thio-adducts to stable ring-open conjugates. The latter has been accomplished through the development of new self-hydrolyzing maleimides, which exhibit superior pharmacokinetic properties [28].

In recent efforts, new generations of thiol-targeting modification reagents have been developed, including electron-deficient alkynes [29], 3-arylpipropionitrile [30,31], allenamides [32], the thiol-yne reactions [33], and carbonylacrylic reagents [34,35]. For example, Wagner et al. reported an amine-to-thiol coupling reagent, sodium 4-((4-(cyanoethyl)benzoyloxy)-2,3,5,6-tetrafluorobenzensulfonate (CBTF), which contains an arylpropinitrile functional group instead of a maleimide (Fig. 3b). The resulting conjugates exhibit superior stability in plasma compared to that of maleimide conjugates [31]. In a recent study, Bernardes et al. rationally designed carbonylacrylic reagents, which undergo thiol-Michael addition using the Cys residues from the POIs [34]. Multiple proteins, including antibodies, were modified using...

1.1.3. Cysteine modifications

Cys is another residue of choice for chemical modifications due to the high nucleophilicity of its thiol and its low natural abundance in proteins, both of which favor single-site modification in POIs [1,5,19]. Under well-controlled pH conditions, selective modification of a Cys residue over other nucleophilic residues such as Lys and His can be obtained [20]. The Cys thiol can be modified using electrophilic α-halocarbonyls (iodoacetyl, bromoacetyl, or chloroacetyl) [21,22], maleimides [23,24], methanesulfonate, or phenylthiosulfonate (Fig. 3a) [25,26]. Maleimides have been widely used to synthesize ADCs, including the FDA-approved ADCs, e.g., brentuximab vedotin, trastuzumab emtansine, and certolizumab pegol [2]. However, the thioether of maleimide conjugates can undergo cleavage via the thiol-exchange or hydrolysis-driven reactions. The resulting products from thiol exchange lead to premature loss of the drug’s efficacy and increased toxicity in vivo [27]. These issues can be addressed by minimizing the thiol exchange reaction or by converting the maleimide-thio-adducts to stable ring-open conjugates. The latter has been accomplished through the development of new self-hydrolyzing maleimides, which exhibit superior pharmacokinetic properties [28].

In recent efforts, new generations of thiol-targeting modification reagents have been developed, including electron-deficient alkynes [29], 3-arylpipropionitrile [30,31], allenamides [32], the thiol-yne reactions [33], and carbonylacrylic reagents [34,35]. For example, Wagner et al. reported an amine-to-thiol coupling reagent, sodium 4-((4-(cyanoethyl)benzoyloxy)-2,3,5,6-tetrafluorobenzensulfonate (CBTF), which contains an arylpropinitrile functional group instead of a maleimide (Fig. 3b). The resulting conjugates exhibit superior stability in plasma compared to that of maleimide conjugates [31]. In a recent study, Bernardes et al. rationally designed carbonylacrylic reagents, which undergo thiol-Michael addition using the Cys residues from the POIs [34]. Multiple proteins, including antibodies, were modified using...
this approach. The modified antibodies were not only homogenous, but also resistant to degradation in plasma. Additional approaches for the Cys-selective protein modification via vinyl/alkyl pyridine, aza-norbornadiene bromovinyl sulfone, and diazo reagents, have been reported by the Bernardes Group [35–37].

In cases where the POIs lack a thiol functional group, disulfide modification serves as an alternative target. Multiple reagents including bissulfones, allyl sulfones, alkynes, and 3,4-disubstituted maleimides have been developed to site-selectively modify the disulfides of proteins as summarized in recent reviews [5,38]. For instance, using 2,3-dibromo maleimide with a C-2 (glycine derived) linker, Doxorubicin (DOX)–antibody conjugates were produced through a bis-alkylation reaction (Fig. 3c). This approach results in homogenous ADCs with enhanced pharmacological properties [39]. Oxetane, an oxygen (ether) containing four-membered ring, has also been used to modify protein disulfides via a site-selective bis-alkylation reaction [40,41]. In one of the reports, oxetane was installed onto a genetically detoxified diphtheria toxin (CRM197 protein) and the resulting modified protein exhibits increased immunogenicity in vivo [40].

1.1.4. Aromatic residue modifications

In addition to Cys, the relatively low natural abundance of aromatic residues, including His, Tyr, Trp, and Phe, offer alternative targets for site-specific modifications. However, obtaining a site-specific modification for one aromatic residue over another remains challenging. The reactivity of the ionizable side chain of Tyr is dependent on its protonation state, which allows the reactivity of Tyr to be modulated by controlling the pH of the reactions. Under acidic conditions, the aromatic ε-carbons adjacent to the hydroxyl group may undergo diazonium couplings (Fig. 4a) [42]. In a recent study, salmon’s calcitonin was conjugated to linear monomethoxy PEG using this approach. The resulting conjugates maintain the ability to reduce the concentration of...
calcium ions in the plasma. In conditions where the pH approaches the $pK_a$ of the phenol side chain of the Tyr residue ($pK_a$ of ~10), alklylation or acylation reactions of the oxygen can occur [43-45]. Another approach for site-selective Tyr modification is to use the $\pi$-allylpalladium complexes (Fig. 4b) [43]. Using this method, a novel polarity-sensitive fluorescent probe was installed onto Tyr108 of a bovine Cu/Zn superoxide dismutase (SOD) which allowed the conformational changes of the Tyr-containing domain to be monitored. This example might serve as a general approach for studying protein conformational changes if the domain of interest has a Tyr residue.

In the case of Trp, modifications of its side chain via metal-catalyzed reactions have been developed (Fig. 5), including alknylations [46], C–H arylation [47], and photoinduced cycladditions of tetrazoles [48]. However, these approaches often involve non-biocompatible reaction conditions such as the use of organic solvents and at high temperatures. An alternative strategy for selective modification of Trp consists of an $N$-oxy radical, 9-azabicyclo[3.3.1]nonane-3-one-$N_2$ in the presence of copper complexes (Fig. 5) [49]. Using this approach, Kanai and co-workers achieved Trp-selective modifications on myoglobin, lysozyme, BSA, and the $\alpha$-microglobulin antibody. This study opens new avenues for Trp modifications without using toxic metals under biocompatible conditions, further paving the way for functional modulation of therapeutic targets such as antibodies. In addition to the aforementioned cases, efforts for site-specific modifications of other proteinogenic residues are well summarized in recent reviews [50-52].

1.2. Protein modifications using ncAAs

Protein modifications using chemical reactions are powerful approaches to selectively install the desired chemical handles on the POIs. Manipulating the protein translation machinery for direct incorporation of ncAAs is another desirable approach, which enables the introduction of novel functionalities and reactivities beyond those offered by the 20 cAAs. Incorporating ncAAs into proteins has been widely applied in biocatalysis to enhance the activity and selectivity of enzymes, and even allows researchers to obtain novel catalytic reactions that are naturally unavailable [53-55]. This section presents some of the recent examples of using the genetic code expansion method to incorporate ncAAs into the POIs.

1.2.1. Genetic code expansion

The natural genetic code is comprised of 64 codons encoded by four nucleobases. Among these, three codons serve as stop codons, while the remaining 61 are recognized by transfer RNAs (tRNAs), and are charged by their cognate tRNA synthetases (aaRSs) with one of the cAAs. All organisms encode the same 20 amino acids (AAs) with the exceptions of pyrrolysinie [56] and selenocysteine [57]. The ability to employ and modify the existing protein translation machinery to incorporate ncAAs into proteins has opened new avenues for protein modifications. One of the strategies relies on the promiscuity of some tRNAs, which can be charged with ncAAs that are structurally analogous to their cognate amino acids (Fig. 6a) [58]. In this method, an auxotrophic bacterial host cannot produce the cAA and relies on the exogenously supplied ncAA. This approach results in a global incorporation of the ncAA to the proteome. One of the prominent examples is the substitution of Met with Se-Met to incorporate heavy atoms into the POIs for crystallographic phasing [59]. This strategy is straightforward and does not require genetic manipulation of the protein translation components. Instead, this method relies on the polyspecificity of the aaRS-tRNA pair, meaning that only ncAAs with structural similarities to the cognate amino acids can be incorporated, limiting the number of modifications. Additionally, global incorporation of ncAAs might alter protein folding, which could result in a perturbation of the activity and/or stability of the POIs.

A powerful alternative approach to site-specifically incorporate ncAAs into the POIs is via genetic code expansion (Fig. 6b). For example, the tyrosyl-tRNA synthetase TyrRS-tRNA$_{CUA}$ pair from Methanocaldococcus jannaschii is orthogonal in E. coli and other bacteria; the TyrRS-tRNA$_{CUA}$ and LeuRS-tRNA$_{CUA}$ pairs from E. coli are orthogonal in eukaryotic cells; the pyrrolyl-tRNA synthetase PylRS-tRNA$_{CUA}$ pairs from Methanosarcina barkeri and Methanosarcina mazei are orthogonal in both bacteria and eukaryotic cells [54,60,61]. Site-specifically modified POIs can be obtained, but the production yield is normally limited by the expression level of the exogenous aaRS-tRNA pairs and the presence of release factor 1 (RF-1), which recognizes the UAG triplet and terminates translation. Recently, an E. coli host has been engineered by removing RF-1 from the E. coli genome. Additionally, 95 out of the 273 amber stop codons were replaced with other more frequently used stop codons.

Fig. 4. Methods for Tyr modification. (a) Chemical modification of Tyr via diazonium couplings. (b) Conjugation of Tyr in SOD to a fluorescent probe using $\pi$-allylpalladium complexes. The blue spheres represent the remaining functional group.

Fig. 5. Metal-free chemical modifications of Trp residues using the $N$-oxy radical. The pink spheres represent other functional groups in the probes.
technology to evolve a PylRS variant with 45-fold higher catalytic efficiency compared to that of the wild-type [65]. Through this technology, Liu et al. employed the phage-assisted continuous evolution (PACE) engineering approaches have been developed to address these challenges. Most importantly, the ncAA incorporation efficiency is >98% in this engineered host strain, allowing a scalable production of the target ncAA-containing protein.

1.2.2. Next-generation genetic code expansion
To date, more than 200 ncAAs have been incorporated into POIs using the amber suppression method, thereby expanding the chemical functionalities and reactivities of proteins [54,64]. Thus far, the vast majority of studies employing this technology are restricted to the incorporation of single ncAAs into the POIs. The ability to incorporate multiple ncAAs into a protein might offer new opportunities for advanced biophysical studies and the synthesis of enhanced protein-based therapeutics. To achieve such goals, the enhanced specificity and orthogonality of aaRS-tRNA pairs is essential. Orthogonal aaRS-tRNA pairs can charge multiple ncAAs during the translation process. Thus, to site-selectively incorporate multiple ncAAs, the orthogonal aaRS-tRNA pairs must be highly selective. Several engineering approaches have been developed to address these challenges. Liu et al. employed the phage-assisted continuous evolution (PACE) technology to evolve a PylRS variant with 45-fold higher catalytic efficiency compared to that of the wild-type [65]. Through this technology, a TyrRS variant with improved selectivity toward p-iodo-D-phenylalanine was also obtained. Engineering of the aaRS-tRNA pairs has been proven useful in incorporating ncAAs at multiple sites. In a recent study, Chatterjee and co-workers simultaneously incorporated three distinct bioconjugation handles onto the GFP protein by employing three orthogonal pairs to decode three different codons. These pairs include the E. coli-derived tryptophanyl pair, the archael tyrosyl, and pyrolysyl pairs. This work allows facile multiple-site labeling of proteins without an exogenous chemical catalyst. However, the incorporation efficiency is only 2% of the wild-type yield [66,67].

Besides improving the specificity of the aaRS-tRNA pair, evolving other translation machinery components has also been important for the efficient incorporation of ncAAs into multiple sites of the POIs. To facilitate the incorporation of ncAAs with bulky side-chains or altered backbones, ribosome engineering has been attempted. The small subunit of the ribosome contains 16S rRNA and binds mRNA, while the large subunit contains 23S rRNA. Both subunits are involved in coordinating the translation process. There are two major challenges in ribosome engineering. First, the activity and fidelity of ribosomes are essential for cellular survival. Second, the 23S rRNA subunit can freely exchange between the native and the orthogonal 16S rRNAs in the cell.

To overcome the former issue, orthogonal (O-)ribosome-mRNA pairs were evolved [68,69]. In these pairs, an O-ribosome containing O–16S rRNA and mutations in the anti-Shine-Dalgarno (ASD) sequence was introduced in E. coli (Fig. 7a). This O-ribosome can selectively translate an orthogonal mRNA (O-mRNA) containing the O-ribosome-binding site, but not native mRNA transcripts. The O-ribosome was further engineered to incorporate different ncAAs into proteins [68,70]. The endogenous and orthogonal ribosomes share the 23S rRNA. Therefore, further efforts on 23S rRNA engineering have also been attempted. These efforts amplified the contribution of new 23S rRNA to O-mRNA translation and minimized the lethal effect of new 23S rRNA in endogenous translation processes. To achieve this goal, Chin et al. reported an orthogonal ribosome with the two subunits linked through an optimized orthogonal mRNA (Fig. 7a). This technology opens avenues for further improvement of the orthogonal ribosomes to incorporate more ncAAs [71].

Other exciting technologies for genetic code expansion extend beyond the canonical 64 codons. The first glimpse of quadruplet codon usage was reported from Salmonella typhimurium, in which a tRNA<sup>56</sup> containing the frameshifted CCCC anticodon was observed [72]. However, the efficiency of natural ribosomes for decoding quadruplet codons is extremely low. To overcome this challenge, Chin and co-workers engineered translation components that make use of quadruplet codons, instead of the traditional triplet ones (Fig. 7b) [70]. In this work, Chin and co-workers evolved an orthogonal ribosome (ribo-Q1) and a variant Seryl-tRNA synthetase/tRNA pair that can incorporate ncAAs at

---

After this engineering, the growth defects of the E. coli host were minimized when it was used to overexpress ncAA-containing proteins [62,63]. Most importantly, the ncAA incorporation efficiency is >98% in this engineered host strain, allowing a scalable production of the target ncAA-containing protein.
Fig. 7. Engineering of other translational machinery components. (a) Aggregated ribosome engineering efforts enable the incorporation of a wide variety of ncAAs. (b) Engineering of the ribosome to efficiently decode quadruplet codons.

Fig. 8. Modifications of ncAAs using click chemistry. (a) Copper-catalyzed conjugation of azide and alkyne. (b) Examples of azide- and alkyne-bearing ncAAs used for click chemistry. (c) Use of click chemistry to label newly synthesized proteins (shown in yellow) using BONCAT technology. (d) Proteins with strained alkyne-containing amino acids can be further modified via SPAAC. The red and blue spheres represent other functional groups in the probes.
the quadruplet codon. This system was then used to incorporate azide and alkyne-containing amino acids to a calcium-binding messenger protein, calmodulin.

Another important piece of work is a semisynthetic bacteria with an expanded genetic code [73]. In this study, Ronesberg et al. introduced an unnatural base pair (UBP) comprised of two deoxynucleoside triphosphates (dNaM and dTPT3). Theoretically, the addition of these two UBPs enables the incorporation of 152 new unnatural codons. In this study, the authors identified nine unnatural codons that can be used to efficiently incorporate ncAAs. Among these, three codons are orthogonal and can be used for three distinct ncAAs, which results in the first 67-codon semisynthetic organism.

1.2.3. Chemical modifications of the incorporated ncAAs

Genetic code expansion allows for the site-specific incorporation of a variety of ncAAs into a POI, providing useful chemical handles for biorthogonal reactions [1,74]. Several functional groups including azides, alkynes, alkenes, and tetrazines, can be incorporated through the amber codon suppression method, which are subsequently modified by appropriate reagents. For instance, the azides or alkynes of incorporated ncAAs can undergo a copper-catalyzed azide-alkyne cycloaddition (CuAAC), which is termed the “CuAAC click chemistry reaction” (Fig. 8a) [75,76]. Due to its high specificity and having a reasonable reaction rate, the click chemistry has been applied in the analysis of the cellular proteome via biorthogonal non-canonical amino acid tagging (BONCAT) [77,78]. By replacing Met in the cell culture media with an azide- or alkyne-containing ncAA such as azidohomoalanine (AHA) or homopropargylglycine (HPG), the newly synthesized proteins are marked with azide or alkyne functionality and are distinguishable from the pool of preexisting proteins (Fig. 8b and c). The labeling time with AHA should be adjusted according to the protein synthesis rate of the target cell type. Subsequently, the AHA- or HPG-containing proteins are covalently attached to an affinity tag such as disulfide biotin alkyne tag (DIS-alkyne) via click chemistry. The newly synthesized and affinity-tagged proteins can be analyzed by conventional biochemical studies or using high-resolution mass spectrometry (Fig. 8c).

In a recent study, an alkyne ε-(propargyloxycarbonyl)-lysine (AlkK) was incorporated into the proteome of murine tissue slices and the brains of live mice expressing an orthogonal pyrrolysyl pair introduced by viral infection [78]. The alkyne-bearing proteins are covalently linked to streptavidin beads via diazobenzene linkers, which are then cleaved to release the labeled proteome for analysis by tandem mass spectrometry. This method has been applied to characterize newly synthesized proteins after pharmacological treatments for a particular cell type [79,80].

The CuAAC reaction involves the use of toxic metal catalysts and exogenous ligands. As an alternative, the strain-promoted azide-alkyne cycloaddition (SPAAC) has emerged as a powerful copper-free click chemistry for the modification of the POIs (Fig. 8d) [1,81]. Several strained alkynes have been developed, including difluorinated cyclooctyne, dibenzocyclooctynol, and biarylazacyclooctynone. Some of these functional groups can be incorporated into the POIs via the amber codon suppression system [82–85]. The copper-free click chemistry has been implemented in a variety of applications, which are well summarized in recent reviews [80,86,87]. Another emerging strategy for ncAA modifications is the photo-click chemistry, which involves a photo-inducible dipolar [3 + 2]-cycloaddition reaction between an alkenyl-ncAA and a nitrile imine [88,89]. This reaction results in the formation of fluorescent adducts, allowing not only for the in vitro modification of isolated proteins, but also the visualization of proteins in living cells [89–91]. These examples illustrate the general applicability of the biorthogonal reactions in the site-specific modifications of the POIs under complex biological environments, including living cells and multi-cellular organisms with minimal disruption of the biological functions and activities of the POIs.

1.3. Ligand-directed modifications of proteins

In some cases, proteins possess high binding affinities for their respective binding partners. The interactions between proteins and their binders allow for affinity-driven site-specific modifications. A variety of recognition-driven chemical modification strategies have been developed, including ligand-directed (LD) modifications [1,2,92]. The LD chemical modifications of proteins have been developed for endogenous protein labeling by harnessing the specific interactions between the POIs and small molecules. In this approach, the labeling reagent contains an affinity ligand, a reporter tag, and a reactive moiety (Fig. 9a). The binding of the ligand to the POI places the reactive unit of the labeling reagent in proximity to residues close to the ligand binding site, which enhances the selective modification. The LD approach has been applied in the labeling of carbonic anhydrase (CA) [93]. In this study, the specific ligand, benzensulfonylamide, is conjugated to a synthetic probe, which includes a fluorophore via the phenylsulfonate linkage (Fig. 9a).

Fig. 9. Affinity-driven protein chemical modifications. (a) The LD-based protein labeling approach. (b) The LDNASA-based covalent inhibitor developed through LD chemical modification. Red and yellow star each represents a probe. Blue and pink spheres each represent a functional group.
Upon binding to the target protein, a nucleophilic side chain in the POI reacts with the electrophilic phenylsulphonate ester group. This reaction leads to the labeling of the POI, while at the same time, the ligand may dissociate. Therefore, the functions of the POI are not disrupted after the chemical modification. Additionally, the LD approach used in the labeling of CA could be applied to both in vitro and in vivo studies [93].

Non-covalent and reversible ligand binders are often useful in many applications. However, in some cases, the ability to form a stable covalent adduct might be important, such as in the development of inhibitors. In a recent study, Hamachi et al. reported a new ligand-directed N-acetyl-N-acyl sulfonamide (LDNSA)-based covalent inhibitor, which targets chaperone Hsp90 (Fig. 9b) [94,95]. LDNSA enabled the site-specific modification of Lys58, which is a residue near the ligand’s binding site in Hsp90. This reaction leads to a covalent adduct between Hsp90 and its inhibitor (Fig. 9b). Treatment of cancer cells with an LDNSA-linked inhibitor results in irreversible covalent modification of Hsp90 at Lys58 by the inhibitor and reduces Hsp90’s molecular chaperone activity in cancer cells.

2. Utilizing protein modifications in the studies of metalloenzymes

Chemical modifications and the expansion of the genetic code allow the incorporation of structurally and chemically diverse ncAAs into the POIs, enabling the study of the biological functions and regulation of proteins including metalloenzymes. In addition to these approaches, protein modifications can also be achieved by installing an abiotic metal cofactor to a protein scaffold, resulting in an artificial metalloenzyme [96–98]. These technologies have opened new avenues for investigating enzymatic reaction mechanisms and engineering of metalloenzymes with new activities. Among these metalloenzymes, non-heme iron (NHFe) and heme enzymes catalyze a remarkable range of chemical transformations, including hydroxylation, endoperoxidation, and desaturation at the expense of molecular oxygen as the final electron acceptor [99–102]. Recent efforts in utilizing these strategies in metalloenzyme engineering and their biochemical applications are summarized in several recent reviews [97,103–106]. This section presents selected case studies, from recent years, of NHFe, heme, and Cu-containing enzymes where ncAAs are used as novel tools. We summarize the discussed cases in Table 2.

| Table 2 | Applications of protein modification in metalloenzyme studies. |
|---------|---------------------------------------------------------------|
| Enzyme | Applications                                                                 |
| Cysteine dioxygenase (CDO) | Investigate the role of the active site Tyr in the Cys-Tyr crosslink biogenesis. |
| Sulfoxide synthase OvoA | Determine the function of the active site Tyr in the oxidative C-S bond formation reaction. |
| Myoglobin and its variants | Investigate the function of Tyr-His crosslink in HCO-catalysis. |
| Heme copper oxidase (HCO) | Engineer swMb variants with NO reduction activity. |
| | Introduce an artificial cofactor to improve enantioselectivity in thioanisole sulfonation reaction. |
| | Engineer the HCO variants with enhanced dioxygen reduction activity. |
| Cytochrome c P450 | Introduce novel activity (cyclopropanation) to engineered enzyme. |
| Azurin (Az) | Determine the function of axial Met ligand in modulating the metalcopter properties and functions. |

CDO, the octahedral ferrous iron is coordinated by a 3-His motif (Fig. 10b). The Cys-Tyr crosslink biogenesis in CDO is an autocatalytic oxidation reaction, in which the oxygen-activated iron center oxidizes the residues at the active site (Cys and Tyr) rather than the substrate, L-Cys [108]. Despite the important role that the Cys-Tyr crosslink plays in the catalytic activity, the mechanistic details of the biogenesis of this crosslink remain unclear, partly due to the challenges of obtaining a homogeneous population of either cross- or uncross-linked CDO. In a recent study, Liu and co-workers employed the amber codon suppression method to replace Tyr157 with a halogen-substituted Tyr in human CDO [109]. The fluoride- and chlorine-containing CDO variants remained active, albeit with a lower activity (2–10% of the wild-type activity). Intriguingly, the Cys93-Tyr157 crosslink is also observed among these variants. Given the high bond dissociation energy, C-F bond cleavage was unexpected in the Cys-Tyr crosslink biogenesis (Fig. 10c). The oxidative C-F bond cleavage might take place during the oxidation of Tyr157 by the iron center or the Cys93 radical. This study represents the first reported case of an oxidative C-F bond cleavage mediated by an NHFe enzyme [109].

The uncross-linked structure of F2-Tyr containing CDO reveals that Cys93 assumes two conformations prior to the Cys-Tyr cofactor formation (Fig. 10d) [109]. During the crosslink biogenesis, the sulfur atom of Cys93 might rotate toward the iron center, which is supported by a recently reported structure of the uncross-linked F2-TyrCDO-L-Cys-NO ternary complex (Fig. 10e) [110]. This structure revealed an interaction between NO and one conformer of Cys93, suggesting that cysteine oxidation may occur prior to the crosslink formation. The oxidation of Cys93 by the iron center is further supported by a computational study, which suggests that an iron-bound oxygen species oxidizes Cys93 as the first step of the crosslink biogenesis, instead of the Tyr157 oxidation, as previously proposed [108,111]. The studies of the CDO cofactor biogenesis highlight the general applicability of site-specific modifications of POIs in mechanistic studies. Due to the important roles of Tyr in enzymatic catalysis, the ability to control the structural features and chemical properties (reduction potential and the pK, of active Tyr residue has been sought after in several cases [55,112–116].

Ergothioneine is a potent antioxidant and has been proposed to be a longevity vitamin [117,118]. Due to its potential benefits, ergothioneine biosynthesis has received considerable interest. In the mycobacterial ergothioneine biosynthetic pathway, a NHFe sulfoxide synthase, EgtB, mediates the oxidative coupling between herycine-γ-glutamyl-cysteine (γ-Glu-Cys, Fig. 11b) [119]. Recently, the fungal ergothioneine biosynthetic pathway was reported, in which a sulfoxide synthase, Egt1, catalyzes the C-S bond formation between herycine and L-Cys (Fig. 11b) [120,121]. A similar reaction has also been observed in the biosynthesis of another potent antioxidant, ovothiol A. In the ovothiol biosynthesis, a sulfoxide synthase, OvoA, catalyzes the oxidative coupling between L-His and L-Cys (Fig. 11b) [122,123]. These ergothioneine and ovothiol sulfoxide synthase reactions differ in terms of their substrate selectivity and their products’ C-S bond regioselectivity. In EgtB- and Egt1-catalysis, the C-S bond is formed at the ε-carbon of herycine’s imidazole ring, while OvoA-catalysis incorporates the C-S bond at the δ-carbon of L-His’ imidazole ring (Fig. 11b). The structure of Mycobacterium thermo resistible’s EgtB reveals a 3-His iron coordination site similar to that of CDO (Fig. 12a) [124]. Mutating the active site Tyr377 to Phe suppressed the sulfoxide synthase activity and the oxidation of γ-Glu-Cys is the dominant activity in the mutant. Recently, the crystal structure of a sulfoxide synthase from Candidatus Chloracidobacterium thermophilum has also been reported and its iron center also has a 3-His iron coordination ligand environment, too [125,126]. Based on the EgtB crystal structure, three different mechanistic models have been proposed based on calculations using the density functional theory (DFT) or the quantum mechanics/molecular mechanics (QM/MM) methods (Fig. 12c) [127–129]. The key questions are whether the reaction involves a sulfinic acid intermediate and what role the active site Tyr377 plays in EgtB-catalysis. In Fig. 12c, two...
Fig. 10. Using ncAAs in the study of the Cys-Tyr cofactor biogenesis in CDO. (a) Cysteine oxidation mediated by CDO. (b) The 3-His motif and the Cys-Tyr crosslink in wild-type CDO (shown in green) in complex with the L-Cys substrate (shown in purple). The iron center is shown in a yellow sphere (PDB ID: 6BGF). (c) C-F bond cleavage in CDO-catalysis. (d) The structure of F2-Tyr CDO•L-Cys without crosslink (shown in blue, PDB ID: 6BPS) overlays with the structure of mature F2-Tyr CDO•L-Cys (shown in pink, PDB ID: 6BPV), revealing two conformations of Cys93. The fluorine atoms are shown in light blue. (e) A ternary complex of F2-Tyr CDO•L-Cys•NO (shown in magenta) reveals that one conformation of Cys93 is 3.1 Å from the NO and primed to be oxidized by the metallocenter (PDB ID: 6BPR).

Fig. 11. Chemical properties of Tyr analogs and the biosynthesis of ergothioneine and ovothiol A. (a) The non-canonical Tyrs that can be incorporated into the POIs using the amber codon suppression method possess different reduction potentials and pKas. (b) The oxidative C-S bond formation reactions mediated by sulfoxide synthases (EgtB/Egt1/OvoA) in the biosyntheses of ergothioneine and ovothiol A.
independent theoretical studies reported by Tian and Wei suggested that the sulfur’s oxidation to sulfenic acid is the first half of the reaction (Path IA-IC), while another study by Faponle et al. proposed that the oxidative C–S bond formation is the first half of EgtB-catalysis (Path II) [127–129]. In these two mechanistic models (pathway I vs pathway II), different functions have been proposed for the active site Tyr377. In the Tian model, Tyr377 functions as a Lewis acid/base (Path I, Fig. 12 c). In the Faponle model, Tyr377 plays a key role in a proton-coupled electron transfer (PCET) process (Path II, Fig. 12 c). Due to the similarities of these reactions, all of the mechanistic discussions regarding EgtB-catalysis are extended to OvoA-studies [112].

To unravel the mechanism of OvoA-catalysis, Liu and co-workers harnessed the unique chemical properties of the non-canonical Tyrs [112]. Notably, the biochemical characterizations also revealed the cysteine oxidation activity in OvoA-catalysis [130]. The computational studies and biochemical results suggest that the sulfoxidation and cysteine oxidation activities might be two pathways branching out from a common intermediate in OvoA-catalysis. Due to this reason, the isotopically sensitive branching method was used to measure the kinetic isotope effect (KIE) of the OvoA reaction [112,131]. Tyr417 in OvoA is the EgtB’s Tyr377 counterpart (Fig. 12 a and b). The Tyr417 of OvoA was replaced with 2-amino-3-(4-hydroxy-3-(methylthio)-phenyl)propanoic acid (MtTyr) using the amber codon suppression system (Fig. 11 a) [112]. The substrate KIE of the OvoA Y417MtTyr variant is close to unity, which is comparable to that of the wild-type. This experimental result contradicts the model suggested by Wei et al., in which they predicted a primary substrate deuterium KIE as high as 5.7 [128]. To further investigate the role of Tyr417, Liu and co-workers replaced Tyr417 with

![Figure 12](image-url)
3-methoxytyrosine (MeOTyr) [132], which possesses a similar pK_a but a much lower reduction potential than that of canonical Tyr [131]. The OvoA_Y417MeOTyr variant exhibited an inverse KIE, while the wild-type has a KIE close to unity. These results imply that Tyr417 might be part of redox-chemistries in OvoA-catalysis.

The model proposed by Faponle et al. [127] is more consistent with the experimental results described above. In this computational model, an inverse deuterium KIE was predicted if the δ-hydrogen of the imidazole ring was replaced with deuterium, which was indeed observed in the OvoA_Y417MeOTyr variant. Given the similarities between OvoA- and EgtB-catalysis, the OvoA mechanism will be discussed in the context of EgtB mechanistic models (Fig. 12c). In the model proposed by Faponle and co-worker, the oxygen binding and activation results in an Fe^{III}-superoxo species (f). Then, the Tyr377 in EgtB participates in the PCET process through the active site’s water network, resulting in the Fe^{III}-hydroperoxo species (g) along with a Tyr377-based radical species (h). The Fe^{III}-superoxo species then abstracts a hydrogen atom from the imidazole ring to regenerate the aromaticity and results in an Fe^{II}-hydroperoxo intermediate (i), which undergoes oxidation to afford the sulfoxide product (Fig. 12c).

Our MeOTyr-containing OvoA studies are consistent with the EgtB model suggested by Faponle et al. [127,131]. However, the OvoA

---

Fig. 13. The endoperoxidation reaction mediated by FtmOx1. (a) The products of the FtmOx1 reactions under single-turnover conditions in the absence of extra reductants (e.g., ascorbate). (b) The distal-type αKG binding configuration in the FtmOx1•αKG binary complex (PDB ID: 4Y5S). αKG is shown in purple and the iron center is shown as a yellow sphere. Water is presented as a red sphere. (c) A proposed FtmOx1 mechanistic model involving the Tyr224 radical as one of the key species (shaded in green) under single-turnover conditions. The Tyr224 radical is also responsible for initiating the oxidation of the C_{13}-hydroxyl group to a keto-product (13). Therefore, there are two reaction cycles in FtmOx1-catalysis under single-turnover conditions.
structure is not available and it is possible that OvoA and EgtB might follow different mechanistic models. Additional kinetic and spectroscopic studies are needed to uncover the mechanistic details of these sulfoxide synthases.

There exists another enzymatic system where the role of Tyr in catalysis is under a heated debate. In the biosynthesis of a tremorgenic mycotoxin verruculogen (2), an α-ketoglutarate-dependent (αKG) NHFe enzyme, FtmOx1, mediates the endoperoxidation of fumitremorgin B (1) (Fig. 13a) [133,134]. Biochemical characterization under single-turnover reaction conditions in the absence of extra reductants (e.g., ascorbate) revealed that each cycle of FtmOx1-catalysis consumes two molecules of molecular oxygen and involves two distinct reactions: the endoperoxidation reaction to afford verruculogen (2) and the oxidation of verruculogen’s C13-hydroxyl group to form a keto-product (3) (Fig. 13a) [113]. Compound 3 is the dominant product under single-turnover conditions. In the reported crystal structure of the FtmOx1•αKG binary complex, αKG coordinates to the metallocenter through a distal-type configuration, and Tyr224 is adjacent to the putative oxygen binding site (Fig. 13b) [113]. Structural, biochemical, and spectroscopic characterizations led to a mechanistic model involving the Tyr224 radical as one of the key species in FtmOx1-catalysis (Fig. 13c). However, in a recent report, Bollinger and co-workers proposed that an alternative tyrosine residue (Tyr68) might be the tyrosyl radical site (Fig. 14) [135]. The authors reported that in the reaction catalyzed by the FtmOx1Y68F variant, a new major, uncharacterized product was formed. In FtmOx1-catalysis, under single-turnover conditions and in the absence of other reductants, the keto-product (3) is the dominant product. In this Tyr68 radical model (Fig. 14), Tyr68 is on the protein surface and fully solvent exposed, and the Fe-center is far from the substrate’s C13–OH group. It will be challenging to explain formation of the keto-product (3) as the dominant product using this Tyr68 radical model [135]. Therefore, Tyr224 remains the best candidate for the proposed tyrosyl radical in FtmOx1-catalysis. To provide further evidence to differentiate between the two proposed mechanistic models, the use of non-canonical Tyr will be beneficial in characterizing the catalytic roles of the two Tyr residues in FtmOx1-catalysis, which are on-going efforts in our laboratory.

2.2. Heme enzymes engineering

Another common type of iron-dependent enzyme is the heme-containing enzymes, in which the iron center is coordinated to a porphyrin as a tetradentate ligand. In addition to heme cofactors, these enzymes often possess additional metals and they catalyze a diverse range of reactions, e.g., copper (CuB) in heme copper oxidase (HCO) and products were characterized [135]. Therefore, Tyr224 remains the best candidate for the proposed tyrosyl radical in FtmOx1-catalysis. To provide further evidence to differentiate between the two proposed mechanistic models, the use of non-canonical Tyr will be beneficial in characterizing the catalytic roles of the two Tyr residues in FtmOx1-catalysis, which are on-going efforts in our laboratory.

2.2. Heme enzymes engineering

Another common type of iron-dependent enzyme is the heme-containing enzymes, in which the iron center is coordinated to a porphyrin as a tetradentate ligand. In addition to heme cofactors, these enzymes often possess additional metals and they catalyze a diverse range of reactions, e.g., copper (CuB) in heme copper oxidase (HCO) and
Iron (Fe$_b$) in nitric oxide reductase (NOR) [136–140]. Designing artificial heme enzymes is a powerful tool not only for facilitating in-depth mechanistic studies, but also for developing novel biocatalysts. Both rational enzyme design and directed evolution methods have been applied for heme-enzyme engineering where metal-ligands and the secondary coordination shell residues are modulated. The readers are referred to several recent reviews on artificial metalloenzyme engineering [103,104,141–144]. The cytochrome c oxidase (CcO) is a membrane-bound protein from the HCO family. It catalyzes the reduction of O$_2$ to water without the generation of toxic reactive oxygen species (ROS). In this reaction, the catalytic functions of the additional Cu$_b$ site remain unclear [145]. Due to the difficulty of obtaining large amounts of membrane-bound CcO, an artificial model for this enzyme has been generated through engineering the heme-containing oxygen carrier protein, myoglobin from sperm whale *Physeter macrocephalus* (swMb, Fig. 15a). Based on the structural comparison between swMb and the native bovine CcO, residues His64, Leu29, and Phe34 from the swMb active site were selected to create the Cu$_b$ binding site [145]. Both Leu29 and Phe34 were mutated to His to create the swMb L29H F34H variant, which is referred to as Cu$_b$Mb. This engineered Cu$_b$Mb variant exhibited biophysical properties similar to the heme-copper center from the HCO family. In addition, spectroscopic studies of the engineered
Cu\textsubscript{4}Mb indicated that the Cu\textsubscript{4} center enhances the molecular oxygen binding affinity to the heme center in the HCO reaction.

The creation of the Cu\textsubscript{4}Mb protein has enabled the subsequent efforts to improve its natural activities and even for creating non-native activities. In Cu\textsubscript{4}Mb, ROS formation was inhibited, while H\textsubscript{2}O production did not change significantly relative to the wild-type swMb. HCOs contain a conserved Tyr-His crosslink in their active site and this crosslink has been proposed to be important for water production (Fig. 15b). Lu and co-workers replaced Gly65, a residue adjacent to the copper-binding site, with a Tyr residue in Cu\textsubscript{4}Mb. The water production rate of the Cu\textsubscript{4}Mb\textsubscript{G65Y} variant was dramatically enhanced compared to that of native Cu\textsubscript{4}Mb [146]. The importance of Tyr in Cu\textsubscript{4}Mb\textsubscript{G65Y}-catalysis motivated the authors to further investigate the contribution of a Tyr-His crosslink in HCO-catalysis using a ncAA incorporated by the amber-codon suppression method. In this study, Lu and co-workers chose (S)-2-Amino-3-(4-hydroxy-3-(1H-imidazole-1-yl)phenyl)propionic acid (imiTyr) to mimic the Tyr-His crosslink [116]. The resulting imiTyr-Cu\textsubscript{4}Mb variant not only mediated the O\textsubscript{2} reduction reaction with more than 1000 turnovers, but also minimized the ROS formation relative to the Cu\textsubscript{4}Mb\textsubscript{G65Y} variant. This study highlights the crucial role of the Tyr-His crosslink in HCO for selective H\textsubscript{2}O production with minimized ROS formation. The enhanced catalytic activity and selectivity of the engineered HCO reveals the value of protein modifications in developing robust biocatalysts.

NORs are similar to HCOs in that they both contain an additional iron-binding site near the heme-center. However, NORs catalyze the two-electron reduction of NO to N\textsubscript{2}O. Additionally, the creation of the NOR activity from a heme template (i.e. swMb) is challenging due to the lack of a NOR crystal structure. Based on their success in the Cu\textsubscript{4}Mb studies, Lu and co-workers created a swMb variant, Fe\textsubscript{4}Mb, with the NO reduction activity [140]. Starting from the Cu\textsubscript{4}Mb (swMb L29H F43H variant) scaffold, an additional Val68 to Glu mutation was introduced to mimic the Glu residue in NOR (Fig. 15a). The Cu\textsubscript{4}Mb\textsubscript{V68E} variant had the desired Fe\textsubscript{4} site and also exhibited the NO reduction activity at a level comparable to that of the native NOR.

In these heme enzymes engineering efforts, besides mutating the metal-binding site, a heme analog can replace the native heme; a second metal-binding site and also exhibited the NO reduction activity at a level comparable to that of the native NOR.

In these heme enzymes engineering efforts, besides mutating the metal-binding site, a heme analog can replace the native heme; a second metal-binding site and also exhibited the NO reduction activity at a level comparable to that of the native NOR.

3. Conclusions

In this review, we summarize some of the recent advances in protein modifications through both chemical and biological approaches, which are widely applied in biochemistry, chemical biology, cell biology, and drug discovery. In recent years, there are more cases of the application of these tools in enzyme mechanistic studies and functional engineering. We select some recent examples where introducing ncAAs or non-native cofactors either served as new mechanistic study tools or to introduce novel catalytic activities that do not exist in the wild-type enzymes.

Author contributions

Conceptualization, P.L.; writing of the original draft, N.N. and R.H.; review and editing, all authors. All authors have read and agreed to the published version of the manuscript.
Funding
This research was partially supported by National Science Foundation (CHE-2004109 to P. L.)

Declaration of competing interest
The authors declare no conflict of interest.

References
[1] Sakamoto S, Hamachi I. Recent progress in chemical modification of proteins. Anal Biochem 2018;535:180–90.
[2] Shadish JA, DeForest CA. Site-selective protein modification: from functionalized proteins to functional biomaterials. Matter 2020;2(1):50–77.
[3] Nishan N, Hackenberger CP. Site-specific PEylation of proteins: recent developments. J Org Chem 2014;79(22):10727–33.
[4] Warn F, Steinbach T, Klok HA. One-pot squaric acid diester mediated aqueous protein conjugation. Chem Commun 2013;49(71):7815–7.
[5] Spicer CD, Davis BG. Selective chemical protein modification. Nat Commun 2014;5(1):5740.
[6] DeForest CA, Tuttell DA. A photoresponsive protein-patterning approach for guiding stem cell fate in three-dimensional gels. Nat Mater 2015;14(5):523–31.
[7] Matos MJ, Oliveira BL, Martínez-Saez N, Guerreiro A, Cal P, Bertoldo J, et al. Chemoselective and region-selective lysine modification on native proteins. J Am Chem Soc 2018;140(11):4004–17.
[8] Nanna AR, Li X, Walseng E, Pedzida L, Goydel RS, Hymel D, et al. Harnessing a catalytic lysine residue for the one-step preparation of homogeneous antibody-drug conjugates. Nat Commun 2017;8(1):1112.
[9] Dawson PE, Muir TW, Clark-Lewis I, Kent S. Synthesis of proteins by native chemical ligation. Chem Commun 2003;72(12):249–49.
[10] Muir TW. Semisynthesis of proteins by expressed protein ligation. Annu Rev Biochem 2002;71(1):47–70.
[11] Vila-Perello M, Liu Z, Shah NH, Willis JA, Idoyaga J, Muir TW. Streamlined expressed protein ligation using split inteins. J Am Chem Soc 2013;135(1):286–92.
[12] Mininni EC, Yokoyama K, Stubbe J. Unnatural amino acids: better than the real things? F1000 Biol Rep 2009;1:88.
[13] David Y, Vila-Perello M, Verma S, Muir TW. Chemical tagging and customizing of cellular chromatins states using ultrafast trans-splicing inteins. Nat Chem 2015;7(5):394–402.
[14] Gilmores JM, Scheck RA, Ersken-Khan AP, Joshi NS, Francis MB. N-terminal protein modification through a biomimetic transamination reaction. Angew Chem Int Ed 2006;45(32):5307–11.
[15] Scheck RA, Francis MB. Region-selective labeling of antibodies through N-terminal transamination. ACSSynth Biol 2004;2(4):247–51.
[16] Scheck RA, Dedo MT, Lavaroni AT, Francis MB. Optimization of a biomimetic transamination reaction. J Am Chem Soc 2008;130(35):11762–70.
[17] Christman KL, Broyer RM, Tolstyk ZP, Maynard HD. Site-specific protein immobilization through N-terminal oxime linkages. J Mater Chem 2007;17(19):2021–7.
[18] Jie W, Peng C. Development and applications of bioorthogonal cleavage tagging things? F1000 Biol Rep 2009;1:88.
[19] Seki Y, Ishiyama T, Sasaki D, Abe J, Soma Y, Otsuki K, et al. Transition metal-free tryptophan-selective biocoupling of proteins. J Am Chem Soc 2016;138(34):10798–801.
[20] deGruyter JN. Malins LR, Baran PS. Residue-specific peptide modification: a chemist’s guide. Biochemistry 2017;56(30):3663–73.
[21] Hoyt EA, Cal PM, Oliveira BL, Bernardes GJ. Contemporary approaches to site-specific protein conjugation. Chem Commun 2013;49(71):7815–7.
[22] Seim KL, Obermeyer AC, Francis MB. Oxidative modification of native protein residues using cerium (IV) ammonium nitrate. J Am Chem Soc 2011;133(42):16970–7.
[23] Tilley SD, Francis MB. Tyrosine-selective protein alkylation using x-allylcycloproplcium complexes. J Am Chem Soc 2006;128(4):1080–1.
[24] Ruan Z, Saurermann N, Manoni E, Ackermann L. Manganese-catalyzed C-H alkylation: expedient peptide synthesis and modification. Angew Chem Int Ed 2017;56(12):3172–6.
[25] Schischko A, Ren H, Kaplaniou N, Ackermann L. Bioorthogonal diversification of peptides through selective ruthenium(ii)-catalysed c-h activation. Angew Chem Int Ed 2017;56(5):1576–80.
[26] Siti W, Khan AK, de Hoog HP-M, Liedberg B, Nallani M. Photo-induced conjugation of tetrazoles to modified and native proteins. Org Biomol Chem 2015;13(11):3202–6.
[27] Seki Y, Ishiyama T, Sasaki D, Abe J, Soma Y, Otsuki K, et al. Transition metal-free tryptophan-selective biocoupling of proteins. J Am Chem Soc 2016;138(34):10798–801.
[28] deGruyter JN. Malins LR, Baran PS. Residue-specific peptide modification: a chemist’s guide. Biochemistry 2017;56(30):3663–73.
[29] Hoyt EA, Cal PM, Oliveira BL, Bernardes GJ. Contemporary approaches to site-specific protein modification. Nat Rev Chem 2019;3(3):147–71.
[30] Conibear AC. Deciphering protein post-translational modifications using chemical biology tools. Nat Rev Chem 2020;1:22.
[31] Drienskova I, Alonso-Cotchico L, Vidošić P, Lledó A, Mareš-Cale J-D, Roelfs G. Design of an enanionselective artificial metallo-hydrazine enzyme containing an unnatural metal-binding amino acid. Chem Sci 2017;8(10):7228–35.
[32] Bzostek E, Myśliwiez-Morales H, Chmielewska K. Synthesis of unnatural amino acids. Nat Chem 2020:1–22.
[33] Drienskova I, Roelfs G. Expanding the enzyme universe with genetically encoded unnatural amino acids. Nat Catal 2020;1:1–20.
[34] Won Y, Pagar AD, Patil MD, Dawson PE, Yun H. Recent advances in enzyme engineering through incorporation of unnatural amino acids. Biotechnol Bioeng 2019;124(1):1–13.
[35] Srinivasan G, James CM, Kryzicki JA. Pyrroline encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science 2002;296(5572):1459–62.
[36] Rock A, Forchhammer K, Heider J, Leinfelder W, Sapers G, Veprek B, et al. Selenocystine: the 21st amino acid. Mol Microbiol 1997;24(1):151–20.
[37] Munier R, Cohen G. Incorporation of structural analogues of amino acids into proteins. J Am Chem Soc 1992;114(12):5159–67.
[38] Kuan SL, Wang T, Weil T, Ecken E, Koniev O, Leriche G, Nothisen M, Remy J-S, Strub J-M, Schaeffer-Reiss C, et al. Stoichiometric and irreversible cysteine-selective protein modification using carboxylicylic reactants. Nat Commun 2016;7(1):1–9.
[39] Bernardin B, Cal PM, Matos MJ, Oliveira BL, Martinez-Saez N, Albuquerque IS, et al. Synthetic and Systems Biotechnology 6 (2021) 32–49
[93] Tsukiji S, Miyagawa M, Takaoka Y, Tamura T, Hamachi I. Ligand-directed tosyl propargylamine labeling of proteins: from test tubes to complex biological systems. J Am Chem Soc 2018;141(15):6204–12.

[94] Tamura T, Ueda T, Goto T, Tsukidate T, Shapira Y, Nishikawa Y, et al. Rapid bioorthogonal click labeling of endogenous proteins: from test tubes to complex biological systems. J Am Chem Soc 2018;141(15):6204–12.

[95] Rackham O, Chin JW. A network of orthogonal ribosome- mRNA pairs. Nat Chem Biol 2005;1(3):159–66.

[96] Hui A, De Boer HA. Specialized ribosome system: preferential translation of a single mRNA species by a population of mutated ribosomes in Escherichia coli. Proc Natl Acad Sci U S A 1987;84(14):4762–7.

[97] Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 2010;464(7287):441–4.

[98] Schmid WH, Tsimov Z, Uattamapinant C, Rae CD, Fried SD, Chin JW. Controlling orthogonal ribosome subunit interactions enables evolution of new function. Nature 2018;564(7736):444–8.

[99] Riddle DL, Carbon J. Frame-shift suppression: a nucleotide addition in the anticodon of a glycine transfer RNA. Nat New Biol 1973;242(121):230–4.

[100] Fischer EC, Hashimoto K, Zhang Y, Feldman AW, Dien VT, Karadeema RJ, et al. Cycloadditions in bioorthogonal chemistry: Springer. 2016.

[101] Lang K, Chin JW. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. Chem Rev 2014;114(9):4764–806.

[102] Rostovtsev VV, Green LG, Fokin VV, Sharpless KB. A stepwise Huisgen cycloaddition protocol for the near-native (1+3)-azaalkyne “ligation” of azides and terminal alkynes. Angew Chem Int 2002;41(14):2596–9.

[103] Tornoe CW, Christensen G, Meldal M. Peptidotriazoles on solid phase: 1,2,3-triazoles by region-specific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. J Org Chem 2002;67(9):3057–64.

[104] Dietrich DC, Link AJ, Graumann J, Tielrooij DA, Schuman EM. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). Proc Natl Acad Sci U S A 2010;107(54):9182–7.

[105] Kregger TP, Ernst RJ, Elliott TS, Calo L, Berendonk T, Davis BM, et al. Correlating crosslink formation with enzymatic activity in cysteine dioxygenase. Biochemistry Biophys Acta (BBA) - Mol Basis Dis 2012;1822(5):784–93.

[106] Casquete T, Müller S, Rodríguez R. Visualizing biologically active small molecules in cells using click chemistry. Nat Rev Chem 2018;2(9):202–15.

[107] Thirumurugan P, Matosiuk D, Jozwiak K. Click chemistry for drug development and diverse chemical biology applications. Chem Rev 2013;113(7):4095–709.

[108] Agard NJ, Prescher JA, Bertozzi CR. A strain-promoted [3 + 2] azide–alkyne cycloaddition for covalent modification of biomolecules in living systems. J Am Chem Soc 2004;126(4):15046–7.

[109] Gadella JA, Bayley H, Schmalz RA, Bertozzi CR. Second-generation difluorocyclooctynes for copper-free click chemistry. J Am Chem Soc 2008;130(34):11486–93.

[110] Ni Q, Guo J, Wollert MA, Booms G-J. Visualizing metabolically labeled glycoconjugates of mammalian cells by copper-free and fast haemoglobin cycloaddition. Angew Chem 2008;47(12):2253–5.

[111] Hewitt JC, Sletten EM, Bertozzi CR. Rapid Cu-free click chemistry with readily photoclickable terminal alkynes to azides. J Am Chem Soc 2018;140(13):4604–12.

[112] Yan W, Song H, Song F, Guo Y, Wu C-H, Sae H, et al. Endoperoxide formation by an α-ketoglutarate-dependent mononuclear non-haem iron enzyme. Nature 2015;527(7579):539–43.

[113] Yu Y, Li X, Li J, Zhou Q, Cui H, Housein-zadeh P, et al. Defining the role of tyrosine and phenylalanine residues by genetic incorporation of unnatural tyrosine analogs. J Am Chem Soc 2015;137(4):1494–7.

[114] Zhou Q, Hu M, Zhang W, Jiang L, Shang Z, Stipanuk MH. Synthesis of aminocarboxyl siderophores in cysteine dioxygenase is regulated by substrate and forms a novel post-translational regulation of activity. J Biol Chem 2008;283(18):12188–201.

[115] Xu H, Yang J, Huang C, Huesman M, et al. Probing the function of the Tyr-Cys cross-link in metalloenzymes by the genetic incorporation of C-5-chlorotyrosine. Angew Chem Int 2013;52(40):12037–7.

[116] Chen L, Naowarojna N, Song H, Wang S, Wang J, Deng Z, et al. Use of a tyrosine analogue to modulate the two activities of a nonheme iron enzyme ovoA in ovothiol biosynthesis, cysteine oxidation versus oxidative C-S bond formation. J Am Chem Soc 2018;140(13):4604–12.

[117] Chpak IH, Halliwell B. Ergothioneine, antioxidant potential, physiological function and role in disease. Biochim Biophys Acta (BBA) - Mol Basis Dis 2012;1822(5):784–93.

[118] Ames BN. Prolonging healthy aging: longevity vitamins and proteins. Proc Natl Acad Sci U S A 2015;112(31):9816–44.

[119] Seebeck FP. In vitro reconstitution of mycobacterial ergothioneine biosynthesis. J Am Chem Soc 2010;132(19):6652–3.

[120] Hu W, Song H, Sue A, Bak DW, Naowarojna N, Elliott SJ, et al. Bioinformatic and biochemical characterization of C-S bond formation and degradative enzymes in the fungus Neurospora crassa ergothioneine biosynthetic pathway. Org Lett 2014;16(20):5382–5.

[121] Ioannides KN, Hua R, Chu Y, Liebisch G, et al. Structure of the ergothioneine synthase BioE from the eukaryote Neurospora crassa. Angew Chem Int 2015;54(9):2821–4.

[122] Naowarojna N, Ioannides KN, Hua R, Chu Y, Liebisch G, et al. Crystal structure of the ergothioneine synthase BioE from the eukaryote Neurospora crassa. ACS Catal 2019;9(8):6955–61.
Synthetic and Systems Biotechnology 6 (2021) 32–49

N. Naowarojna et al.

Appendix A: Abbreviation List

2-PCA: 2-pyridinecarbaldehyde
AA: Amino acid
adcR: tRNA synthetase
ADCT: Antibody-drug conjugates
AHA: Azidohomoalanine
AKR: Alkyne (propargyloxy)carbonyl-1-hydroxyamine
BONCAT: Bioorthogonal non-canonical amino acid labeling
caAA: Canonical amino acid
CBBT: 4-(cyanoethyl)benzoylloxo)-2,3,5,6-tetrafluorobenzensulfonate
Cox: Cytochrome c oxidase
CDO: Cytochrome d oxidase
cAAC: Copper-catalyzed azide-alkyne cycloaddition
cCox: Copper binding site
cMtb: Mtb-L29H F437H variant
cST: alkylidene bisazide biotinyl alkane tag
EPL: Expressed protein ligation
Fe-DPX: Iron-deuteroporphyrin IX
K20Ub: Ubiquitinated H2B
HArom: Hydrogen-atom transfer
HCO: Hemocyanin-oxidase
HPG: Homopropargyglycine
imTyr: (S)-2-amino-3-(4-hydroxy-3-(1H-imidazol-1-yl)phenyl)propanoic acid
KIE: Kinetic isotope effect
LD: Ligand-directed
LDNASA: Ligand-directed N-acetyl-N-alkylsulfonamide
McTyr: 3-methoxytyrosine
Mn(salen): Manganese-salen complex
MeTyr: 2-amino-3-(4-hydroxy-3-(methylthio)phenyl)propanoic acid
nAA: Noncanonical amino acid
NCL: Native chemical ligation
NHE: Non-heme iron enzyme
NHS: N-hydroxysuccinimide
NOR: Nitric oxide reductase
PACE: Phage-assisted continuous evolution
PCET: Proton-coupled electron transfer
PCS: Primary coordination shell
PEG: Polyethylene glycol
PLP: Pyridoxal 5'-phosphate
POE: Protein of interest
PTC: Pepidyl transferase center
PTM: Post-translational modification
PyRs: Pyrrole-H2N synthetase
QM/MM: Quantum mechanics/mechanical mechanics
RF-1: Release factor 1
ROS: Reactive oxygen species
SCS: Secondary coordination shell
SD: Shine-Dalgarno
SOD: Superoxide dismutase
SPACC: Strain-promoted azide-alkyne cycloaddition
swMtb: Physterine microcephalophus Myoglobin
tRNA: Transfer RNA
UBP: Unnatural base pair
ωK: α-ketoglutarate