


REVIEW

Applications of genetic code expansion technology in eukaryotes

Qiao-ru Guo¹, Yu J. Cao^{1,2,*} ¹State Key Laboratory of Chemical Oncogenomic, Guangdong Provincial Key Laboratory of Chemical Genomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China²Institute of Chemical Biology, Shenzhen Bay Laboratory, Shenzhen 518132, China*Correspondence: joshuacao@pku.edu.cn (Y. J. Cao)

Abstract

Unnatural amino acids (UAAs) have gained significant attention in protein engineering and drug development owing to their ability to introduce new chemical functionalities to proteins. In eukaryotes, genetic code expansion (GCE) enables the incorporation of UAAs and facilitates posttranscriptional modification (PTM), which is not feasible in prokaryotic systems. GCE is also a powerful tool for cell or animal imaging, the monitoring of protein interactions in target cells, drug development, and switch regulation. Therefore, there is keen interest in utilizing GCE in eukaryotic systems. This review provides an overview of the application of GCE in eukaryotic systems and discusses current challenges that need to be addressed.

Keywords genetic code expansion, unnatural amino acid, eukaryotes, basic research, therapeutic applications

Introduction

Genetic code expansion (GCE) has gained considerable interest in protein engineering, as it allows the site-specific incorporation of unnatural amino acids (UAAs) into proteins, resulting in the introduction of novel properties and functions. Proteins are composed of 20 natural amino acids, which display a wide range of functions. The limited number of building blocks constrains the ability to modify their function and explore their structure and interactions. The use of UAAs through GCE offers a promising avenue for expanding the diversity and functionality of proteins.

To overcome the limitations of the 20 natural amino acids, modified UAAs have been introduced to alter the physicochemical and biological properties of peptides and proteins. UAAs were first reported to be incorporated into dipeptides in 1983 (Heckler et al., 1983) and subsequently, in 1989, they were introduced into proteins (Noren et al., 1989). Since then, the application of GCE technology has steadily expanded to include enzyme engineering, drug discovery, conditional control

of biological processes, biophysical probing, and post-transcriptional modification (PTM), among others, by utilizing UAAs incorporated into multiple proteins to meet diverse research needs. At present, most GCE techniques are conducted primarily in prokaryotic systems such as *Escherichia coli* due to their advantages of cost-efficiency, high yield, and ease of operation. However, there is a growing need to extend these techniques to eukaryotic systems because prokaryotic systems are incapable of introducing PTMs to proteins and thus cannot fully replicate the original functions of proteins in eukaryotes. Moreover, restricting the introduction of UAA exclusively to prokaryotic systems would severely limit the application of GCE in critical areas such as live imaging in eukaryotes and the development of novel eukaryotic cell-based therapeutic strategies.

Since the successful incorporation of UAAs into the genetic code of *Saccharomyces cerevisiae* by Chin et al. in 2003 (Chin et al., 2003), UAA use has gradually been adopted in eukaryotic cells. Compared to prokaryotic expression systems, eukaryotic expression systems can

produce more complex glycosylated proteins, including full-length antibodies (Tian et al., 2014). The expansion of the genetic code in eukaryotic cells has greatly broadened the scope of UAA applications, including novel antibody conjugates, vaccines, living cell imaging (Peng and Hang, 2016), and other areas. In the following sections, we will provide a detailed overview of the application of GCE technology in eukaryotic platforms and explore the implications of this technology in eukaryotic systems.

Eukaryotic platform for gene expression systems

The first successful incorporation of UAAs into a protein was achieved in *E. coli* in 2001 (Wang et al., 2001). Since then, *E. coli* has become the most common host organism for GCE because of its simplicity and efficiency (Smolskaya and Andreev, 2019). However, as technology advances and the needs of scientific research evolve, there is an increasing demand for GCE in eukaryotic cells and animals. Over the past few decades, researchers have made significant progress in applying GCE in eukaryotes, such as yeasts, mammalian cells, stem cells, invertebrates, and vertebrates, as host organisms, enabling the resolution of various scientific problems (Fig. 1).

Technical elements

Unnatural amino acids

UAAs are typically chemically synthesized or derived from secondary metabolites of bacteria, fungi, plants, or marine organisms (Blaskovich, 2016). In xenobiology, UAAs are broadly categorized into two classes: (i) UAAs with structures similar to those of natural amino acids, often referred to as analogs or derivatives (Agostini et al., 2017) and (ii) UAAs with structures significantly different from those of natural amino acids, known as surrogates (Lepthien et al., 2008). Two main approaches are employed to synthesize UAAs: chemical synthetic routes and biocatalytic routes.

Chemical synthesis offers the advantages of synthesizing UAAs with diverse structures and wide applicability. Chemical synthetic routes are employed to obtain not only individual UAA structures but also dipeptide structures. The propeptide strategy stands out as an effective method to enhance the uptake efficiency of UAAs. The chemical synthesis of dipeptide structures facilitates the transport of poorly bioavailable UAAs into cells through ATP-binding box transporters (Fickel and Gilvarg, 1973; Parrish et al., 2012). Luo et al. prepared two dipeptides, H₂N-lysine-phosphotyrosine-COOH (Lys-pTyr) and H₂N-lysine-4-phosphomethyl-L-phenylalanine-COOH (Lys-Pmp), which led to an increase in the cytoplasmic concentration of these two UAAs in *E. coli* (Luo et al., 2017a). However, chemical synthesis lacks enantioselectivity and stereoselectivity, involves tedious steps and results in moderate yields (Narancic et al., 2019). In contrast, enzymatic asymmetric synthesis of chiral amino acids has the benefits of simple reaction steps and high enantioselectivity, conversion, and yield (Xue et al., 2018). Protein engineering has enabled the application of enzymes in UAA synthesis, overcoming their low stability and limited substrate range found in nature. Large-scale biosynthesis of UAAs is currently limited by its dependence on cofactors, but this challenge could be addressed by improving cofactor recycling methods (Narancic et al., 2019).

Blank codons and orthogonal aaRS/tRNA pair

Using auxotrophic strains to reduce the concentration of natural amino acids can help prevent the incorporation of nonspecific amino acids, even if they are not targeted to a specific site in the protein of interest (POI). However, even in auxotrophic strains, amino acids may still be present in the cellular pool and can be preferentially incorporated by aminoacyl-tRNA synthetase (aaRS)/tRNA pairs, leading to a lower efficiency of UAA incorporation. To achieve site-specific incorporation of UAAs into target proteins, a strategy using nonsense

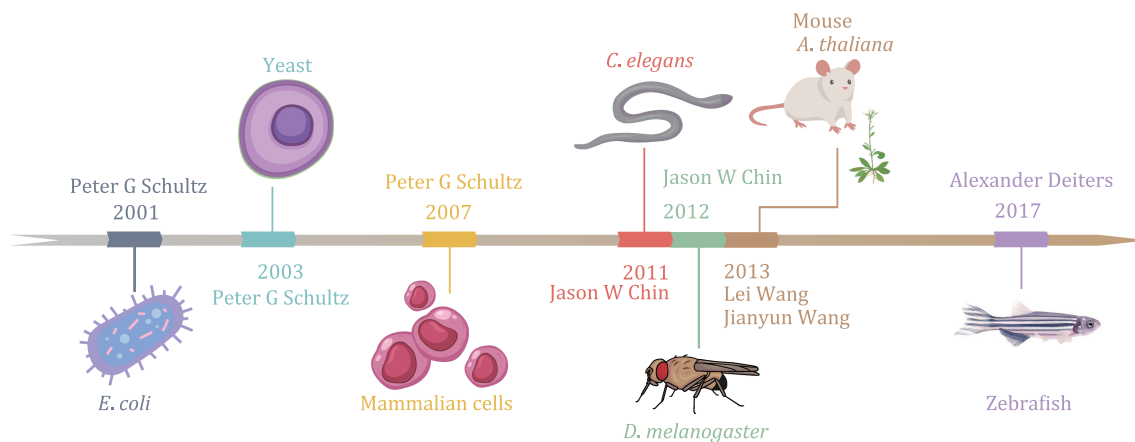


Figure 1. Timeline of the development of genetic code expansion in various organisms.

suppression is needed. This involves mutating the target site to a suppressor codon and introducing orthogonal aaRS/tRNA pairs into cells (Fig. 2).

The most widely used suppressor codon today is the amber nonsense codon (TAG), which has proven to be an effective tool for studying protein structure and manipulating protein function (Wang et al., 2001; Chemla et al., 2018). Additionally, the emergence of quadruplet codons in recent years has provided new ideas for incorporating UAAs (Neumann et al., 2010; Wang et al., 2012; Hankore et al., 2019). Regarding orthogonal aaRS/tRNA pairs, this means that each specific aaRS is specific only to its homologous amino acid and only recognizes its cognate isoacceptor tRNA (iso-tRNA) set, while the iso-tRNA set is only recognized by the given aaRS (Chin, 2006). The orthogonality of the suppression system in GCE ensures that site-specific incorporation of UAAs cannot be disturbed by the host's endogenous aaRS/tRNA pair or other natural amino acids, greatly improving the accuracy and reliability of UAA incorporation into proteins (Mukai et al., 2017).

In eukaryotic systems, the common orthogonal aaRS/tRNA pairs include PylRS/tRNA^{Pyl}, EcTyrRS/tRNA^{Tyr}, EcLeuRS/tRNA^{Leu}, and EcTrpRS/tRNA^{Trp}. The typical UAAs recognized by each orthogonal pair are listed in Table 1. PylRS/tRNA^{Pyl} and EcTyrRS/tRNA^{Tyr} are the two most frequently used aaRS/tRNA pairs in eukaryotic systems. Owing to its high substrate side chain promiscuity and

high tolerance for mutations in the substrate binding pocket (Wan et al., 2014), PylRS, derived from archaea of the family *Methanosarcinaceae*, possesses flexible active sites and has been widely shown to be orthogonal in bacteria and eukaryotes (Tharp et al., 2018). Therefore, PylRS is one of the most frequently applied aaRSs in eukaryotes. The EcTyrRS/tRNA^{Tyr} pair is also an aaRS pair with good orthogonality and was first developed by the Schultz group in 2003 (Chin et al., 2003). TyrRS has a spacious active site and can be mutagenized relatively easily to favor the incorporation of bulky UAAs (Crmković et al., 2019). EcLeuRS/tRNA^{Leu} and EcTrpRS/tRNA^{Trp} were developed in 2004 and 2017, respectively, for the incorporation of derivatives of leucine, methionine and cysteine (Wu et al., 2004) or tryptophan derivatives (Italia et al., 2017a). It is worth mentioning that in a recent study, Osgood et al. reported that the EcTrpRS/tRNA^{Trp} pair, unlike the other three common aaRS/tRNA pairs, exhibited a higher suppression efficiency for the TGA codon than for the TAG codon in mammalian cells. This property makes it an excellent tool for establishing a dual-UAA incorporation system (Osgood et al., 2023).

Challenges and optimization

Biosynthesis of UAAs

One of the key challenges in biosynthesizing UAAs is the difficulty of the heterologous expression of certain

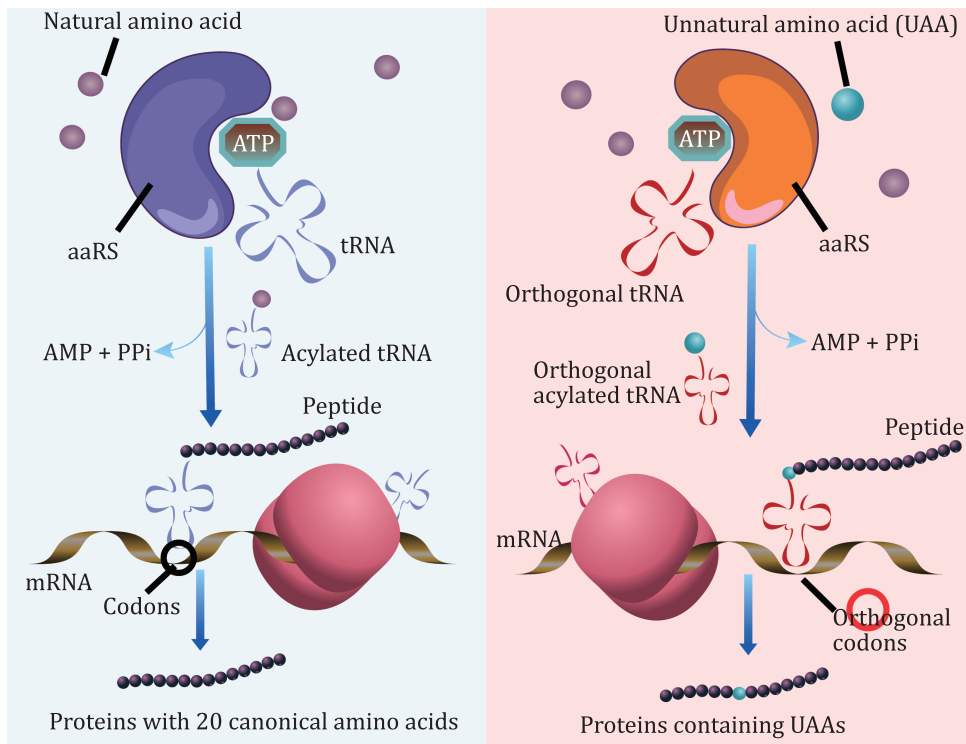


Figure 2. Overview of the working process and key components of the suppression system. Orthogonal tRNA synthetases (aaRSs) are responsible for aminoacylating their corresponding orthogonal tRNAs through UAAs. The aminoacylated tRNA then recognizes the blank codon on the ribosome, leading to the site-specific incorporation of UAAs into target proteins.

Table 1. Four common orthogonal aaRS/tRNA pairs and the typical UAAs in eukaryotes.

Orthogonal system	Origins	Codons	Typical UAAs	References
PylRS/tRNA ^{Pyl} pair	Mj. Mm.	UAG/GCA/AGC/CUG/AUG/ UAA/GUG/AAG/GAG/UAC/ UGG/AGGA/CUAG/UAGA	Lysine derivatives Phenylalanine derivatives	Mukai et al. (2008), Chatterjee et al. (2013b), Niu et al. (2013), Xiao et al. (2013), Elliott et al. (2014), Mills et al. (2021), and Xi et al. (2022)
TyrRS/tRNA ^{Tyr} pair	E. coli	UAG	Tyrosine derivatives Phenylalanine derivatives	Chen et al. (2007), Parrish et al. (2012), Palzer et al. (2013), Shao et al. (2015), and Mattheisen et al. (2023)
LeuRS/tRNA ^{Leu} pair	E. coli	UAG	Methionine derivatives Cysteine derivatives Lysine derivatives Alanine derivative	Wu et al. (2004), Takimoto et al. (2010), Wu et al. (2022b), and Mattheisen et al. (2023)
TrpRS/tRNA ^{Trp} pair	E. coli, Bs.	UAG/UGA	Tryptophan derivatives	Zhang et al. (2004), Italia et al. (2017a), and Osgood et al. (2023)

enzymes required for enzymatic catalysis in surrogate hosts. However, the cell-free protein system (CFPS) offers a new solution for the biosynthesis of UAAs (Tian et al., 2022). In a recent study, five enzymes in a biosynthetic pathway were successfully expressed using a CFPS, leading to the catalytic conversion of the lysine substrate to various lysine derivatives (Chen et al., 2023).

Utilization of UAAs

Numerous studies have reported methods to enhance the transmembrane efficiency of UAAs, such as utilizing a propeptide strategy in which the target UAA and its nonhydrolyzable analogs are incorporated into a dipeptide. This dipeptide is transported into cells and then hydrolyzed by nonspecific intracellular peptidases, thereby increasing the transmembrane efficiency of UAAs (Luo et al., 2017a). Another approach involves engineering bacterial organisms with an improved UAA uptake capacity by screening periplasmic binding protein mutants for UAA recognition (Ko et al., 2019).

When incorporating UAAs into eukaryotic cells or multicellular organisms, it is a common practice to administer high concentrations of exogenous UAAs to compensate for their poor integration efficiency in these systems. Autonomous biosynthesis systems for UAAs were initially developed for bacterial cells (Mehl et al., 2003; Chen et al., 2018). More recently, similar systems have been reported to replace the need for high concentrations of exogenous UAAs in eukaryotic cells and zebrafish (Wu et al., 2022a). Tyrosine sulfation, with the formation of sulfotyrosine (sTyr), is a PTM of proteins with low membrane permeability. To address the issue of low membrane permeability, the Xiao group utilized bioinformatics and computational methods to identify a sulfotransferase for the biosynthesis of sTyr. After further optimization of this sulfotransferase, cells capable of autonomous production of sTyr were constructed, which ameliorated the problem of inefficient incorporation

caused by the low membrane permeability of sTyr (Chen et al., 2022a). This kind of autonomous biosynthetic UAA incorporation may facilitate the application of GCE in multicellular organisms. While there are various methods to enhance the utilization of UAAs, there is currently no universal and convenient approach to improve UAA uptake efficiency in eukaryotes.

Directed evolution of aaRSs and tRNAs

To achieve orthogonality, aaRS evolution is required to remove the affinity of aaRSs for cognate amino acids and to improve the acylation efficiency of UAAs. The engineering of aaRSs includes complex analysis and redesign of the aaRS active sites, as well as mutagenesis of more than 15 residues around the active sites to evolve aaRSs with better UAA specificity (Crnković et al., 2019). Most evolved aaRS mutants contain only a few mutations that can enhance their activity against specific UAAs. Therefore, it is feasible to generate aaRSs capable of aminoacylating specific UAAs by building a library of aaRS mutants and identifying small improvements in aaRSs after high-throughput screening (Liu and Schultz, 2010; Crnković et al., 2019). The evolution of aaRSs using reporter assays, in which successful incorporation of a UAA into proteins serves as a marker of successful UAA aminoacylation, has been widely used (Liu et al., 1997; Santoro and Schultz, 2003). In addition, a two-step process involving positive and negative rounds to further evaluate and improve the orthogonality of an aaRS/tRNA pair toward a UAA has been extensively used (Santoro et al., 2002; Melançon and Schultz, 2009). The phage-assisted continuous evolution (PACE) technique has also been applied to the directed evolution and screening of aaRSs, enabling the rapid identification of PylRS mutants with improved kinetic parameters (Esvelt et al., 2011).

Functional expression of orthogonal tRNAs in eukaryotic systems is challenging owing to differences in tRNA transcription between prokaryotes

and eukaryotes. Orthogonal tRNAs are typically derived from prokaryotes, wherein tRNA transcription is driven by upstream promoters. In contrast, in eukaryotes, transcription is initiated by promoter elements, known as A- and B-boxes within a tRNA gene, which are absent in prokaryotes. Therefore, it is difficult to express prokaryotic-derived orthogonal tRNAs in eukaryotes (Wang, 2017). In 2007, the Wang group reported that type-3 Pol III mammalian promoters (such as H1, 7SK, U6 snRNA, etc.) in combination with 3'-flanking sequences could drive prokaryotic tRNA expression in mammalian cells (Wang et al., 2007). They also identified two Pol III yeast promoters for the expression of prokaryotic tRNAs in yeasts (Wang and Wang, 2008, 2012). Additionally, the expression level of a tRNA can be increased by repeating these tRNA expression cassettes (Coin et al., 2011).

In vitro eukaryotic platform

Yeast

The successful incorporation of UAAs in *E. coli* has led to further exploration of the applications of the GCE strategy in eukaryotic cells for the manipulation or study of eukaryotic cellular processes. In 2003, Chin and his colleagues overcame the obstacles posed by the translation machinery used in *E. coli*, which could not be applied to eukaryotic expression systems. They first incorporated five UAAs in *S. cerevisiae* (Chin et al., 2003). However, the low efficiency of UAA incorporation in yeast limits the effective application of this technology to the eukaryotic expression platform. The nonsense-mediated mRNA decay (NMD) pathway in yeast always leads to the rapid degradation of an mRNA that contains premature stop codons (Mitchell and Tollervey, 2003). The incorporation of UAAs into proteins can be affected when NMD occurs, leading to a shorter lifetime of the target mRNA and a lower protein yield. To address this problem, Wang et al. (2008) produced an NMD-deficient yeast strain and obtained a high yield compared to that in wild-type yeast in 2008. In addition, they optimized the tRNAs by applying internal leader Pol III yeast promoters to improve the efficiency of UAA incorporation.

The generation of the synthetic yeast *S. cerevisiae* genome 2.0 (Sc2.0) has provided a more diverse and promising eukaryotic expression platform for GCE. Sc2.0 enables the rapid generation of genetic variants with a large number of genomic rearrangements and modifications by using the Synthetic Chromosome Rearrangement and Modification by LoxPsym-mediated Evolution (SCRaMbLE) system mediated by loxP sites (Wu et al., 2018a). This facilitates rapid optimization of the genomic background, thereby improving the efficiency of UAA integration. Sc2.0 also provided a useful model for exploring the limitations of translation in eukaryotes (Sanders et al., 2022).

Mammalian cell lines

Mammalian cellular models are widely used research objects in the study of physiological and pathological processes. GCE is a powerful tool that can be applied in mammalian cells to dynamically monitor these processes, offering valuable insights for researchers.

The incorporation of UAAs into mammalian cells faces two major obstacles. The first is the low expression efficiency of prokaryotic-derived tRNAs in mammalian cells (Wang, 2017), while the second is the limited availability of engineered aaRS/tRNA pairs screened by the selection system, which restricts their incorporation into cells (Italia et al., 2017b). To overcome these difficulties, researchers have employed several strategies, including optimizing the expression of tRNAs by introducing the type-3 polymerase III promoter (Wang et al., 2007) and using transfer strategies (Zheng et al., 2017), increasing the affinity of aaRS/tRNA pairs (Takimoto et al., 2009), and constructing a special selection system for aaRS/tRNA pairs known as ATM (Italia et al., 2017a). Stem cells are characterized by their ability to differentiate into distinct cell lineages through a complex signaling network that regulates self-renewal and pluripotency. This feature has made stem cells a valuable tool for developing several therapeutic strategies (Hoang et al., 2022). However, in the case of terminally differentiated, nondividing cells, such as neurons, introducing UAAs into proteins can modulate their protein function and monitor their physiological or pathological cellular processes. Obtaining such mature cells through conventional methods, which involve isolation from living tissues, can be challenging as it is difficult to stably introduce exogenous material into them (Shen et al., 2011). To address this, researchers have explored incorporating UAAs into proteins in stem cells and inducing their differentiation into mature cells for research purposes.

This idea raises new questions as achieving stable expression in stem cells is challenging, and commonly used transient transfection methods are not applicable. Transient transfection of stem cells is known to be inefficient (Wang et al., 2007), and the differentiation process of stem cells is often lengthy, leading to a loss of the orthogonal aaRS/tRNA pair if transient transfection is used. This creates uncertainty regarding the successful incorporation of UAAs into terminally differentiated nondividing cells (Shen et al., 2011). Therefore, stable expression of exogenous aaRS/tRNA pairs in stem cell chromosomes is necessary to study the entire differentiation process of stem cells using UAAs. Shen and colleagues (2011) utilized a lentiviral vector to stably incorporate fluorescent UAAs into the voltage-dependent membrane lipid phosphatase of the neural stem cell line HCN-A94 and induced stem cell differentiation. The study also showed that UAA incorporation did not affect the differentiation process of stem cells. Furthermore, the application of

GCE in this study helped reveal conformational changes in voltage-sensitive domains during depolarization of the cell membrane.

In vivo eukaryotic platforms

Invertebrate animals

Caenorhabditis elegans

Caenorhabditis elegans has been an important model system for understanding mutant phenotypes at the cellular level since the 1970s because of its sequenced genome and mapped lineage for each cell from embryogenesis through postembryonic development (Sulston and Horvitz, 1977; Kimble and Hirsh, 1979). With its excellent optical transparency (Wang and Audhya, 2014; Kinser and Pincus, 2017), *C. elegans* has been extensively utilized in studies of photocaging (Alexpandi et al., 2019; Prasanth et al., 2020), photocrosslinking (Burnett et al., 2018), and fluorescent UAAs (Chang et al., 2013). In 1985, researchers demonstrated that *C. elegans* tolerated amber stop codon suppression (Hodgkin, 1985). Furthermore, *C. elegans* is small and easy to culture, making it an ideal candidate for GCE. Until recently, *C. elegans* was used as the primary multicellular biological platform for GCE.

To incorporate UAAs into *C. elegans*, researchers used a type 3 Pol III promoter (Chang et al., 2013) or the RNase P RNA (*rpr-1*) promoter (Davis et al., 2021) to express orthogonal tRNAs from *E. coli*. Previous studies have shown that orthogonal aaRS/tRNA pairs and reporters were transiently expressed on extrachromosomal arrays in *C. elegans*, and the expression was maintained through the use of antibiotics (Frøkjær-Jensen et al., 2008). However, this approach leads to a UAA-independent amber codon read-through and a low success rate in constructing mutant *C. elegans* (Wang, 2017). Genomic integration has significantly increased the success rate of creating transgenic *C. elegans* that exhibits a high degree of UAA dependence (Parrish et al., 2012). This phenomenon highlights that GCE in *C. elegans* requires the stabilization of orthogonal aaRS/tRNA pairs in the genome, ensuring genetic uniformity and inheritance.

A study conducted by Parrish et al. (2012) demonstrated the incorporation of various UAAs in response to the amber stop codon into several proteins in *C. elegans* by directly evolving the *E. coli*-derived tyrosyl and amber leucyl suppressor aaRS/tRNA pairs. The NMD pathway is a major factor that impedes efficient UAA incorporation in *C. elegans*. To overcome this issue, researchers disrupted the NMD pathway by knocking down *smg-1*, which resulted in efficient incorporation of UAAs in *C. elegans* (Parrish et al., 2012). Another study aimed to eliminate the interference from the NMD pathway by crossing NMD-deficient *smg-2(e2008)* worms with worms expressing a reporter, which led to an increased GFP reporter expression signal (Greiss and Chin, 2011).

Drosophila melanogaster

Similar to *C. elegans*, *D. melanogaster*, as the most established metazoan model organism, has a sequenced genome (Adams et al., 2000) and was shown to be tolerant of amber stop codon suppression in a previous study (Laski et al., 1989). Additionally, it has advantages as a potential platform for expanding the genetic code, including a short generation time, an established tool for creating transgenic flies, and control of transgene expression (Laski et al., 1989; Garza et al., 1990; Washburn and O'Tousa, 1992). Consequently, in 2012, Bianco et al. (2012) investigated the incorporation of UAAs into *D. melanogaster* by expressing PylT from a type 3 Pol III promoter and using an orthogonal PylRS/tRNA system to successfully integrate two UAAs into *D. melanogaster*.

Vertebrate animals

Zebrafish

Zebrafish is a widely used model organism in the study of vertebrate development and disease owing to its high fecundity and short generation time. Its excellent optical transparency during embryonic stages enables noninvasive *in vivo* imaging with high spatial and temporal resolutions, making zebrafish a powerful tool for investigating gene regulation and networks. In summary, zebrafish is an outstanding noninvasive optical model for biological research.

To expand the genetic code in zebrafish, the PylRS system was introduced, and photocaged UAAs were incorporated into zebrafish embryos to optically control mitogen-activated protein kinase kinase 1 (MEK1) activation, affecting the MEK/extracellular signal-regulated kinase (ERK) pathway and observing the role of the activation of the MEK/ERK pathway at different stages of zebrafish development (Liu et al., 2017). Subsequently, some investigators successfully incorporated several UAAs that had never been used in the zebrafish expression system (Syed et al., 2019). It has been suggested that the application of the zebrafish expression system as a relatively new platform for GCE has gradually matured.

Mouse

Mus musculus is a commonly used laboratory mouse that is widely employed as a mammalian model in the field of life sciences for physiological, pathological, and pharmacological studies (Chen et al., 2017). In 2013, the GCE technique was first applied to the embryonic mouse neocortex *in vivo* to genetically incorporate a photoreactive UAA into Kir2.1, an inwardly rectifying potassium channel, thereby regulating K⁺ channel switching to control neuronal firing (Kang et al., 2013). A major challenge for UAA incorporation in mice, as in other *in vivo* platforms, is to find an efficient way to deliver and express the necessary components for UAA incorporation. In

the aforementioned work, researchers employed uterine electroporation to deliver the *E. coli*-derived LeuRS/tRNA_{CUA} pair into the mouse neocortex.

However, there is still a need to develop widely applicable and efficient gene delivery methods for *in vivo* platforms. In 2016, Ernst et al. (2016) reported the use of adeno-associated viral (AAV) vectors to deliver the components of the suppression system to live mouse brains. This delivery method offers the advantage of site-specific and efficient incorporation of different UAAs through the injection of AAV vectors. However, stable gene expression cannot be achieved with this approach, and the expression of aaRS/tRNA pairs is transient. One year later, a transgenic mouse capable of stably expressing AcKRS/tRNA pairs was developed to temporally and spatially control protein acetylation in different mouse organs (Han et al., 2017). In this study, two transgenic mice that expressed PylRS/tRNA and GFP39TAG/tRNA were developed. Furthermore, they were crossed to generate double-heterozygous animals carrying the components of the suppression system and the reporter gene GFPuv. When these double-heterozygous mice were crossed with a transgenic mouse carrying an amber stop codon, a UAA was successfully incorporated into multiple organs through daily intraperitoneal injection alone (Han et al., 2017). This strategy provides a convenient system for the expansion of the genetic code *in vivo*, but establishing a stable mouse line is time consuming.

Plants

Using plants instead of microbial fermentation or animal cell cultures to produce therapeutic proteins has been proposed as an attractive approach. Expanding the genetic code in plants could lead to safer, more productive, and less costly recombinant therapeutic processes. Arabidopsis, a model organism in plant biology, was the first plant species to have its genome sequenced. In 2013, Li et al. attempted to introduce UAAs into *Arabidopsis thaliana*. To circumvent the NMD pathway, researchers selected a mutant lacking the UPF1 gene for their study, as UPF1 is a key regulator of the NMD pathway in *A. thaliana* (Li et al., 2013). Furthermore, the authors chose the type 3 Pol III Arabidopsis promoter 7SL4 to express tRNA_{CUA}^{Pyl}. In this work, the UAA *N*- ϵ -acrylylsine (AcrK) was site-specifically incorporated into a reporter in *A. thaliana* (Li et al., 2013). This approach could be useful for expanding genetic coding in other plants in the future.

Applications of GCE

Basic research

Protein activity switch

Optical control

To manipulate biological processes naturally and precisely, scientists require means to regulate protein

activity. Traditional methods of regulating protein activity include using the tetracycline (Tet) system (Kozlova and Berens, 2012), gene recombination systems such as Cre-loxP (Kim et al., 2018), or siRNAs (Stokman et al., 2010; Friedrich and Aigner, 2022). However, optical control of protein function provides noninvasive and precise spatiotemporal regulation compared to these approaches. The incorporation of light-responsive amino acids into proteins is an important way to optically regulate protein functions. Over the past few decades, several studies have demonstrated the effectiveness of optical control of protein function. With recent research advancements, significant breakthroughs have been made in the application of this approach in eukaryotes.

One approach to optically regulate protein function is to incorporate UAAs containing a photolabile protecting group (photocage) at specific sites of proteins. The activation state of proteins can be modulated by light of specific wavelengths, depending on the nature of the photocage group. The *O*-nitrobenzyl (ONB) group is a common photocage group (Wu et al., 2004), and the ONB-caged versions of several UAAs are stable under physiological conditions. However, they require high-intensity UV light (250–365 nm) to remove the cage group, which can cause cellular damage. To overcome this limitation, modifications to the structure of cage groups have been made. For instance, dimethoxy-nitrobenzyl (DMNB)-caged UAAs have high quantum yields and can be deprotected in living cells by visible blue light (Lemke et al., 2007). UAA modified with these photocage groups has been incorporated into mammalian cells, *C. elegans* and zebrafish to precisely regulate the function of various proteins, including proteases, nucleases, kinases, and the Cas9 protein. Table 2 summarizes the applications of these photocaged UAAs.

Although the incorporation of photocaged amino acids has shown advantages in rapidly and precisely regulating protein function, their irreversible activation based on the decaging reaction limits their application in studying protein activation-deactivation cycles under physiological conditions (Courtney and Deiters, 2018). Moreover, targeting a single residue to modulate protein function has limited effects, as protein function often involves the structure of protein domains and conformation (Beharry and Woolley, 2011). To overcome these limitations, photoswitchable amino acids (PSAAs) have been developed. PSAAs containing an azobenzene moiety can covalently bind to a cysteine residue in the protein domain to form a photocontrollable bridge *in situ*, allowing light-inducible control of protein conformation (Hoppmann et al., 2015; Hoppmann and Wang, 2019). Furthermore, the azobenzene moiety in a PSAA can reversibly change its conformation under two wavelengths of light, making the regulation of protein activity reversible (Aemissegger and Hilvert, 2007). In 2006, a photoswitchable azobenzene

Table 2. List of photocaged amino acids incorporated in eukaryotes.

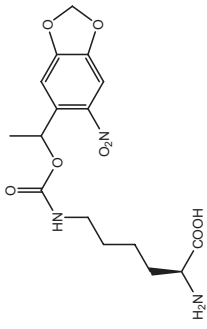
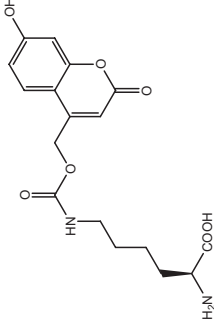
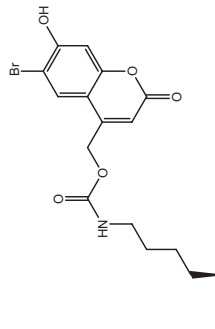
Amino acid	Structure	Eukaryotes	aaRS/tRNA pair	Application	References
Caged lysine		HEK293T	PyIRS/tRNA ^{PyI}	Cre recombinase T7 RNA polymerase MEK1 Isocitrate dehydrogenase Nuclear localization signal (NLS) of the transcription factor SATB1 NLS of nucleoplasmin and the tumor suppressor p53	Luo et al. (2016a, b) Hemphill et al. (2013) Gautier et al. (2011) Walker et al. (2016) Engelke et al. (2014) Gautier et al. (2010)
		HEK-TCR cell (Created by HEK293T and stably express 1G4 TCRαβ chains and CD3γδϵζ chains) <i>C. elegans</i>		CRISPR associated (Cas) proteins A tyrosine kinase LCK	Hemphill et al. (2015) Liaunardy-Jopeace et al. (2017)
		HEK293T		Cre recombinase Cre recombinase EGFP and Fluc SH2 domain-containing protein tyrosine phosphatase-2 MAPK phosphatase	Davis et al. (2021) Xi et al. (2022) Luo et al. (2014) Ryan et al. (2022) Courtney and Deiters (2019)
		Zebrafish embryos		MEK1 Cre recombinase EGFP and Fluc	Liu et al. (2017) Brown and Deiters (2019) Luo et al. (2014)
		HEK293T			

Table 2. Continued

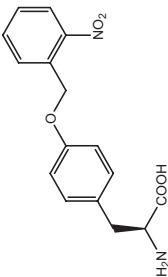
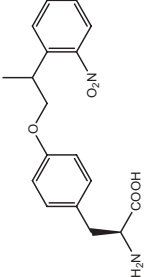
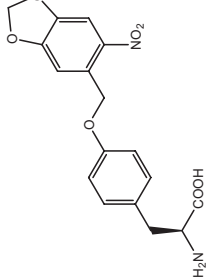
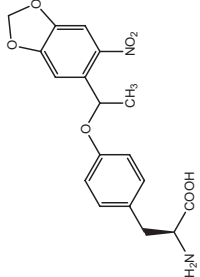
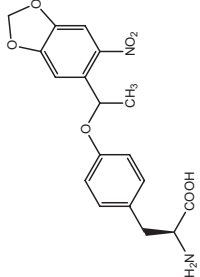
Amino acid	Structure	Eukaryotes	aaRS/tRNA pair	Application	References
Caged tyrosine		HEK293T	PyIRS/tRNA ^{PyI}	EGFP and Fluc Cre recombinase T7 RNA polymerase Zinc-Finger nucleases	Luo et al. (2017b) Edwards et al. (2009) Chou et al. (2010) Chou and Deiters (2014)
		HEK293T, PC12 cell line and SH-SY5Y cell		STAT1 phosphorylation Toll-like receptor 8 Tyrosine kinase A	Arbely et al. (2012) Yang et al. (2022) Zhao et al. (2020)
		HEK293T		EGFP and Fluc	Luo et al. (2017b)
		HEK293T		EGFP and Fluc SH2 domain-containing protein tyrosine phosphatase-2 Toll-like receptor 8	Luo et al. (2017b) Ryan et al. (2022) Yang et al. (2022)
		HEK293T		EGFP and Fluc Cre recombinase	Luo et al. (2017b) Luo et al. (2016a)

Table 2. Continued

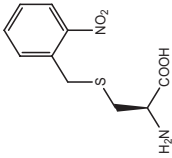
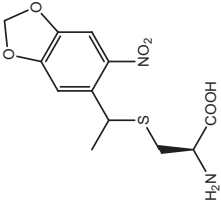
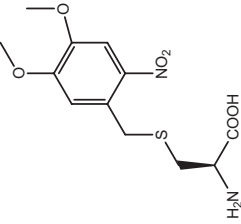
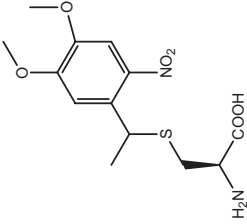
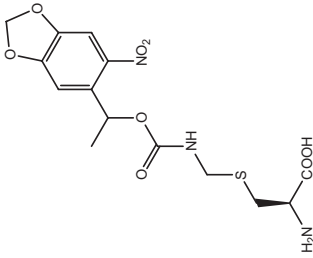
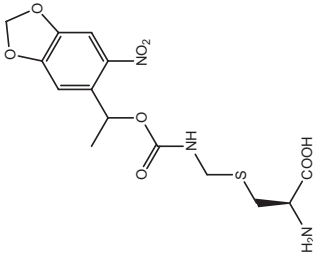
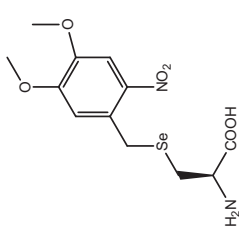
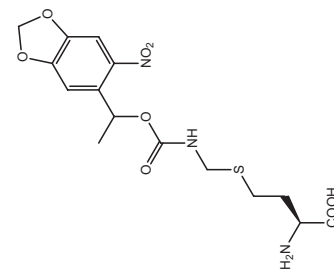
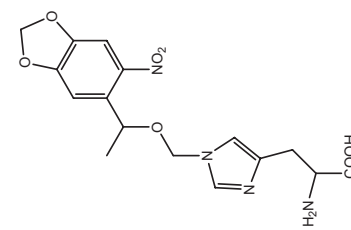
Amino acid	Structure	Eukaryotes	aaRS/tRNA pair	Application	References
Caged cysteine/ selenocysteine/ homocysteine		HEK293T	PyIRS/tRNA ^{Pyl}	TEV protease	Nguyen et al. (2014)
		HEK293T <i>C. elegans</i>	PyIRS/tRNA ^{Pyl} PyIRS/tRNA ^{Pyl}	TEV protease Casepase-3	Nguyen et al. (2014) Xi et al. (2022)
		HEK293T	LeuRS/tRNA ^{Leu}	MAPK phosphatase	Courtney and Deiters (2019)
		Mouse		Src Kir2.1	Ren et al. (2015a) Kang et al. (2013)
		HEK293T	LeuRS/tRNA ^{Leu}	Src	Ren et al. (2015a)
		HEK293T	PyIRS/tRNA ^{Pyl}	EGFP and Renilla luciferase	Uprety et al. (2014)

Table 2. Continued

Amino acid	Structure	Eukaryotes	aaRS/tRNA pair	Application	References
		HEK293T	LeuRS/tRNA ^{Leu}	EGFP	Peeler et al. (2020)
		HEK293T	PyIRS/tRNA ^{Pyl}	EGFP and Renilla luciferase	Uprety et al. (2014)
Caged histidine		Zebrafish embryos	PyIRS/tRNA ^{Pyl}	Cre recombinase and the single-chain variant viral protease from the hepatitis C virus	Brown et al. (2023a)

amino acid was successfully encoded in *E. coli* (Bose et al., 2006). Earlier studies incorporated amino acids with azobenzene moieties into proteins such as catabolite activator protein (Bose et al., 2006) and horseradish peroxidase (Muranaka et al., 2002), which played well-controlled roles. This strategy has also been applied to eukaryotic systems. The Chin group used this strategy for their research and reported the rapid and selective regulation of specific MEK isoenzymes in mammalian cells (Tsai et al., 2015). The Wang group subsequently achieved reversible photoswitching control of glutamate receptors in mammalian cells, thereby proving the feasibility of reversible optical control of neuronal receptors (Klippenstein et al., 2017). Nonetheless, there are several limitations to the widespread application of PSAAs. One limitation is the difficulty in determining suitable PSAA incorporation sites, which requires extensive screening. Current technology cannot guarantee that such sites can be found in all target proteins (Nödling et al., 2019).

Small-molecule regulation

In addition to incorporating light-responsive UAAs, protein function can also be controlled by introducing small molecule-cleavable caging groups. Currently, there are three main approaches for removing protecting groups from UAAs using small-molecule modulation in eukaryotic systems. The first method involves using palladium catalysts to deprotect propargyloxycarbonyl-caged lysine or allene-caged tyrosine derivatives (Li et al., 2014b; Wang et al., 2016a) (Fig. 3A). While this approach is highly biocompatible, its slow kinetics and low decaging yields limit its generalizability.

The *trans*-cyclooctenyloxycarbonyl (TCO) group is cleaved by tetrazine via Diels–Alder (invDA) cycloaddition, leading to the release of lysine (Li et al., 2014a) (Fig. 3B). Li et al. were the first to investigate this invDA reaction between the TCO-caged Fmoc-lysine analog (Fmoc-TCOK) and various tetrazine derivatives. They identified the optimal caging–decaging pair as TCOK- α and Me2Tet, which they used to successfully incorporate and activate the GFP protein in 293T cells (Li et al., 2014a). This work offers a general approach for rapidly switching protein function through small molecules, which could potentially be extended to the decaging of other natural amino acids. Additionally, the high biocompatibility and deep tissue penetration make this method suitable for use in intact animals. In a subsequent study, researchers applied this UAA-decaging toolkit in mice to manipulate kinase activity *in vivo* (Zhang et al., 2016).

The third strategy involves removing the azidobenzoyloxycarbonyl (ABZ) group via the Staudinger reaction, which is induced by phosphine and reduces the azide to an amine, leading to self-immolation and exposure of lysine (Fig. 3C). The Staudinger reduction of aryl azides by phosphines has been demonstrated to

be highly biocompatible and nontoxic in living cells or organisms in several studies (Agard et al., 2006; Sletten and Bertozzi, 2011; Fang et al., 2019; Fottner et al., 2019). Based on this reaction, the ABZ-lysine derivative *ortho*-azidobenzoyloxycarbonyl lysine (OABK) was first used to control protein function in mammalian cells and regulate multiple cellular processes, including bioluminescence (luciferase), protein translocation (nuclear localization sequence) and DNA recombination (Cre) (Luo et al., 2016b). However, OABK has several limitations that hinder its broad application, the most significant of which is its slow kinetics, which makes it challenging to control protein function on short timescales (Luo et al., 2016b). Consequently, the research team improved the structure of OABK by relocating the azide from the *ortho* position to the *para* position to accelerate the self-immolation reaction of substrates (Alouane et al., 2015), and developed a lysine derivative, PABK, with rapid decaging properties (Wesalo et al., 2020). PABK triggered nuclear translocation 81% faster than OABK and successfully achieved small-molecule control of SUMOylation in mammalian cells (Wesalo et al., 2020).

In a recent study, researchers performed small molecule-triggered UAA-decaging reactions in zebrafish embryos (Brown et al., 2023b). They selected two caging–decaging pairs, *N*^ε-*p*-azidobenzoyloxycarbonyl lysine (PABK)/phosphine 2-(diphenylphosphino) benzamide (2DPBM) and TCOK/methyl tetrazine (MeTz), to investigate their incorporation and decaging efficiencies, as well as their effects on the growth of zebrafish. The study showed that although TCOK/MeTz exhibited a good incorporation efficiency and a faster decaging process, most embryos died at a high MeTz concentration (100 μ mol/L) (Brown et al., 2023b). Therefore, the researchers further explored PABK/2DPBM. The results demonstrated that the small molecule-triggered decaging of PABK enabled tunable protein activation in DNA recombination and nuclear translocation events via viral recombinase and viral protease. This system was also used to investigate the molecular mechanism of heart defects caused by a RASopathy mutant of NRAS (Brown et al., 2023b).

Overall, the advantage of small molecule-triggered UAA-decaging strategies over traditional small-molecule protein control methods is their greater universality. This is because small-molecule protein activators are usually identified through high-throughput screening or accidental discovery, which is not universally applicable and cannot be directly transferred to other proteins. In contrast to protein function control based on photocaged UAAs, small-molecule-triggered decaging reactions often do not require special photostimulation equipment. Additionally, UV irradiation of live cells or organisms may cause cytotoxicity and alter intracellular signaling pathways. The application of small molecules in decaging reactions tends to lead to better penetration

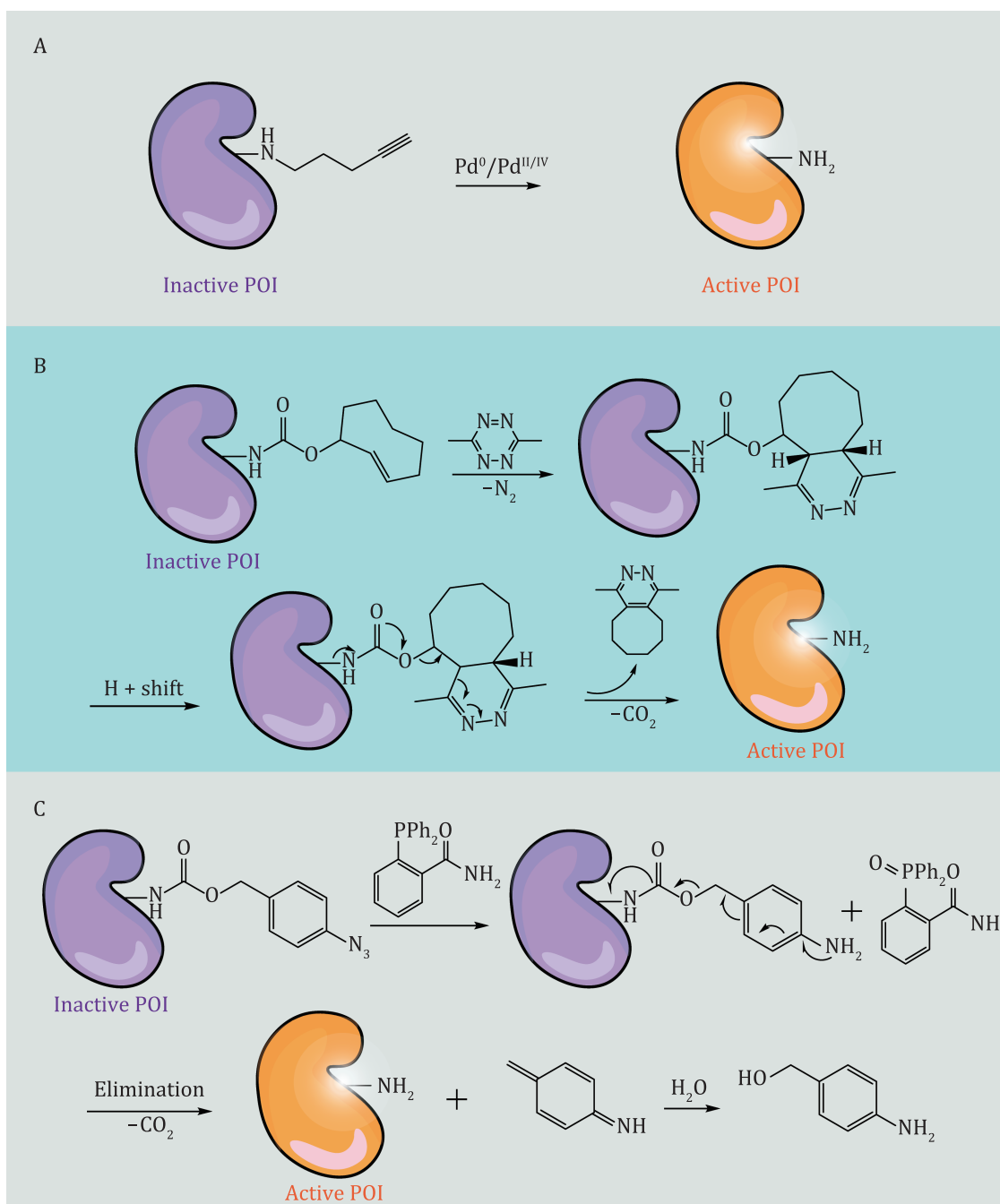


Figure 3. Three primary reactions for removing protective groups from UAAs through small molecule modulation. (A) Utilizing a palladium catalyst to decage propargyloxycarbonyl-caged lysine or allene-caged tyrosine derivatives. (B) Employing the Diels–Alder (invDA) cycloaddition reaction to eliminate the *trans*-cyclooctenyloxycarbonyl (TCO) group. (C) Applying the Staudinger reaction to decage the azidobenzyloxycarbonyl (ABZ) group.

into deep tissues than that observed upon optical activation, making it more feasible for use in intact animals.

Posttranscriptional modifications

PTMs cause covalent attachments of specific chemical groups to side chains of nucleophilic amino acids and can be enzymatic or nonenzymatic (Lee et al., 2023). These modifications greatly enrich the structural and functional diversity of the proteome in cells. PTMs are

mediated by enzymes and can affect protein functions, including subcellular localization, stability, activity, and interactions (Barber and Rinehart, 2018). Aberrant PTMs are closely related to many human diseases, and understanding their functional significance in disease can deepen our understanding of diseases and help develop effective therapeutic strategies. Various approaches have been applied to study PTMs, such as replacing serine with glutamate in proteins (Chang et al., 2014), but these

methods cannot mimic the stereoscopic effects of PTMs on proteins (Chen and Tsai, 2022). Although enzymes can be used to regulate the PTM process in cells, achieving site-specific modifications can be challenging (Zhang et al., 2016). While chemical synthesis can achieve site-specific modifications, it is difficult, and delivering the modified protein into cells is also a challenge (Siman and Brik, 2012).

GCE appears to be a promising approach for studying PTMs, as the incorporation of UAAs into proteins allows site-specific modification of PTMs and the generation of homogeneously modified proteins. The applications of UAAs in PTM research can be divided into two main categories: direct incorporation and indirect incorporation (incorporating precursors and converting them to the desired modifications in cells). Additionally, there have been studies that used photocaged UAAs or other methods to regulate the occurrence of PTMs by switching the activity of PTM-related enzymes (Mangubat-Medina and Ball, 2021). For example, a recent study incorporated photocaged lysine derivatives into O-GlcNAc transferase (OGT) to regulate the activation of OGT by photodecaging, enabling spatiotemporal precision control of cellular O-GlcNAc levels (He et al., 2022). This information will not be repeated here, and interested readers can refer to Section “A switch for controlling protein activity”.

Previous studies have shown that direct incorporation of UAAs in eukaryotic platforms can regulate PTMs such as Ser phosphorylation, Tyr sulfation, and Lys acetylation and acylation. However, for specific PTMs such as ubiquitylation that require UAAs with large modification groups (e.g., ubiquitylation, etc.), evolving the aaRS can be challenging. In addition, delivering ubiquitin-lysine into cells can also be challenging (Chen and Tsai, 2022). Therefore, researchers have used indirect incorporation methods, in which the precursor of the modified amino acid is incorporated into proteins and enzymatic reactions in cells are used to produce the protein with the desired modifications. This approach has been used to study the methylation, ubiquitination and ubiquitin-like modification of lysine and cysteine. In eukaryotic platforms, lysine-SUMOylation has been achieved using this method (Fottner et al., 2019). Table 3 summarizes the studies of modified UAAs in PTM research in eukaryotic systems.

Protein interaction analysis

GCE-based photo- or chemical-crosslinking enables residue-selective crosslinking, making it a useful tool for studying protein interactions with other biomolecules. By incorporating UAAs carrying chemical groups with photocrosslinking or chemical-crosslinking abilities, the limitations of traditional chemical crosslinking methods, such as high false-positive identification rates, can be overcome. Additionally, this approach can capture weak

and transient interaction partners, which may be missed by immunoprecipitation methods. In other areas, such as nanopores, which known for their high-throughput, cost-effectiveness, and rapid response as biosensors, the integration of UAAs can significantly enhance the chemical environment within the nanopore and increase its structural and functional diversity (Wu et al., 2023; Yang et al., 2023). Leveraging these advantages, these crosslinked UAAs not only serve as excellent tools for investigating protein interaction mechanisms, but also hold tremendous potential in antibody selection.

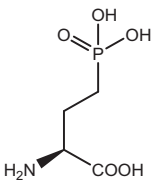
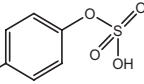

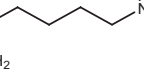
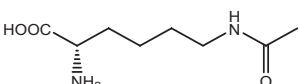

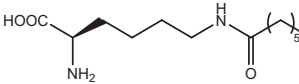
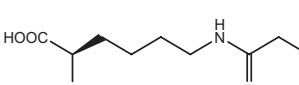
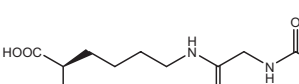
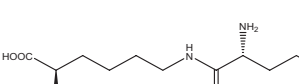
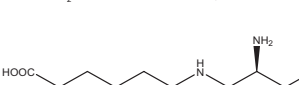
Photocrosslinking

To ensure the stability and functionality of photocrosslinked UAAs in cells, the photocrosslinking groups must be chemically stable and orthogonal to all the functional groups present in cells. Moreover, the excitation wavelength used for crosslinking should be a longer wavelength and not cause any harm to cells. These factors together can lead to a higher crosslinking efficiency and more stable complex formation.

Three major photoreactive groups, namely, aryl azides, benzophenone, and diazine, have been incorporated into UAAs (Coin, 2018) (Fig. 4). Aryl azides are activated at a wavelength of 250 nm to form the reactive nitrene intermediate that inserts into nearby C–H and heteroatom–H bonds (Tanaka et al., 2008) [Fig. 4B (1)]. Benzophenones, which are activated at wavelengths of 350–365 nm, produce ketyl diradicals upon activation, showing a preference for γ -CH bonds (Coin, 2018) [Fig. 4B (2)]. Although benzophenones are widely used because of their high crosslinking efficiency, they have a bulky and rigid structure with a hydrophobic core and side reactivity with amines (Young et al., 2011). On the other hand, diazoxides, which are smaller in size than the previous two photocrosslinkers, are activated at wavelengths of 350–365 nm to produce a reactive carbene intermediate [Fig. 4B (3)]. The carbene intermediate readily inserts into C–H and heteroatom–H bonds (Coin, 2018). Diazoxides are characterized by rapid crosslinking, but the formation of diazo-isomers in the reaction can affect the efficiency of crosslinking. In a 2017 study, aryl azides and diazirines were found to be more effective than benzophenones in reducing off-target crosslinking and background noise (Kleiner et al., 2017). To function stably in cells, the photocrosslinking groups on UAAs need to be chemically stable and orthogonal to all the functional groups in cells. Additionally, the longer light wavelength used to excite the crosslinking activity should not be harmful to cells, thereby allowing a higher crosslinking efficiency and more stable complex formation.

In 2002, the successful incorporation of *p*-benzoyl-L-phenylalanine (pBpa), a photocrosslinkable UAA containing a benzophenone group, into the dimeric protein glutathione S-transferase in *E. coli* through the

Table 3. PTM modified UAAs in eukaryotic platform.

PTM	Amino acid	Structure	aaRS/tRNA	Eukaryotes	Action mode	References
Phosphorylation	Ser		SerRS/tRNA	HEK293T	Direct incorporation	Beránek et al. (2018)
Sulfation	Tyr		TyrRS/tRNA	HEK293T	Direct incorporation	Italia et al. (2020)
	Tyr		TyrRS/tRNA	HEK293T	Direct incorporation	He et al. (2020)
	Tyr		TyrRS/tRNA	HEK293T	Direct incorporation	Chen et al. (2022a)
Acetylation	Lys		PylRS/tRNA	HEK293T	Direct incorporation	Mukai et al. (2008)
	Lys		PylRS/tRNA	Mouse embryonic stem cell	Direct incorporation	Elsässer et al. (2016)
Acylation	Lys		PylRS/tRNA	HEK293T	Direct incorporation	Wang et al. (2016b)
	Lys		PylRS/tRNA	HEK293T	Direct incorporation	Wang et al. (2019)
SUMOylation	Lys		PylRS/tRNA	HEK293T	Indirect incorporation	Fottner et al. (2019)
Ubiquitination	Lys		PylRS/tRNA	HEK293T	Indirect incorporation	Zang et al. (2023)
						

MjTyrRS/tRNA pair, resulted in efficient crosslinking of protein subunits after irradiation. This work was the first application of UAAs for photocrosslinking control of proteins (Chin et al., 2002). However, since the archaeal TyrRS is not orthogonal in mammalian cells, Hino et al. (2011) introduced EcTyrRS variants to achieve the incorporation of pBpa in mammalian cells. To further expand the specific synthetase available for photocrosslinkable UAAs in eukaryotic systems, Lacey et al. (2013) constructed highly diverse synthetase libraries and evolved MmPylRS for pBpa and other phenylalanine derivatives with multiple conjugated aromatic rings to be incorporated in mammalian cells.

A significant portion of photocrosslinkable UAAs are derivatives of phenylalanine. These derivatives

often have short, rigid side chains, and their photoactive groups are located close to the protein backbone, resulting in the formation of short covalent bridges that can capture only short-range interaction partners (Yanagisawa et al., 2012). This property limits the study of long-range protein crosslinking. To overcome this limitation, researchers have attempted to introduce lysine derivatives with photocrosslinkable groups using MbPylRS/tRNA or MmPylRS/tRNA pairs for long-range UAA photocrosslinking, resulting in the achievement of “wide-range” photocrosslinking of proteins in yeast and mammalian cells (Hancock et al., 2010; Yanagisawa et al., 2012).

Photocrosslinkable UAAs have been utilized to investigate various types of protein interactions, such as

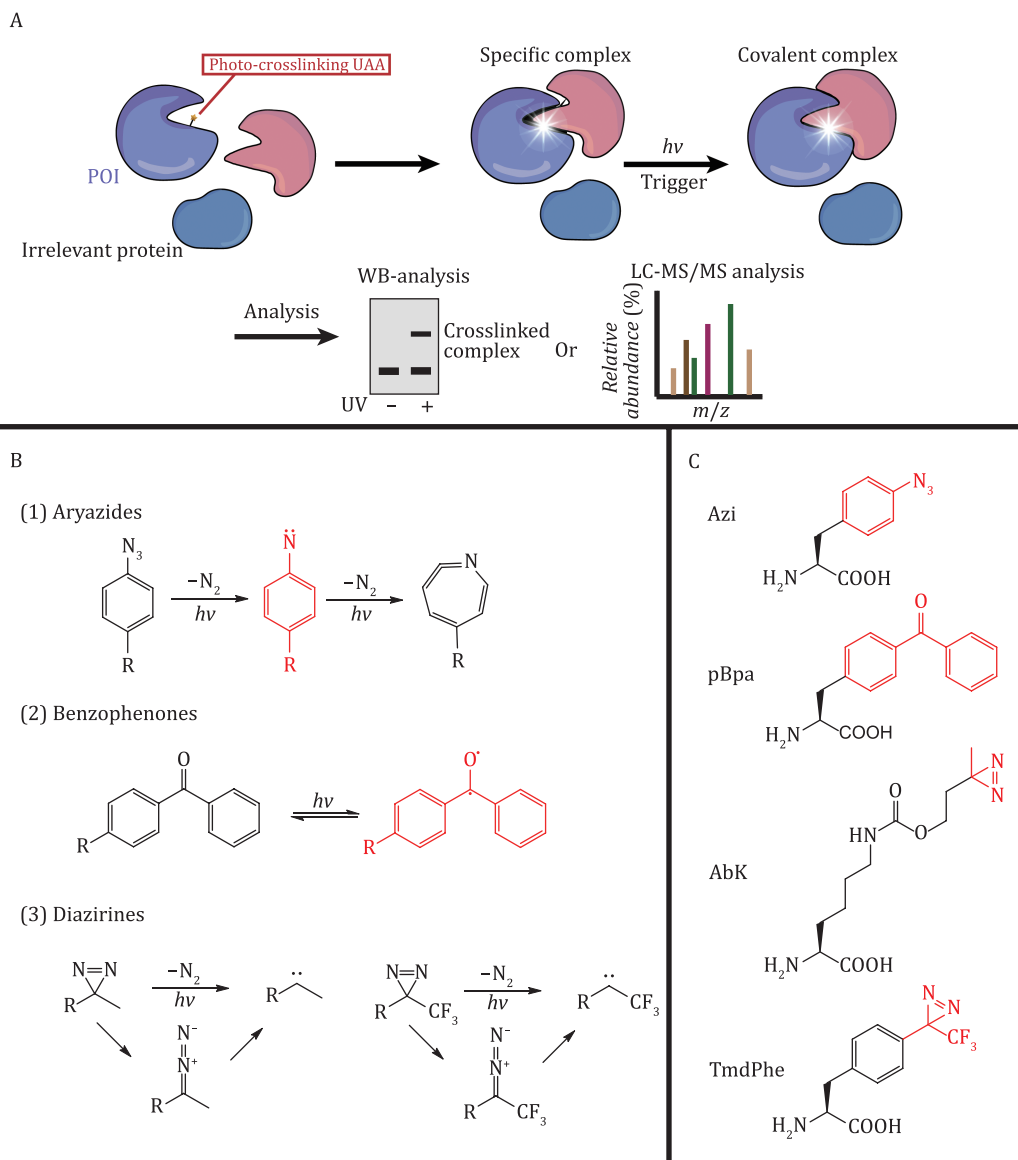


Figure 4. Schematic diagram of UAA photocrosslinking. (A) Mode of action for UAA photocrosslinking: upon incorporation of a UAA into the POI, the UAA can covalently bind to the target protein when exposed to specific wavelengths of irradiation. This method enables downstream analysis through techniques such as western blotting (WB) or liquid chromatography–mass spectrometry (LS-MS). (B) Activation mechanism of the three main photoreactive groups, including arylazides, benzophenone and diazirine. (C) Structures of four representative photocrosslinkable UAAs in eukaryotes.

receptor–ligand interactions in cell signaling pathways, protein interactions in the chromatin, and protein interactions in living cells under different physiological conditions (Nguyen et al., 2018). Furthermore, photocrosslinkable UAAs play a significant role in antibody selection. In the early stages of antibody selection, preserving a broader sequence diversity is crucial for the development of therapeutic antibodies. Traditional phage display strategies often impose limitations on the diversity of identified antibody sequences and binding epitopes (Ledsgaard et al., 2022). By incorporating photocrosslinkable UAAs at antigen-specific sites, antibodies from the library can be readily crosslinked with UAAs,

and subsequently the antibody binding to specific antigen epitopes can be screened. Currently, this technique was successfully implemented in *E. coli*, where Chen et al. (2020) incorporated pBpa into specific epitopes of IL-1 β and complement 5a to select epitope-directed antibodies from phage display libraries. While there have been no reports of incorporating UAA into antigens in eukaryotic systems, the challenge is meaningful. After all, antigens expressed in eukaryotic systems undergo PTM, making them more capable of replicating the true interaction between antigens and antibodies *in vivo*. Figure 4C depicts some representative applications along with the photocrosslinkable UAAs used.

Chemical crosslinking and photoactivatable chemical crosslinking

The reaction of photocrosslinkable UAAs is characterized by the nonspecific insertion of reactive groups into C–H or heteroatom–H bonds of neighboring molecules after photoactivation. However, this reaction mode has a low yield as intermediates generated by the crosslinker are easily quenched by nucleophiles or water in the reaction environment, which can seriously affect the enrichment of interacting partners in subsequent studies (Coin, 2018). Therefore, a chemical-crosslinking modality with a different concept was proposed. The strategy to achieve chemical-crosslinking is to install an electrophilic group on the UAA, which will react with the nucleophilic proximal amino acid residues, such as the thiol group of Cys, to generate covalent linkages (Nguyen et al., 2018) (Fig. 5A). This mode of crosslinking does not require external triggering and provides residue specificity. Moreover, this chemical-crosslinking mode provides a higher crosslinking yield than photocrosslinking.

The electrophilicity of the UAA bearing a chemical crosslinker needs to be carefully adjusted to ensure that the reactive moiety has sufficient reactivity, high stability in cells, and high selectivity for the target amino acid residues (Liu et al., 2023). Currently, a variety of chemical crosslinkers have been installed on UAAs to achieve proximity reactions and target different residues (Nguyen et al., 2018). These representative structures and reactions are summarized in Fig. 5B.

Under physiological conditions, UAAs bearing chemical crosslinkers are unable to react with free natural amino acids, but when they are incorporated into target proteins, their crosslinked groups can react with adjacent natural amino acid residues. The first crosslinking reaction attempted was the reaction of halides with the thiol group of Cys to form a covalent crosslink. The first UAA used for chemical protein crosslinking was *p*-2'-fluoroacetylphenylalanine (Ffact), which was designed in 2013 to introduce covalent bonds into proteins and was crosslinked with Cys, making an affibody irreversibly bind to its protein substrate, thereby achieving peptide–protein interaction capture in mammalian cells (Xiang et al., 2013). In addition to capturing and identifying protein–protein interactions, covalent crosslinking of UAAs helps stabilize low-affinity protein complexes on eukaryotic platforms (Cigler et al., 2017). Similar to the utilization of photocrosslinkable UAAs for antibody selection described earlier, chemically crosslinkable UAAs also have applications in antibody selection. Zhu et al. (2023) incorporated UAAs with crosslinking capabilities into antigen-specific epitopes to select antibodies produced by immunized mice that could form covalent crosslinks with the antigen. Nevertheless, the potential application of this technique in eukaryotic expression systems has yet to

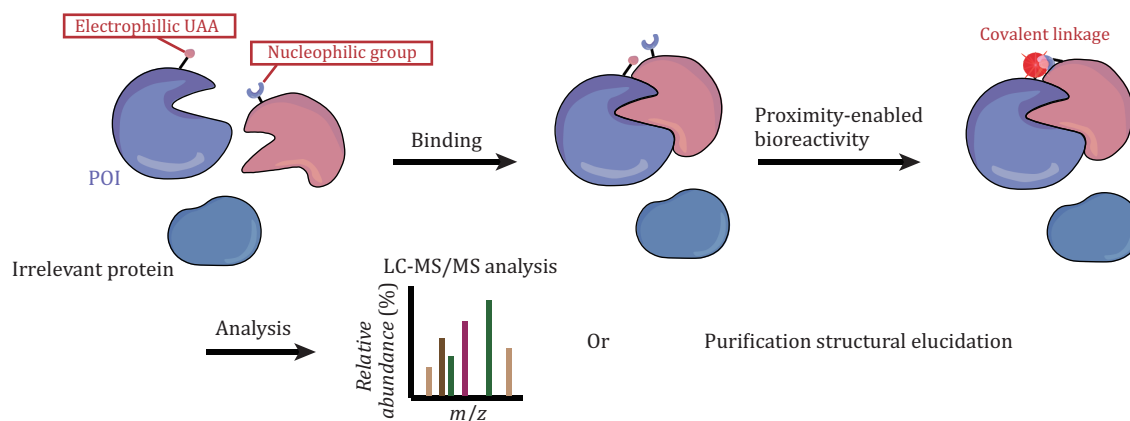
be investigated. Achieving covalent binding between proteins through chemical-crosslinking in eukaryotic systems is of great interest. Currently, this method has become an important strategy for the development of covalent protein drugs (Li et al., 2020; Wang and Wang, 2022; Yu et al., 2022). This topic will be covered in detail in Section “Covalent protein drugs”.

By combining photocrosslinking with chemical-crosslinking, a photoactivatable chemical crosslinker on a UAA has been developed. The photoactivated chemical group on the UAA releases a nonradical functional group under UV irradiation and selectively reacts with the targeted amino acid residue, thus improving the residue selectivity of photocrosslinking and the crosslinking efficiency, and avoiding potential harm from chemical-crosslinking agents in living systems. Tetrazole is a common photoreactive group, and after UAA photolysis, the nitrile imine is released from the tetrazole group and reacts with the nearby Glu residue to achieve protein crosslinking (Herner et al., 2016). For example, *N*-methylpyrroletetrazole lysine (mPyTK), a tetrazole-containing lysine derivative, was successfully incorporated into the growth factor receptor-bound protein 2 (Grb2) protein and crosslinked with epidermal growth factor receptor (EGFR) in mammalian cells (Tian et al., 2017) (Fig. 5C). To identify unknown protein interactions, crosslinkers that can target multiple amino acid residues have been developed. Quinone methide (QM), a photoactivated crosslinker that acts as a Michael acceptor for various nucleophiles (Toteva and Richard, 2011), can crosslink with 9 amino acid residues (Cys, His, Lys, Tyr, Trp, Asn, Gln, Asp, Glu, and Met). (2*R*)-2-Amino-3-fluoro-3-(4-((2-nitrobenzyl)oxy) phenyl) propanoic acid (FnBY), which installs QM, has been developed and successfully incorporated into mammalian cells (Liu et al., 2019) (Fig. 5C).

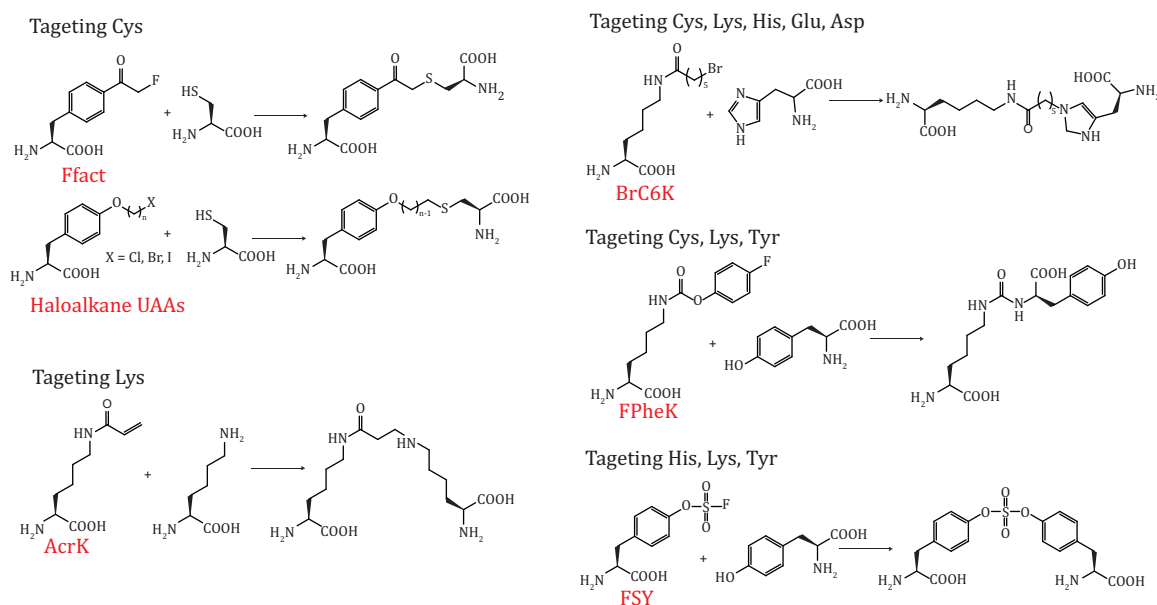
Fluorescent labeling of proteins

Fluorescent labeling of proteins enables the accurate and even real-time detection of the protein conformation, interactions, and dynamics. Conventional fluorescent labeling methods include the fusion of a fluorescent protein (such as GFP) with the target protein or the addition of a fluorescent dye to react with the corresponding group of the protein. However, these methods are often not site-specific, and their development is limited by the size and photophysical properties of fluorescent proteins. The large size of a fluorescent protein can easily perturb protein function. Early attempts were made to achieve fluorescent labeling of proteins by modifying natural amino acids, such as methionine or alanine analogs, to be recognized by endogenous aaRS/tRNA pairs and react with fluorophores (Beatty et al., 2006; Dieterich et al., 2010). The disadvantage of this approach is that no specific protein can be tagged, and any endogenous

A



B



C

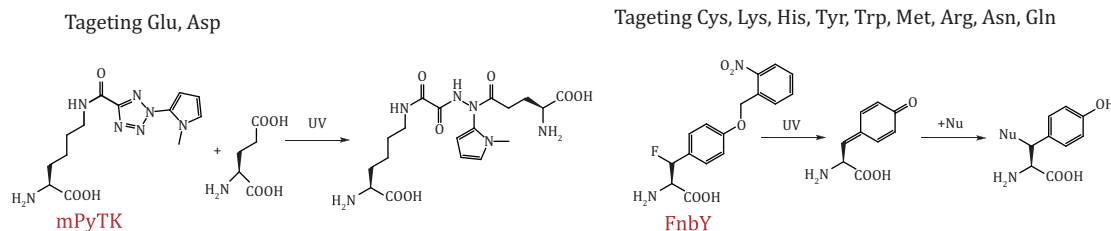


Figure 5. Schematic diagram of chemical-crosslinking of UAAs. (A) Mode of action for chemical-crosslinking of UAAs: UAAs containing electrophilic groups are incorporated into a POI and react with nearby amino acid residues containing nucleophilic groups, forming covalent bonds. (B) Reaction of chemical-crosslinking of UAAs with targeted AAs. (C) Reaction of photoactivatable chemical-crosslinking UAAs with targeted AAs.

methionine or alanine may be replaced by these modified amino acids.

The use of GCE facilitates the site-specific fluorescent labeling of a POI. There are two ways to fluorescently label proteins using GCE. The first method

involves directly incorporating a UAA that has a fluorophore into the POI (Fig. 6A). However, UAAs with fluorophores are often bulky and inefficient. The second and commonly used method is to incorporate a UAA with an orthogonal handle into the POI using

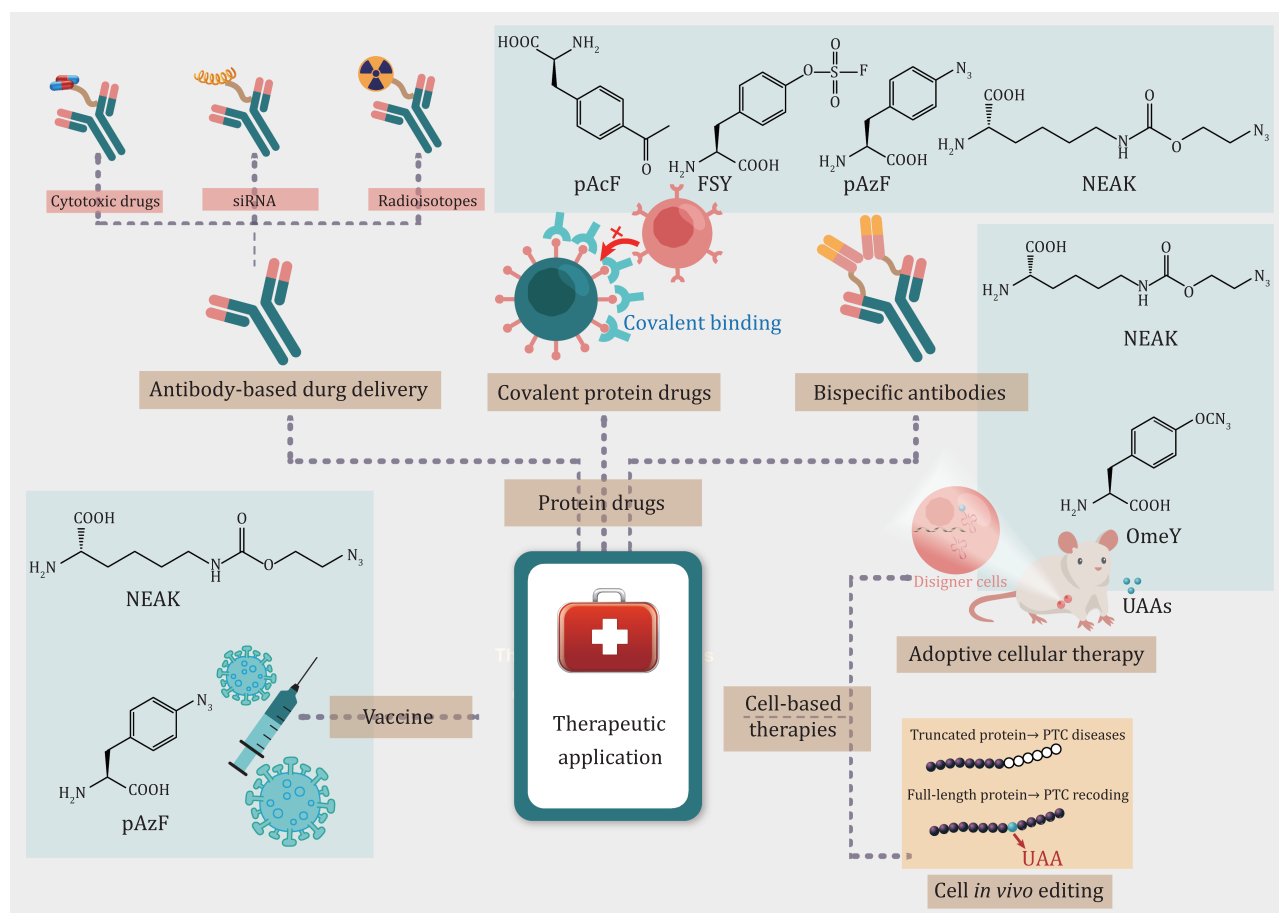


Figure 7. Application of genetic code expansion technology in eukaryotic expression systems for drug development and disease treatment, along with representative UAA structures.

group of a fluorescent dye through a click reaction, offering several advantages over fAAs. First, fluorescent dyes are smaller than fluorescent proteins, reducing the possibility of perturbing POI function and avoiding the restriction that fluorescent proteins can only attach to the N- or C-terminus of POI (Peng and Hang, 2016). Second, fluorescent dyes have excellent photophysical properties and are less phototoxic to cells than fluorescent proteins (Elia, 2021). Third, the click reaction is stoichiometric, with the fluorescent dye binding to a UAA in a 1:1 ratio (Serfling et al., 2019). This property facilitates the quantification of proteins by fluorescence intensity. Moreover, the slow incorporation reaction of a fAA limits its application, whereas click labeling between a UAA and fluorescent dye is rapid and facilitates quantitative site-specific protein labeling in a timely manner (Lang et al., 2012a).

In 2012, the Chin laboratory achieved rapid fluorescent labeling of a POI in mammalian cells by designing a lysine derivative loaded with norbornene through an inverse electron demand Diels–Alder cycloaddition reaction (IEDDA) with a tetrazine probe (Lang et al., 2012b). Further optimization was performed by designing a fast reaction of BCN- and *trans*-cyclooctene-containing

amino acids with a tetrazine probe (Lang et al., 2012a). This reaction has the advantage of producing a single product of defined stereochemistry compared to the previous structure. These works laid the foundation for a subsequent series of bioorthogonal chemical labeling with small fluorophores on proteins. However, bioorthogonal chemical fluorescence labeling using tetrazine-based cycloaddition reactions mostly uses trained hydrophobic alkenes or alkynes, which limits their broad application. The Xiao laboratory has further developed UAAs containing small and hydrophilic isocyanate or isonitrile groups for incorporation into mammalian cells. Through isonitrile–tetrazine coupling reactions, these methods provide good alternatives to hydrophobic alkenes or alkynethines (Chen et al., 2019) (Fig. 6D).

Based on the development and optimization of bioorthogonal chemical fluorescence labeling methods, GCE has found numerous applications for protein labeling in eukaryotic systems. This strategy has been successfully utilized for imaging various receptors and channels in living cells (Lang et al., 2012a; Nikić and Lemke, 2015; Gupta et al., 2019; Segal et al., 2020). Notably, this method has played a crucial role in super-resolution

microscopy (Nikić et al., 2016) and single molecule labeling (Segal et al., 2020) techniques.

Therapeutic applications

Protein drugs

Antibody-based drug delivery

The lack of specificity of chemotherapeutic drugs often leads to systemic side-effects. Antibody drug conjugates (ADCs) are a promising drug development strategy that has received considerable attention in recent years. A functional ADC consists of a highly selective humanized or human monoclonal antibody (mAb) that binds to a potent cytotoxic compound (payload) with a specific drug-antibody ratio (DAR) via a chemical linker (Leong et al., 2023). This approach combines the pharmacological potency of cytotoxic drugs with the specificity of therapeutic antibodies to develop biologic antibody therapies. ADCs have the advantage of delivering cytotoxic drugs to the tumor site with greater precision and efficiency, reducing the minimum effective dose and increasing the tolerated dose of the drug compared to conventional cytotoxic drugs (Schumacher et al., 2016).

Site-specific conjugation, as opposed to the conventional antibody conjugation method, avoids heterogeneous mixtures resulting from random reactions, which can significantly impact analytical characterization and manufacturing (Hao et al., 2021; Walsh et al., 2021). Additionally, site-specific conjugation greatly improves the homogeneity of the DAR, further reducing the ADC toxicity and enhancing its efficacy. GCE technology-produced ADCs are a representative example of third-generation ADCs for site-specific conjugation, which provides an efficient method for site-specific antibody modification, allowing the incorporation of UAAs into antibody sequences (Axup et al., 2012).

Recombinant antibodies for ADC preparation by GCE can be generated using either cell-based or cell-free systems (Walsh et al., 2021). While prokaryotic expression systems, such as *E. coli*, have high expression yields, they can only express simple proteins without glycosylation, which often fail to meet the requirements of antibody modification. Therefore, it is common to incorporate UAAs in mammalian cell platforms, particularly CHO cells, which ensures proper protein folding and PTM. It has been reported that the keto group of *para*-acetyl phenylalanine (pAcF) can be coupled with alkoxyamine-derived drugs to generate stable oxime bonds (Dirksen et al., 2006; Hutchins et al., 2011a, 2011b). Based on this principle, Axup et al. incorporated pAcF at specific positions in constant regions of trastuzumab (an anti-Her2 antibody) and coupled a C-terminal alkoxy-amine auristatin derivative with a modified trastuzumab mutant for the treatment of mammary fat pad tumors. These ADCs showed excellent antitumor efficiency in a xenograft

model (Axup et al., 2012). In another study, researchers developed a CHO cell-based EuCODE expression system for the stable production of antibodies, incorporating UAAs with titers over 1 g/L. This expression platform was used to incorporate pAcF into anti-5T4 and anti-Her2 antibodies for site-specific coupling with the tubulin inhibitor monomethyl auristatin D (MMAD). The ADC developed using this expression platform showed a great efficacy (Tian et al., 2014). Additionally, Roy et al. developed a CHO cell-based UAA incorporation platform for producing UAA-containing ADCs. The ability of this platform to express UAA-containing IgG at high titers was demonstrated using perfusion bioreactors that allowed continuous addition of fresh media and nutrients and removal of cellular waste to maintain optimal growth and expression conditions. Furthermore, the researchers developed a scale-up process compatible with established upstream platform perfusion processes for future commercialization (Roy et al., 2020).

It is a common approach to construct ADCs by conjugating cytotoxic agents to antibodies that target tumor-associated antigens (TAAs). In addition, conjugating noncytotoxic agents such as kinase inhibitors or agonists to antibodies that target immune cells can help reduce off-target toxicity (Wang et al., 2015). The development of immunosuppressive ADCs has expanded the application of ADCs. Site-specific conjugation of drugs to antibodies helps improve the stability and pharmacokinetic properties of ADCs. Lim et al. incorporated pAcF into the heavy-chain constant region of anti-CD11a IgG in CHO cells. The pAcF on anti-CD11a IgG reacted with terminal aminoxy groups of liver X receptors (LXRs) agonists, resulting in LXR agonists targeting THP-1 monocytes/macrophages for treating atherosclerosis (Lim et al., 2015). Such agonists better target LXR on macrophages than traditional LXR agonists, contributing to the suppression of an array of inflammatory genes and inhibition of macrophage proliferation without activating LXR on hepatocytes, which has been reported to increase hepatic triglyceride levels (Lim et al., 2015).

In addition to delivering cytotoxic compounds, UAAs in antibodies can be utilized to deliver siRNAs or radioisotopes. Traditional tissue-specific siRNA delivery systems have drawbacks, including selective binding to tissue-associated antigens, which can result in siRNA degradation or off-target gene silencing. For example, antibody-protamine fusion proteins and scFv-9R conjugates have been used for selective tissue-specific siRNA delivery (Song et al., 2005; Kumar et al., 2008; Yao et al., 2012), but their effectiveness is limited. Similarly, nanostructures coated with targeted ligands also have drawbacks such as a poor binding affinity and low ligand density (Kim et al., 2008; Lee et al., 2012). To overcome these limitations, pAcF has been incorporated into an anti-Her2 Fab or the full-length IgG using *E. coli* or CHO

expression systems, respectively, (Lu et al., 2013). The resulting antibodies can be conjugated to amino-oxo-derivatized cationic block copolymers to generate antibody-polymer conjugates (APCs) with a more homogeneous structure and better affinity, resulting in low nonspecific cytotoxicity (Lu et al., 2013). These APCs were able to effectively target siRNA to Her2+ cells and silence genes at both the mRNA and protein levels.

Radioimmunotherapy is a therapeutic approach that involves labeling antibodies with radioisotopes using a bifunctional linker to specifically deliver the conjugate to cells expressing a specific antigen for imaging or treatment of lesions (Wu, 2009). Previously, isotopically labeled antibodies were generated by coupling bifunctional linkers with the thiol group of cysteine or the ϵ -amino group of lysine on mAbs. However, this approach resulted in heterogeneous products, which affected the activity of the conjugate (Adumeau et al., 2016). Wu et al. used a site-specific approach to introduce the UAA N ϵ -2-azidoethylloxycarbonyl-L-lysine (NEAK) into the anti-CD20 antibody rituximab using a 293F expression system. This modification allowed specific binding of the bifunctional linker 1,4,7,10-tetraazacyclotetradecane-1,4,7,10-tetraacetic acid (DOTA) to the antibody, and the rituximab conjugates were labeled with the radioisotope ^{64}Cu or ^{177}Lu (Wu et al., 2016). The conjugates produced in this way were highly homogeneous, and radiolabeling did not alter the binding affinity of rituximab to the antigens. This study provides a great idea for the development of radioimmunotherapy drugs.

Covalent protein drugs

Conventional targeted covalent inhibitors (TCIs) are small-molecule compounds that bind covalently to specific proteins, thereby inhibiting their biological functions. Covalent small molecule drugs typically contain functional groups, such as acrylamide, that can form covalent bonds with specific amino acid residues in targeted proteins (Baillie, 2016; Gehringer and Laufer, 2019). TCIs offer several advantages, including a prolonged drug efficacy and an increased therapeutic index. However, the off-target effects of covalent small molecules remain a major hurdle for their development (Singh et al., 2011; Sibaud et al., 2020).

In recent years, covalent protein drugs have garnered increasing attention owing to their higher specificity and greater potential for modification (Wang and Wang, 2022). However, achieving covalent binding between proteins has proven to be a challenge, primarily because proteins mainly interact through noncovalent bonds, and there are no naturally occurring amino acids that can form covalent bonds with one another (Darby and Creighton, 1995). Furthermore, amino acids are widely distributed in cells and cannot be selectively bound to target proteins (Wang and Wang, 2022). To address this problem, Xiang

et al. proposed the concept of proximity-enabled bioactivity (PERx) in 2013. This approach involves introducing potentially bioactive UAAs into proteins. UAAs can then react with proximal cysteine or lysine residues of targeted proteins, resulting in the formation of a new covalent bond that allows for irreversible binding of the utility-to-target proteins (Xiang et al., 2013). Initially, cysteine or lysine residues were used as suitable targets for covalent crosslinking, as the sulfhydryl group of cysteine is highly nucleophilic (Xiang et al., 2013), and lysine residues are relatively abundant on the protein surface and at interaction interfaces (Hacker et al., 2017). Based on these two targets of covalent crosslinking, a variety of UAAs have been developed and introduced into proteins for various applications, including increasing the thermostability of protein junctions (Xiang et al., 2014), performing specific optical modulation of protein structure and function (Hoppmann et al., 2014, 2015), and modifying antibodies to increase their stable binding to specific antigens on cancer cells (Furman et al., 2014).

The development of covalent protein drugs using PERx requires careful selection of the UAA, which must be chemically inert *in vivo* and able to efficiently react with neighboring target residues prior to drug-target dissociation. Additionally, it is essential to ensure that the UAA itself and its metabolic byproducts are nontoxic. Fluorosulfate-L-tyrosine (FSY) has been identified as a suitable UAA for the development of covalent protein drugs targeting programmed death-1 (PD-1) (Li et al., 2020). In this study, FSY was incorporated into PD-1 through PERx. When PD-1 interacted with PD-L1, FSY selectively reacted with a histidine residue near the contact site of programmed death ligand-1 (PD-L1). The resulting UAA-modified PD-1 was irreversibly bound to PD-L1, exhibiting potent antitumor activity (Li et al., 2020). Although FSY was integrated into PD-1 in *E. coli* in this study, FSY has the potential to be incorporated into proteins in eukaryotic systems. Wang et al. reported that FSY could also participate in proximity-enabled sulfur-fluoride exchange (SuFEx) reactions in mammalian cells, efficiently reacting with proximal lysine, histidine and tyrosine residues of target proteins. Therefore, FSY also holds the potential for incorporation into therapeutic proteins in eukaryotic platforms (Wang et al., 2018).

Previous studies have shown that FSY has great potential as an aryl fluorosulfate-containing UAA for the development of covalent protein drugs. However, because of its relatively rigid and short side chain, FSY has a relatively short reaction radius, which limits the diversity of proteins that can be targeted (Liu et al., 2021). To address this issue, Liu et al. developed a novel UAA, fluorosulfonyloxybenzoyl-L-lysine (FSK), which also contains an aryl fluorosulfate group (Liu et al., 2021). Studies have demonstrated that covalent nanobodies with incorporated FSY irreversibly bind

to EGFR on cells. FSK and FSY target different sites on EGFR, counteracting the potential mutational resistance of EGFR. Researchers have also confirmed the protein-crosslinking ability of FSK in mammalian cells and demonstrated that FSK could be successfully incorporated into EGFP in HeLa cells and into proteins in 293T cells (Liu et al., 2021).

Bispecific antibodies

GCE has also been widely employed in the development of antibody drugs. This method allows the site-specific incorporation of UAAs with reactive groups, facilitating site-specific modification of antibody drugs and extending their application in immunotherapy. For reasons of speed and cost-effectiveness, the incorporation of UAAs into antibody fragments is typically achieved through bacterial expression systems, whereas incorporation into a full-length IgG is more dependent on eukaryotic hosts because of the requirement of PTM, such as glycosylation.

The development of bispecific antibodies (bsAbs) has addressed the issue of inefficacy in patients with low expression of target antigens (Lewis Phillips et al., 2008). As they bind not only to TAAs but also to an invariant component of T-cell receptors (TCRs) such as CD3 epsilon, bsAbs have been developed to enhance the efficacy of antibody drugs against tumors with low expression of Her2 (Rader, 2020).

Researchers have demonstrated that incorporating UAAs into specific sites of the Fab fragment of bsAbs can be beneficial for rational control of the binding site and the linker length between antibodies (Kim et al., 2012). To investigate the effects of different forms of bsAbs on cancer cells expressing different levels of Her2, Cao et al. designed four different structures of bsAbs. Using a CHO cell expression system, *para*-azidophenylalanine (*pAzF*) was incorporated into a specific site of the anti-Her2 IgG of trastuzumab and also into anti-CD3 Fab fragment using an *E. coli* expression system (Cao et al., 2015). BsAbs were developed based on IgG or Fab to bind either monovalent or bivalent CD3 Fab. In another study, the same approach was used to incorporate UAAs into specific sites of trastuzumab IgG in a CHO cell expression system for the production of bsAbs (Cao et al., 2021).

In addition to the CHO cell expression system, yeasts are one of the simplest and most cost-effective eukaryotic expression systems. *Pichia pastoris* is a promising host for recombinant protein production owing to its high secretion capacity and low levels of endogenous protein secretion (Choi et al., 2003). Additionally, *P. pastoris* is more amenable to high-cell-density cultivation than *S. cerevisiae*, leading to higher product titers (Young et al., 2009). Tir et al. employed *P. pastoris* to optimize the conditions for site-specific incorporation of the *pAzF* UAA into the trastuzumab Fab and full-length IgG.

Through selection of appropriate fusion secretion signals and modulation of genes involved in protein folding and oxidative stress regulation, the authors were able to generate a full-length trastuzumab IgG with the *pAzF* modification in the light chain, with a titer of up to 15 mg/L. Furthermore, by knocking out YPT7 to prepare strains with disrupted vacuole sorting, they were able to improve the yield of *pAzF* incorporation into the trastuzumab Fab (Tir et al., 2022).

Cell-based therapies

Adoptive cellular therapy

With the advancement of designer cell-based therapies, researchers are exploring ways to treat diabetes by modifying β cells (Kitada et al., 2018) because the current treatment for patients with type 1 diabetes is still limited to blood glucose monitoring throughout the day and insulin injections. Small molecule and light-inducible therapeutic systems cannot accurately control blood glucose and achieve rapid insulin expression in diabetes treatment (Zaykov et al., 2016; Teixeira and Fussenegger, 2019). Researchers have developed engineered electrosensitive β -cells for the rapid release of insulin by implanting a specific bioelectronic device (Krawczyk et al., 2020). However, the regular maintenance of the bioelectronic device limits the wide application of this therapy. To achieve precise and rapid control of insulin expression, Chen et al. (2022b) developed a UAA-based therapeutic switch (NATS) system based on GCE. The NATS system is a modified mammalian cell that contains a bacterial aARS/tRNA pair and an insulin gene carrying an ectopic amber codon. Microencapsulated cells can be implanted into mice, followed by oral intake of a UAA-containing inducer to achieve precise control of insulin expression. Diabetic model mice implanted with microencapsulated cells harboring the NATS system were able to achieve the relief of hyperglycemia within 90 minutes after the intake of the UAA inducer (Chen et al., 2022b). This method provides a new approach to diabetes treatment.

As a type of a living drug for cancer treatment, chimeric antigen receptor (CAR)-engineered T (CAR-T) cells have demonstrated a significant efficacy in the treatment of a variety of cancers. CAR-T-cell therapy enables T cells to recognize specific antigens on the surface of cancer cells by artificially modifying the T-cell receptor. CARs are hybrid receptors composed of an extracellular antigen-binding domain and an intracellular signaling domain (Rong et al., 2022). This structure helps T cells selectively target and eliminate cancer cells expressing the antigen of interest.

The activation of CAR-T cells needs to be precisely controlled, and the development of switchable CAR-T (sCAR-T) cells allows regulation of CAR-T-cell activity, to

a certain extent. The use of GCE has been described in the development of sCAR-T cells (Ma et al., 2016). However, the efficacy of CAR-T-cell therapy against solid tumors is still insufficient (Maalej et al., 2023). In subsequent studies, researchers applied this method to develop sCAR-T cells targeting breast cancer cells with high Her2 expression; the developed sCAR-T cells showed a potent efficacy and safety, further expanding the application of CAR-T cells in the treatment of solid tumors (Cao et al., 2016). Rong et al. (2022) also reported the use of single-chain variable fragment (scFv) antibodies that incorporated the UAA pAcF to redirect dinitrobenzyl (DNP)-mediated universal CAR-T cells, which can improve the safety of CAR-T-cell therapy. However, because of technical bottlenecks, UAA incorporation in eukaryotic expression systems for sCAR-T-cell research still needs to be explored.

In vivo cell editing

Duchenne muscular dystrophy (DMD) is a disease caused by nonsense mutations resulting from deletions, duplications or point mutations in the DMD gene (Karijolic and Yu, 2014; Politano, 2021). The DMD gene encodes dystrophin, and the introduction of premature termination codons (PTCs) causes progressive muscle weakness and a shortened lifespan (Foster et al., 2012). Current treatments for nonsense mutation diseases include the use of aminoglycoside antibiotics or non-aminoglycoside drugs, which induce ribosomes to read through PTCs by randomly binding amino acids and produce full-length functional proteins (Asiful Islam et al., 2017). In addition to using PTC readthrough drugs, researchers have employed CRISPR-Cas9 genome editing to restore dystrophin (Long et al., 2014). Furthermore, Shi et al. (2022) applied the PylRS-tRNA^{Pyl} system to treat dystrophin expression disorders caused by point mutations in the DMD gene. In this study, the feasibility of using this UAA incorporation system to read through the endogenous PTC was verified in differentiated myoblasts derived from mdx mice with DMD point mutations and from DMD patients (Shi et al., 2022). Additionally, to deliver this UAA incorporation system *in vivo*, researchers constructed an AAV carrying this UAA incorporation system to provide a research basis for the clinical application of this method.

Vaccines

Protein-based vaccines

Vaccines developed for different types of diseases can be divided into two categories: protein-based vaccines and virus vaccines. The principle of protein-based vaccines is to elicit a strong immune response against specific epitopes or self-proteins that are weakly immunogenic or induce immunotolerance in patients (Fok and Mayer, 2020). Common strategies include nonselective chemical modification of endogenous proteins or introduction of chimeric antigens with foreign T-helper cell epitopes

to elicit antibodies that are selective for both self-proteins and modified proteins and break immunotolerance (Weigle, 1965; Zuany-Amorim et al., 2004). However, these methods cannot precisely modify the target proteins, whereas GCE can achieve controllable and precise protein modification (Grünwald et al., 2009). In recent years, the GCE strategy has been applied to the development of protein-based vaccines for the prevention of autoimmune disorders or cancer.

For example, nitroaryl groups have been demonstrated to be highly immunogenic haptens, as they can attach to proteins and induce a strong immune response, which can help overcome autologous immune tolerance (Jones, 2015; Fok and Mayer, 2020). Therefore, proteins that are modified with immunogenic UAAs, such as *p*-sulfo-L-phenylalanine (SO₃Tyr), 3-nitro-L-tyrosine(3NO₂Tyr) or *p*-nitro-L-phenylalanine (pNO₂Phe), when injected into the body, can produce high titer antibodies that are cross-reactive to the wild-type proteins (Grünwald et al., 2008). Vaccines, such as a recombinant RANKL vaccine, which utilizes site-specific incorporation of pNO₂Phe into murine tumor necrosis factor- α (mTNF- α), have been developed for the prevention of collagen-induced arthritis and ovariectomy (OVX)-induced bone loss (Li et al., 2018; Wu et al., 2018b). However, the vast majority of these UAA incorporations were performed in prokaryotes. For the eukaryotic platform, DNP use has been reported for the incorporation of UAAs, which can be applied in HEK293T cells. Ren et al. (2015b) demonstrated that the DNP-containing UAA N(6)-(2-(2,4-dinitrophenyl)acetyl)lysine (DnpK) was unstable in *E. coli* but more stable in HEK293T cells and induced selective interaction with anti-DNP antibodies. This study suggests the possibility of future incorporation of DNP-containing UAAs in eukaryotic platforms for the development of protein-based vaccines.

Virus vaccines

Virus vaccines are generally classified into two types: live-attenuated virus (LAV) vaccines and inactivated whole virus particle vaccines. The latter are safer than the former, but require multiple booster shots to sustain the protection. LAVs typically elicit strong immune responses, but carry a risk of causing disease in immunocompromised individuals. The conventional approach to generating LAVs involves passaging the virus through an abnormal host or under non-physiological conditions to reduce its virulence and infectivity. However, both types of virus vaccines have limitations that hinder their wide application.

To address safety concerns associated with live-attenuated viruses, researchers have begun using molecular biology techniques to attenuate viruses in a more controlled manner. One such technique involves introducing PTCs into the viral genome, which results in incomplete translation of

viral proteins and weakens the virus replication ability and virulence (Mayer, 2019; Wang et al., 2022). This approach was first applied to the HEK293T cell line in 2016 using GCE technology (Si et al., 2016). Researchers transferred an orthogonal translation system (*MbPylRS/tRNA_{CUA}*) into transgenic cells, inserted an amber stop codon into the genome of the influenza A virus, and analyzed mutations in multiple genes to produce PTC-harboring viruses. They observed the infectivity and stability of the progeny virus and found that PTC-harboring vaccines produced by this strategy exhibited strong immune effects on multiple organisms, providing a novel and versatile means for generating live virus vaccines (Si et al., 2016). Moreover, Yuan et al. (2017) designed an unnatural genetic switch using GCE technology in HEK293T cells to control the replication of live-attenuated human immunodeficiency virus type 1 (HIV-1) viruses. Based on the strategy of live virus vaccines using GCE technology, researchers have subsequently developed PTC-harboring vaccines for foot-and-mouth disease virus and pseudorabies virus (Hao et al., 2021; Wang et al., 2022).

Prospectives

GCE has become an attractive tool for modifying protein properties. Incorporating UAAs in eukaryotic platforms has significantly expanded the applications of GEC, allowing site-specific attachment of new chemical properties to target proteins, such as fluorescence, crosslinking or photocaging, to serve various research purposes. Currently, GCE is not limited to basic research and is gradually employed in the development of new disease treatment strategies (Fig. 7).

Although many bottlenecks have been addressed, GCE in eukaryotes still presents challenges. For instance, while efficient incorporation of UAAs has been achieved in some readily transfected cell lines, it remains challenging in other cell lines, with a low transfection efficiency of aaRS/tRNA pairs. This issue may necessitate further optimization of viral vectors and the suppression system. Moreover, researchers have focused on optimizing the structure and enhancing the incorporation efficiency of UAAs. Notably, release factor engineering has been reported to improve the UAA incorporation efficiency in mammalian cells. Additionally, the poor efficiency of heterologous tRNAs can be overcome through directed evolution of tRNAs in foreign host cells. However, this strategy has thus far only been implemented in bacterial expression systems and not in eukaryotic systems. Directed evolution of translation components (such as aaRS, tRNA, or EF-Tu) is typically a time-consuming task, and it allows the evolution of individual components in isolation rather than multiple components simultaneously. This limitation poses

one of the most pressing challenges that needs to be addressed. Overall, with advancements in genetic manipulation, artificial intelligence, structural biology and interdisciplinary collaboration, the limitations of GCE technology encountered in eukaryotes will gradually be overcome, leading to further advancements in biotherapeutic applications.

Acknowledgements

This work was supported by the National Key R&D Program of China (Nos. 2019YFA0904200 and 2019YFA0906100), the National Natural Science Foundation of China (No. 32171464) and Shenzhen Science and Technology Innovation Program (JCYJ20180504165501371).

Conflict of interest

All the authors of this paper declare that they have no conflicts of interest.

Author contributions

Q.G. and Y.J.C. conceived this review and discussed the structure of the manuscript. Q.G. wrote the manuscript and Y.J.C. edited the manuscript. All authors approved the final version of the manuscript.

References

- Adams MD, Celniker SE, Holt RA et al. The genome sequence of *Drosophila melanogaster*. *Science (New York, N.Y.)* 2000;**287**:2185–95.
- Adumeau P, Sharma SK, Brent C et al. Site-specifically labeled immunoconjugates for molecular imaging—part 1: cysteine residues and glycans. *Mol Imaging Biol* 2016;**18**:1–17.
- Aemissegger A, Hilvert D. Synthesis and application of an azobenzene amino acid as a light-switchable turn element in polypeptides. *Nat Protocols* 2007;**2**:161–7.
- Agard NJ, Baskin JM, Prescher JA et al. A comparative study of bioorthogonal reactions with azides. *ACS Chem Biol* 2006;**1**:644–8.
- Agostini F, Völler J-S, Kokschi B et al. Biocatalysis with unnatural amino acids: enzymology meets xenobiology. *Angew Chem Int Ed Engl* 2017;**56**:9680–703.
- Alexpandi R, Prasanth MI, Ravi AV et al. Protective effect of neglected plant *Diplocyclos palmatus* on quorum sensing mediated infection of *Serratia marcescens* and UV-A induced photoaging in model *Caenorhabditis elegans*. *J Photochem Photobiol B* 2019;**201**:111637.
- Alouane A, Labruère R, Le Saux T et al. Self-immolative spacers: kinetic aspects, structure–property relationships, and applications. *Angew Chem Int Ed Engl* 2015;**54**:7492–509.

- Arbely E, Torres-Kolbus J, Deiters A *et al.* Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J Am Chem Soc* 2012;**134**:11912–5.
- Asiful Islam M, Alam F, Kamal MA *et al.* Therapeutic suppression of nonsense mutation: an emerging target in multiple diseases and thrombotic disorders. *Curr Pharm Des* 2017;**23**:1598–609.
- Axup JY, Bajjuri KM, Ritland M *et al.* Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc Natl Acad Sci USA* 2012;**109**:16101–6.
- Baillie TA. Targeted covalent inhibitors for drug design. *Angew Chem Int Ed Engl* 2016;**55**:13408–21.
- Barber KW, Rinehart J. The ABCs of PTMs. *Nat Chem Biol* 2018;**14**:188–92.
- Beatty KE, Liu JC, Xie F *et al.* Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew Chem Int Ed Engl* 2006;**45**:7364–7.
- Beharry AA, Woolley GA. Azobenzene photoswitches for biomolecules. *Chem Soc Rev* 2011;**40**:4422–37.
- Beránek V, Reinkemeier CD, Zhang MS *et al.* Genetically encoded protein phosphorylation in mammalian cells. *Cell Chem Biol* 2018;**25**:1067–1074.e5.
- Bianco A, Townsley FM, Greiss S *et al.* Expanding the genetic code of *Drosophila melanogaster*. *Nat Chem Biol* 2012;**8**:748–50.
- Blaskovich MAT. Unusual amino acids in medicinal chemistry. *J Med Chem* 2016;**59**:10807–36.
- Bose M, Groff D, Xie J *et al.* The incorporation of a photoisomerizable amino acid into proteins in *E. coli*. *J Am Chem Soc* 2006;**128**:388–9.
- Brown W, Deiters A. Light-activation of Cre recombinase in zebrafish embryos through genetic code expansion. *Methods Enzymol* 2019;**624**:265–81.
- Brown W, Galpin JD, Rosenblum C *et al.* Chemically Acylated tRNAs are functional in zebrafish embryos. *J Am Chem Soc* 2023a;**145**:2414–20.
- Brown W, Wesalo J, Tsang M *et al.* Engineering small molecule switches of protein function in zebrafish embryos. *J Am Chem Soc* 2023b;**145**:2395–403.
- Burnett K, Edsinger E, Albrecht DR. Rapid and gentle hydrogel encapsulation of living organisms enables long-term microscopy over multiple hours. *Commun Biol* 2018;**1**:73.
- Cao Y, Axup JY, Ma JSY *et al.* Multiformat T-cell-engaging bispecific antibodies targeting human breast cancers. *Angew Chem Int Ed Engl* 2015;**54**:7022–7.
- Cao Y, Rodgers DT, Du J *et al.* Design of switchable chimeric antigen receptor T cells targeting breast cancer. *Angew Chem Int Ed Engl* 2016;**55**:7520–4.
- Cao YJ, Wang X, Wang Z *et al.* Switchable CAR-T cells outperformed traditional antibody-redirected therapeutics targeting breast cancers. *ACS Synth Biol* 2021;**10**:1176–83.
- Chang H, Han M, Huang W *et al.* Light-induced protein translocation by genetically encoded unnatural amino acid in *Caenorhabditis elegans*. *Protein Cell* 2013;**4**:883–6.
- Chang S-W, Liu W-C, Liao K-Y *et al.* Phosphorylation of HPV-16 E2 at serine 243 enables binding to Brd4 and mitotic chromosomes. *PLoS One* 2014;**9**:e110882.
- Chatterjee A, Guo J, Lee HS *et al.* A genetically encoded fluorescent probe in mammalian cells. *J Am Chem Soc* 2013a;**135**:12540–3.
- Chatterjee A, Xiao H, Bollong M *et al.* Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc Natl Acad Sci USA* 2013b;**110**:11803–8.
- Chemla Y, Ozer E, Algov I *et al.* Context effects of genetic code expansion by stop codon suppression. *Curr Opin Chem Biol* 2018;**46**:146–55.
- Chen J, Tsai Y-H. Applications of genetic code expansion in studying protein post-translational modification. *J Mol Biol* 2022;**434**:167424.
- Chen S, Schultz PG, Brock A. An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. *J Mol Biol* 2007;**371**:112–22.
- Chen Y, Ma J, Lu W *et al.* Heritable expansion of the genetic code in mouse and zebrafish. *Cell Res* 2017;**27**:294–7.
- Chen Y, Loreda A, Gordon A *et al.* A noncanonical amino acid-based relay system for site-specific protein labeling. *Chem Commun (Cambridge, England)* 2018;**54**:7187–90.
- Chen Y, Wu K-L, Tang J *et al.* Addition of isocyanide-containing amino acids to the genetic code for protein labeling and activation. *ACS Chem Biol* 2019;**14**:2793–9.
- Chen L, Zhu C, Guo H *et al.* Epitope-directed antibody selection by site-specific photocrosslinking. *Sci Adv* 2020;**6**:eaaz7825.
- Chen Y, Jin S, Zhang M *et al.* Unleashing the potential of noncanonical amino acid biosynthesis to create cells with precision tyrosine sulfation. *Nat Commun* 2022a;**13**:5434.
- Chen C, Yu G, Huang Y *et al.* Genetic-code-expanded cell-based therapy for treating diabetes in mice. *Nat Chem Biol* 2022b;**18**:47–55.
- Chen Y, Liu W-Q, Zheng X *et al.* Cell-free biosynthesis of lysine-derived unnatural amino acids with chloro, alkene, and alkyne groups. *ACS Synth Biol* 2023;**12**:1349–57.
- Chin JW. Modular approaches to expanding the functions of living matter. *Nat Chem Biol* 2006;**2**:304–11.
- Chin JW, Martin AB, King DS *et al.* Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci USA* 2002;**99**:11020–4.
- Chin JW, Cropp TA, Anderson JC *et al.* An expanded eukaryotic genetic code. *Science (New York, N.Y.)* 2003;**301**:964–7.
- Choi B-K, Bobrowicz P, Davidson RC *et al.* Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. *Proc Natl Acad Sci USA* 2003;**100**:5022–7.
- Chou C, Deiters A. Light-activated gene editing with a photocaged zinc-finger nuclease. *Angew Chem Int Ed Engl* 2011;**50**:6839–42.
- Chou C, Young DD, Deiters A. Photocaged t7 RNA polymerase for the light activation of transcription and gene function in pro- and eukaryotic cells. *Chembiochem* 2010;**11**:972–7.
- Cigler M, Müller TG, Horn-Ghetko D *et al.* Proximity-triggered covalent stabilization of low-affinity protein

- complexes *in vitro* and *in vivo*. *Angew Chem Int Ed Engl* 2017;**56**:15737–41.
- Coin I. Application of non-canonical crosslinking amino acids to study protein-protein interactions in live cells. *Curr Opin Chem Biol* 2018;**46**:156–63.
- Coin I, Perrin MH, Vale WW *et al*. Photo-cross-linkers incorporated into G-protein-coupled receptors in mammalian cells: a ligand comparison. *Angew Chem Int Ed Engl* 2011;**50**:8077–81.
- Courtney T, Deiters A. Recent advances in the optical control of protein function through genetic code expansion. *Curr Opin Chem Biol* 2018;**46**:99–107.
- Courtney TM, Deiters A. Optical control of protein phosphatase function. *Nat Commun* 2019;**10**:4384.
- Crnković A, Vargas-Rodriguez O, Söll D. Plasticity and constraints of tRNA aminoacylation define directed evolution of aminoacyl-tRNA synthetases. *Int J Mol Sci* 2019;**20**:2294.
- Darby N, Creighton TE. Disulfide bonds in protein folding and stability. *Methods Mol Biol (Clifton, N.J.)* 1995;**40**:219–52.
- Davis L, Radman I, Goutou A *et al*. Precise optical control of gene expression in *C. elegans* using improved genetic code expansion and Cre recombinase. *ELife* 2021;**10**:e67075.
- Dieterich DC, Hodas JLL, Gouzer G *et al*. In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat Neurosci* 2010;**13**:897–905.
- Dirksen A, Hackeng TM, Dawson PE. Nucleophilic catalysis of oxime ligation. *Angew Chem Int Ed Engl* 2006;**45**:7581–4.
- Edwards WF, Young DD, Deiters A. Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS Chem Biol* 2009;**4**:441–5.
- Elia N. Using unnatural amino acids to selectively label proteins for cellular imaging: a cell biologist viewpoint. *FEBS J* 2021;**288**:1107–17.
- Elliott TS, Townsley FM, Bianco A *et al*. Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat Biotechnol* 2014;**32**:465–72.
- Elsässer SJ, Ernst RJ, Walker OS *et al*. Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nat Methods* 2016;**13**:158–64.
- Engelke H, Chou C, Uprety R *et al*. Control of protein function through optochemical translocation. *ACS Synth Biol* 2014;**3**:731–6.
- Ernst RJ, Krogager TP, Maywood ES *et al*. Genetic code expansion in the mouse brain. *Nat Chem Biol* 2016;**12**:776–8.
- Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. *Nature* 2011;**472**:499–503.
- Fang Y, Judkins JC, Boyd SJ *et al*. Studies on the stability and stabilization of *trans*-Cyclooctenes through radical inhibition and silver (I) metal complexation. *Tetrahedron* 2019;**75**:4307–17.
- Fickel TE, Gilvarg C. Transport of impermeant substances in *E. coli* by way of oligopeptide permease. *Nat New Biol* 1973;**241**:161–3.
- Fok JA, Mayer C. Genetic-code-expansion strategies for vaccine development. *Chembiochem* 2020;**21**:3291–300.
- Foster H, Popplewell L, Dickson G. Genetic therapeutic approaches for Duchenne muscular dystrophy. *Hum Gene Ther* 2012;**23**:676–87.
- Fottner M, Brunner A-D, Bittl V *et al*. Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat Chem Biol* 2019;**15**:276–84.
- Friedrich M, Aigner A. Therapeutic siRNA: state-of-the-art and future perspectives. *BioDrugs* 2022;**36**:549–71.
- Frøkjær-Jensen C, Davis MW, Hopkins CE *et al*. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet* 2008;**40**:1375–83.
- Furman JL, Kang M, Choi S *et al*. A genetically encoded azamichael acceptor for covalent cross-linking of protein-receptor complexes. *J Am Chem Soc* 2014;**136**:8411–7.
- Garza D, Medhora MM, Hartl DL. *Drosophila* nonsense suppressors: functional analysis in *Saccharomyces cerevisiae*, *Drosophila* tissue culture cells and *Drosophila melanogaster*. *Genetics* 1990;**126**:625–37.
- Gautier A, Nguyen DP, Lusich H *et al*. Genetically encoded photocontrol of protein localization in mammalian cells. *J Am Chem Soc* 2010;**132**:4086–8.
- Gautier A, Deiters A, Chin JW. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J Am Chem Soc* 2011;**133**:2124–7.
- Gehring M, Laufer SA. Emerging and re-emerging warheads for targeted covalent inhibitors: applications in medicinal chemistry and chemical biology. *J Med Chem* 2019;**62**:5673–724.
- Greiss S, Chin JW. Expanding the genetic code of an animal. *J Am Chem Soc* 2011;**133**:14196–9.
- Grünewald J, Tsao M-L, Perera R *et al*. Immunochemical termination of self-tolerance. *Proc Natl Acad Sci USA* 2008;**105**:11276–80.
- Grünewald J, Hunt GS, Dong L *et al*. Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids. *Proc Natl Acad Sci USA* 2009;**106**:4337–42.
- Gupta K, Toombes GE, Swartz KJ. Exploring structural dynamics of a membrane protein by combining bioorthogonal chemistry and cysteine mutagenesis. *ELife* 2019;**8**:e50776.
- Hacker SM, Backus KM, Lazear MR *et al*. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat Chem* 2017;**9**:1181–90.
- Han S, Yang A, Lee S *et al*. Expanding the genetic code of *Mus musculus*. *Nat Commun* 2017;**8**:14568.
- Hancock SM, Uprety R, Deiters A *et al*. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J Am Chem Soc* 2010;**132**:14819–24.
- Hankore ED, Zhang L, Chen Y *et al*. Genetic incorporation of noncanonical amino acids using two mutually orthogonal quadruplet codons. *ACS Synth Biol* 2019;**8**:1168–74.
- Hao R, Ma K, Ru Y *et al*. Amber codon is genetically unstable in generation of premature termination codon

- (PTC)-harbouring Foot-and-mouth disease virus (FMDV) via genetic code expansion. *RNA Biol* 2021;**18**:2330–41.
- He X, Chen Y, Beltran DG *et al.* Functional genetic encoding of sulfotyrosine in mammalian cells. *Nat Commun* 2020;**11**:4820.
- He J, Fan Z, Tian Y *et al.* Spatiotemporal activation of protein O-GlcNAcylation in living cells. *J Am Chem Soc* 2022;**144**:4289–93.
- Heckler TG, Zama Y, Naka T *et al.* Dipeptide formation with misacylated tRNAPheS. *J Biol Chem* 1983;**258**:4492–5.
- Hemphill J, Chou C, Chin JW *et al.* Genetically encoded light-activated transcription for spatiotemporal control of gene expression and gene silencing in mammalian cells. *J Am Chem Soc* 2013;**135**:13433–9.
- Hemphill J, Borchardt EK, Brown K *et al.* Optical control of CRISPR/Cas9 gene editing. *J Am Chem Soc* 2015;**137**:5642–5.
- Herner A, Marjanovic J, Lewandowski TM *et al.* 2-Aryl-5-carboxytetrazole as a new photoaffinity label for drug target identification. *J Am Chem Soc* 2016;**138**:14609–15.
- Hino N, Oyama M, Sato A *et al.* Genetic incorporation of a photo-crosslinkable amino acid reveals novel protein complexes with GRB2 in mammalian cells. *J Mol Biol* 2011;**406**:343–53.
- Hoang DM, Pham PT, Bach TQ *et al.* Stem cell-based therapy for human diseases. *Signal Transduct Target Ther* 2022;**7**:272.
- Hodgkin J. Novel nematode amber suppressors. *Genetics* 1985;**111**:287–310.
- Hoppmann C, Wang L. Genetically encoding photoswitchable click amino acids for general optical control of conformation and function of proteins. *Methods Enzymol* 2019;**624**:249–64.
- Hoppmann C, Lacey VK, Louie GV *et al.* Genetically encoding photoswitchable click amino acids in *Escherichia coli* and mammalian cells. *Angew Chem Int Ed Engl* 2014;**53**:3932–6.
- Hoppmann C, Maslennikov I, Choe S *et al.* In Situ formation of an Azo Bridge on proteins controllable by visible light. *J Am Chem Soc* 2015;**137**:11218–21.
- Hutchins BM, Kazane SA, Staflin K *et al.* Site-specific coupling and sterically controlled formation of multimeric antibody fab fragments with unnatural amino acids. *J Mol Biol* 2011a;**406**:595–603.
- Hutchins BM, Kazane SA, Staflin K *et al.* Selective formation of covalent protein heterodimers with an unnatural amino acid. *Chem Biol* 2011b;**18**:299–303.
- Italia JS, Addy PS, Wrobel CJJ *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat Chem Biol* 2017a;**13**:446–50.
- Italia JS, Zheng Y, Kelemen RE *et al.* Expanding the genetic code of mammalian cells. *Biochem Soc Trans* 2017b;**45**:555–62.
- Italia JS, Peeler JC, Hillenbrand CM *et al.* Genetically encoded protein sulfation in mammalian cells. *Nat Chem Biol* 2020;**16**:379–82.
- Jones LH. Recent advances in the molecular design of synthetic vaccines. *Nat Chem* 2015;**7**:952–60.
- Kang J-Y, Kawaguchi D, Coin I *et al.* In vivo expression of a light-activatable potassium channel using unnatural amino acids. *Neuron* 2013;**80**:358–70.
- Karijolic J, Yu Y-T. Therapeutic suppression of premature termination codons: mechanisms and clinical considerations (review). *Int J Mol Med* 2014;**34**:355–62.
- Kim SH, Jeong JH, Lee SH *et al.* LHRH receptor-mediated delivery of siRNA using polyelectrolyte complex micelles self-assembled from siRNA-PEG-LHRH conjugate and PEI. *Bioconjug Chem* 2008;**19**:2156–62.
- Kim CH, Axup JY, Dubrovskaya A *et al.* Synthesis of bispecific antibodies using genetically encoded unnatural amino acids. *J Am Chem Soc* 2012;**134**:9918–21.
- Kim H, Kim M, Im S-K *et al.* Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab Anim Res* 2018;**34**:147–59.
- Kimble J, Hirsh D. The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev Biol* 1979;**70**:396–417.
- Kinsler HE, Pincus Z. High-throughput screening in the *C. elegans* nervous system. *Mol Cell Neurosci* 2017;**80**:192–7.
- Kitada T, DiAndreth B, Teague B *et al.* Programming gene and engineered-cell therapies with synthetic biology. *Science (New York, N.Y.)* 2018;**359**:eaad1067.
- Kleiner P, Heydenreuter W, Stahl M *et al.* A whole proteome inventory of background photocrosslinker binding. *Angew Chem Int Ed Engl* 2017;**56**:1396–401.
- Klippenstein V, Hoppmann C, YePaoletti, SP *et al.* Optocontrol of glutamate receptor activity by single side-chain photoisomerization. *ELife* 2017;**6**:e25808.
- Ko W, Kumar R, Kim S *et al.* Construction of bacterial cells with an active transport system for unnatural amino acids. *ACS Synth Biol* 2019;**8**:1195–203.
- Kozlova EN, Berens C. Guiding differentiation of stem cells in vivo by tetracycline-controlled expression of key transcription factors. *Cell Transplant* 2012;**21**:2537–54.
- Krawczyk K, Xue S, Buchmann P *et al.* Electrogenetic cellular insulin release for real-time glycemic control in type 1 diabetic mice. *Science (New York, N.Y.)* 2020;**368**:993–1001.
- Kumar P, Ban H-S, Kim S-S *et al.* T cell-specific siRNA delivery suppresses HIV-1 infection in humanized mice. *Cell* 2008;**134**:577–86.
- Lacey VK, Louie GV, Noel JP *et al.* Expanding the library and substrate diversity of the pyrrolysyl-tRNA synthetase to incorporate unnatural amino acids containing conjugated rings. *Chembiochem* 2013;**14**:2100–5.
- Lang K, Davis L, Wallace S *et al.* Genetic encoding of bicyclic nonynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels–Alder reactions. *J Am Chem Soc* 2012a;**134**:10317–20.
- Lang K, Davis L, Torres-Kolbus J *et al.* Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* 2012b;**4**:298–304.
- Laski FA, Ganguly S, Sharp PA *et al.* Construction, stable transformation, and function of an amber suppressor

- tRNA gene in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 1989;**86**:6696–8.
- Ledsgaard L, Ljungars A, Rimbault C et al. Advances in antibody phage display technology. *Drug Discov Today* 2022;**27**:2151–69.
- Lee HS, Guo J, Lemke EA et al. Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. *J Am Chem Soc* 2009;**131**:12921–3.
- Lee J, Yun K-S, Choi CS et al. T cell-specific siRNA delivery using antibody-conjugated chitosan nanoparticles. *Bioconjug Chem* 2012;**23**:1174–80.
- Lee JM, Hammarén HM, Savitski MM et al. Control of protein stability by post-translational modifications. *Nat Commun* 2023;**14**:201.
- Lemke EA, Summerer D, Geierstanger BH et al. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat Chem Biol* 2007;**3**:769–72.
- Leong S, Lam HPJ, Kirkham Z et al. Antibody drug conjugates for the treatment of multiple myeloma. *Am J Hematol* 2023;**98**:S22–34.
- Lepthien S, Hoesl MG, Merkel L et al. Azatryptophans endow proteins with intrinsic blue fluorescence. *Proc Natl Acad Sci USA* 2008;**105**:16095–100.
- Lewis Phillips GD, Li G, Dugger DL et al. Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody-cytotoxic drug conjugate. *Cancer Res* 2008;**68**:9280–90.
- Li F, Zhang H, Sun Y et al. Expanding the genetic code for photoclick chemistry in *E. coli*, mammalian cells, and *A. thaliana*. *Angew Chem Int Ed Engl* 2013;**52**:9700–4.
- Li J, Jia S, Chen PR. Diels–Alder reaction-triggered bioorthogonal protein decaging in living cells. *Nat Chem Biol* 2014a;**10**:1003–5.
- Li J, Yu J, Zhao J et al. Palladium-triggered deprotection chemistry for protein activation in living cells. *Nat Chem* 2014b;**6**:352–61.
- Li F, Li H, Zhai Q et al. A new vaccine targeting RANKL, prepared by incorporation of an unnatural amino acid into RANKL, prevents OVX-induced bone loss in mice. *Biochem Biophys Res Commun* 2018;**499**:648–54.
- Li Q, Chen Q, Klauser PC et al. Developing covalent protein drugs via proximity-enabled reactive therapeutics. *Cell* 2020;**182**:85–97.e16.
- Liaunardy-Jopeace A, Murton BL, Mahesh M et al. Encoding optical control in LCK kinase to quantitatively investigate its activity in live cells. *Nat Struct Mol Biol* 2017;**24**:1155–63.
- Lim RKV, Yu S, Cheng B et al. Targeted delivery of LXR agonist using a site-specific antibody-drug conjugate. *Bioconjug Chem* 2015;**26**:2216–22.
- Liu CC, Schultz PG. Adding new chemistries to the genetic code. *Annu Rev Biochem* 2010;**79**:413–44.
- Liu DR, Magliery TJ, Pastrnak M et al. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. *Proc Natl Acad Sci USA* 1997;**94**:10092–7.
- Liu J, Hemphill J, Samanta S et al. Genetic code expansion in zebrafish embryos and its application to optical control of cell signaling. *J Am Chem Soc* 2017;**139**:9100–3.
- Liu J, Li S, Aslam NA et al. Genetically encoding photocaged quinone methide to multitarget protein residues covalently *in vivo*. *J Am Chem Soc* 2019;**141**:9458–62.
- Liu J, Cao L, Klauser PC et al. A genetically encoded fluorosulfonyloxybenzoyl-L-lysine for expansive covalent bonding of proteins via SuFEx chemistry. *J Am Chem Soc* 2021;**143**:10341–51.
- Liu J, Yang B, Wang L. Residue selective crosslinking of proteins through photoactivatable or proximity-enabled reactivity. *Curr Opin Chem Biol* 2023;**74**:102285.
- Long C, McAnally JR, Shelton JM et al. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science (New York, N.Y.)* 2014;**345**:1184–8.
- Lu H, Wang D, Kazane S et al. Site-specific antibody-polymer conjugates for siRNA delivery. *J Am Chem Soc* 2013;**135**:13885–91.
- Luo J, Uprety R, Naro Y et al. Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation. *J Am Chem Soc* 2014;**136**:15551–8.
- Luo J, Arbely E, Zhang J et al. Genetically encoded optical activation of DNA recombination in human cells. *Chem Commun (Cambridge, England)* 2016a;**52**:8529–32.
- Luo J, Liu Q, Morihiro K et al. Small-molecule control of protein function through Staudinger reduction. *Nat Chem* 2016b;**8**:1027–34.
- Luo X, Fu G, Wang RE et al. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat Chem Biol* 2017a;**13**:845–9.
- Luo J, Torres-Kolbus J, Liu J et al. Genetic encoding of photocaged tyrosines with improved light-activation properties for the optical control of protease function. *Chembiochem* 2017b;**18**:1442–7.
- Ma JSY, Kim JY, Kazane SA et al. Versatile strategy for controlling the specificity and activity of engineered T cells. *Proc Natl Acad Sci USA* 2016;**113**:E450–8.
- Maalej KM, Merhi M, Inchakalody VP et al. CAR-cell therapy in the era of solid tumor treatment: current challenges and emerging therapeutic advances. *Mol Cancer* 2023;**22**:20.
- Mangubat-Medina AE, Ball ZT. Triggering biological processes: methods and applications of photocaged peptides and proteins. *Chem Soc Rev* 2021;**50**:10403–21.
- Mattheisen JM, Wollowitz JS, Huber T et al. Genetic code expansion to enable site-specific bioorthogonal labeling of functional G protein-coupled receptors in live cells. *Protein Sci* 2023;**32**:e4550.
- Mayer C. Selection, addiction and catalysis: emerging trends for the incorporation of noncanonical amino acids into peptides and proteins *in vivo*. *Chembiochem* 2019;**20**:1357–64.
- Mehl RA, Anderson JC, Santoro SW et al. Generation of a bacterium with a 21 amino acid genetic code. *J Am Chem Soc* 2003;**125**:935–9.

- Melançon CE, Schultz PG. One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorg Med Chem Lett* 2009;**19**:3845–7.
- Mills EM, Barlow VL, Jones AT et al. Development of mammalian cell logic gates controlled by unnatural amino acids. *Cell Rep Methods* 2021;**1**:100073.
- Mitchell P, Tollervey D. An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'→5' degradation. *Mol Cell* 2003;**11**:1405–13.
- Mukai T, Kobayashi T, Hino N et al. Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem Biophys Res Commun* 2008;**371**:818–22.
- Mukai T, Lajoie MJ, Englert M et al. Rewriting the genetic code. *Annu Rev Microbiol* 2017;**71**:557–77.
- Muranaka N, Hohsaka T, Sisido M. Photoswitching of peroxidase activity by position-specific incorporation of a photoisomerizable non-natural amino acid into horseradish peroxidase. *FEBS Lett* 2002;**510**:10–2.
- Narancic T, Almaboub SA, O'Connor KE. Unnatural amino acids: production and biotechnological potential. *World J Microbiol Biotechnol* 2019;**35**:67.
- Neumann H, Wang K, Davis L et al. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 2010;**464**:441–4.
- Nguyen DP, Mahesh M, Elsässer SJ et al. Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J Am Chem Soc* 2014;**136**:2240–3.
- Nguyen T-A, Cigler M, Lang K. Expanding the genetic code to study protein–protein interactions. *Angew Chem Int Ed Engl* 2018;**57**:14350–61.
- Nikić I, Lemke EA. Genetic code expansion enabled site-specific dual-color protein labeling: superresolution microscopy and beyond. *Curr Opin Chem Biol* 2015;**28**:164–73.
- Nikić I, Estrada Girona G, Kang JH et al. Debugging eukaryotic genetic code expansion for site-specific click-PAINT super-resolution microscopy. *Angew Chem Int Ed Engl* 2016;**55**:16172–6.
- Niu W, Schultz PG, Guo J. An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem Biol* 2013;**8**:1640–5.
- Nödling AR, Spear LA, Williams TL et al. Using genetically incorporated unnatural amino acids to control protein functions in mammalian cells. *Essays Biochem* 2019;**63**:237–66.
- Noren CJ, Anthony-Cahill SJ, Griffith MC et al. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science (New York, N.Y.)* 1989;**244**:182–8.
- Osgood AO, Zheng Y, Roy SJS et al. An efficient opal-suppressor tryptophanyl pair creates new routes for simultaneously incorporating up to three distinct noncanonical amino acids into proteins in mammalian cells. *Angew Chem Int Ed Engl* 2023;**62**:e202219269.
- Palzer S, Bantel Y, Kazenwadel F et al. An expanded genetic code in *Candida albicans* to study protein–protein interactions in vivo. *Eukaryot Cell* 2013;**12**:816–27.
- Park S-H, Ko W, Lee HS et al. Analysis of protein–protein interaction in a single live cell by using a FRET system based on genetic code expansion technology. *J Am Chem Soc* 2019;**141**:4273–81.
- Park S-H, Ko W, Park S-H et al. Evaluation of the interaction between Bax and Hsp70 in cells by using a FRET system consisting of a fluorescent amino acid and YFP as a FRET pair. *Chembiochem* 2020;**21**:59–63.
- Parrish AR, She X, Xiang Z et al. Expanding the genetic code of *Caenorhabditis elegans* using bacterial aminoacyl-tRNA synthetase/tRNA pairs. *ACS Chem Biol* 2012;**7**:1292–302.
- Peeler JC, Falco JA, Kelemen RE et al. Generation of recombinant mammalian selenoproteins through genetic code expansion with photocaged Selenocysteine. *ACS Chem Biol* 2020;**15**:1535–40.
- Peng T, Hang HC. Site-specific bioorthogonal labeling for fluorescence imaging of intracellular proteins in living cells. *J Am Chem Soc* 2016;**138**:14423–33.
- Politano L. Read-through approach for stop mutations in Duchenne muscular dystrophy an update. *Acta Myol* 2021;**40**:43–50.
- Prasanth MI, Gayathri S, Bhaskar JP et al. Understanding the role of p38 and JNK mediated MAPK pathway in response to UV-A induced photoaging in *Caenorhabditis elegans*. *J Photochem Photobiol B* 2020;**205**:111844.
- Rader C. Bispecific antibodies in cancer immunotherapy. *Curr Opin Biotechnol* 2020;**65**:9–16.
- Ren W, Ji A, Ai H-w. Light activation of protein splicing with a photocaged fast intein. *J Am Chem Soc* 2015a;**137**:2155–8.
- Ren W, Ji A, Wang MX et al. Expanding the genetic code for a dinitrophenyl hapten. *Chembiochem* 2015b;**16**:2007–10.
- Rong L, Lim RM, Yin X et al. Site-specific dinitrophenylation of single-chain antibody fragments for redirecting a universal CAR-T cell against cancer antigens. *J Mol Biol* 2022;**434**:167513.
- Roy G, Reier J, Garcia A et al. Development of a high yielding expression platform for the introduction of non-natural amino acids in protein sequences. *MAbs* 2020;**12**:1684749.
- Ryan A, Janosko CP, Courtney TM et al. Engineering SHP2 phosphatase for optical control. *Biochemistry* 2022;**61**:2687–97.
- Sanders J, Hoffmann SA, Green AP et al. New opportunities for genetic code expansion in synthetic yeast. *Curr Opin Biotechnol* 2022;**75**:102691.
- Santoro SW, Schultz PG. Directed evolution of the substrate specificities of a site-specific recombinase and an aminoacyl-tRNA synthetase using fluorescence-activated cell sorting (FACS). *Methods Mol Biol (Clifton, N.J.)* 2003;**230**:291–312.
- Santoro SW, Wang L, Herberich B et al. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat Biotechnol* 2002;**20**:1044–8.
- Schumacher D, Hackenberger CPR, Leonhardt H et al. Current status: site-specific antibody drug conjugates. *J Clin Immunol* 2016;**36**:100–7.

- Segal I, Nachmias D, Konig A *et al.* A straightforward approach for bioorthogonal labeling of proteins and organelles in live mammalian cells, using a short peptide tag. *BMC Biol* 2020;**18**:5.
- Serfling R, Seidel L, Bock A *et al.* Quantitative single-residue bioorthogonal labeling of G protein-coupled receptors in live cells. *ACS Chem Biol* 2019;**14**:1141–9.
- Shandell MA, Quejada JR, Yazawa M *et al.* Detection of Nav15 conformational change in mammalian cells using the noncanonical amino acid ANAP. *Biophys J* 2019;**117**:1352–63.
- Shao N, Singh NS, Slade SE *et al.* Site specific genetic incorporation of Azidophenylalanine in *Schizosaccharomyces pombe*. *Sci Rep* 2015;**5**:17196.
- Shen B, Xiang Z, Miller B *et al.* Genetically encoding unnatural amino acids in neural stem cells and optically reporting voltage-sensitive domain changes in differentiated neurons. *Stem Cells (Dayton, Ohio)* 2011;**29**:1231–40.
- Shi N, Yang Q, Zhang H *et al.* Restoration of dystrophin expression in mice by suppressing a nonsense mutation through the incorporation of unnatural amino acids. *Nat Biomed Eng* 2022;**6**:195–206.
- Si L, Xu H, Zhou X *et al.* Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science (New York, N.Y.)* 2016;**354**:1170–3.
- Sibaud V, Beylot-Barry M, Protin C *et al.* Dermatological toxicities of Bruton's tyrosine kinase inhibitors. *Am J Clin Dermatol* 2020;**21**:799–812.
- Siman P, Brik A. Chemical and semisynthesis of post-translationally modified proteins. *Org Biomol Chem* 2012;**10**:5684–97.
- Singh J, Petter RC, Baillie TA *et al.* The resurgence of covalent drugs. *Nat Rev Drug Discov* 2011;**10**:307–17.
- Sletten EM, Bertozzi CR. From mechanism to mouse: a tale of two bioorthogonal reactions. *Acc Chem Res* 2011;**44**:666–76.
- Smolskaya S, Andreev YA. Site-specific incorporation of unnatural amino acids into *Escherichia coli* recombinant protein: methodology development and recent achievement. *Biomolecules* 2019;**9**:255.
- Song E, Zhu P, Lee S-K *et al.* Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. *Nat Biotechnol* 2005;**23**:709–17.
- Stokman G, Qin Y, Rácz Z *et al.* Application of siRNA in targeting protein expression in kidney disease. *Adv Drug Deliv Rev* 2010;**62**:1378–89.
- Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol* 1977;**56**:110–56.
- Syed J, Palani S, Clarke ST *et al.* Expanding the zebrafish genetic code through site-specific introduction of azido-lysine, bicyclononyne-lysine, and diazirine-lysine. *Int J Mol Sci* 2019;**20**:2577.
- Takimoto JK, Adams KL, Xiang Z *et al.* Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol Biosyst* 2009;**5**:931–4.
- Takimoto JK, Xiang Z, Kang J-Y *et al.* Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. *Chembiochem* 2010;**11**:2268–72.
- Tanaka Y, Bond MR, Kohler JJ. Photocrosslinkers illuminate interactions in living cells. *Mol Biosyst* 2008;**4**:473–80.
- Tharp JM, Ehnobom A, Liu WR. tRNAPyl: structure, function, and applications. *RNA Biol* 2018;**15**:441–52.
- Teixeira AP, Fussenegger M. Engineering mammalian cells for disease diagnosis and treatment. *Curr Opin Biotechnol* 2019;**55**:87–94.
- Tian F, Lu Y, Manibusan A *et al.* A general approach to site-specific antibody drug conjugates. *Proc Natl Acad Sci USA* 2014;**111**:1766–71.
- Tian Y, Jacinto MP, Zeng Y *et al.* Genetically encoded 2-Aryl-5-carboxytetrazoles for site-selective protein photocross-linking. *J Am Chem Soc* 2017;**139**:6078–81.
- Tian X, Liu W-Q, Xu H *et al.* Cell-free expression of NO synthase and P450 enzyme for the biosynthesis of an unnatural amino acid L-4-nitrotryptophan. *Synth Syst Biotechnol* 2022;**7**:775–83.
- Tir N, Heisteringer L, Grünwald-Gruber C *et al.* From strain engineering to process development: monoclonal antibody production with an unnatural amino acid in *Pichia pastoris*. *Microb Cell Fact* 2022;**21**:157.
- Toteva MM, Richard JP. The generation and reactions of quinone methides. *Adv Phys Org Chem* 2011;**45**:39–91.
- Tsai Y-H, Essig S, James JR *et al.* Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nat Chem* 2015;**7**:554–61.
- Uprety R, Luo J, Liu J *et al.* Genetic encoding of caged cysteine and caged homocysteine in bacterial and mammalian cells. *Chembiochem* 2014;**15**:1793–9.
- Walker OS, Elsässer SJ, Mahesh M *et al.* Photoactivation of mutant isocitrate dehydrogenase 2 reveals rapid cancer-associated metabolic and epigenetic changes. *J Am Chem Soc* 2016;**138**:718–21.
- Walsh SJ, Bargh JD, Dannheim FM *et al.* Site-selective modification strategies in antibody-drug conjugates. *Chem Soc Rev* 2021;**50**:1305–53.
- Wan W, Tharp JM, Liu WR. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim Biophys Acta* 2014;**1844**:1059–70.
- Wang L. Engineering the genetic code in cells and animals: biological considerations and impacts. *Acc Chem Res* 2017;**50**:2767–75.
- Wang L, Audhya A. In vivo imaging of *C. elegans* endocytosis. *Methods (San Diego, Calif.)* 2014;**68**:518–28.
- Wang Q, Wang L. New methods enabling efficient incorporation of unnatural amino acids in yeast. *J Am Chem Soc* 2008;**130**:6066–7.
- Wang Q, Wang L. Genetic incorporation of unnatural amino acids into proteins in yeast. *Methods Mol Biol (Clifton, N.J.)* 2012;**794**:199–213.
- Wang N, Wang L. Genetically encoding latent bioreactive amino acids and the development of covalent protein drugs. *Curr Opin Chem Biol* 2022;**66**:102106.

- Wang L, Brock A, Herberich B et al. Expanding the genetic code of *Escherichia coli*. *Science (New York, N.Y.)* 2001;**292**:498–500.
- Wang W, Takimoto JK, Louie GV et al. Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat Neurosci* 2007;**10**:1063–72.
- Wang K, Schmied WH, Chin JW. Reprogramming the genetic code: from triplet to quadruplet codes. *Angew Chem Int Ed Engl* 2012;**51**:2288–97.
- Wang RE, Liu T, Wang Y et al. An immunosuppressive antibody-drug conjugate. *J Am Chem Soc* 2015;**137**:3229–32.
- Wang J, Zheng S, Liu Y et al. Palladium-triggered chemical rescue of intracellular proteins via genetically encoded Allene-Caged tyrosine. *J Am Chem Soc* 2016a;**138**:15118–21.
- Wang WW, Zeng Y, Wu B et al. A chemical biology approach to reveal Sirt6-targeted Histone H3 sites in nucleosomes. *ACS Chem Biol* 2016b;**11**:1973–81.
- Wang N, Yang B, Fu C et al. Genetically encoding Fluorosulfate-l-tyrosine to react with lysine, histidine, and tyrosine via SuFEx in proteins *in vivo*. *J Am Chem Soc* 2018;**140**:4995–9.
- Wang WW, Angulo-Ibanez M, Lyu J et al. A click chemistry approach reveals the chromatin-dependent histone H3K36 Deacylase nature of SIRT7. *J Am Chem Soc* 2019;**141**:2462–73.
- Wang T-Y, Sang G-J, Wang Q et al. Generation of premature termination codon (PTC)-harboring pseudorabies virus (PRV) via genetic code expansion technology. *Viruses* 2022;**14**:572.
- Washburn T, O'Tousa JE. Nonsense suppression of the major rhodopsin gene of *Drosophila*. *Genetics* 1992;**130**:585–95.
- Weigle WO. The induction of autoimmunity in rabbits following injection of heterologous or altered homologous thyroglobulin. *J Exp Med* 1965;**121**:289–308.
- Wesalo JS, Luo J, Morihiro K et al. Phosphine-activated lysine analogues for fast chemical control of protein subcellular localization and protein SUMOylation. *Chembiochem* 2020;**21**:141–8.
- Wu AM. Antibodies and antimatter: the resurgence of immunopET. *J Nucl Med* 2009;**50**:2–5.
- Wu N, Deiters A, Cropp TA et al. A genetically encoded photocaged amino acid. *J Am Chem Soc* 2004;**126**:14306–7.
- Wu Y, Zhu H, Zhang B et al. Synthesis of site-specific radiolabeled antibodies for radioimmunotherapy via genetic code expansion. *Bioconjug Chem* 2016;**27**:2460–8.
- Wu Y, Zhu R-Y, Mitchell LA et al. *In vitro* DNA SCRaMble. *Nat Commun* 2018a;**9**:1935.
- Wu T, Li F, Sha X et al. A novel recombinant RANKL vaccine prepared by incorporation of an unnatural amino acid into RANKL and its preventive effect in a murine model of collagen-induced arthritis. *Int Immunopharmacol* 2018b;**64**:326–32.
- Wu K-L, Moore JA, Miller MD et al. Expanding the eukaryotic genetic code with a biosynthesized 21st amino acid. *Protein Sci* 2022a;**31**:e4443.
- Wu D, Zhang Y, Tang Z et al. Creation of a yeast strain with co-translationally acylated nucleosomes. *Angew Chem Int Ed Engl* 2022b;**61**:e202205570.
- Wu X-Y, Li M-Y, Yang S-J et al. Controlled genetic encoding of unnatural amino acids in a protein nanopore. *Angew Chem Int Ed Engl* 2023;**62**:e202300582.
- Wulf M, Pless SA. High-sensitivity fluorometry to resolve ion channel conformational dynamics. *Cell Rep* 2018;**22**:1615–26.
- Xi Z, Davis L, Baxter K et al. Using a quadruplet codon to expand the genetic code of an animal. *Nucleic Acids Res* 2022;**50**:4801–12.
- Xiang Z, Ren H, Hu YS et al. Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity. *Nat Methods* 2013;**10**:885–8.
- Xiang Z, Lacey VK, Ren H et al. Proximity-enabled protein crosslinking through genetically encoding haloalkane unnatural amino acids. *Angew Chem Int Ed Engl* 2014;**53**:2190–3.
- Xiao H, Chatterjee A, Choi S-hyun et al. Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angew Chem Int Ed Engl* 2013;**52**:14080–3.
- Xue Y-P, Cao C-H, Zheng Y-G. Enzymatic asymmetric synthesis of chiral amino acids. *Chem Soc Rev* 2018;**47**:1516–61.
- Yanagisawa T, Hino N, Iraha F et al. Wide-range protein photocrosslinking achieved by a genetically encoded N(ϵ)-(benzyloxycarbonyl)lysine derivative with a diazirinyl moiety. *Mol Biosyst* 2012;**8**:1131–5.
- Yang Y, Luo S, Huang J et al. Photoactivation of innate immunity receptor TLR8 in live mammalian cells by genetic encoding of photocaged tyrosine. *Chembiochem* 2022;**23**:e202100344.
- Yang J, Wang K, Zhang S et al. Site-specific introduction of bioorthogonal handles to nanopores by genetic code expansion. *Angew Chem Int Ed Engl* 2023;**62**:e202216115.
- Yao Y-d, Sun T-meng, Huang S-yin et al. Targeted delivery of PLK1-siRNA by ScFv suppresses Her2+ breast cancer growth and metastasis. *Sci Transl Med* 2012;**4**:130ra48.
- Young TS, Ahmad I, Brock A et al. Expanding the genetic repertoire of the methylotrophic yeast *Pichia pastoris*. *Biochemistry* 2009;**48**:2643–53.
- Young TS, Young DD, Ahmad I et al. Evolution of cyclic peptide protease inhibitors. *Proc Natl Acad Sci USA* 2011;**108**:11052–6.
- Yu B, Li S, Tabata T et al. Accelerating PERx reaction enables covalent nanobodies for potent neutralization of SARS-CoV-2 and variants. *Chem* 2022;**8**:2766–83.
- Yuan Z, Wang N, Kang G et al. Controlling multicycle replication of live-attenuated HIV-1 using an unnatural genetic switch. *ACS Synth Biol* 2017;**6**:721–31.
- Zang J, Chen Y, Liu C et al. Genetic code expansion reveals aminoacylated lysine ubiquitination mediated by UBE2W. *Nat Struct Mol Biol* 2023;**30**:62–71.
- Zaykov AN, Mayer JP, DiMarchi RD. Pursuit of a perfect insulin. *Nat Rev Drug Discov* 2016;**15**:425–39.
- Zhang Z, Alfonta L, Tian F et al. Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc Natl Acad Sci USA* 2004;**101**:8882–7.
- Zhang G, Li J, Xie R et al. Bioorthogonal chemical activation of kinases in living systems. *ACS Cent Sci* 2016;**2**:325–31.

- Zhao S, Shi J, Yu G *et al.* Photosensitive tyrosine analogues unravel site-dependent phosphorylation in TrkA initiated MAPK/ERK signaling. *Commun Biol* 2020;**3**:706.
- Zheng Y, Lewis TL, Igo P *et al.* Virus-enabled optimization and delivery of the genetic machinery for efficient unnatural amino acid mutagenesis in mammalian cells and tissues. *ACS Synth Biol* 2017;**6**:13–8.
- Zhu C, Xu L, Chen L *et al.* Epitope-directed antibody elicitation by genetically encoded chemical cross-linking reactivity in the antigen. *ACS Cent Sci* 2023;**9**:1229–40.
- Zuany-Amorim C, Manlius C, Dalum I *et al.* Induction of TNF- α autoantibody production by AutoVac TNF106: a novel therapeutic approach for the treatment of allergic diseases. *Int Arch Allergy Immunol* 2004;**133**:154–63.