Therapeutic applications of genetic code expansion

Yujia Huang, Tao Liu*
State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing, 100191, China

ARTICLE INFO
Keywords:
Genetic code expansion
Noncanonical amino acids
Biotherapeutics
Antibody-drug conjugates
Live-attenuated vaccines

ABSTRACT
In nature, a limited, conservative set of amino acids are utilized to synthesize proteins. Genetic code expansion technique reassigns codons and incorporates noncanonical amino acids (ncAAs) through orthogonal aminoacyl-tRNA synthetase/ tRNA pairs. The past decade has witnessed the rapid growth in diversity and scope for therapeutic applications of this technology. Here, we provided an update on the recent progress using genetic code expansion in the following areas: antibody-drug conjugates (ADCs), bispecific antibodies (BsAb), immunotherapies, long-lasting protein therapeutics, biosynthesized peptides, engineered viruses and cells, as well as other therapeutic related applications, where the technique was used to elucidate the mechanisms of bio-therapeutics and drug targets.

1. Introduction

While DNA and RNA are two ways how creatures on Earth store hereditary information, it is protein that constitutes most of cells’ dry mass and carry out most information-bearing functions. Natural proteins are synthesized by 20 canonical amino acids. The different chemical properties of these building blocks equip proteins with multifarious structures and functions, however, at the same time, restrict the development of protein properties. Post-translational modifications (PTM) such as methylation, phosphorylation and glycosylation, as well as the binding of co-enzyme factors are different strategies natural proteins take to overcome these limitations at a certain extent. To unleash the full power of amino acid diversities and meet the challenges for protein related applications, scientists have devised a variety of ways to incorporate noncanonical amino acids (ncAAs) into proteins, over the past few decades, greatly expanding protein properties.

Genetic code expansion is one of these techniques, which allows site-specific incorporation of ncAAs into any protein of interest in cells. This is achieved with engineered aminoacyl-tRNA synthetase/tRNA (aaRS/tRNA) pairs which can specifically aminoacylates its cognate tRNA with a given ncAA, devoted to decoding reassigned codons, which are typically nonsense codons (UAG/UGA/UAA). The aminoacylated suppressor tRNA charged with the provided ncAA is then recognized by ribosome allowing the ncAA to be site-specifically incorporated into the synthesizing peptides during translation in response to assigned codons. This system must be orthogonal with the native translation system, including endogenous tRNA, canonical amino acids as well as endogenous aaRS (Fig. 1(a)).

Several aaRS/tRNA pairs have been developed to incorporate more than 200 different ncAAs in different species. For recombinant expression therapeutic proteins in E. coli, the most widely used aaRS/tRNA pairs are the Archean TyrRS/tRNA^{Tyr} pair from *Methanocaldococcus jannaschii* (MfTyrRS/tRNA^{Tyr}) [1–4] and the Archean pyrrolysine aaRS/tRNA^{Pyl} pair (Mb/MmPylRS/tRNA^{Pyl}) [5–9]. The MfTyrRS/tRNA^{Tyr} pair was the first genetic code expansion system developed for site-specific ncAAs incorporation in living cells in 2000 [1] and has now been widely used in prokaryotic genetic code expansion. This system is highly efficient in incorporation of aromatic ncAAs with a wide range of novel chemical, physical and biological properties. While the MfTyrRS/tRNA^{Tyr} pair is commonly used in *E. coli*, it is not orthogonal in higher organisms. Unlike other systems, *Mb/MmPylRS/tRNA^{Pyl} pair* was a gift from nature. The discovery of this naturally existing orthogonal system was first described in 2002 and it allows the genetic incorporation of a natural ncAA, pyrrolysine, which was then referred as the 22nd amino acid [10,11], site-specifically into the proteome of living cells in response to amber stop codon. The pyrrolysine system has shown a great orthogonality in both prokaryotes and eukaryotic cells. Such orthogonality enables the evolution of PylRS/tRNA^{Pyl} pairs in *E. coli* and the evolved PylRSs can be directly applied in mammalian cells. In addition, the evolved PylRS mutants can incorporate not only pyrrolysine analogs but also a diverse side chain structures [12].

Systems most widely used in eukaryotes nowadays are *E. coli* TyrRS/tRNA^{Tyr} ([EcTyrRS/tRNA^{Tyr}]) [13–17], *E. coli* LeuRS/tRNA^{Leu} ([EcLeuRS]}
tRNALeu) [16,18,19] and Mb/Mm PylRS/tRNAPyl pair [17,20–28]. The incorporation of ncAAs in Saccharomyces cerevisiae was first achieved with E. coli TyrRS/tRNATyr pair by Chin, J. W. et al. [14]. The E. coli TyrRS has also been cooperated with the Bacillus stearothermophilus tRNATyr [13,29] to achieve more efficient suppression. These pairs successfully introduced phenylalanine derivatives containing chemical structures like benzophenone [15], ketone [30], iodide [30] and azide [30] groups. E. coli LeuRS/tRNALeu pair was first developed in 2004 [18]. A variety of structures can be incorporated by this pair through mult-site mutations for its tolerance for structural diversity. Mb/Mm PylRS/tRNAPyl pair has also been extensively applied in mammalian cells as previously described. Other platforms like TrpRS/tRNA^Trp [7,31–34], PheRS/tRNA^Phe [35,36], LysRS/tRNA^Lys [37], SepRS/tRNA^Cys [38,39] have also been developed and were further explained elsewhere [40]. The orthogonality feature of these pairs determines the disadvantage that two distinct platforms need to be developed for applications of one single ncAA in both bacterial and eukaryotic cells. Italia, J. S. et al. recently reported a novel strategy to expand the genetic code of E. coli by replacing its endogenous pair with an eukaryotic-archaeal counterpart, achieving a general way to develop new aaRS/tRNA pairs for both E. coli and eukaryotes [34,41].

Genetic code expansion has been commonly used to study protein structures and functions both in vitro and in vivo (Fig. 1(b)). Chemical cross-linkable ncAAs were used to catch proximity-based protein-protein interactions in live cells [42,43]. Fluorescent ncAAs [44] or fluorophores that react with ncAAs through bioorthogonal ligation reactions [45,46] were used in live cell imaging studies. Through incorporation of photo-switchable [47,48] or chemical-switchable ncAAs [49], a new field in controllable protein function has also been opened. In addition to the above applications, expanding the genetic code has shown great advantages in the study of PTM, like serine phosphorylation [39,50], threonine phosphorylation [51], tyrosine phosphorylation [52], lysine acetylation [6], lysine crotonylation [53] and lysine hydroxyisobutyrylation [54].

The site-specific modification characteristic of genetic code expansion technique makes it a powerful tool in the improvement of pharmacological properties of protein therapeutics as well as in developing novel therapeutic methods. In this review, we will be focused on the therapeutic applications of genetic code expansion in recent years.

2. Antibody-drug conjugate (ADC)

Antibody-drug conjugate (ADC) is a class of novel biopharmaceutical drugs with a significantly enhanced therapeutic window against cancer. Four ADCs have been approved by FDA and EMA in clinical use [55]. Conventional method of ADC production is through non-selectively coupling to lysine or cysteine residues, yielding heterogeneous products, which decreases ADC potency, circulation half-life, stability, tolerability as well as antigen binding [56]. In 2010, the first approved ADC, gemtuzumab ozogamicin (Mylotarg™) was voluntarily withdrawn from the market due to toxicity and a lack of efficacy [57]. To avoid the problem of heterogeneity, different ways have been exploited to engineer ADCs. Reactive cysteine residues were introduced into antibody-Fab, which is called THIOMAB (the antibodies with engineered reactive cysteine residues), exhibiting an improved
DIBO-DOTA was site-specific acid; DUPA: 2-[3-(1,3-dicarboxy propyl)-ureido] pentanedioic acid. The inherent cysteine-maleimide instability and the multi-step process precluded it from clinical applications as well. Other site-specific methods utilizing enzymatic modifications like β,4-galactosyltransferase (Gal T), α,2,6-sialyltransferase (Sial T) [60], transglutaminase (TG) [61], formylglycine-generating enzyme (FGE) or anaerobic sulfatase-maturating enzyme (anSME) [62] have also been demonstrated to successfully introduce specific residues into antibodies. Nevertheless, the positions that can be targeted on antibodies are limited. The incorporation of ncAAs using genetic expression systems like auxotrophic strain or selenocysteine have provided new insights. The incorporation of ncAAs using genetic expression systems like auxotrophic strain or selenocysteine have provided new insights.

Genetic code expansion owns the advantage of synthesizing homogenous ADCs through incorporating ncAAs with different moieties at one or more specific sites, such as ketone, azide, alkyne and alkene groups [57,65]. To produce antibodies with excellent potency, pharmacokinetics and safety, p-acetylphenylalanine (pAcF) was genetically incorporated into two sites on an anti-Her2 antibody Fab fragment and coupled with the microtubule toxin auristatin F (AF) (Fig. 2) [66]. The synthesis of chemical defined CCKR4-auristatin ADC, anti-CD11a IgG conjugated with liver X receptor (LXR) agonist or phosphodiesterase 4 (PDE4) embraced the same strategy [67–69]. These ADCs not only showed significant effects in in vivo disease models but also reduced unwanted side-effects due to the improvement of drug specificity. To produce more complex proteins like full-length antibodies, Tian, F et al. utilized CHO cells to develop a stable expression system and promoted the titer of antibodies containing ncAAs to over 1 g/L [70]. The genetic code expansion has also been used in radioimmunotherapy. Through incorporation of an azido group-bearing amino acid, Nε-2-azideoethoxyoxycarbonyl-L-lysine; MMAE: monomethyl auristatin E; DIBO: 4-dibenzocyclooctynol; DOTA: 1,4,7,10-tetraazacyclodecane-1,4,7,10-tetraacetic acid; DUPA: 2-[3-(1,3-dicarboxy propyl)-ureido] pentanedioic acid.

expression systems that efficiently incorporated a cyclopropene derivative of lysine (CypK) into trastuzumab through genetic code expansion [72]. Above all else, this technique has also been applied to immunosuppressive drugs. This was exemplified by the site-specific conjugation of dasatinib with humanized antibody HLCX, selectively delivering this Lck inhibitor to human T lymphocytes [73].

3. Bispecific antibody (BsAb)

Bispecific antibody is a type of bioengineered antibodies with the ability to simultaneously recognize two different types of epitopes. In 2009, catumaxomab (Removab) was approved in the European Union for the treatment of malignant ascites [74]. The FDA granted accelerated approval of blinatumomab (BlinCyto) in December, 2014 for the treatment of Philadelphia chromosome-negative relapsed or refractory precursor B-cell acute lymphoblastic leukemia (R/R ALL) [75]. However, the development of bispecific antibodies is hampered by the production of correctly assembled structures (i.e., correct assembly of heavy and light chains). One way to produce bispecific antibodies is through chemical conjugations using lysine or cysteine residues within the antibody, which often yields heterogenous products unsuitable for therapeutic applications.

Through site-specific encoding pAcF into anti-Her2 and anti-CD3 Fabs and subsequently coupling the Fabs to bifunctional ethylene glycol linkers with an azoxy-amine on one end and an azide or cyclooctyne group on the other, an anti-Her2/anti-CD3 bispecific antibody was synthesized by a copper-free click chemistry, effectively crosslinking HER2+ cells and CD3+ cells (Fig. 3) [76,77]. Similar strategy has also been applied to the synthesis of a novel bispecific antibody, cCLL1–cCD3, which completely eliminated established tumors in a cell line xenograft model, showing potential as a treatment for acute myeloid leukemia (AML) [78]. A general and straightforward approach using Watson-Crick base pairing interactions has been demonstrated for producing homodimeric, heterodimeric and multimeric Fabs with defined composition, valency and geometry. The chemical reactive ncAA in this platform was used to site-specifically couple oligonucleotides or...
lymphoma [84]. The conventional CAR-T cell therapy is facing a CAR that recognizes anti-tag CAR technology which was adaptable and capable of regulating cell activities when necessary. In 2012, Tamada, K et al. developed an cytokine release syndrome (CRS). Hence, it is important to control T cell number of challenges. One of the most important obstacles is the severe refractory pre-B cell acute lymphoblastic leukemia and di therapy targeting CD19 was approved by FDA for the treatment of re- limitations of the need for MHC expression. In August 2017, a CAR-T of an anti-tumor single-chain variable fragment (scFv), overcoming the cells derived from patients to recognize cancer cells through equipment conjugate [82] and a switchable α

peptide nucleic acids (PNAs) to antibodies, with the assemblies showing enhanced anti-tumor activity [79]. Different types of chemical reactions have been constantly exploited to develop novel bispecific antibodies. In 2015, Ramadoss, N. S et al. utilized the reaction between tetrazine (TET) and bicyclononyne (BCN) to develop anti-BCMA and anti-CS1 bispecific antibodies (Fig. 2). The resulting BiFab-BCMA robustly activated T cells and mediated rapid tumor regression [80].

Apart from bispecific antibodies, bispecific targeting therapeutics consisting of a small molecule 2-[3-(1, 3-dicarboxy propyl)-ureido]pentanedioic acid (DUPA), that selectively targets a tumor-associated antigen, prostate-specific membrane antigen (PSMA) were also developed (Fig. 2). With DUPA conjugating to pAcF site-specifically expressed on anti-CD3 Fab, the EC_{50} of the conjugate has been reduced to 100pM, and serum half-life has been improved to 5–6 h, resulting effective in a treatment in prostate cancer xenograft mouse models [81]. Same strategy was also employed to synthesize an anti-CD3 Fab-folate conjugate [82] and a switchable αGCN4-Fab conjugate [83].

4. Chimeric antigen receptor T (CAR-T) cell immunotherapy

Chimeric antigen receptor T (CAR-T) cell therapy is a promising immunotherapy for cancer treatment. CARs endow the engineered T cells derived from patients to recognize cancer cells through equipment of an anti-tumor single-chain variable fragment (scFv), overcoming the limitations of the need for MHC expression. In August 2017, a CAR-T therapy targeting CD19 was approved by FDA for the treatment of refractory pre-B cell acute lymphoblastic leukemia and diffuse large B cell lymphoma [84]. The conventional CAR-T cell therapy is facing a number of challenges. One of the most important obstacles is the severe cytokine release syndrome (CRS). Hence, it is important to control T cell activities when necessary. In 2012, Tamada, K et al. developed an anti-tag CAR technology which was adaptable and capable of regulating CAR T-cell functions. T cells were genetically engineered to express a CAR that recognizes fluorescein isothiocyanate (FITC) molecules and various antibodies conjugated with FITC served as an intermediary [85]. However, antibodies in this study were conjugated with FITC using nonspecific labeling methods, with an average of 3 FITC molecules conjugating per one molecule of antibody. FITC-labelled Abs-induced in vivo anti-FITC immune responses might shorten the half-life of these intermediary molecules and thus weakened curative effects. In this case, a site specifically modified “switch” molecule is needed to minimize the potential for immunogenicity. Ma, J. S. Y et al. and Cao, Y et al. later developed switchable CAR-T (sCAR-T) systems based on genetic code expansion, in which anti-CD19, anti-CD22 and anti-Her2 antibody fragments were site-specifically modified with FITC through para-azidophenylalanine (pAzF) and iterated specific spatial interactions between sCAR-T cells and cancer cells (Fig. 3) [86,87]. The strictly switch dose-dependency of this sCAR-T system enabled CAR-T cell activity to become potent and controllable, thus improving efficacy and safety. Furthermore, this system is able to target multiple antigens without further genetic engineering of a patient’s T cells, making it a versatile immunotherapy.

5. Increasing therapeutic agent serum half-lives

PEGylation is a well-established technique for improving the pharmacokinetics and biocompatibility of therapeutic agent with short in vivo half-lives. The first unspecific multi-PEGylated drug, Adagen, was approved by FDA in 1990 [88]. PEGylation is usually achieved through chemical conjugation to reactive side chains [89]. The acquired heterogeneous mixture gave rise to difficulties in batch-to-batch reproducibility and quality control, making site-specific PEGylation attractive.

In spite of the efficacy of human growth hormone (hGH) therapy for treatment for pathological short stature, it requires daily subcutaneous injection due to its short circulating half-life [90]. Through site-specific conjugation of hGH containing pAcF with PEG, the homogeneous hGH variant was demonstrated to be as effective and safe as the native hGH. Meanwhile, the injection frequency was reduced to once a week [90]. In addition, site-specific multi-PEGylation of hGH was achieved by Wu, L et al., which showed further reduced immunogenicity and improved pharmacokinetic profiles [91]. Apart from hGH, site-specific PEGylation using genetic code expansion has also been expanded to the optimization of IFN-α2b [92], fibroblast growth factor 21 (FGF21) [93] as well as AAV vectors [94]. Similar strategy was developed to achieve site-specific Poxylation of Interleukin-4 (IL-4), either [95].

In addition to site-selective conjugation, ncAAs themselves are able to enhance protein stability through formation of more stable bonds. ncAAs containing long side-chain thios were incorporated into β-lactamase, forming unnatural disulfide bond and exhibiting significantly enhanced thermostability both in vitro and in vivo [96]. Genetically incorporation of p-fluorophenylalanine (pFF) [97] or para-isothiocyanate phenylalanine (pNCSP) [98] was also capable of enhancing protein stability at elevated temperature.

6. Biosynthesized therapeutic peptides

Genetic code expansion was used in combination with split inmune catalyzed ligation of proteins and peptides (SICLOPPS) to construct a large cyclic peptide library, which was subsequently harnessed to evolve an HIV protease inhibitor [99]. More recently, co-translational and site-specific incorporation of ncAAs has been applied to the biosynthesis of ribosomally synthesized and post-translationally modified peptide (RIPP) natural products to expand their chemistry and structure repertoire. For example, incorporation of ncAAs at discrete positions on Nisin expanded building blocks that can be biosynthetically incorporated into lanthipeptide and diversified macrocyclic topologies of this antibacterial peptide [100]. Likewise ncAAs have been successfully introduced into other lanthipeptides [101,102], lasso peptides [103,104], macrocyclic peptides [105], thiopeptides [106] as well as in the natural product cinnamycin, recently [107].
7. Site specifically modified viruses

Adeno-associated virus (AAV) is a small non-enveloped human virus that has proven to be a viral vector for gene therapy with its attractive feature of lacking apparent pathogenicity, replication deficiency, persistent transgene expression as well as the capability to infect nondividing and dividing cells [108]. Itsicosahedral capsid is made up of three capsid proteins, VP1, VP2, and VP3 [109]. However, the serum instability, immunogenicity and nonspecific cell infectivity remain major obstacles for effective gene transfer. Although reengineering of altered capsid proteins [110,111], fusion of high-affinity ligands to capsid proteins [112] and chemical conjugation of small molecule effector functionality [113] were applied to overcome these challenges, few methods can produce heterogeneous vectors for specific targets.

To alleviate the immunogenicity and instability, polyethylene glycols (PEG) was covalently conjugated to AAV-2 site-selectively through genetic code expansion (Fig. 4) [94]. In this study, NAEEK was introduced to a series of chosen sites on the capsid protein VP1. PEGylation at these sites showed protective effect and improvement in stability. Selective tropism towards αβ3 integrin receptors overexpressing tumor cells was enhanced through bioorthogradically and stoichiometrically attachment of AAV2 with a cyclic-RGD peptide precisely, constructing chimeric viruses as simple and viable targeting delivery vehicles (Fig. 4) [114,115]. This method greatly expanded the versatility of AAV vectors as it could be easily applied to different ligands and to other types of viral vectors for gene therapy. Recently, Erickson, S. B et al. combined AAV with photocaged derivatives of lysine to precisely perturb virus-host interaction with high in vivo spatial and temporal control [116]. A photocaged amino acid NBK was site-specifically incorporated to a functionally important site on the virus capsid protein VP1, abrogating the positive charge on the lysine residue and subsequently undergoing efficient decaging to non-toxic products upon irradiation with 365 nm light (Fig. 4). With this methodology, it is able to synchronize the invading viruses at a specific stage. Fluorescent molecules were also lately labelled to AAV2 for visualization and tracking of virus vectors [72].

8. Therapeutic vaccines

8.1. Protein-based vaccines

Immunization of patients with self-proteins or specific epitopes like tumor antigens is difficult due to weakly immunogenicity and immunotolerance. A number of strategies including adjuvants, the introduction of foreign immunodominant T-helper (Th) epitopes into chimeric antigens [117,118], chemical derivatization of self-antigens [119], and DNA vaccines [120] have been exploited to break immunological self-tolerance. Autologous tumor cells modified with a hapten, dinitrophenyl (DNP), was demonstrated to elicit patients’ immune responses cross-reacting with unmodified, native tumor-cell antigens, likely due to the formation of strong stacking and van der Waals interaction [121]. Therefore, genetic code expansion provided a way to introduce immunogenic ncAAs into autologous proteins and induce immune responses to these self-tolerated epitopes. Site-specific incorporation of p-nitrophenylalanine (pNO2Phe) to murine tumor necrosis factor-α (mTNF-α) successfully generated a sustained high-titer antibody response to both parental and antigenically distinct mTNF-α variants without the need for strong adjuvants [122]. Another DNP-containing ncAA, N(2)-(2,4-dinitrophenyl)acetylyllysine (DnpK) was later incorporated into HEK 293T cells, although unstable in Escherichia coli [123]. Sulfotyrosine (SO3Tyr) and 3-nitrotyrosine (3NO2Tyr), two PTM-related ncAAs were introduced to specific sites in mTNF-α and EGF, indicating that PTM might play a role in autoimmune disorders [124]. This methodology could also be applied to a self-protein unrelated to immune function, murine retinol-binding protein 4 (RBP4) [125] and might as well be applied to other weakly immunogenic antigens.

8.2. Organism-based vaccines

Although traditional live-attenuated vaccines (LAVs) are potent, the potential pathogenic consequences have aroused safety concerns [126]. Genetic code expansion strategy was first coupled with intact and infectious virus engineering using hepatitis D virus (HDV) as a model system [127]. The resultant HDV virions maintained near wild-type viability and infectivity. Then this strategy was expanded to influenza A viruses and human immunodeficiency virus type 1 (HIV-1) virus for the development of LAVs and could be potentially adapted to almost any viral vaccines. Live but replication-incompetent influenza A virus vaccine was generated in a special transgenic cell line harboring MbpYlRS/tRNA^{Asn}_{Lam}/AcylA, eliciting robust immunity against both wild type and ncAA-containing strains (Fig. 5(a)) [128]. By means of introducing blank codons on HIV-1 essential proteins within the genome, all-in-one live-attenuated HIV-1 mutants were constructed, and HIV-1 viability was precisely controlled by ncAAs [129,130]. This live-attenuated HIV-1 vaccine constructed through genetic code expansion marked an important step towards combating AIDS pandemic worldwide.

Genetic code expansion showed its enormous potential in developing organisms strictly depending on ncAA for growth. After ncAA incorporation had been achieved in the intracellular pathogen Mycobacterium tuberculosis (MtB), which facilitated studies of tuberculosis (TB) vaccine development [131], increasing number of genetically modified organisms (GMOs) were developed. Unlike viruses, evolutionary pressure compels cellular organisms to impose mutagenic drift, environmental supplementation and horizontal gene transfer (HGT), resulting in high escape frequencies [132]. Computationally redesign of essential enzymes and the combination of multiple redesigned enzymes resulted in remarkable reduced escape rate, alleviating this problem to some degree [132]. Using multiplex automated genome engineering, TAG codons were introduced into permissive locations of 22 essential genes in an Escherichia coli strain that lacked all TAG codons and release factor 1, creating a notable strain with robust growth and undetectable escape frequencies (Fig. 5(b)) [133]. To addic the modified organisms...
to ncAAs, a selection method was developed using TEM-1 β-lactamase as a model system [134].

Inspired by PTM, scientists developed a general strategy to create genetically modified organisms with stringent dependency on the presence of N-ε-acetyl-L-Lys (AcK), which blocked a functionally essential lysine residue required for cell growth and can be rescued by endogenous deacetylases. The combination of this approach with a bar-nase-based suicide switch finally gave a low escape frequency [135]. Other ncAAs with specific chemical properties have also been utilized for biological containment. The hydrophobic amino acid p-benzoyl-l-phenylalanine (pBzF) was introduced to a homodimeric chorismate mutase (CM) to disrupt protein-protein interfaces for its large nonpolar side chain which cannot be replaced by 20 canonical amino acids in theory. Through screening of ncAA surrounding residues, CM was evolved to an enzyme that could only retain active in presence of pBzF, thus developing Escherichia coli strain dependent on ncAAs [136]. However, this deficiency might be complemented by aromatic amino acids in the host. Afterwards, replacing key histidine (His) residues in essential metalloproteins with 3-methyl-l-histidine (MeH) provided a straightforward strategy to solve this problem [137]. Due to the requirement of two single nucleotide mutations to substitute UAG codon (CRF1R) [139], the binding sites between antidepressant drugs and human serotonin transporter (hSERT) [140] and the interactions between exendin-4 peptide and the glucagon-like peptide-1 receptor (GLP-1R) [141], which may contribute to the pursuit of structure-based drug design.

Apart from these applications, it is worth mentioning that clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system have also been combined with a genetically installed photocaged lysine (PCK), achieving opto-chemical controllable gene editing [142]. Although the CRISPR system has not yet been applied in clinic, it has shown a great potential in future gene therapy and cell therapy. We think that the genetic code expansion technology has not yet shown its full potential in the gene editing field.

10. Challenges and prospects

Since the first ncAA, O-methyl-L-tyrosine was incorporated into proteins in Escherichia coli, expansion of the genetic code has allowed over 200 ncAAs to be site-specifically incorporated into proteins in bacteria, yeast [14,16,18,20,142–147], plant [23], worm [21,148,149], insect [22,25,27,150], fish [151] and mammal [13,15,17,19,24–26,28,31,34,144,151–153], including stem cells [154,155], thus affording novel avenues for biotherapeutic applications both in vitro and in vivo. Drug candidates produced with genetic code expansion like cHER2 ADC (Ambrx, ARX7988) and PEG-FGF21 (Ambrx, ARX618) are in phase 1 and phase 2 clinical trial respectively.

Although the competence of precise and highly tailored structural perturbations to proteins provides abundant advantages, an expanded genetic code is far from reaching its full potential. While ncAAs were incorporated at high levels in E. coli and several cell lines (e.g. HEK293T cells), one of the major challenges of this technique is the low suppression efficiency in more challenging cell lines and species due to its competition with endogenous termination system. This was partly solved by engineering components in the expression system like eRF1 [156–158], the elongation factor Tu (EF-Tu) [159], ribosome [160] and the nonsense-mediated mRNA decay (NMD) pathway [16,28], codon reassignment [161,162] or by improving expression vectors [163]. Up to 30 ncAA residues incorporated with high yields and accuracy into one recombinant protein can now be achieved in E. coli [4]. Optimization of tRNA structure have also been applied in bacteria. However, the evolution of the tRNA sequence in mammalian cells is still challenging [164]. This might also arouse the suspicion that suppressor tRNAs would have toxic effects to cell fitness due to uncontrolled proteome-wide ncAA incorporation. Quadruplet codons offer an optional route to avoiding this problem [3,37]. Another issue is the limited ncAA core structure motif owing to the limited orthogonal pairs. Many natural PtsMs still cannot be encoded either in natural or mimic forms. This can only be resolved by exploiting novel aaRS/tRNA system and higher throughput screening methods. Recently, a phage-assisted continuous evolution (PACE) [165] and a phage-assisted non-continuous evolution (PANCE) [166] have been developed, achieving rapid production of orthogonal aaRS variants that are of highly activity and amino acid specificity. Although PyrRS does not recognize the tRNA^Pyl anticodon, the fact that many aaRS use the anticodon as an identity element [167] and tRNA^Pyl, along with tRNA^Pyl display much lower suppression efficiency than tRNA^Pyl largely limited the incorporation of different ncAAs at the same time. Willis, J. C. et al. recently created new PyrRS/tRNA^Pyl pairs that are orthogonal with MmpYrRS/tRNA^Pyl pairs [169].

Future directions, in terms of biotherapeutic applications, include genetic encoding of more building blocks with multifarious chemical, physical and biological characteristics, enhancing suppression efficiencies in therapeutic related species, and to apply this system to relevant therapeutic agents including proteins, cells and even organisms. Merging between genetic code expansion with other disciplines like synthetic biology may force it to a low-cost and available technique to a great extent. Currently, the genetic code expansion technology can
allow a maximum of two additional nCAs to be site-specifically incorporated into one protein [9,24,170,171]. Recent advances in cell engineering [132,133], codon reassignment [172] and expansion of the genetic alphabet [173,174] would theoretically allow the creation of organisms with multiple nCAs in the proteome. On the other hand, more mutually orthogonal systems and structurally diverse nCAs need to be developed to accommodate these engineered organisms with free codons or expanded codons. And more importantly, more researchers need to apply the genetic code expansion technique in their research to exploit new applications and functions of nCAs, which will eventually benefit human therapy. As this technique develops, we are sure that new generations of biotherapeutics will emerge.

Acknowledgments

We thank the following agencies for financial support on related ongoing projects in our lab: National Key Research and Development Program of China (No. 2016YFA0201400), the National Natural Science Foundation of China (21778005), Peking University Health Science Center (BMU20160537 and BMU2017QQ006) and Youth-Talents Program of China.

References

[1] Wang L, Magliery TJ, Liu DR, Schultz PG. A new functional suppressor tRNA/arnzyme ribozyme pair for the in vivo incorporation of unnatural amino acids into proteins. J Am Chem Soc 2000;122:5010–1.
[2] Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of Escherichia coli. Science 2001;292:498–500.
[3] Neumann H, Peak-Chew SY, Chin JW. Genetically encoding N(epsilon)-acetamido-lysine in mammalian cells. Nat Biotechnol 2004;22:353–5.
[4] Neumann H, Peak-Chew SY, Chin JW. Genetically encoding phosphoserine and its nonhydrolyzable analog. Nat Biotechnol 2015;33:1272–9.
[5] Blight SK, Larue RC, Mahapatra A, Longsta...
Hydroxyisobutyryl-lysine into recombinant histones. ACS Chem Biol 2015;10:1599-603.

[55] Hoffmann RM, Coutube BGT, Josephs DH, Miele S, Illieva KM, Cheung A, et al. Antibody structure and engineering considerations for the design and function of Antibody Drug Conjugates (ADCs). Oncosensation 2018;7:1599127.

[56] Sievers EL, Senter PD. Antibody-drug conjugates in cancer therapy. Annu Rev Med 2013;64:15-29.

[57] Kim CH, Aupx JT, Schultz PG. Protein conjugation with genetically encoded un- site-specific amino acids. Annu Rev Biochem 2013;82:171-212.

[58] Panowski S, Bhakta S, Raab H, Polakis P, Junutula JR. Site-specific antibody drug conjugates for cancer therapy. Mabs-Austin 2014;6:34-45.

[59] Gomuz N, Vinson AR, Ouyang J, Nguyen MDH, Chen XN, Sharma VK, et al. Triple light chain antibodies: factors that influence its formation in cell culture. Biotechnol Bioeng 2010;105:748-60.

[60] Zhou Q, Stefano JE, Manning C, Kyzarice J, Bich, B, Giansallo DA, et al. Site-specific antibody-drug conjugation through glyceroengineering. Bioconjugate Chem 2014;25:520-60.

[61] Farias SE, Strop P, Delaria K, Casas MG, Dorywalska M, Shelton DL, et al. Mass spectrometric characterization of transglutaminase based site-specific antibody-drug conjugates. Bioconjugate Chem 2014;25:246-50.

[62] Appel MJ, Bertozzi CR. Formylglycine, a post-translationally generated residue with unique catalytic capabilities and biotechnology applications. ACS Chem Biol 2015;10:73-82.

[63] Hofer T, Shepherd LR, Chapman CM, Rader C. Molecularly defined antibody conjugation through a selenocysteine interface. Biochemistry 2009;48:12047-57.

[64] Thyer R, Shroff R, Klein DR, D'Oenisteis S, Cotham VC, Byrom M, et al. Custom selenoprotein production enabled by laboratory evolution of recombed strains. Nat Biotechnol 2018;36:7624-37.

[65] Wang L, Zhang ZW, Brock A, Schultz PG. Addition of the kuto-functional group to the genetic code of Escherichia coli. Proc Natl Acad Sci USA 2003;100:5907-12.

[66] Aupx JT, Bajuri KM, Riehd M, Hutchins BM, Kim CH, Kozane, SA, et al. Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. Proc Natl Acad Sci USA 2012;109:16101-6.

[67] Kularatne SA, Deshmukh V, Ma J, Tardif V, Lim RK, Pugh HM, et al. A CXCR4-targeted site-specific antibody-drug conjugate. Angew Chem Int Ed Eng 2014;53:11860-77.

[68] Lim RKV, Yu S, Cheng B, Li S, Kim NJ, Cao Y, et al. Targeted delivery of LXR agonist using a site-specific antibody drug conjugate. Bioconjugate Chem 2015;26:2216-22.

[69] Yu S, Pearson AD, Lim RKV, Rodgers DT, LU S, Parker HB, et al. Targeted delivery of an anti-inflammatory PDE4 inhibitor to immune cells via an antibody-drug conjugate. Mol Ther 2016;24:2078-89.

[70] Tian F, Lu YC, Manibusan A, Sellsa A, Han T, Sun Y, et al. A general approach to the genetic code expansion of bispecific antibodies. Nat Biotechnol 2014;32:11766-71.

[71] Wu YM, Zhu H, Zang B, Liu F, Chen JX, Wang YF, et al. Synthesis of site-specific radiolabeled antibodies for radioimmunotherapy via genetic code expansion. Bioconjugate Chem 2016;27:2460-8.

[72] Zhang C, Zhou X, Yao T, Tian Z, Zhou D. Precision fluorescent labeling of an adeno-associated virus vector to monitor the viral infection pathway. Biotechnol J 2018;13:1700734.

[73] Shi YX, Jie L, Wu Y, Cao Y, DU JT, Luo XZ, et al. An immunosuppressive antibody-drug conjugate. J Am Chem Soc 2015;137:3229-32.

[74] Holmes D. Buy buy bispecific antibodies. Nat Rev Drug Discov 2011;10:798–800.

[75] Prasipori, D, Ko DW, Deiserosk R, Ancey CL, Chaudron-Chacou R, Chiu JH, et al. Site-specific antibody-drug conjugates. Nat Prod Sci 2014;31:11766-71.

[76] Wu YM, Zhu H, Zang B, Liu F, Chen JX, Wang YF, et al. Synthesis of site-specific radiolabeled antibodies for radioimmunotherapy via genetic code expansion. Bioconjugate Chem 2016;27:2460-8.

[77] Shi YX, Jie L, Wu Y, Cao Y, DU JT, Luo XZ, et al. An immunosuppressive antibody-drug conjugate. J Am Chem Soc 2015;137:3229-32.

[78] Lopatniuk M, Myronovskyi M, Luzhetskyy A. Streptomyces albus: a new cell fac- tors that in-...
therapeutic approach for the treatment of allergic diseases. Int Arch Allergy Immunol 2004;133:154–63.

[119] Weigle WO. The induction of autoimmunity in rabbits following injection of heterologous or altered homologous thyroglobulin. J Exp Med 1965;121:289–308.

[120] Leitner WW, Huang LN, deVeir MJ, Zhou A, Silverman RH, Williams BT, et al. Alphavirus-based DNA vaccine breaks immunological tolerance by activating innate antiviral pathways. Nat Med 2003;9:33–9.

[121] Berd D. M-Vax: an autologous, hapten-modified vaccine for human cancer. Expert Rev Vaccines 2004;3:521–7.

[122] Grunewald J, Tsao ML, Perera R, Dong L, Niessen F, Wen BG, et al. Immunochemical termination of self-tolerance. Proc Natl Acad Sci USA 2008;105:11276–80.

[123] Ren W, Ji A, Wang MX, Al HW. Expanding the genetic code for a dinofuran-hapten. ChemBioChem 2015;16:2007–10.

[124] Gamba V, Grunewald J, Gorney V, Deiters A. Optical control of receptor (GLP-1R). J Biol Chem 2017;292:7131–7.

[125] Coin I, Katritch V, Sun T, Xiang Z, Siu FY, Beyermann M, et al. Genetically encoded protein-protein interface with a noncanonical amino acid. J Am Chem Soc 2017;139:1261–7.

[126] Bryson DI, Fan C, Guo LT, Miller C, Soll D, Liu DR. Continuous directed evolution of original and derivative Methanosarcina mazei pyrrolysyl-tRNA synthetase-tRNA expression and engineered eRF1. J Am Chem Soc 2014;136:15577–83.

[127] Johnson DBF, Xu JF, Shao N, Singh NS, Slade SE, Jones AM, et al. A strategy for creating organisms dependent on noncanonical amino acid incorporation in Mycobacterium tuberculosis. PLoS One 2010;5:e9354.

[128] Johnson DBF, Xu JF, Zhou A, Silverman RH, Williams BR, et al. Addicting unnatural amino acids to the genetic code of mammalian cells. Biochemistry 2009;48:2643–53.

[129] Palzer S, Bantel Y, Kazenwadel F, Berg M, Rupp S, Sohn K. An expanded genetic code in Candida albicans to study protein-protein interactions in vivo. Eukaryot Cell 2013;12:816–27.

[130] Shao N, Singh NS, Slade SE, Jones AM, Balasubramanian MK. Site specific genetic incorporation of azidophenylalanine in Schizosaccharomyces pombe. Sci Rep 2015;5:17196.

[131] Parrish AR, She XY, Xiang Z, Coin I, Shao N, Briggs SP, et al. Expanding the genetic code of Caenorhabditis elegans using bacterial aminoacyl-tRNA synthetase/tRNA pairs. ACS Chem Biol 2012;7:1292–302.

[132] Chang H, Han M, Huang W, Wei G, Chen J, Chen PR, et al. Light-induced protein translocation by genetically encoded unnatural amino acid in Caenorhabditis elegans. Protein Cell 2013;4:883–6.

[133] Muki A, Wakiyama M, Sakamoto K, Yokoyama S. Genetic encoding of non-natural amino acids in Drosophila melanogaster Schneider 2 cells. Protein Sci 2010;19:440–6.

[134] Chen YT, Ma J, Lu WQ, Tian ML, Thauvin M, Yuan CG, et al. Heritable expansion of the genetic code in mouse and zebrafish. Cell Res 2017;27:294–7.

[135] Kang YJ, Kawaguchi D, Coin I, Xiang Z, O’Leary DDM, Slesinger PA, et al. In vivo expression of a light-activatable potassium channel using unnatural amino acids. Nature 2013;8:357–80.

[136] 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.

[137] Shen B, Xiang Z, Miller B, Louie G, Wang WY, Noel JP, et al. Genetically encoding unnatural amino acids in neural stem cells and optically reporting voltage-sensing domain changes in differentiated neurons. Stem Cell 2011;29:1231–40.

[138] Ellasser SJ, Ernst RJ, Walker OS, Chin JW. Genetic code expansion in stable cell lines enables encoded chromatin modification. Nat Methods 2016;13:158–64.

[139] Schmid W, Ellasser SJ, Uttamapinant C, Chin JW. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrroaryl tRNA synthetase/tRNA expression and engineered eRF1. J Am Chem Soc 2014;136:15577–83.

[140] Johnson DBF, Xu JF, Shao N, Takimoto JK, Schultz MD, Schmitz RJ, et al. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 2011;7:779–86.

[141] Johnson DBF, Wang C, Xu JF, Schultz MD, Schmitz RJ, Ecker JR, et al. Release factor one is nonessential in Escherichia coli. ACS Chem Biol 2012;7:1337–44.

[142] Schrader JM, Chapman SJ, Uhlenbeck OC. Tuning the affinity of aminocyl-tRNA to elongation factor Tu for optimal decoding. Proc Natl Acad Sci USA 2011;108:5215–20.

[143] Wang K, Neumann H, Peak-Chew SY, Chin JW. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol 2017;35:275–7.

[144] Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, et al. Genomically encoded organisms expand biological functions. Science 2015;349:341–45.

[145] Ostrov N, Landon M, Guehl M, Kuznetsov G, Teramoto J, Cervantes N, et al. Design, synthesis, and testing toward a 57-codon genome. Science 2016;353:819–22.

[146] Zheng Y, Lewis JR, Igo P, Polleux F, Chatterjee A. Virus-enabled optimization of mammalian cells with noncanonical amino acids. Angew Chem Int Ed Engl 2017;56:9170–3.

[147] Mandell DJ, Lajoie MJ, Mei MT, Takeuchi R, Kuznetsov G, Norville JE, et al. Biocontainment of genetically modified organisms by synthetic protein design. Nat Methods 2015;12:55–60.

[148] Ray-Saha S, Huber T, Sakmar TP. Antibody epitopes on gp105 expressed in plants by genetically encoded photoactivatable cross-linkers. Biochemistry 2017;53:1302–10.

[149] Coint I, Katrich V, Sun T, Xiang Z, Siu FY, Beyerman M, et al. Genetically encoded chemical probes in cells reveal the binding path of urocortin 1 to CRF class B GPCRs. Cell 2013;155:1258–69.

[150] Rannarsson H, Andersen J, Sorensen L, Bang-Andersen B, Park M, Huber T, et al. Genetically encoded photocrosslinkers locate the high-affinity binding site of an antidepressant drug in the human serotonin transporter. Nat Commun 2016;7.

[151] Koo CE, Reynolds CA, Mobarec JC, Hick C, Sexton PM, Sakmar TP. Genetically encoded photocross-linkers determine the biological binding site of exondin-4 peptide in the N-terminal domain of the intact human glucagon-like peptide-1 receptor (GLP-1R). J Biol Chem 2017;292:711–44.

[152] Hennpl J, Borcherdt ER, Brown K, Asokan A, Deiters A. Optical control of CRISPRi/Cas9 gene editing. J Am Chem Soc 2015;137:5642–5.

[153] Shen B, Schultz PG, Guo J. Functional replacement of histidine in proteins to generate noncanonical amino acid-dependent organisms. J Am Chem Soc 2018;140:3829–32.

[154] Suzuki T, Miller C, Guo LT, Ho JML, Bryson DI, Wang YS, et al. Crystal structures reveal an elusive functional domain of pyrroaryl-tRNA synthetase. Nat Chem Biol 2017;13:1261–6.

[155] Willis JCW, Chin JW. Mutually orthogonal pyrroaryl-tRNA synthetase/tRNA pairs. Nat Chem Biol 2018;10:831–7.

[156] Wang W, Huang Y, Wang ZY, Russell WK, Pai JS, Russell DH, et al. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. Angew Chem Int Ed Engl 2010;49:3211–4.

[157] Zheng Y, Addy PS, Mukherjee R, Chatterjee A. Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells. Chem Sci 2017;8:7211–7.

[158] Wang K, Fredens J, Brunner SF, Kim SH, Chia T, Chin JW. Defining synonymous codon compression schemes by genome reordering. Nature 2016;539:59–64.

[159] Zhang Y, Ptacin JL, Fischer EC, Aerni HR, Caffaro GE, Jose KS, et al. A semi-synthetic organism that stores and retrieves increased genetic information. Nature 2017;551:644–7.

[160] Zhang Y, Lamb BM, Feldman AW, Zhou AX, Lavergne T, Li L, et al. A semi-synthetic organism engineered for the stable expansion of the genetic alphabet. Proc Natl Acad Sci U S A 2017;114:1317–22.

Synthetic and Systems Biotechnology 3 (2018) 150–158