Preprints.org

Review

Advances in Biosynthesis of Non-canonical Amino Acids (ncAAs) and the Methods of ncAAs Incorporation into Proteins

Liang Chen [*] , Xiulan Xin , Yuning Zhang , Shunyao Li , Xiaoman Zhao , Song Li , Zhaochu Xu

Review

Advances in Biosynthesis of Non-Canonical Amino Acids (ncAAs) and the Methods of ncAAs Incorporation into Proteins

Liang Chen *, Xiulan Xin, Yuning Zhang, Shunyao Li, Xiaoman Zhao, Song Li and Zhaochu Xu

College of Bioengineering, Beijing Polytechnic, Beijing 100176, China

* Correspondence: chenliang@bpi.edu.cn; Tel.: +00-86-010-87226006

Abstract: The emergence of non-canonical amino acids (ncAAs) has enriched the functional pool inherent in canonical amino acids (cAAs). ncAAs are important building blocks in the production of various pharmaceuticals. The biosynthesis of ncAAs has emerged as an alternative to traditional chemical synthesis because of its environmental friendliness and high efficiency. The breakthrough genetic code expansion (GCE) technique developed in recent years has allowed for the incorporation of ncAAs into target proteins, giving them special functions and biological activities. Biosynthesis of ncAAs and their incorporation into target proteins within a single microbe has become an enticing application of such molecules. Based on that, in this study we first review the biosynthesis methods for ncAAs and analyze the difficulties in biosynthesis. We then summarize the GCE methods and analyze their advantages and disadvantages. Further, we review the application progress of ncAAs and anticipate the challenges and future development directions of ncAAs.

Keywords: non-canonical amino acid; genetic code expansion; biosynthesis; incorporation; tailor-made protein

1. Introduction

Proteins are essential to life processes, playing important roles in transport, catalysis, and regulation. Most proteins are composed of only 20 canonical amino acids (cAAs). While the encoding of cAAs is sufficient for fundamental growth and metabolic functions, proteins require additional chemical group such as phosphate, methyl, hydroxyl, and acetyl group to perform more complicated and diverse biological functions[1–3]. The requirement for these extra functional groups demonstrates that the 20 cAAs severely restrict the types and functional applications of proteins, which are no longer sufficient to meet the research needs of fields such as biological science, chemistry, and medicine[4,5].

Non-canonical amino acids (ncAAs) are derivatives of cAAs, also known as non-standard amino acids (nsAAs), unnatural amino acids (unAAs or uAAs) or non-proteinogenic amino acids (npAAs). Some examples of ncAAs include L-homoserine, which serves as a precursor for the synthesis of essential amino acids and other chemical products[6,7]. 5-Aminolevulinic acid is an important precursor for the biosynthesis of heme, porphyrin, chlorophyll, and vitamin B12, which are biodegradable, non-toxic, and residue-free[8]. NcAAs contain various functional groups such as ketyl, alkynyl, azide, nitro, phosphate, sulfonate, among others, which allow them to modify proteins through various reactions. So far, more than 200 ncAAs can be inserted proteins by using genetic code expansion (GCE) technology[9,10], which offers new opportunities for theoretical research and applications for protein structures and functions.

Despite the growing diversity and range of applications of ncAAs, their synthesis remains challenging[11–13]. The traditional chemical synthesis methods are limited by harsh reaction conditions[14], highly volatile toxic substances[15,16], environmental pollution and high raw material costs[17]. Metabolic engineering is an emerging technique that could potentially solve these issues and enable green and efficient production of ncAAs. Based on that, this study reviews the methods, applications, difficulties, and solutions of metabolic engineering biosynthesis of ncAAs.

The methods for incorporating ncAAs into proteins by using GCE methods are summarized as well. The potential applications of tailor-made proteins are also discussed and the future applications are prospected.

2. Biosynthesis of ncAAs

As mentioned above, the chemical synthesis of ncAAs has significant drawbacks. In addition, ncAAs have high technical barriers to synthesis due to the limitations of key technologies such as the screening and preparation of chemical catalysts, the construction of process routes, and the regulation of catalytic processes[18]. Development of a green and efficient synthesis method for ncAAs is crucial. Metabolic engineering offers promising solutions for synthesizing ncAAs by understanding the catalytic mechanisms of related enzymes. The key heterologous enzymes can be recombined, modified, and optimized in the engineered microbes. These create the possibility of establishing a production platform for ncAAs. The usage of metabolic engineering offers a promising path toward creating a sustainable and cost-effective method for producing ncAAs.

2.1. 5-Hydroxytryptophan

5-Hydroxytryptophan (5-HTP) is an important compound with medicinal value used to treat depression, insomnia, and other diseases. Wang *et al.* have successfully constructed a recombinant strain to biosynthesize 5-HTP[19]. The biosynthesis pathway was constructed on two plasmids containing three functional modules, substrate L-Trp biosynthesis modules, hydroxylation module, and cofactor regeneration module (Figure 1). Besides, the human tryptophan hydroxylase I (TPH I) was introduced into *E. coli* BL21Δ*tnaA* strain to hydroxylate L-Trp to produce 5-HTP. As a result, the engineered strain produced 5.1 g/L 5-HTP through fed-batch fermentation. By further inserting the tryptophan synthesis pathway into the genome, the yield of 5-HTP in shake-flask fermentation was increased to 1.61 g/L while reducing the accumulation of precursor L-Trp, which was beneficial for the subsequent separation and purification of 5-HTP[20]. After that, Lin *et.al* first engineered the phenylalanine-4-hydroxylase from *Xanthomonas campestris* (*Xc*P4H) and introduced into a L-Trp-producing *E. coli* strain with co-factor regeneration pathway. The engineered strain produced 1.2 g/L 5-HTP [21]. Mora-Villalobos utilized sequence analysis, phylogenetic analysis and functional differential analysis tools to predict, screen and design the specific mutations of substrate-specific sites of aromatic amino acid hydroxylase from *Cupriavidus taiwanensis* (*Ct*AAAH). The substrate preference of the *Ct*AAAH was transferred from L-Phe to L-Trp, enabling the generation of 5-HTP with L-Trp as substrate[22,23]. These studies show that microbial-based metabolic engineering has achieved the green, efficient, and low-cost production of 5-HTP.
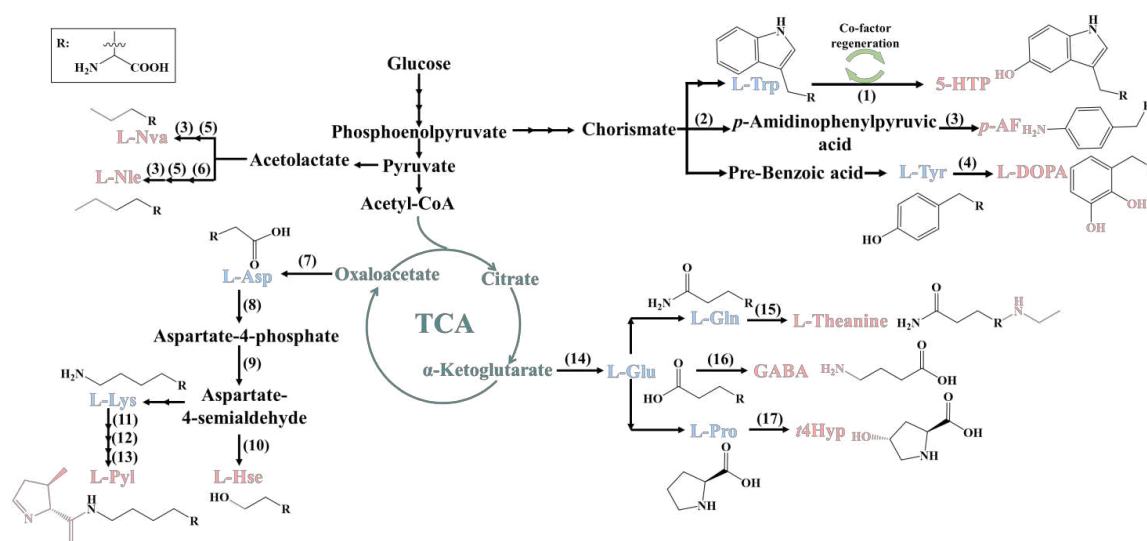

Figure 1. Biosynthetic pathways of some ncAAs derived from glucose. The blue color indicates the cAAs. The pink color indicates the ncAAs. Some key enzymes in the pathways: (1) Tryptophan

hydroxylase or aromatic amino acid hydroxylase; (2) PapA, PapB and PapC; (3) Branched-chain amino acid transaminase; (4) Tyrosine phenol-lyase; (5) LeuABCD; (6) IlvCD; (7) Aspartate amino transferase; (8) Aspartokinase; (9) Aspartate-semialdehyde dehydrogenase; (10) Homoserine dehydrogenase; (11) Radical SAM enzyme PylB; (12) ATP-dependent PylC; (13) PylD for oxidation; (14) Glutamate dehydrogenase; (15) Theanine synthetase; (16) Glutamate decarboxylase; (17) Proline-4-hydroxylase (P4H). Abbreviation: 5-HTP: 5-Hydroxy tryptophan; *p*-AF: *p*-amino-phenylalanine; L-DOPA: Levodopa (3,4-dihydroxy-L-phenylalanine); L-Nva: L-norvaline; L-Nle: L-norleucine; L-Pyl: L-pyrrolysine; L-Hse: L-homoserine; GABA: Gamma-aminobutyric acid; *t*4Hyp: *Trans*-4-hydroxyproline.

2.2. L-Homoserine

L-Homoserine (L-Hse), also known as 2-amino-4-hydroxybutyric acid, is a valuable platform chemical that has been widely used in various fields such as medicine, agriculture, cosmetics, and spices. The microbial fermentation method has great potential for large-scale production of L-Hse. Recent studies have focused on using *E. coli* and *Corynebacterium glutamate* (*C. glutamate*) to achieve the high-level production of L-Hse[24–27]. For example, Cai *et al.* enhanced the production of L-Hse by using a non-auxotrophic deficient and plasmid-free *E. coli* chassis[28]. They first constructed a *E. coli* chassis host strain via knock-down of the L-Hse degradation pathway[29]. Then, they optimized the metabolic flux of L-Hse biosynthesis by overexpressing the *ppc*, *aspC*, *aspA* , *thrA*fbr, and *lysC*fbrcgl (Figure 1). Additionally, they promoted L-Hse efflux by modifying the transport system and introduced a strategy of synergistic utilization of co-factors to promote the regeneration of NADPH and coordinated the level of redox co-factors by incorporating a heterologous dehydrogenase. As a result, the engineered strain was able to produce 85.29 g/L of L-Hse in 5-L fermenter, which was the highest titer of the plasmid-free and non-auxotrophic strains reported to date. This study demonstrates the effectiveness of optimizing L-Hse production using metabolic engineering strategies and the potential for microbial fermentation as a primary method for producing valuable platform chemicals. Although *E. coli* as an amino acid-producing chassis has achieved high-level production of L-Hse, large amounts of by-products such as acetate are also produced during the fermentation process[24,30]. To address this issue, *C. glutamate* which is known for its ability to synthesize useful compounds using cheap feedstock, was used to produce L-Hse. Through overexpressing key kinase genes, disrupting competing and degrading pathways and promoting the synthetic flux, the engineered *C. glutamate* produced 8.8 g/L L-Hse[26].

2.3 Trans-4-hydroxyproline

Trans-4-hydroxyproline (*t*4Hyp) as a value-added amino acid, has been widely used in medicine, food and cosmetics, especially in the field of chiral synthetic material. *t*4Hyp is Traditionally produced by the acidic hydrolysis of collagen, but the process has some drawbacks such as low productivity and complicated process. Metabolic engineering has been used to efficiently construct microbial cell factories of *E. coli* or *C. glutamate* in order to biosynthesize *t*4Hyp[31–33]. Introduction of a heterologous proline 4-hydroxylase from *Alteromonas mediterranea* (*Al*P₄H) into *E. coli* enabled an accumulation of 45.83 g/L *t*4Hyp within 36 hours in a 5-L fermenter without the addition of proline[34]. Knockout of the genes of *putA*, *proP*, *putP* and *aceA* in competing pathway and mutation of ProB to D107N/E143A ProB in order to alleviate the feedback inhibition of L-Pro maximized the production of L-Pro, so as to enhance the biosynthesis of *t*4Hyp. Subsequently, Rao *et.al.* optimized the enzyme activity of L-Pro hydroxylase by using genome mining technology and rational design. After systematic modification, an engineered strain can use glycerol and glucose as carbon sources to produce 54.8 g/L *t*4Hyp in 60 hours[33]. The microbial metabolic network is large and complex. Genome modification methods such as gene knockout may lead to slow cell growth, stagnation or even death, which may not be suitable for blocking some competing pathways. So far, CRISPR interference (CRISPRi) that can reduce the transcription of target genes by up to 1000-fold inhibition without miss effect[35,36] has been appeared as an alternative to down-regulate the

expression of enzymes, which may be employed to repress the expression of the *putA* gene to further increase *t*4Hyp production in the future.

2.4 Other ncAAs

L-Pyrrolysine (L-Pyl) is the 22nd amino acid that has been discovered so far to insert into proteins[37]. Krzycki *et al.* reported that L-Lys is the only precursor of L-Pyl. By providing isotopically labeled L-Lys to methanogenic *Archaea* with the *pylTSBCD* gene cluster, methylamine methyltransferase with L-Pyl incorporation was obtained by mass spectrometry analysis and purification. Further, the biosynthetic process of the converting two L-Lys molecules into one L-Pyl molecule were revealed[38]. The *pylBCD* genes are for L-Pyl synthesis with tRNA-independent[39], while the *pylT* gene can produce tRNA$_{CUA}$ (also called tRNA^{Pyl}), and the *pylS* gene can encode pyrrolysyl-tRNA synthase[40]. Further, introduction of *pylTSBCD* genes into *E. coli* can enable the incorporation of endogenously biosynthesized pyrrolysine into proteins. The L-Pyl production capacity of *E. coli* was improved by Ho *et al.* via rational engineering and directed evolution of the whole biosynthetic pathway of L-Pyl. They also developed Alternating Phage Assisted Non-Continus Evolution (Alt-PANCE), alternating mutagenesis and selective phage growth, to accommodate the toxicity of L-Pyl biosynthetic genes[41]. The evolutionary pathway enabled a 32-fold increase in pyl-incorporating protein yield compared to rationally modified pathway. The evolved PylB mutant had a 4.5-fold increase in intracellular levels and a 2.2-fold increase in protease resistance.

Gamma-aminobutyric acid (GABA) which has high nutritional value, is produced by lactic acid bacteria[42–44]. The GABA production capacity of *Lactiplantibacillus plantarum* was improved by changing crucial fermentation parameters. Optimization of the inoculum percentage, initial pH, inorganic ions and nutrients concentration significantly improved the production of GABA[45].

Selenium is an essential micronutrient that can be incorporated into the active site of specific selenocysteine proteins in the organism through the form of selenocysteine. Selenium-containing proteins play an important role in the regulation of organisms and can be used as research targets for the treatment of some diseases. Normally, biosynthetic ncAAs are formed in the cytoplasmic matrix, which is then linked by aaRS to the corresponding tRNA, thereby completing the incorporation into protein. The biosynthetic pathway of selenocysteine is different from that of ordinary ncAAs. Selenocysteine has a homologous tRNA^{Sec}, but there is no free selenocysteine in the cytoplasmic matrix and no corresponding selenocysteinyl-tRNA synthase. The synthesis of selenocysteine does not begin with the ligation of selenocysteine to homologous tRNA^{Sec}, but rather the seryl-tRNA synthase first attaches L-Ser to non-homologous tRNA^{Sec} to form seryl-tRNA^{Sec}. In bacteria, selenocysteine synthase (SelA) acts directly on seryl-tRNA^{Sec} and removes hydroxyl group from seryl group to generate an intermediate. The intermediate then receives the activated selenophosphate to eventually form selenocysteinyl-tRNA^{Sec}. Subsequently, selenocysteinyl-tRNA^{Sec} is paired with UGA codon to complete the incorporation of selenocysteine into the protein[46–48].

Enome-scale Models (GEMs) of metabolism as a new technology is composed of the full inventory of metabolic reactions encoded by the genome of an organism, and has been used to achieve the efficient production of ncAAs[49]. GEMs can explore trade-offs between growth rate and production, while computer simulations can be used to analyze metabolic pathways and identify strategies for improving production. For a specific example, the *papBAC* gene cluster from *Pseudomonas fluorescens* was introduced into *E. coli* strain EcNR2 to produce *p*-amino-phenylalanine (*p*AF), but there was a trade-off between *p*AF production and growth rate[50]. To increase *p*AF production, a GEM of *E. coli* metabolism with computer design was used to identify metabolic pathways and determine the recombinant strain metabolism. Upregulating the metabolic flux in the chorismate biosynthetic pathway by eliminating feedback inhibition was the most effective strategy for increasing *p*AF production[51]. This study demonstrates the power of GEMs and computational analysis for optimizing metabolic pathways and improving the production of valuable compounds.

The construction of efficient microbial cell factories for ncAAs has become popular in recent years. These cell factories are mainly created by reconstructing synthesis pathways, designing and modifying key enzymes, coordinating precursor regulation, knocking out competing bypass

pathways, constructing cofactor regeneration systems, and intelligently regulating the fermentation process. So far, only a few biosynthetic pathways of ncAAs have been confirmed[11]. Advancements in synthetic and computational biology technologies, as well as multidisciplinary collaborations, have begun to shed light on ncAAs biosynthesis, both in the laboratory and industry. In the future, the precise design of ncAAs synthesis pathways may be accomplished with advanced bioinformatics or biosynthesis simulation tools. Additionally, more chassis with high tolerance to specific ncAAs must be engineered or screened to achieve compatibility between microbes and heterologous pathways, target ncAAs.

3. Methods for ncAAs incorporation into proteins

3.1 GCE methods for ncAAs incorporation into proteins

The biosynthesis of ncAAs and their incorporation into target proteins within a single microbe has become an attractive application of such molecules. This approach allows for the site-specific labeling or modification of proteins with ncAAs that have unique chemical properties, which can be used to improve the proteins properties and interactions. Recent advancements in genetic engineering techniques have enabled the production of the enzymes with incorporation of ncAAs by replacing cAAs with non-canonical ones at specific sites. In addition, optimization of the efficiency and selectivity of orthogonal reactions can enable the selective labeling of tailor-made proteins with fluorophores or other probes. The ability to produce and incorporate ncAAs into proteins within the same microbe provides significant advantages over traditional incorporation method based on chemical synthesis, enzyme conjugation and *in vitro* translation, as these methods are time-consuming and expensive, and often require post-translational modifications (PTMs) to proteins. The biosynthesis of ncAAs and the incorporation into target proteins within the same microbe relies on a technology of genetic code extension (GCE).

The process of GCE for integrating ncAAs into proteins requires a rigorous and efficient translation mechanism that resembles the natural translation process. For GCE to work, each amino acid is matched with its corresponding codon and an orthogonal translation system (OTS). This OTS contains exogenous tRNA and exogenous aminoacyl-tRNA synthetases (aaRS) that adhere to strict orthogonality criteria. There are two major routes to expand the amino acid repertoire: chemical aminoacylation of tRNA[52,53] and enzyme-mediated (aaRS) aminoacylation of tRNA, which can generate an aminoacyl-tRNA (aa-tRNA) that is active for protein synthesis on the ribosome. The aminoacylated tRNA carries the ncAA along with the reassigned and desired codons, thereby facilitating the incorporation of the ncAA into the extended peptide chain (Figure 2). It is important that orthogonal system do not cross-react with the endogenous translation system, as this will lead to confusion in the translation system. To prevent false cross-reactions, each orthogonal translation system must go through multiple rounds of positive and negative screening. The specific codons used in GCE are typically obtained from 'redundant' codons, reassigned stop codons, sense codons, quadruplet codons, and others.

3.1.1. GCE based on stop codon suppression (SCS)

In the natural translation system, there are three types of stop codons: UAG (amber codon), UAA (ochre codon), and UGA (opal codon). These codons do not encode any amino acids but mediate the termination of the translation process. Release factors (RFs) are proteins that recognize and bind to these stop codons, leading to the release of the synthesized protein. Interestingly, there is some redundancy in the genetic code when it comes to stop codons. For example, RF1 can recognize both UAA and UAG, while RF2 can recognize both UAA and UGA. This means that there is at least one 'redundant' codon that can potentially be repurposed to encode a different amino acid. In *E. coli*, Schultz and co-workers chemically acylated a suppressor tRNA to encode a ncAA in response to the 'sense codons' position which was replaced by a stop codon UAG[54]. This method known as stop codon suppression (SCS) takes advantage of the degeneracy of the three stop codons and the low abundance of the stop codon UAG (Figure 2, XYZ = Stop codon). As an example, Pastore *et al.* used

GCE based on SCS to replace a single tryptophan residue in a copper protein amicyanin with ncAA 5-HTP. The incorporation of 5-HTP changed the fluorescence emission maximum of amicyanin from 318 to 331 nm. Besides, the fluorescence quantum yield of 5-HTP-containing amicyanin was much less than that of native amicyanin. The incorporation of ncAAs significantly alters protein properties[55].

Both RFs and suppressor tRNAs have the ability to recognize stop codons. However, competitive recognition between these factors can result in premature termination of translation, leading to protein truncation and reducing the incorporation efficiency of ncAAs, and protein yield and purity. To address this issue, Lajoie *et al.* utilized multiplex automated genome engineering (MAGE) and conjugative assembly genome engineering (CAGE) techniques to replace all 321 UAG codons with synonymous UAA codons in the *E. coli* genome[56]. Additionally, they deleted the *prfA* gene which encodes RF1 without affecting the strain growth. A genomically recoded organism (GRO) called *E. coli* C321.ΔA was created, which no longer recognized UAG as a stop codon. This strain has higher efficiency for incorporation of ncAAs as compared to wild-type strain[57,58]. So far, only genome-recoded *E. coli* was obtained and used as chassis. It is crucial to develop other chassis with modified genomes to expand the range of applicable hosts for incorporating ncAAs into proteins based on this principle. Overall, SCS allows for the introduction of ncAAs into specific positions in a protein, expanding the possibilities for protein engineering and modification.

3.1.2. GCE based on synonymous codon compression

Eighteen of the 20 cAAs have two or more codons. L-Leu, L-Arg and L-Ser are encoded by up to six synonymous codons, while only L-Met and L-Trp are encoded by one codon. Hence, utilizing the degeneracy of codons to replace synonymous codons with a different codon throughout the genome-wide scale and deleting the tRNAs that decode them, can release the replaced codons from the standard genetic code and reassign the replaced codons to ncAAs. This method is called synonymous codon compression[59]. By reducing the number of synonymous codons encoding cAAs, the recoded codons can be freed up for the incorporation of ncAAs or other non-amino acid molecules (Figure 2, XYZ = Sense codon). As an example, Fredens *et al.* replaced three L-Ser codons (UCG, UCA) and a stop codon (UAG) with their synonymous codons (AGC, AGU, UAA) and deleted the corresponding tRNA and the RF1 to construct an *E. coli* strain called Syn61[60]. This allowed the incorporation of three ncAAs into a single protein in the engineered strain[61]. However, the reassignment of sense codons at the genome level is still a major technical challenge. Completely eliminating a large number of codons from the genome is not a straightforward task, and the replacement of synonymous codons can have various effects on gene expression, cell fitness[62], translation speed[63,64], and cellular homeostasis. Additionally, the current approach is primarily limited to bacterial hosts and not readily applicable to more complex eukaryotic systems.

3.1.3. GCE based on other approaches

Another method for incorporating ncAAs into proteins is to use quadruplet or quintuplet codons, which are called 'frameshift suppression'. Unlike the redistribution or reassignment of codons, this method allows for the expansion of the genetic code without extensive modifications to the genome. 'Frameshift suppression' provides the potential to expand from the 64 codons of the natural genetic code to $4^4 = 256$[65] or $4^5 = 1024$[66] 'blanking' codons. The use of quadruplet codons offers more flexibility, as it allows for the incorporation of ncAAs without the need for large-scale genome modifications. This method also helps to reduce potential cross-decoding effects of endogenous triplet codons[67–69]. Quadruplet codons have been primarily used in single-celled systems such as the nematode *Caenorhabditis elegans*. The application range is currently limited and the efficiency is currently low. Efforts to improve the efficiency of quadruplet codon decoding have focused on directed evolution techniques, including the evolution of anticodon loops of tRNAs and the evolution of the ribosome decoding center in the chassis[70]. The use of quadruplet codons is an interesting and promising method for incorporating ncAAs into proteins. Further work needs to concentrate on expand its applicability and improve decoding efficiency.

Besides, unnatural base pairs (UBPs) provide another alternative for expanding the genetic codon library. For example, Bain *et al.* incorporated the ncAA L-iodotyrosine into proteins using an UBPs (*iso*-C)AG-tRNA$_{CU (iso-dG)}$. The (*iso*-C)AG was called the 65th codon[71]. After that, several UBPs were also developed[72–74]. Malyshev created a synthetic base pair called X-Y, where X represents dNaM and Y represents d5SICS[72]. UBPs involve the introduction of new codon-anticodon interactions, enabling the site-specific incorporation of ncAAs into proteins through ribosomal-mediated translation. In a six-letter unnatural bases system, it is theoretically possible to generate 152 new codons for ncAAs[75]. The use of UBPs allows for the expansion of the genetic code in a novel form, providing additional opportunities for the incorporation of ncAAs into proteins. This technology holds promise for developing new functions and properties of proteins.
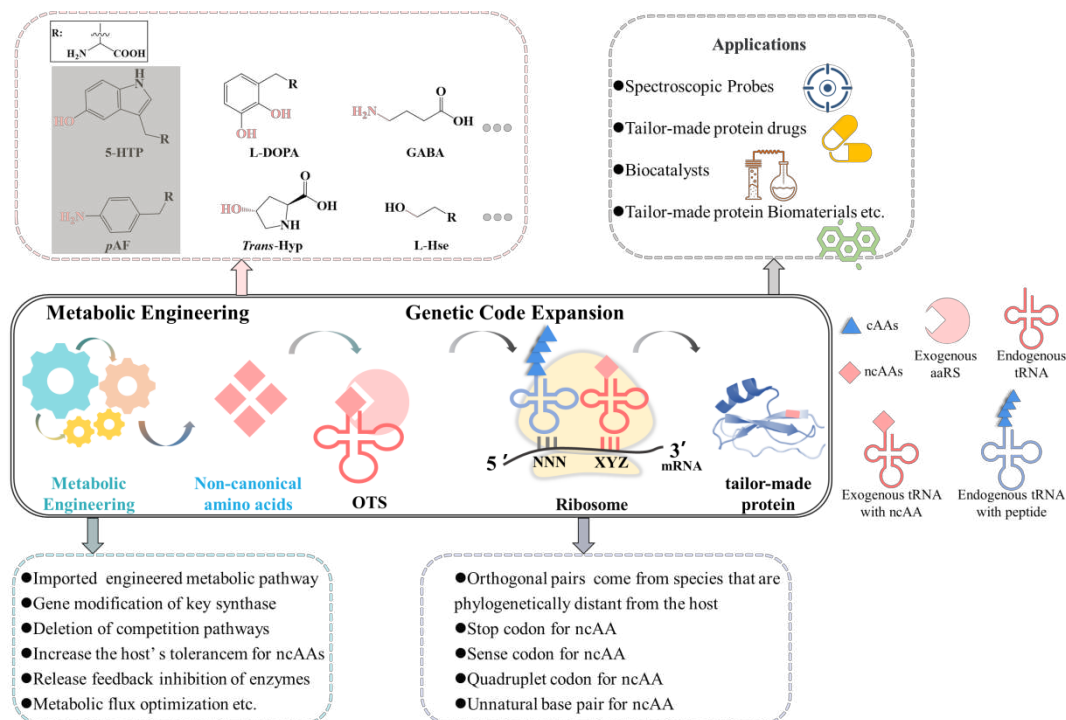

Figure 2. A diagram of the combination of metabolic engineering and GCE. The diagram includes the metabolic biosynthesis of ncAAs, incorporation into proteins by GCE and applications of tailor-made proteins. 5-HTP[5] and *p*AF[76] has been biosynthesized *in vivo*, incorporated into proteins and implemented for applications.

3.2. Selective pressure incorporation (SPI)

Selective pressure incorporation (SPI), also known as 'residue-specific incorporation', is a method that relies on auxotrophic hosts which are deficient in one or more specific cAAs[77]. This approach is based on the substrate tolerance of cellular systems[78]. Endogenous aaRS play a crucial role in replacing the cAAs in protein sequences with one or more amino acid analogs that have similar structure and chemistry[79]. The auxotrophic hosts are cultured in a medium that initially contains a limited amount of nAAs. As the cells grow and begin to express recombinant proteins, the cAAs are gradually depleted. At an appropriate growth state, exogenous ncAA analogs are supplemented in the medium. The expression of target protein gene is regulated by inducible promoter, ensuring that protein translation is dependent on the availability of ncAAs in the medium[80] (Figure 3). This method is commonly used to modify ribosomal synthetic and post-translationally modified peptides by incorporating ncAAs. Through adding ncAAs in the medium, unique functional polypeptides can be produced. SPI enables the specific modification of residues to generate novel functions and chemical diversity in peptides or proteins[81–83].

SPI allows for the incorporation of multiple ncAAs throughout the proteome. This can lead to a cumulative or synergistic effect of the ncAAs, resulting in more significant changes in the properties

of the target proteins compared to a single substitution. However, SPI is more suitable for ncAAs that have similar properties to cAAs. This limitation restricts the application of SPI for ncAAs with distinct or unique properties. Additionally, SPI does not guarantee the specific incorporation of a single site, which is not conducive to study fine modification of protein sites or precise control of protein modification. Despite these limitations, SPI remains a valuable tool for introducing ncAAs into proteins and exploring their effects on protein structure and function. It offers a versatile approach for creating tailor-made proteins with expanded chemical diversity and potential applications in various fields.
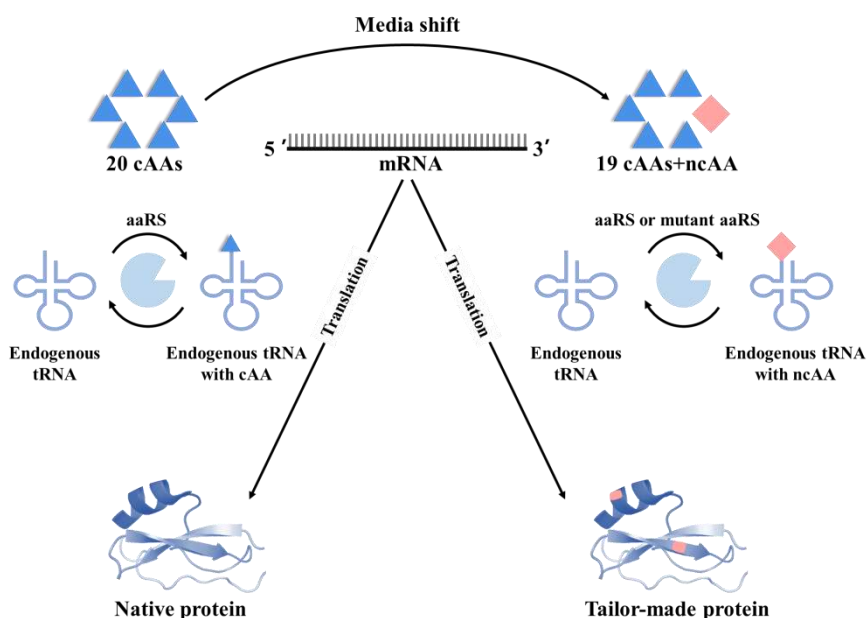

Figure 3. SPI. The incorporation of ncAAs into proteins is realized by an endogenous orthogonal system. Exogenous ncAA analogs are supplemented in the medium to induce the protein translation based on the availability of ncAAs with assistance of the wild-type or mutated aaRS.

3.3. Solid-phase peptide synthesis (SPPS)

In vitro methods can be used to incorporate ncAAs into proteins, which is advantageous when an endogenous expression system for tailor-made proteins is unavailable or impractical. One commonly used method for *in vitro* protein synthesis is solid-phase peptide synthesis (SPPS) (Figure 4). SPPS involves the stepwise assembly of peptide chains attached to an insoluble resin support. Multiple amino acids are linked together through peptide bonds to produce peptides. In SPPS, the C-terminal of the first amino acid which is protected by an amino-protecting group, is connected to the resin via a covalent bond. The protective group is then removed to expose a free amino group, which can be coupled with the next amino acid. This process is repeated, extending the length of the peptide chain. The general principle of SPPS involves a repeated cycle of coupling-flushing-deprotection-flushing-coupling. The polypeptide remains immobilized on the solid phase throughout the synthesis. After assembly, the polypeptide is separated from the resin using trifluoroacetic acid, and the protective group is also removed. Commonly used amino-protecting groups include 9-fluorenylmethoxycarbonyl (Fmoc) and *t*-butyloxycarbonyl (*t*Boc)[84,85]. SPPS has practical applications in the chemical synthesis of recombinant proteins using ncAAs[84,86,87]. It can be used to synthesize the peptides that are difficult to be endogenously expressed, such as those composed of D-amino acids. SPPS is also useful for modifying the main chain of peptides or proteins. It is particularly suitable for incorporating amino acid analogs that may be toxic to cells or incompatible with cellular translation mechanisms. The chemical synthesis of peptides is limited by the length of peptide (＜50~60 amino acids)[88,89] and the speed of synthesis. Overall, SPPS is a valuable tool for incorporating ncAAs into proteins *in vitro*, allowing for the synthesis of peptides

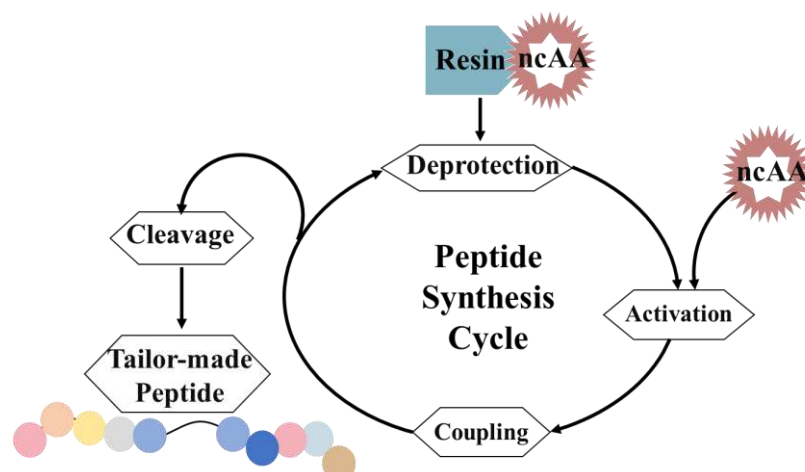and proteins that are challenging or impossible to produce using traditional endogenous expression system.

Figure 4. SPPS. In SPPS, the C-terminal of the first amino acid protected by an amino-protecting group needs to connect to the resin via a covalent bond. After that, the protective group is removed to expose a free amino group, which can be coupled with the next amino acid. The general principle of SPPS consists of a repeated cycle of coupling-flushing-deprotection-flushing-coupling.

3.4. Cell-free protein synthesis (CFPS)

Cell-free protein synthesis (CFPS) is another method to synthesize non-canonical proteins in vitro. CFPS is a rapid and high-throughput expression technique that allows the production of proteins using exogenous DNA or mRNA as a template [90,91] (Figure 5). This method requires the use of multiple enzymes provided by cell lysates to provide substrate and energy. CFPS systems are available based on different cell lysates, including *E. coli* lysate, yeast extract, rabbit reticulocyte lysate, or wheat germ extract[92–94]. Each system offers unique advantages and may be preferred depending on the specific requirements of the protein synthesis. The *in vitro* CFPS systems allow for flexibility and freedom in protein design, making it a powerful tool for protein engineering and biopharmaceutical production. CFPS has been successfully scaled up for manufacturing purpose. However, CFPS also have some limitations. For instance, the reaction duration is relatively short, which can result in low protein yields. The cost of CFPS is expensive due to the need of specialized reagents and enzymes. Despite these limitations, CFPS remains a valuable tool for the rapid synthesis of tailor-made proteins *in vitro*, offering advantages such as flexibility and high-throughput capability. Ongoing research continues to address the limitations of CFPS and improve its efficiency and scalability for broader applications.
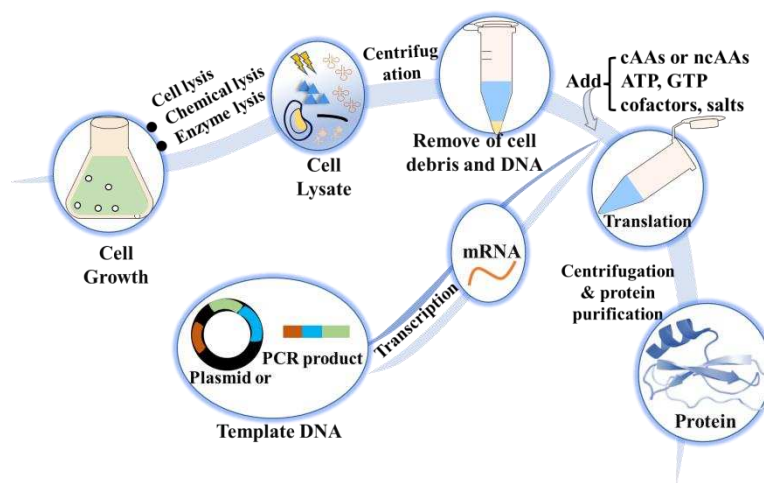
Figure 5. CFPS. Mixing exogenous DNA or mRNA, cell lysate, ncAAs, ATP, cofactors salts could achieve the synthesis of tailor-made proteins with ncAAs incorporation.

3.5 Other methods for the incorporation of ncAAs

In addition to the above methods, post-translational mutagenesis and cysteine functionalization are also the methods for incorporating ncAAs into proteins. Natural PTMs form bonds with heteroatoms (non-carbon) at the γ (CysSγ, ThrOγ, SerOγ) or ω (LysNω, TyrOω) position of the side chain, while post-translational mutagenesis uses free radical chemistry to form carbon-carbon bonds between amino acid residues and selected functional groups, which can introduce functions or labels into a wide range of proteins[95]. Unpaired free cysteine often exhibits high nucleophilicity and is relatively rare among proteins, making it an ideal site for protein modifications. Cysteine residues can be easily introduced into proteins by site-specific mutagenesis[96]. Cysteine-specific modification is usually achieved via the reaction of sulfhydryl groups with electrophilic reagents such as iodoacetamides, pyridine disulfide, maleimides, and alkyl halides[97]. Cysteine modification is a practical method to produce functional proteins with a wide range of biomedical applications.

4. The applications of tailor-made proteins

4.1. Tailor-made protein materials

Due to the overproduction of special mussel adhesion proteins, marine mussels are able to firmly adhere to various wet surfaces. One of the key amino acids involved in special mussel foot proteins is L-3,4-dihydroxyphenylalanine (DOPA) containing a catechol group[98–100]. The adhesive property of these proteins relies on post-translational hydroxylation of L-DOPA. This modification can only be completed in eukaryotic cells. Therefore, recombinant production of these proteins in prokaryotic cells remains a significant challenge. Overcoming this challenge would have important implications for the engineering and design of new adhesives and coatings with improved underwater performance[101]. One approach to address this challenge is to develop an efficient orthogonal aaRS/tRNA pairs. Using the orthogonal aaRS/tRNA pairs to incorporate the photoactivatable ncAAs (O-nitrobenzyl DOPA) into proteins can achieve the production of mussel adhesion proteins *in vivo*[102]. Incorporation of DOPA-like ncAAs into proteins using an orthogonal system offers the possibility of developing bio-adhesives and coatings that mimic the adhesive properties of marine mussels.

4.2. Tailor-made protein probes

Proteomics presents challenges in analyzing protein-protein interactions within cells. Current methods often rely on affinity purification, but it is difficult to isolate and purify the complete protein complexes from cells. This can result in incomplete protein-protein interactions, as some proteins may dissociate during cell lysis or affinity purification. Alternative methods that utilize fluorescent protein photocrosslinking probes to target proteins of interest (POIs) have been developed[103–107]. These probes have become valuable tools for studying intracellular protein-protein interactions. By incorporating fluorescent ncAAs into proteins through site-specific incorporation, microorganisms can be modified to allow for live imaging and tracking[46]. For example, Praveschotinunt *et al.* incorporated *p*-azido-L-phenylalanine (*p*AzF) into *E. coli* in response to the UAG stop codon. They engineered the encoding gene of cell surface protein (CsgA) with a UAG codon. When *p*AzF was incorporated into CsgA, the strain could be covalently labeled with a Cy5 dye using a dibenzocyclooctyl (DBCO) group through a copper-free orthogonal reaction. This labeling strategy was also validated using a murine model for *in vivo* imaging. Fluorescent protein probes have also become powerful tools for exploring protein conformational changes, localization, and molecular interactions[108,109]. These probes have fluorescent properties that offer several advantages over traditional methods, including small size, ease of use, ability to provide various colors, and improved photochemical properties. The synthesis of fluorescent probes provides numerous benefits for in-depth study of protein properties and interactions.

4.3. Tailor-made protein drugs

The development of modern drugs mainly focuses on the screening of chemical compounds, with limited emphasis on the design and modification of protein drugs. So far, the tailor-made protein drugs have emerged, and the special functional groups of ncAAs have greatly expanded the design space for protein drugs, bringing new developments to the biomedical field[110,111]. One strategy for the development of covalent protein drugs is based on proximity-enabled reactivity (PERx)[112,113]. This approach needs to introduce an aromatic fluorosulfate group into the side chain of tyrosine and lysine residues in proteins, such as human programmed cell death protein-1 (PD-1). The ncAAs FSY (fluorosulfate-L-tyrosine) and FSK (fluorosulfonyloxy-benzoyl-L-lysine) are designed based on sulfur-fluorine replacement reactions with proximate lysine, histidine, and tyrosine residues. The covalent modification of PD-1 with FSY and FSK can result in stronger anti-tumor effects compared to wild-type PD-1 in immunized mice. This strategy highlights the potential of ncAAs in developing covalent protein drugs with enhanced therapeutic efficacy. Another important development in targeted anticancer drugs is the use of antibody-drug conjugates (ADCs) [114–116]. ADCs involve the site-specific conjugation of ncAAs, providing a new approach that overcomes the limitations of conventional modifications dependent on cysteine residues[117]. This site-specific conjugation of ncAAs allows ADCs to exhibit improved pharmacokinetics, higher titers, and enhanced antigen binding characteristics. These advancements have expanded the possibilities for the development of targeted therapies for cancer. Overall, the incorporation of ncAAs into protein drugs offers new avenues for drug design and modification, enabling the development of covalent protein drugs with enhanced therapeutic effects and the creation of more effective targeted therapies such as ADCs.

4.4. Applications of ncAAs in vaccine

GCE technology offers a groundbreaking solution for the development of therapeutic and prophylactic vaccines. For example, PTMs of L-tyrosine residues in proteins lead to the formation of nitrotyrosine and sulfonyl tyrosine, which have been found to be associated with the breaking of immune tolerance and the induction of autoimmune diseases. The nitro and sulfonyl groups possess strong immunogenicity, which is crucial for breaking self-immune tolerance[118]. Besides, Grünewald *et al.* utilized GCE to mutate the 11th L-Lys or the 86th L-Tyr residue in murine tumor necrosis factor-alpha (mTNF-α) into pNO$_2$Phe [119]. These modifications introduced a new antigenic determinant in mTNF-α, effectively breaking autoimmune tolerance and generating high titers of antibodies that could recognize both the mutant and wild-type mTNF-α[120,121]. This demonstrates the potential of GCE in developing vaccines that target specific PTMs and break immune tolerance. GCE technology has also been applied in the development of preventive attenuated vaccines. For example, Zhou *et al.* incorporated the ncAA Nε-2-azidoxycarbonyl-L-lysine (NAEK) into the NP protein in influenza virus using the *Mb*PylRS/tRNA$_{CUA}$ pairs in transgenic 293T cells to construct a replication-incompetent influenza virus with a low escape rate[122]. This defective attenuated vaccine demonstrated stronger immunogenicity compared to commercial vaccines, eliciting a more robust and extensive immune response in the host immune system. These examples highlight the versatility and potential of ncAAs in vaccine development. GCE technology allows for the precise modification of proteins to generate novel antigenic determinants, enhancing immunogenicity and leading to effective vaccines. The broad application of ncAAs in vaccines holds promise for improving preventive and therapeutic strategies against various diseases.

5. Perspectives

Although NCAA shows exciting potential, there are still some challenges in its application and implementation. One challenge is the lack of clear biosynthesis pathways for some ncAAs and the limited knowledge of critical enzymes in the pathways. Enzymes are crucial for constructing microbial cell factories of ncAAs. Currently, the properties, structure, and interactions of enzymes with ncAA biosynthesis pathways or microbes are poorly understood[123]. Therefore, simulation

prediction of enzyme structures and biosynthesis pathways is needed to reveal and elucidate the metabolic pathways and key enzymes of ncAAs biosynthesis. Understanding these pathways will help combine the production and incorporation of ncAAs to generate desired proteins. Another challenge is the development of a diverse set of codon-aaRS-tRNA orthogonal systems that exhibit sufficient efficiency and reading fidelity. The orthogonal systems are crucial for expanding the application of ncAAs. The design and modification of orthogonal systems are essential for achieving precise and reliable incorporation of ncAAs into proteins. The multiple strategies of GCE have challenged the 'frozen accident hypothesis'[124] of the genetic codon and demonstrated the plasticity and evolvability of the standard genetic code. This serves new possibilities for the design of the advanced life forms and the incorporation of ncAAs.

Due to the paucity of ncAAs biosynthetic pathways[125] and the frequent incompatibilities between GCE techniques and pathways[126], only a few ncAAs[5,76,127] are currently able to achieve biosynthesis and incorporation into proteins *in vivo*. By leveraging synthetic biology and evolutionary engineering, it is anticipated that more pathways for ncAAs production can be constructed in suitable chassis in the future. This will enable the biosynthesis and incorporation of desired ncAAs into target protein within the same microbe. We can expect the discovery of more ncAAs, synthesis pathways, physiochemical properties, and structures related to tailor-made proteins. Generally, addressing the challenges associated with ncAAs will require interdisciplinary efforts involving biochemistry, molecular biology, synthetic biology, and systems biology.

References

1. Wang L. Genetically encoding new bioreactivity[J]. N Biotechnol, 2017, 38: 16-25.
2. Des Soye B J, Patel J R, Isaacs F J, et al. Repurposing the translation apparatus for synthetic biology[J]. Curr Opin Chem Biol, 2015, 28: 83-90.
3. Macek B, Forchhammer K, Hardouin J, et al. Protein post-translational modifications in bacteria[J]. Nat Rev Microbiol, 2019, 17(11): 651-664.
4. Gao W, Bu N, Lu Y. Recent advances in cell-free unnatural protein synthesis[J]. Sheng Wu Gong Cheng Xue Bao, 2018, 34(9): 1371-1385.
5. Chen Y, Tang J, Wang L, et al. Creation of Bacterial cells with 5-Hydroxytryptophan as a 21st Amino Acid Building Block[J]. Chem, 2020, 6(10): 2717-2727.
6. Huang J-F, Liu Z-Q, Jin L-Q, et al. Metabolic engineering of *Escherichia coli* for microbial production of L-methionine[J]. Biotechnol Bioeng, 2017, 114(4): 843-851.
7. Hong K-K, Kim J H, Yoon J H, et al. *O*-Succinyl-L-homoserine-based C4-chemical production: succinic acid, homoserine lactone, γ-butyrolactone, γ-butyrolactone derivatives, and 1,4-butanediol[J]. J Ind Microbiol Biotechnol, 2014, 41(10): 1517-1524.
8. Kang Z, Zhang J, Zhou J, et al. Recent advances in microbial production of δ-aminolevulinic acid and vitamin B$_{12}$[J]. Biotechnol Adv, 2012, 30(6): 1533-1542.
9. Zhu H Q, Tang X L, Zheng R C, et al. Recent advancements in enzyme engineering via site-specific incorporation of unnatural amino acids[J]. World J Microbiol Biotechnol, 2021, 37(12): 213.
10. Smolskaya S, Andreev Y A. Site-Specific Incorporation of Unnatural Amino Acids into *Escherichia coli* Recombinant Protein: Methodology Development and Recent Achievement[J]. Biomolecules, 2019, 9(7): 255.
11. Narancic T, Almahboub S A, O'connor K E. Unnatural amino acids: production and biotechnological potential[J]. world j microbiol biotechnol, 2019, 35(4): 1-11.
12. Zou H, Li L, Zhang T, et al. Biosynthesis and biotechnological application of non-canonical amino acids: complex and unclear[J]. Biotechnol Adv, 2018, 36(7): 1917-1927.
13. Karbalaei-Heidari H R, Budisa N. Combating antimicrobial resistance with new-to-nature lanthipeptides created by genetic code expansion[J]. Front Microbiol, 2020, 11: 590522.
14. Tsubogo T, Kano Y, Ikemoto K, et al. Synthesis of optically active, unnatural α-substituted glutamic acid derivatives by a chiral calcium-catalyzed 1,4-addition reaction[J]. Tetrahedron Asymmetry, 2010, 21(9): 1221-1225.
15. Li B, Zhang J, Xu Y, et al. Improved synthesis of unnatural amino acids for peptide stapling[J]. Tetrahedron Lett, 2017, 58(24): 2374-2377.

16. Wang Z, Huang D, Xu P, et al. The asymmetric alkylation reaction of glycine derivatives catalyzed by the novel chiral phase transfer catalysts[J]. Tetrahedron Lett, 2015, 56(9): 1067-1071.

17. Anderhuber N, Fladischer P, Gruber-Khadjawi M, et al. High-level biosynthesis of norleucine in *E. coli* for the economic labeling of proteins[J]. J Biotechnol, 2016, 235: 100-111.

18. Bornscheuer U T, Huisman G W, Kazlauskas R J, et al. Engineering the third wave of biocatalysis[J]. Nature, 2012, 485(7397): 185-194.

19. Wang H, Liu W, Shi F, et al. Metabolic pathway engineering for high-level production of 5-hydroxytryptophan in *Escherichia coli*[J]. Metab Eng, 2018, 48: 279-287.

20. Xu D, Fang M, Wang H, et al. Enhanced production of 5-hydroxytryptophan through the regulation of L-tryptophan biosynthetic pathway[J]. Appl Microbiol Biotechnol, 2020, 104(6): 2481-2488.

21. Lin Y, Sun X, Yuan Q, et al. Engineering bacterial phenylalanine 4-hydroxylase for microbial synthesis of human neurotransmitter precursor 5-hydroxytryptophan[J]. ACS Synth Biol, 2014, 3(7): 497-505.

22. Mora-Villalobos J A, Zeng A P. Synthetic pathways and processes for effective production of 5-hydroxytryptophan and serotonin from glucose in *Escherichia coli*[J]. J Biol Eng, 2018, 12(1): 1-12.

23. Mora-Villalobos J A, Zeng A P. Protein and pathway engineering for the biosynthesis of 5-hydroxytryptophan in *Escherichia coli*[J]. Eng Life Sci, 2017, 17(8): 892-899.

24. Mu Q, Zhang S, Mao X, et al. Highly efficient production of L-homoserine in *Escherichia coli* by engineering a redox balance route[J]. Metab Eng, 2021, 67: 321-329.

25. Sun B Y, Wang F Q, Zhao J, et al. Engineering *Escherichia coli* for L-homoserine production[J]. J Basic Microbiol, 2022.

26. Li N, Xu S, Du G, et al. Efficient production of L-homoserine in *Corynebacterium glutamicum* ATCC 13032 by redistribution of metabolic flux[J]. Biochem Eng J, 2020, 161: 107665.

27. Li N, Zeng W, Zhou J, et al. *O*-Acetyl-L-homoserine production enhanced by pathway strengthening and acetate supplementation in *Corynebacterium glutamicum*[J]. Biotechnol Biofuels Bioprod, 2022, 15(1): 27.

28. Cai M, Zhao Z, Li X, et al. Development of a nonauxotrophic L-homoserine hyperproducer in *Escherichia coli* by systems metabolic engineering[J]. Metab Eng, 2022, 73: 270-279.

29. Huang J-F, Zhang B, Shen Z-Y, et al. Metabolic engineering of *E. coli* for the production of *O*-succinyl-L-homoserine with high yield[J]. 3 Biotech, 2018, 8(7): 310.

30. Li H, Wang B, Zhu L, et al. Metabolic engineering of *Escherichia coli* W3110 for L-homoserine production[J]. Process Biochem, 2016, 51(12): 1973-1983.

31. Yi Y, Sheng H, Li Z, et al. Biosynthesis of *trans*-4-hydroxyproline by recombinant strains of *Corynebacterium glutamicum* and *Escherichia coli*[J]. BMC Biotechnol, 2014, 14: 44.

32. Zhang Y, Zhang Y, Shang X, et al. Reconstruction of tricarboxylic acid cycle in *Corynebacterium glutamicum* with a genome-scale metabolic network model for *trans*-4-hydroxyproline production[J]. Biotechnol Bioeng, 2019, 116(1): 99-109.

33. Long M, Xu M, Ma Z, et al. Significantly enhancing production of *trans*-4-hydroxy-l-proline by integrated system engineering in *Escherichia coli*[J]. Sci Adv, 2020, 6(21): eaba2383.

34. Wang X-C, Liu J, Zhao J, et al. Efficient production of *trans*-4-hydroxy-L-proline from glucose using a new *trans*-proline 4-hydroxylase in *Escherichia coli*[J]. J Biosci, 2018, 126(4): 470-477.

35. Qi L S, Larson M H, Gilbert L A, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression[J]. Cell, 2013, 152(5): 1173-83.

36. Ghavami S, Pandi A. CRISPR interference and its applications[J]. Prog Mol Biol Transl Sci, 2021, 180: 123-140.

37. James C M, Ferguson T K, Leykam J F, et al. The Amber Codon in the Gene Encoding the Monomethylamine Methyltransferase Isolated from *Methanosarcina barkeri* Is Translated as a Sense Codon*[J]. J Biol Chem, 2001, 276(36): 34252-34258.

38. Gaston M A, Zhang L, Green-Church K B, et al. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine[J]. Nature, 2011, 471(7340): 647-650.

39. Longstaff D G, Larue R C, Faust J E, et al. A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine[J]. Proc Natl Acad Sci USA, 2007, 104(3): 1021-6.

40. Srinivasan G, James C M, Krzycki J A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA[J]. Science, 2002, 296(5572): 1459-62.

41. Ho J M L, Miller C A, Smith K A, et al. Improved pyrrolysine biosynthesis through phage assisted non-continuous directed evolution of the complete pathway[J]. Nat Commun, 2021, 12(1): 3914.

42. Liu W, Li H, Liu L, et al. Screening of gamma-aminobutyric acid-producing lactic acid bacteria and the characteristic of glutamate decarboxylase from Levilactobacillus brevis F109-MD3 isolated from kimchi[J]. J Appl Microbiol, 2022, 132(3): 1967-1977.

43. Li Y, Chen X, Shu G, et al. Screening of gamma-aminobutyric acid-producing lactic acid bacteria and its application in Monascus-fermented rice production[J]. Acta Sci Pol Technol Aliment, 2020, 19(4): 387-394.

44. Kim N Y, Kim S K, Ra C H. Evaluation of gamma-aminobutyric acid (GABA) production by Lactobacillus plantarum using two-step fermentation[J]. Bioprocess Biosyst Eng, 2021, 44(10): 2099-2108.

45. Diez-Gutiérrez L, Vicente L S, Sáenz J, et al. Biosynthesis of gamma-aminobutyric acid by Lactiplantibacillus plantarum K16 as an alternative to revalue agri-food by-products[J]. Sci Rep, 2022, 12(1): 18904.

46. Praveschotinunt P, Dorval Courchesne N-M, Den Hartog I, et al. Tracking of Engineered Bacteria In Vivo Using Nonstandard Amino Acid Incorporation[J]. ACS Synth Biol, 2018, 7(6): 1640-1650.

47. Hatfield D L, Tsuji P A, Carlson B A, et al. Selenium and selenocysteine: roles in cancer, health, and development[J]. Trends Biochem Sci, 2014, 39(3): 112-20.

48. Schmidt R L, Simonović M. Synthesis and decoding of selenocysteine and human health[J]. Croat Med J, 2012, 53(6): 535-50.

49. Long M R, Ong W K, Reed J L. Computational methods in metabolic engineering for strain design[J]. Curr Opin Biotechnol, 2015, 34: 135-41.

50. Masuo S, Zhou S, Kaneko T, et al. Bacterial fermentation platform for producing artificial aromatic amines[J]. Sci Rep, 2016, 6: 25764.

51. Zomorrodi A R, Hemez C, Arranz-Gibert P, et al. Computational design and engineering of an *Escherichia coli* strain producing the nonstandard amino acid *para*-aminophenylalanine[J]. iScience, 2022, 25(7): 104562.

52. Hecht S M, Alford B L, Kuroda Y, et al. "Chemical aminoacylation" of tRNA's[J]. J Biol Chem, 1978, 253(13): 4517-20.

53. Gamper H, Hou Y M. A Label-Free Assay for Aminoacylation of tRNA[J]. Genes, 2020, 11(10).

54. Noren C J, Anthony-Cahill S J, Griffith M C, et al. A general method for site-specific incorporation of unnatural amino acids into proteins[J]. Science, 1989, 244(4901): 182-188.

55. Pastore A J, Ficaretta E, Chatterjee A, et al. Substitution of the sole tryptophan of the cupredoxin, amicyanin, with 5-hydroxytryptophan alters fluorescence properties and energy transfer to the type 1 copper site[J]. J Inorg Biochem, 2022, 234: 111895.

56. Lajoie M J, Rovner A J, Goodman D B, et al. Genomically recoded organisms expand biological functions[J]. Science, 2013, 342(6156): 357-360.

57. Yi H, Zhang J, Ke F, et al. Comparative Analyses of the Transcriptome and Proteome of *Escherichia coli* C321.△A and Further Improving Its Noncanonical Amino Acids Containing Protein Expression Ability by Integration of T7 RNA Polymerase[J]. Front Microbiol, 2021, 12: 744284.

58. Costa S A, Mozhdehi D, Dzuricky M J, et al. Active targeting of cancer cells by nanobody decorated polypeptide micelle with bio-orthogonally conjugated drug[J]. Nano letters, 2019, 19(1): 247-254.

59. Wang K, Fredens J, Brunner S F, et al. Defining synonymous codon compression schemes by genome recoding[J]. Nature, 2016, 539(7627): 59-64.

60. Fredens J, Wang K, De La Torre D, et al. Total synthesis of *Escherichia coli* with a recoded genome[J]. Nature, 2019, 569(7757): 514-518.

61. Robertson W E, Funke L F H, De La Torre D, et al. Sense codon reassignment enables viral resistance and encoded polymer synthesis[J]. Science, 2021, 372(6546): 1057-1062.

62. Plotkin J B, Kudla G. Synonymous but not the same: the causes and consequences of codon bias[J]. Nat Rev Genet, 2011, 12(1): 32-42.

63. Brule C E, Grayhack E J. Synonymous codons: choose wisely for expression[J]. Trends Genet, 2017, 33(4): 283-297.

64. Mitra S, Ray S K, Banerjee R. Synonymous codons influencing gene expression in organisms[J]. Rese Rep Biochem, 2016, 2016(6): 57-65.

65. Wang N, Shang X, Cerny R, et al. Systematic Evolution and Study of UAGN Decoding tRNAs in a Genomically Recoded Bacteria[J]. Sci Rep, 2016, 6: 21898.

66. Hohsaka T, Ashizuka Y, Murakami H, et al. Five-base codons for incorporation of nonnatural amino acids into proteins[J]. Nucleic Acids Res, 2001, 29(17): 3646-3651.

67. Xi Z, Davis L, Baxter K, et al. Using a quadruplet codon to expand the genetic code of an animal[J]. Nucleic Acids Res, 2022, 50(9): 4801-4812.

68. De La Torre D, Chin J W. Reprogramming the genetic code[J]. Nat Rev Genet, 2021, 22(3): 169-184.

69. Gamper H, Masuda I, Hou Y-M. Genome Expansion by tRNA +1 Frameshifting at Quadruplet Codons[J]. J Mol Biol, 2022, 434(8): 167440.

70. Guo J, Niu W. Genetic Code Expansion Through Quadruplet Codon Decoding[J]. J Mol Biol, 2022, 434(8): 167346.

71. Bain J D, Switzer C, Chamberlin R, et al. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code[J]. Nature, 1992, 356(6369): 537-539.

72. Malyshev D A, Dhami K, Lavergne T, et al. A semi-synthetic organism with an expanded genetic alphabet[J]. Nature, 2014, 509(7500): 385-388.

73. Benner S A, Karalkar N B, Hoshika S, et al. Alternative Watson-Crick Synthetic Genetic Systems[J]. Cold Spring Harb Perspect Biol, 2016, 8(11): a023770.

74. Kimoto M, Hirao I. Genetic alphabet expansion technology by creating unnatural base pairs[J]. Chem Soc Rev, 2020, 49(21): 7602-7626.

75.   Kimoto M, Hirao I. Genetic Code Engineering by Natural and Unnatural Base Pair Systems for the Site-Specific Incorporation of Non-Standard Amino Acids Into Proteins[J]. Front Mol Biosci, 2022, 9: 851646.

76.   Mehl R A, Anderson J C, Santoro S W, et al. Generation of a Bacterium with a 21 Amino Acid Genetic Code[J]. J Am Chem Soc, 2003, 125(4): 935-939.

77.   Budisa N. Prolegomena to Future Experimental Efforts on Genetic Code Engineering by Expanding Its Amino Acid Repertoire[J]. Angew Chem Int Ed, 2004, 43(47): 6426-6463.

78.   Hoesl M G, Budisa N. In vivo incorporation of multiple noncanonical amino acids into proteins[J]. Angew Chem Int Ed Engl, 2011, 50(13): 2896-902.

79.   Baumann T, Schmitt F J, Pelzer A, et al. Engineering 'Golden' Fluorescence by Selective Pressure Incorporation of Non-canonical Amino Acids and Protein Analysis by Mass Spectrometry and Fluorescence[J]. J Vis Exp, 2018(134): 57017.

80.   Johnson J A, Lu Y Y, Van Deventer J A, et al. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications[J]. Curr Opin Chem Biol, 2010, 14(6): 774-780.

81.   Deng J, Viel J H, Chen J, et al. Synthesis and Characterization of Heterodimers and Fluorescent Nisin Species by Incorporation of Methionine Analogues and Subsequent Click Chemistry[J]. ACS Synth Biol, 2020, 9(9): 2525-2536.

82.   Baumann T, Nickling J H, Bartholomae M, et al. Prospects of In vivo Incorporation of Non-canonical Amino Acids for the Chemical Diversification of Antimicrobial Peptides[J]. Front Microbiol, 2017, 8: 124.

83.   Nickling J H, Baumann T, Schmitt F J, et al. Antimicrobial Peptides Produced by Selective Pressure Incorporation of Non-canonical Amino Acids[J]. J Vis Exp, 2018(135): 57551.

84.   Münzker L, Oddo A, Hansen P R. Chemical synthesis of antimicrobial peptides[J]. Methods Mol Biol, 2017, 1548: 35-49.

85.   Wang H. Comprehensive Organic Name Reactions[M]. Wiley, 2010.

86.   Tsuchiya K, Numata K. Chemoenzymatic synthesis of polypeptides containing the unnatural amino acid 2-aminoisobutyric acid[J]. Chem Commun, 2017, 53(53): 7318-7321.

87.   Jakas A, Vlahoviček-Kahlina K, Ljolić-Bilić V, et al. Design and synthesis of novel antimicrobial peptide scaffolds[J]. Bioorg Chem, 2020, 103: 104178.

88.   Mueller L K, Baumruck A C, Zhdanova H, et al. Challenges and Perspectives in Chemical Synthesis of Highly Hydrophobic Peptides[J]. Front Bioeng Biotechnol, 2020, 8.

89.   Behrendt R, White P, Offer J. Advances in Fmoc solid-phase peptide synthesis[J]. J Pept Sci, 2016, 22(1): 4-27.

90.   Yue K, Trung T N, Zhu Y, et al. Co-Translational Insertion of Aquaporins into Liposome for Functional Analysis via an *E. coli* Based Cell-Free Protein Synthesis System[J]. Cells, 2019, 8(11): 1325.

91.   Levine M Z, Gregorio N E, Jewett M C, et al. *Escherichia coli*-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology[J]. J Vis Exp, 2019(144).

92.   Zemella A, Thoring L, Hoffmeister C, et al. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems[J]. Chembiochem, 2015, 16(17): 2420-31.

93.   Purkayastha A, Iyappan K, Kang T J. Multiple Gene Expression in Cell-Free Protein Synthesis Systems for Reconstructing Bacteriophages and Metabolic Pathways[J]. Microorganisms, 2022, 10(12): 2477.

94.   Anastasina M, Terenin I, Butcher S J, et al. A technique to increase protein yield in a rabbit reticulocyte lysate translation system[J]. Biotechniques, 2014, 56(1): 36-9.

95.   Wright T H, Bower B J, Chalker J M, et al. Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity[J]. Science, 2016, 354(6312): aag1465.

96.   Zhang Y, Zang C, An G, et al. Cysteine-specific protein multi-functionalization and disulfide bridging using 3-bromo-5-methylene pyrrolones[J]. Nat Commun, 2020, 11(1): 1015.

97.   Chalker J M, Bernardes G J L, Lin Y A, et al. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology[J]. Chemistry, 2009, 4(5): 630-640.

98.   Silverman H G, Roberto F F. Understanding marine mussel adhesion[J]. Mar Biotechnol, 2007, 9(6): 661-81.

99.   Wang N, Li Y, Zhang Y, et al. High-strength photoresponsive hydrogels enable surface-mediated gene delivery and light-induced reversible cell adhesion/detachment[J]. Langmuir, 2014, 30(39): 11823-32.

100.   Yu J, Kan Y, Rapp M, et al. Adaptive hydrophobic and hydrophilic interactions of mussel foot proteins with organic thin films[J]. Proc Natl Acad Sci 2013, 110(39): 15680-5.

101.   Li Y, Cao Y. The molecular mechanisms underlying mussel adhesion[J]. Nanoscale Adv, 2019, 1(11): 4246-4257.

102.   Hauf M, Richter F, Schneider T, et al. Photoactivatable mussel-based underwater adhesive proteins by an expanded genetic code[J]. ChemBioChem, 2017, 18(18): 1819-1823.

103.   Xiang Z, Ren H, Hu Y S, et al. Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity[J]. Nat Methods, 2013, 10(9): 885-8.

104.   Zappala F, Tsourkas A: Site-specific photocrosslinking to immunoglobulin G using photoreactive antibody-binding domains, Bioconjugation: Springer, 2019: 275-286.

105. Elia N. Using unnatural amino acids to selectively label proteins for cellular imaging: a cell biologist viewpoint[J]. Febs J, 2021, 288(4): 1107-1117.

106. Pagar A D, Jeon H, Khobragade T P, et al. Non-Canonical Amino Acid-Based Engineering of (R)-Amine Transaminase[J]. Front Chem, 2022, 10: 839636.

107. Miyazaki R, Akiyama Y, Mori H. A photo-cross-linking approach to monitor protein dynamics in living cells[J]. Biochim Biophys Acta Gen Subj, 2020, 1864(2): 129317.

108. Jones C M, Sungwienwong I, Petersson E J. The development of intrinsically fluorescent unnatural amino acids for in vivo incorporation into proteins[J]. Biophys, 2019, 116(3): 473a.

109. Curnew L J F, Mcnicholas K, Green B, et al. Visualizing HCV core protein via fluorescent unnatural amino acid incorporation[J]. Proceedings, 2020, 50(1): 129.

110. Wang Y, Chen X, Cai W, et al. Expanding the structural diversity of protein building blocks with noncanonical amino acids biosynthesized from aromatic thiols[J]. Angew Chem Int Ed Engl, 2021, 60(18): 10040-10048.

111. Benedini L. Advanced Protein Drugs and Formulations[J]. Curr Protein Pept Sci, 2022, 23(1): 2-5.

112. Li Q, Chen Q, Klauser P C, et al. Developing covalent protein drugs via proximity-enabled reactive therapeutics[J]. Cell, 2020, 182(1): 85-97.e16.

113. Liu J, Cao L, Klauser P C, et al. A Genetically Encoded Fluorosulfonyloxybenzoyl-L-lysine for Expansive Covalent Bonding of Proteins via SuFEx Chemistry[J]. J Am Chem Soc, 2021, 143(27): 10341-10351.

114. Lindberg J, Nilvebrant J, Nygren P, et al. Progress and Future Directions with Peptide-Drug Conjugates for Targeted Cancer Therapy[J]. Molecules, 2021, 26(19): 6042.

115. Fu Z, Li S, Han S, et al. Antibody drug conjugate: the "biological missile" for targeted cancer therapy[J]. Signal Transduct Target Ther, 2022, 7(1): 93.

116. Jin Y, Schladetsch M A, Huang X, et al. Stepping forward in antibody-drug conjugate development[J]. Pharmacol Ther, 2022, 229: 107917.

117. Hallam T J, Wold E, Wahl A, et al. Antibody conjugates with unnatural amino acids[J]. Mol Pharm, 2015, 12(6): 1848-1862.

118. Grünewald J, Tsao M-L, Perera R, et al. Immunochemical termination of self-tolerance[J]. Proc Natl Acad Sci USA, 2008, 105(32): 11276-11280.

119. Grünewald J, Hunt G S, Dong L, et al. Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids[J]. Proc Natl Acad Sci U S A, 2009, 106(11): 4337-42.

120. Wang N, Li Y, Niu W, et al. Construction of a live-attenuated HIV-1 vaccine through genetic code expansion[J]. Angew Chem Int Ed Engl, 2014, 53(19): 4867-71.

121. Yuan Z, Wang N, Kang G, et al. Controlling Multicycle Replication of Live-Attenuated HIV-1 Using an Unnatural Genetic Switch[J]. ACS Synth Biol, 2017, 6(4): 721-731.

122. Si L, Xu H, Zhou X, et al. Generation of influenza A viruses as live but replication-incompetent virus vaccines[J]. Science, 2016, 354(6316): 1170-1173.

123. Chang Z, Liu D, Yang Z, et al. Efficient Xylitol Production from Cornstalk Hydrolysate Using Engineered *Escherichia coli* Whole Cells[J]. J Agric Food Chem, 2018, 66(50): 13209-13216.

124. Crick F H C. The origin of the genetic code[J]. J. Mol. Biol, 1968, 38(3): 367-379.

125. Kim S, Sung B H, Kim S C, et al. Genetic incorporation of l-dihydroxyphenylalanine (DOPA) biosynthesized by a tyrosine phenol-lyase[J]. Chem Commun, 2018, 54(24): 3002-3005.

126. Marchand J A, Neugebauer M E, Ing M C, et al. Discovery of a pathway for terminal-alkyne amino acid biosynthesis[J]. Nature, 2019, 567(7748): 420-424.

127. Zhang M S, Brunner S F, Huguenin-Dezot N, et al. Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing[J]. Nat Methods, 2017, 14(7): 729-736.