Chapter 3

Selective and Site-Specific Incorporation of Nonstandard Amino Acids Within Proteins for Therapeutic Applications

Neil D. Butler and Aditya M. Kunjapur

Abstract

The incorporation of nonstandard amino acids (nsAAs) within protein sequences has broadened the chemical functionalities available for use in the study, prevention, or treatment of disease. The ability to genetically encode the introduction of nsAAs at precise sites of target recombinant proteins has enabled numerous applications such as bioorthogonal conjugation, thrombin inhibition, intrinsic biological containment of live organisms, and immunochemical termination of self-tolerance. Genetic systems that perform critical steps in enabling nsAA incorporation are known as orthogonal translation systems or orthogonal aminoacyl-tRNA synthetase/tRNA pairs. In *Escherichia coli*, several of these have been designed to accept novel nsAAs. Certain endogenous proteins, codon context, and standard amino acid concentrations can affect the yield of recombinant protein, the rate of nsAA incorporation within off-target proteins, and the rate of misincorporation due to near-cognate suppression or misacylation of orthogonal tRNA with standard amino acids. As a result, a significant body of work has been performed in engineering the *E. coli* genome to alleviate these issues. Here, we describe common methods applicable to nsAA incorporation within proteins in *E. coli* for sufficient purity and characterization for downstream therapeutic applications.

Key words Nonstandard amino acids, *Escherichia coli*, Amino-acyl tRNA synthetase, Posttranslational proofreading, Site-specific nsAA incorporation

1 Introduction

Nonstandard amino acids (nsAAs) have the potential to incorporate chemical functionalities into proteins that are distinct from the 20 standard amino acids. In therapeutic applications, these functionalities can be utilized as means for strict modification and control over protein properties that are difficult to achieve through chemical or enzymatic posttranslational modifications of standard amino acids. For example, some functional groups absent in the standard amino acids, such as azides, tetrazines, and alkynes, are orthogonal to biological molecules but are highly reactive handles for click-chemical conjugations [1, 2]. For therapeutic proteins,

Millicent O. Sullivan et al. (eds.), *Therapeutic Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 2720,
https://doi.org/10.1007/978-1-0716-3469-1_3,

this property is particularly advantageous, enabling the linkage of molecules for improved stability and circulation (e.g., PEGylation [3]), for delivery of drug payloads (e.g., antibody-drug conjugates [4]), or for improved antigen recognition (bispecific antibodies [5]) with high specificity. Compared to traditional methods of protein conjugation (e.g., surface-exposed lysine or cysteine conjugation), introduction of nsAAs enables a higher degree of control and reduction in the heterogeneity of final therapeutic products [6], which has led several companies to pursue conjugatable nsAAs in clinical therapeutics (namely, Ambrx [7], AstraZeneca [8], BrickBio [9], Sutro [10], and Synthorx/Sanofi [11]).

Apart from bioorthogonal conjugation, several additional properties imparted by nsAAs have therapeutic relevance. Some nsAAs exhibit properties benefiting therapeutic applications themselves, such as sulfonation (sulfotyrosine) [12, 13], phosphorylation (phosphotyrosine) [14], and nitration (nitrophenylalanine and 3-nitrotyrosine) [15, 16], which have been shown to impact elements of protein interactions with the immune system, like antibody affinity and antigen immunogenicity. In each of these cases, incorporation of an nsAA simulates naturally occurring posttranslational modifications, which can enable researchers to develop therapeutic proteins making use of the unique interactions these natural PTMs have within the body. Other nsAAs have imparted properties such as stimuli responsiveness (photoactivation [17]), labeling (via a single residue fluorescent amino acid [18]), and improved protein stability [19], which each can enable control, tracking, and sustained delivery in the context of therapeutic proteins. Further application of nsAAs in therapeutics include the area of intrinsic biological containment of live organisms. For living therapeutics, controlled implementation of the species limits the risk of off-target proliferation and consequent safety concerns. By engineering essential proteins in these living therapeutics to rely upon nsAAs for full-length translation or proper folding, the proliferation of these therapeutics can be appropriately attenuated, for potential controlled localization of drug payload delivery and reduced safety concern [20–23]. Thus, the efficient, selective incorporation of nsAAs has broad utility in the therapeutic space.

To incorporate nsAAs within proteins, there are two primary methods of ribosomal incorporation: (1) residue-specific incorporation [24] or (2) site-specific incorporation [25]. In both cases, engineered translation machinery, primarily an aminoacyl-tRNA synthetase (AARS)/tRNA pair, is engineered to facilitate incorporation. For residue-specific incorporation, a polyspecific AARS substitutes nsAAs bearing structural similarity to native amino acids across the proteome. This approach is less advantageous for most therapeutic applications, where a select number of nsAAs are desired to be integrated at particular sites within proteins, as all sites of a select native standard amino acid are mutated

to nsAAs. An exception would be therapeutic polypeptide biomaterials, where the global substitution offered by residue-specific incorporation would bestow consistency of monomer units. For site-specific incorporation, an AARS/tRNA pair orthogonal to the native translation machinery and standard amino acids facilitates the translation of nsAAs at a designated nonsense codon (normally a stop codon) or quadruplet codon [26, 27], thereby allowing the specification of the site for incorporation within the genetic code. With suppression of quadruplet codons, the genetic code can be significantly expanded (from 64 to 256 potential codons) but decoding of quadruplet codons can be inefficient without significant engineering of tRNA [28] or engineering of further orthogonal translational machinery such as an orthogonal ribosome [29, 30]. As an alternative decoding mechanism, expanded genetic nucleotide alphabets have been also engineered and applied within living microbes [31–33] for further orthogonality and expansion of the genetic code. Explorations by Synthorx (since acquired by Sanofi) have applied this expanded alphabet for controlled conjugation of clinical therapeutics. While there is a breadth of novel developments in nsAA incorporation for therapeutics, in this chapter, we will focus on the site-specific incorporation of nsAAs using triplet codons, native ribosomes, and standard deoxyribonucleic acids. These techniques are the most widely accessible and well-established for researchers, do not require significant heterologous or chemically synthesized components, and still enable the targeting of nsAA incorporation to specified sites within proteins.

In particular, we will focus on the use of nsAAs for therapeutic applications in *Escherichia coli*, for which several families of orthogonal AARSs have been engineered (Table 1). Using these AARS/tRNA pairs, a broad range of nsAAs have been capable of being incorporated into proteins, mostly consisting of aromatic derivatives of phenylalanine (Phe) with additional AARS/tRNA capable of incorporation of pyrrolysine/lysine (Pyl/Lys) or tryptophan (Trp) derivatives [34–37]. For the objective of molecular attachment to therapeutic protein surfaces, nsAAs that contain functional groups amenable to conjugation have been incorporated using all these families of AARSs. Thus, the choice of AARS when designing a procedure is more dependent upon the degree of activity needed for sufficient protein titer, the selectivity and purity of target protein, and whether multiple distinct nsAAs are desired to be incorporated.

In designing systems for nsAA incorporation, expression cassettes for the orthogonal AARS/tRNA and a mutated reporter/gene of interest are required at a minimum, and these are often expressed on separate plasmids. For the expression of AARS/tRNA, two primary systems have been utilized: (i) the pEVOL plasmid [38] and (ii) the pUltra plasmid [39]. These cassettes differ in their promoter systems (the P_{AraBAD} arabinose-inducible system

Table 1
Families of aminoacyl-tRNA synthetases engineered for use in *E. coli*

AARS	Kingdom of origin	Species of origin	Primary nsAA substrates	Codon assignment
MjTyrRS	Archaea	*Methanococcus jannaschii*	Phe and Tyr derivatives	TAG [55], AGGA [56], TAA [57]
PylRS	Archaea	*Methanosarcina bakeri* and *Methanosarcina mazei*	Pyl/Lys, Phe, His, and Tyr derivatives	TGA [58], TAA [58], TAG [59], AGGA [26], AGG [60]
chPheRS	Archaea/ mammalian	*Homo sapiens* mitochondrial chimera with *M. mazei*	Aromatic Phe derivatives	TAG [61]
EcTrpRS	Bacteria	*E. coli*	Trp derivatives with substitution at 5-position	TGA, TAG [36]

for pEVOL, the P_{tac} IPTG-inducible system for pUltra), origin of replication (p15a for pEVOL, cloDF for pUltra), copies of the AARS gene (one for pEVOL, two for pUltra), and antibiotic resistance marker (chloramphenicol for pEVOL, spectinomycin for pUltra). Both systems result in similar protein titer in direct comparisons with each other and can result in robust nsAA incorporation. Generally, greater consideration is required for how to best clone the therapeutic protein of interest, where structural insights can aid in selection of sites for mutagenesis with low probability of inducing protein misfolding. Oftentimes, the amber stop codon is chosen for suppression and incorporation of an nsAA given its low prevalence in *E. coli* (only 321 instances) [40]. If attempting to incorporate more than one nsAA, other stop codons (or quadruplet codons) must be chosen which will result in greater proteome-wide incorporation or decreased protein titer. In addition, the different AARS/tRNAs, in addition to plasmids used must be orthogonal as well, increasing system complexity.

When selecting an *E. coli* strain for expression of a desired therapeutic protein, there are additional considerations researchers can make regarding expression strain. For protein overexpression generally, the BL21(DE3) strain is often the optimal choice; however, when attempting to perform amber suppression at multiple sites within a protein, AARS/tRNA activity is often outcompeted by native translational termination machinery (release factor 1). Thus, the genomically recoded C321.ΔA strain was created, wherein all the amber codons (TAG) were converted to ochre codons (TAA) and release factor 1 was knocked out [40]. This strain is particularly well-suited to produce proteins that require

multisite suppression at amber codons. However, it can result in lower fidelity with some AARSs given they were originally engineered in strains containing release factor 1. Additionally, for the EcTrpRS strain, an engineered strain (ATMW1) with replacement of native TrpRS/tRNA by that from *S. cerevisiae* must be used as well [36]. If it is desired to use a strain without prior nsAA incorporation applications, it is important to ensure amber suppressor mutations are not present on the genome, which are often integrated for phage-based engineering and can be found in cloning strains like DH5α.

One additional consideration when incorporating nsAAs into therapeutic proteins is how to ensure selective incorporation. Nonselective incorporation of an nsAA results in heterogenous protein production which is difficult to separate and could result in lowered pharmacokinetics for biotherapeutics or altered properties for biomaterials. Additionally, if using nsAA incorporation for biocontainment, AARS promiscuity can increase escape and reduce overall fitness. To enhance selective incorporation, directed evolution protocols are often applied at the AARS level, wherein positive/negative-based selections such as ampicillin/barnase selection [41] or phage-assisted continuous evolution (PACE) [42], or screening protocols like fluorescence-activated cell sorting (FACS) on libraries of AARS variants can enable identification of selective variants. However, these methods can be inaccessible for some researchers or insufficient in identifying a selective enough AARS. In these cases, one may be able to alter the media composition to control the concentration of the standard amino acids that are likely to be misincorporated in the target site. Alternatively, one can increase the stringency of the negative selection step during AARS evolution, for example by increasing the standard amino acid concentrations or by decreasing the number of UAG codons in the negative selection marker.

A related challenge of maintaining selectivity at the step of the positive selection or screen is that conventional procedures only enrich for AARS variants that increase the rate of full-length protein translation, which can occur due to suppression with any amino acid. To address this issue, we developed a screening technique called posttranslational proofreading. This system hijacks a pathway of conditional proteolysis and repurposes it to degrade proteins that contain the incorrect amino acid incorporated at their N-terminus. Target proteins that instead contain the desired nsAA at their N-terminus are not bound by the *E. coli* ClpS adaptor protein, thereby accumulating with decreased false-positive signal. The ability to discriminate among distinct nsAAs and to tune this discrimination was also shown by engineering ClpS variants that have altered binding pockets [43]. However, this system is still limited in that it can increase the complexity and burden associated with nsAA incorporation, which can decrease the titer of reporter

protein. For therapeutic protein production, it also requires incorporation of the nsAA at the N-terminus, though that may be a suitable site for bioconjugation. Finally, the ClpS protein has been engineered to discriminate among only a select number of nsAAs from one another or from standard amino acids. On balance, posttranslational proofreading is most useful to include when trying to discern the true rate of incorporation of a bulky desired nsAA using AARS/tRNA pairs that catalyze nonnegligible levels of misincorporation in its absence.

Here, we describe a general protocol for incorporation of nsAAs within proteins in *E. coli*. We provide options to create strains that exhibit a range of phenotypes that balance nsAA incorporation efficiency and final protein titer for the desired therapeutic application. We describe methods to enhance the selectivity of nsAA incorporation within proteins through the addition of posttranslational proofreading or by altering media conditions (Fig. 1). Lastly, we include techniques to ensure purity and validate identity for subsequent therapeutic applications.

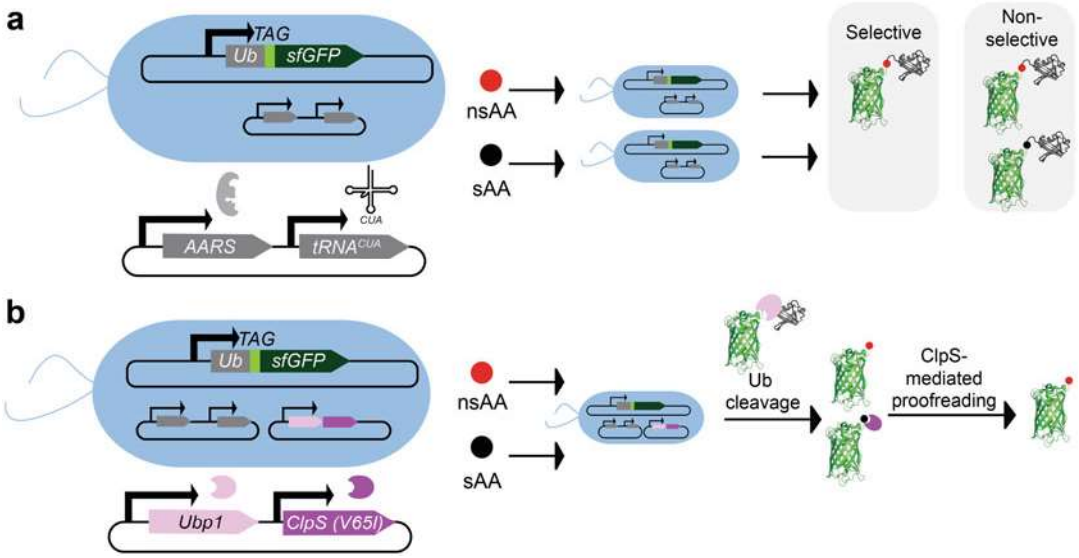

Fig. 1 Graphical representation of screening output with and without posttranslational proofreading. (**a**) Screening of an AARS/tRNA with a genomic ubiquitin-TAG-GFP (Ub-TAG-GFP) reporter system. Selective incorporation is demonstrated if GFP expression can be observed in the presence of a desired nsAA and not from the standard amino acids (sAAs) from metabolism or media components. (**b**) By screening with posttranslational proofreading, Ub is cleaved by Ubp1, exposing the amino acid suppressing the TAG codon at the N-terminus. Engineered N-terminal degradation machinery controlled by ClpS eliminates sAA containing proteins, proofreading for GFP containing the nsAA of interest

2 Materials

1. Stocks of the desired *Escherichia coli* strain (Table 2).
2. Appropriate growth media for *E. coli* (Table 3). For AARS/tRNA systems expressed on pEVOL plasmids, M9-minimal media and MOPS EZ Rich media are supplemented with 1% glycerol as opposed to 1% glucose to allow the arabinose inducible promoter systems to function in the absence of catabolite repression.
3. AARS/tRNA expression plasmid for inducible expression in *E. coli*. Many commonly used AARS/tRNA variants are available on Addgene (for example, pEVOL-pAzF [44]).
4. *E. coli* plasmid containing your desired therapeutic protein with nonsense codon placed in-frame at the target site within the gene sequence.
5. *E. coli* plasmid encoding a reporter protein for screening of AARS/tRNA activity. Examples can be found on Addgene (P_{AraBAD}-sfGFP_150TAG [45]), or others found in literature (pZE21-Ub-UAG-GFP [43]).
6. *E. coli* plasmid expressing engineered ClpS (Addgene, pZE21/UBP1/ClpS_V65I [43]).

Table 2
***E. coli* strains capable of nsAA incorporation**

Name	Growth temperature	Application	Source
BL21	37 °C	Protein overexpression	New England Biolabs
BL21(DE3)	37 °C	Protein overexpression from T7 polymerase	New England Biolabs
C321.ΔA	34 °C	Protein overexpression with incorporation at multiple sites	Addgene [40]
C321.ΔA.Opt	34 °C	Protein overexpression with incorporation at multiple sites. Superior growth rate compared to C321.ΔA	Addgene [62]
C321.ΔClpS. Ub-UAG-sfGFP	34 °C	Screening AARS variants for incorporation efficiency	Addgene [43]
ATMW1	37 °C	Protein overexpression using EcTrpRS/tRNA	Italia, et al. [36]
ClearColi™ BL21(DE3)	37 °C	Protein overexpression with reduced lipopolysaccharide expression	Lucigen

Table 3
Media used for overexpression of proteins for nsAA incorporation or associated protocols

Media	Preparation	Application
Lysogeny broth (LB)	Mix 10 g/L bacto tryptone, 5 g/L sodium chloride, and 5 g/L yeast extract. Autoclave at 121 °C for 30 min	Protein overexpression
Terrific broth (TB)	Mix 20 g/L bacto tryptone, 4 mL/L glycerol, 24 g/L yeast extract, 17 mM KH_2PO_4, and 72 mM K_2HPO_4. Autoclave at 121 °C for 30 min	Protein overexpression
M9-minimal media	Mix 1x M9 salt medium containing 12.8 g/L Na_2HPO_4 $7H_2O$, 3 g/L KH_2PO_4, 1 g/L NH_4CI, and 0.5 g/L NaCl, supplemented with 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1% glucose, and trace elements A (1.60 μg/mL $CuSO_4 \cdot 5H_2O$, 863.00 μg/mL $ZnSO_4 \cdot 7H_2O$, 17.30 μg/mL selenite • 2Na, 1155.10 μg/mL ferric citrate). Sterile filter (0.22 μm)	Protein overexpression for lower rates of off-target incorporation. For C321-based strains, add biotin (4 μM) to enable growth
MOPS EZ rich media	Commercially available (Teknova M2105 and M2136)	Protein overexpression for lower rates of off-target incorporation for C321-based strains, add biotin (4 μM) to enable growth
SOC media	Mix 20 g/L bacto tryptone, 5 g/L yeast extract, 10 mM NaCl, 2. mM NaCl, 10 mM $MgCl_2$, and 10 mM $MgSO_4$. Autoclave at 121 °C for 30 min. Supplement with 20 mM final concentration of sterile filtered glucose	Posttransformation recovery

7. 34 mg/mL Chloramphenicol in ethanol.
8. 30 mg/mL Kanamycin in water.
9. 95 mg/mL Spectinomycin in water.
10. 50 mg/mL Carbenicillin in ethanol.
11. 1 M Isopropyl ß-D-1-thiogalactopyranoside (IPTG) in water.
12. 0.2 μM Anhydrotetracycline (aTc) in ethanol.
13. 20% L-Arabinose in water.
14. Phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4.
15. 100 mM nsAA chemical stocks in water (*see* **Note 1**).
16. Lamelli SDS reducing sample buffer: 62.5 mM Tris–HCl, 1.5% SDS, 8.3% glycerol, 1.5% 2-mercaptoethanol, 0.005% bromophenol blue.

17. Lysis buffer: 25 mM sodium phosphate, 10 mM imidazole, 300 mM NaCl, pH 7.4 (imidazole can be omitted if Ni-NTA purification is not to be performed).
18. Shaking incubator.
19. Standing incubator.
20. 96-well plate reader.
21. Orbital plate shaker.
22. Refrigerated bench top centrifuge.
23. Sonicator (for example, QSonica Q125 model).
24. Vortexer with microcentrifuge attachment.
25. Fast protein liquid chromatography (FPLC) system (Akta Pure from Cytiva).
26. Ni-NTA affinity chromatography resin/columns (either gravity flow column or FPLC; for example: HisTrap™ High Performance from Cytiva).
27. Ion exchange chromatography column for FPLC (for example, HiTrap™ Q HP from Cytiva).
28. Size exclusion chromatography column for FPLC (for example, Sephadex G-25 from Cytiva).
29. Ultrafiltration cassette for protein buffer exchange and protein concentration (for example Amicon™ centrifugal concentrator units).
30. Endotoxin removal chromatography column for FPLC (for example, EndoTrap® HD)[50].
31. HPLC-MS (for example, electrospray ionization quadrupole Time of Flight MS).
32. 96 deep-well plates (1 mL well volume).
33. Black clear bottom 96-well plates.
34. 250 mL and 5 L baffled shake flasks.
35. Glass beads.

3 Methods

3.1 Identification of an AARS and Media Conditions for Optimal, Selective Protein Titer Using nsAA of Interest

1. Identify AARS/tRNA pairs for screening of activity on the nsAA of interest. The choice of AARS/tRNA can be difficult to predict a priori if there is no literature precedent. A number of AARS variants are known to have broad substrate scope (pCNFRS [46], HpRS [47], and CouARS [48] for example) and can be good choices to investigate initially. For other nsAAs (the conjugatable *p*-azidophenylalanine for example), there are several AARSs with demonstrated activity, though often few comprehensive comparisons. In these cases, it is

valuable to compare multiple AARSs for optimization of incorporation, as the literature designated AARS/tRNA for an nsAA may not necessarily be the most active or selective.

2. Clone the AARS/tRNA pairs chosen into expression cassettes such as pEVOL or pUltra if not available from Addgene.
3. Prepare electrocompetent cells of the target *E. coli* strain (*see* **Note 2** for help with strain selection) by initially starting a fresh overnight culture of the target strain (~3 mL in LB) from an isolated colony on a freshly streaked plate with appropriate antibiotics added (1000x dilution from stock concentrations provided in Subheading 2) if needed.
4. Then, the next day, dilute 1:100 in fresh LB media (0.5 mL overnight culture in 49.5 mL LB). Incubate at ideal growth temperature (Table 2) with shaking.
5. When culture has grown to mid-exponential phase (OD_{600} of 0.4–0.6), move the culture to prechilled centrifuge tubes and centrifuge the cells at 4000× *g* for 10 min at 4 °C to pellet.
6. Discard the supernatant and resuspend the cells in 10% glycerol, moving the resuspension to prechilled microcentrifuge tubes on ice (5x). Then, wash the cells 3x by centrifugation at 8000× *g* for 1 min, discarding the supernatant after each wash.
7. Resuspend each tube in 100 μL of 10% glycerol. Aliquot 20 μL into separate prechilled microcentrifuge tubes.
8. Mix 20 μL of 10% glycerol with 1 uL of plasmid DNA for AARS/tRNA plasmid and 1 μL of fluorescent reporter plasmid (pZE21-Ub-UAG-GFP for example) if using a strain without a genomic reporter.
9. Transfer 40 μL to a prechilled electroporation cuvette 10 mm gap and electroporate (1.8 kV, 200 Ω, 25 μF). Recover the cells in 500 μL SOC media and incubate at ideal growth temperature with shaking for 1 h.
10. Then, spread 30 μL and 300 μL of cells on agar plates containing appropriate concentrations of antibiotics (1000x dilution from stock concentrations provided in Subheading 2) and grow at ideal growth temperature overnight. Check plate for colonies after approximately 12–18 h.
11. Prepare overnights of the transformed *E. coli* strains that harbor AARS/tRNA and reporter. Pick a single colony for each condition tested. Overnight in LB media with appropriate antibiotics (0.5–2 mL total). For use of M9-minimal and MOPS EZ Rich media, prepare a separate overnight with 1% glycerol added in LB. Additionally, grow overnight cultures of the following strains as controls: (1) strain expressing a fluorescent reporter lacking a TAG mutation (we suggest a TAC

codon in its place), (2) strain expressing a fluorescent reporter that contains in-frame TAG but lacks AARS/tRNA plasmid.

12. Prepare 96 deep-well plates (1 mL) to test incorporation efficiency (*see* **Note 3**). For each media condition, prepare media (*see* Table 3) with appropriate antibiotics added (1000x dilution from stock concentrations provided in Subheading 2). If multiple antibiotics are used, add half the indicated working concentration. Aliquot 300 μL to each test well, with three wells used for each experimental condition (triplicate measurement). Test nsAA concentrations of 0, 0.5, 1, and 5 mM by supplying these amounts to different wells for each AARS/tRNA being tested.
13. Inoculate cultures at 1:100 dilution in wells. Then, move to an orbital plate with shaking at 1000 RPM in an incubator set to optimal growth temperature and grow until mid-exponential phase.
14. At mid-exponential phase, induce both the AARS and GFP reporter and continue to grow at optimal temperature.
15. After 18 h, centrifuge cultures in the 96 deep-well plate at 4000× *g* for 10 min. Then, remove supernatant.
16. Resuspend cultures in 300 μL PBS and centrifuge again at 4000× *g* in the deep-well plate. Remove supernatant and resuspend in 300 μL PBS.
17. Then, dilute the resuspended cultures 1:3 with PBS in a black clear bottom 96-well plate and measure the fluorescence (depends on fluorescent reporter chosen) and optical density at 600 nm (*see* **Notes 4** and **5**; **step 3.1** summarized in Fig. 2).

3.2 Posttranslational Proofreading for Improving Purity of N-Terminally Incorporated nsAAs or Identification of Alternative AARS

1. Co-transform the C321.ΔClpS.Ub-UAG-sfGFP strain with pZE21/UBP1/ClpS_V65I and AARS/tRNA plasmids (*see* **steps 3.1.3–3.1.10**).
2. Prepare overnights of the transformed *E. coli* strains that harbor AARS/tRNA and proofreading system within plasmids and that harbor fluorescent reporter within the genome (*see* **step 3.1.11**).
3. The following day, set up 96 deep-well plates (1 mL) to test incorporation efficiency (*see* **step 3.1.4**).
4. Inoculate cultures at 1:100 dilution in wells. Then, move to an orbital plate with shaking at 1000 RPM in an incubator set to optimal growth temperature and grow until mid-exponential phase.
5. At mid-exponential phase, induce the AARS and the Ubp1/ClpS proteins GFP reporter and continue to grow at optimal temperature overnight.

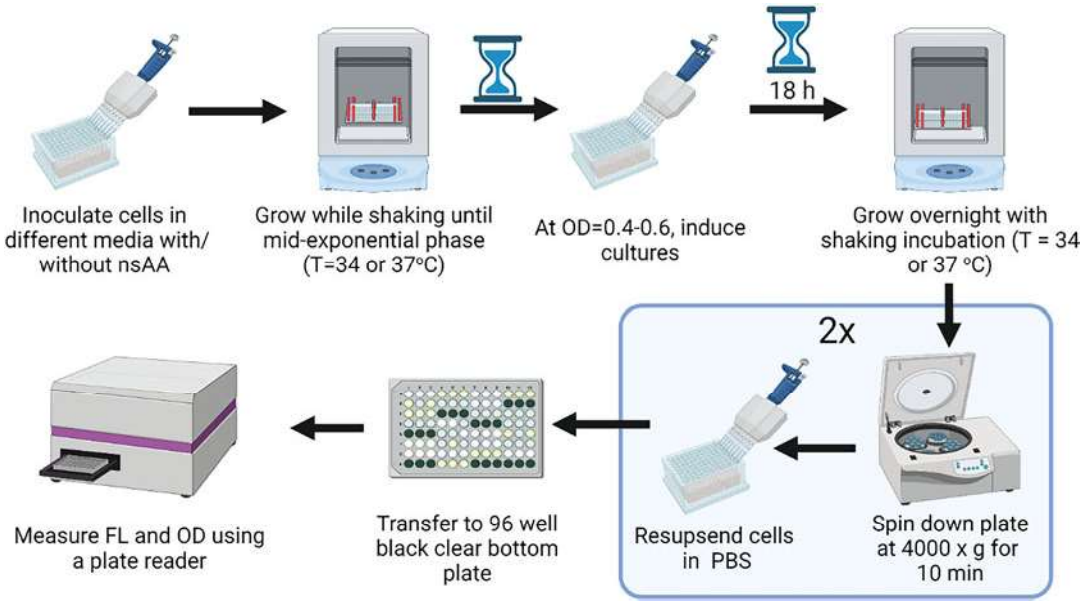

Fig. 2 Graphical depiction of the screening protocol described in **step 3.1**. (This figure was created with BioRender.com)

6. After 18 h, centrifuge cultures and wash 1x in 300 μL PBS.
7. Then, mix 150 μL of culture with 50 μL PBS in a black clear bottom 96-well plate and measure the fluorescence and optical density (*see* **Note 6**).

3.3 Site Selection for nsAA Incorporation Within Protein of Interest

1. After identifying an AARS/tRNA that appears to mediate incorporation of your desired nsAA within a GFP reporter protein, a next step is to then choose your target protein for the desired therapeutic application. Within this protein, you will need to choose one or more sites for nsAA incorporation. If there is flexibility in selecting a site for nsAA incorporation within the protein of interest, first identify which sites within the protein of interest are most amenable to mutation using a computational prediction tool such as SIFT [49] or PROVEAN [50].
2. If possible, target sites with high tolerance for mutation and structural similarity in the wild-type amino acid at that position to the nsAA (*see* **Note 7**). If a crystal structure exists for your protein, visualize the protein to qualitatively determine if the site is in a region desired for mutation (for example on the protein surface).
3. Then, assess the potential mRNA sequence context among potential indices for mutation. If possible, target a site where an adenine will be the first nucleotide following the TAG codon as this has been shown to significantly enhance protein expression.

4. Clone the gene for the therapeutic protein of interest with desired site for nsAA incorporation mutated to a TAG codon into either a plasmid with orthogonal origin of replication to the AARS/tRNA plasmid or in an operon with the AARS (*see* **Note 8**). Clone a 6x histidine tag or other affinity purification tag on a terminus of the protein for downstream purification. If the tag must be removed for downstream applications, additionally clone a corresponding protease cleavage site (TEV for an N-terminal tag or SNAC for a C-terminal tag for example). As a control, a version of the gene in this corresponding plasmid without the mutated TAG codon should be cloned as well.

3.4 Assessment of Efficiency of nsAA Incorporation During Therapeutic Protein Overexpression

1. Transform the plasmid(s) containing orthogonal AARS/tRNA and therapeutic gene of interest into the chosen strain (*see* **Note 9**).
2. Prepare overnights of the strain containing expression systems for AARS/tRNA and desired therapeutic protein with TAG mutation in addition to the positive control strain without TAG mutation in the therapeutic protein.
3. The next day, inoculate 30 mL cultures to initially test the expression of the chosen therapeutic protein in the optimal media from **steps 3.1–3.2** at varying postinduction temperatures. Here, prepare 30 mL of the media previously determined in baffled shake flasks with appropriate antibiotics added (1000x dilution). If multiple antibiotics are used, add half the working concentration. Inoculate cultures at 1:100 dilution. In these flasks, it is recommended to test the positive control (the no TAG codon strain), a condition without nsAA supplementation (TAG codon strain), and a condition with optimal nsAA supplementation from **steps 3.1–3.2** with one flask for each condition at each postinduction temperature.
4. Then, grow cultures at optimal growth temperature until mid-exponential phase. At this time, fully induce the AARS and reporter protein expression.
5. Then, grow cultures at various postinduction temperatures for 18 h (*see* **Note 10**).
6. After 18 h, centrifuge cells at 4000× *g*, dispose of supernatant, and resuspend the pellet in 1 mL of an appropriate lysis buffer. Then, lyse cells using preferred method of cell lysis (*see* **Note 11**).
7. Then, perform an SDS-PAGE gel to approximate protein titer across conditions. To start, measure the protein concentration of each lysate using a Bradford assay. Then, normalize protein concentrations to 1 mg/mL in 20 μL samples to mix with Lamelli SDS reducing sample buffer. Boil samples by incubation for 10 min at 95 °C and run 10 μL of sample on an SDS-

PAGE gel with a prestained protein ladder to identify protein containing fractions and confirm their size.

8. Visualize the SDS-PAGE gel by staining with Coomassie Blue and image the gel using visible light. Use an image processing software to approximate the relative protein titers obtained with the positive control, no nsAA, and nsAA containing conditions (*see* **Note 12**).

3.5 Purification and Characterization of Therapeutic Protein with nsAA

1. Prepare an overnight culture of 10 mL of the strain containing expression system for AARS/tRNA and the therapeutic protein of interest. Also, purify the protein of interest without TAG mutation using the same process as a control.
2. Using the optimal conditions from the previous steps, perform a 1 L overexpression experiment by scaling components appropriately and cycling temperatures as determined from **step 3.3**.
3. After 18 h of growth, centrifuge cells and resuspend in 30 mL of lysis buffer.
4. Lyse cells using preferred method of cell lysis (*see* **Note 13**).
5. Then, centrifuge the cells at over 15,000× *g* for 1 h and remove lysate from the pellet.
6. Next, purify the His-tagged protein using Ni-NTA chromatography (*see* **Note 14**).
7. Concentrate the combined fractions to <1 mL using a centrifugal spin concentrator with a cutoff 3–5 times smaller than the protein of interest. Then, perform buffer exchange by diluting 10:1 in a low-salt (20 mM sodium phosphate, pH 7.4) buffer. Repeat this process at least 3x.
8. Then, evaluate the protein using intact protein MS to ensure incorporation of the nsAA of interest and correct mass (*see* **Notes 15** and **16**).
9. To further purify protein for therapeutic applications, then perform anion exchange chromatography (*see* **Note 17**). If needed, buffer exchange the protein into a buffer with a pH at least 1 unit above the predicted isoelectric point.
10. Then, perform a qualitative SDS-PAGE to determine fractions containing protein of interest and assess protein purity. Combine fractions containing protein and concentrate using a centrifugal spin concentrator to 1 mL.
11. To further remove endotoxins for downstream therapeutic application, submit the concentrated sample to EndoTrap FPLC following manufacturer protocol.
12. Then, perform qualitative SDS-PAGE, combine fractions containing protein and concentrate fractions down to 1 mL in a centrifugal concentrator.

13. Then, perform desalting buffer exchange using size exclusion chromatography to remove any remaining impurities (*see* **Note 18**). Elute the protein in desired downstream buffer.
14. Perform a final SDS-PAGE to ensure protein purity and then pool and concentrate protein containing fractions to the desired product concentration.

4 Notes

1. Add 1 M NaOH as necessary to aid in solubility. If this is required, ensure that the media conditions provide adequate buffering when working concentrations of the nsAA are added. If needed, prepare stocks at lower concentrations.
2. The choice of strain will be dependent on the end application. Table 2 lists the applications for each strain. It may be useful to test both BL21(DE3) and C321.ΔA.opt to see which obtains sufficient purity/titer.
3. In initial screening, it is recommended to use MOPS EZ Rich media with an aromatic amino acid dropout, also available from Teknova.
4. Dilution at this stage is to ensure a lack of saturation of the plate reader and ensure optical density readings with linear response to cell growth (OD > 0.2 and OD < 1.0). If OD is not within this region, adjust dilution accordingly.
5. When analyzing the results from **step 3.1**, target conditions for nsAA incorporation where the fold-change in fluorescence normalized by optical density is at least 5. If those conditions are not met, proceed to **step 3.2**. In addition, use the non-TAG containing reporter expressing strain to evaluate the relative titer of protein produced using orthogonal AARS/tRNA as compared to the wild type.
6. Posttranslational proofreading can decrease the extent of standard amino acid incorporation. In some cases, especially for smaller and polar Phe derivatives, the proofreading machinery will also degrade the nsAA of interest. Completion of **step 3.2** should indicate if that is the case.
7. When targeting sites for mutation, it can be helpful, for example, if a phenylalanine-derived nsAA is to be introduced to target sites of aromatic standard amino acids in the wild-type protein (such as Phe, Tyr, or Trp).
8. When using plasmid systems with the protein of interest expressed in an operon with the AARS or genomically, the rate of variability can be reduced, as either the AARS/target protein transcript ratio is maintained, or the copy of the target protein is held constant.

9. If posttranslational proofreading is desired for use in expressing the modified protein, prepare a strain for testing with genetic inactivation of the ClpS adaptor protein for use without the Ub-UAG-sfGFP reporter. To do so, methods like multiplexed automated genome engineering (MAGE) can be used, for which prior protocols have been written [51, 52]. In addition, if posttranslational proofreading is used, ensure incorporation of the nsAA at the N-terminus is sufficient for the end application and that plasmid constructs are compatible and transformed appropriately.
10. Since many AARS systems are originally derived from thermophilic archaeal species, their optimal temperature for activity is often higher than that optimal for protein overexpression. A range of postinduction temperatures is recommended to find the optimum (for example, maintaining the optimal strain growth temperature, lower temperatures such as 30 °C and 20 °C, and maintenance of the optimal growth temperature for 5 h followed by reduction of temperature to 20 °C overnight.
11. We recommend glass bead lysis for higher throughput at this stage of the protocol. To perform glass bead lysis, mix 1 mL of cells with 0.05 mL of glass beads and then vortex using a Vortex Genie 2 at maximum RPM for 15 min. After this time, centrifuge the lysate at >15,000× *g* at 4 °C for 30 min and then collect the supernatant.
12. If the expression the nsAA containing sample is still quite low as compared to the positive control, we recommend cloning a small library of RBS variants (3–5) for the AARS and the therapeutic gene of interest using software available from De Novo DNA [53] and **repeating step 3.3** with varying RBS and using the optimal temperature from the initial screen.
13. We recommend use of a French Press homogenizer or sonication. For sonication, we recommend cycles of 2 min with 5 s on and 20 s off at 75% amplitude with 5 min rest between each cycle and a total of five cycles. We perform this process on ice and monitor the temperature to ensure it does not rise above 10 °C.
14. We recommend the use of a HisTrap™ FPLC column for this process following manufacturer's protocol. We recommend a final elution buffer containing 25 mM sodium phosphate, 250 mM imidazole, 300 mM NaCl, pH 7.4.
15. Several MS systems are capable of intact protein MS. We recommend UPLC-MS using a quadrupole Time-of-Flight Mass Spectrophotometer with a C18 UPLC column. An example method for intact protein MS using protein in low-salt buffer, inject the protein with an initial mobile phase of solvent A/B = 85/15 (solvent A, water, 0.1% formic acid; solvent B,

acetonitrile, 0.1% formic acid) held at 85/15 for 1 min followed by a gradient elution from (A/B) 85/5 to 5/95 over 5 min. Maintain the flow rated at 0.5 mL min^{-1}. Analyze the spectrum from m/z 500 to 2000 and deconvolute the spectra using maximum entropy in MassLynx software.

16. If a protease site was included for downstream removal, now is an optimal time to facilitate cleavage. Often, the protein of interest can be separated from the protease using a second round of Ni-NTA chromatography.
17. We recommend the use of a HiTrap Q HP FPLC column for anion exchange chromatography following manufacturer's protocols. Anion exchange can be a helpful initial step for endotoxin removal which is often important for therapeutic application [54].
18. We recommend the use of a Sepharose G-25 desalting column following manufacturer's protocols.

Acknowledgments

This work was supported by the National Science Foundation (NSF CBET #2032243).

References

1. Sletten EM, Bertozzi CR (2009) Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. Angew Chem Int Ed 48:6974–6998
2. Kolb HC, Finn MG, Sharpless KB (2001) Click chemistry: diverse chemical function from a few good reactions. Angew Chem Int Ed 40: 2004–2021
3. Cho H, Daniel T, Buechler YJ et al (2011) Optimized clinical performance of growth hormone with an expanded genetic code. Proc Nat Acad Sci 108:9060–9065
4. Hallam TJ, Wold E, Wahl A et al (2015) Antibody conjugates with unnatural amino acids. Mol Pharmaceutics 12(6):1848–1862
5. Kim CH, Axup JY, Dubrovska A et al (2012) Synthesis of bispecific antibodies using genetically encoded unnatural amino acids. J Am Chem Soc 134:9918–9921
6. Agarwal P, Bertozzi CR (2015) Site-specific antibody-drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development. Bioconjugate Chem 26:176–192
7. Tian F, Lu Y, Manibusan A et al (2014) A general approach to site-specific antibody drug conjugates. Proc Natl Acad Sci U S A 111:1766–1771
8. Roy G, Reier J, Garcia A et al (2020) Development of a high yielding expression platform for the introduction of non-natural amino acids in protein sequences. MAbs 12:1684749
9. BrickBio, https://www.brickbio.com/
10. Groff D, Carlos NA, Chen R et al (2022) Development of an E coli strain for cell-free ADC manufacturing. Biotechnol Bioeng 119: 162–175
11. Milla ME, Ptacin JL, Ma L et al (2019) THOR-707, a novel not-alpha IL-2, promotes all key immune system anti-tumoral actions of IL-2 without eliciting vascular leak syndrome (VLS). Ann Oncol 30:v501
12. Liu CC, Schultz PG (2006) Recombinant expression of selectively sulfated proteins in Escherichia coli. Nat Biotechnol 24:1436–1440
13. Liu CC, Choe H, Farzan M, Smider VV, Schultz PG (2009) Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. Biochemistry 48(37):8891–8898
14. Chen S, Maini R, Bai X et al (2017) Incorporation of phosphorylated tyrosine into proteins:

in vitro translation and study of phosphorylated IκB-α and its interaction with NF-κB. J Am Chem Soc 139:14098–14108
15. Grünewald J, Tsao ML, Perera R et al (2008) Immunochemical termination of self-tolerance. Proc Natl Acad Sci U S A 105: 11276–11280
16. Gauba V, Grunewald J, Gorney V et al (2011) Loss of CD4 T-cell-dependent tolerance to proteins with modified amino acids. Proc Nat Acad Sci 108:12821–12826
17. Hemphill J, Borchardt EK, Brown K et al (2015) Optical control of CRISPR/Cas9 gene editing. J Am Chem Soc 137:5642–5645
18. Jones CM, Robkis DM, Blizzard RJ et al (2021) Genetic encoding of a highly photostable, long lifetime fluorescent amino acid for imaging in mammalian cells. Proc Nat Acad Sci 12:11955–11964
19. Thyer R, Shroff R, Klein DR et al (2018) Custom selenoprotein production enabled by laboratory evolution of recoded bacterial strains. Nat Biotechnol 36:624–631
20. Yuan Z, Wang N, Kang G et al (2017) Controlling multicycle replication of live-attenuated HIV-1 using an unnatural genetic switch. ACS Synth Biol 6:721–731
21. Wang N, Li Y, Niu W et al (2014) Construction of a live-attenuated HIV-1 vaccine through genetic code expansion. Angew Chem Int Ed Engl 53:4867–4871
22. Mandell DJ, Lajoie MJ, Mee MT et al (2015) Biocontainment of genetically modified organisms by synthetic protein design. Nature 518: 55–60
23. Kunjapur AM, Napolitano MG, Hysolli E et al (2021) Synthetic auxotrophy remains stable after continuous evolution and in coculture with mammalian cells. Science 7:eabf5851
24. Johnson JA, Lu YY, Van DJA et al (2010) Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. Curr Opin Chem Biol 14:774–780
25. Singh-Blom A, Hughes RA, Ellington AD (2013) Residue-specific incorporation of unnatural amino acids into proteins in vitro and in vivo. In: Methods in molecular biology. Humana Press, Totowa, pp 93–114
26. Anderson JC, Wu N, Santoro SW et al (2004) An expanded genetic code with a functional quadruplet codon. Proc Natl Acad Sci USA 101:7566–7571
27. Magliery TJ, Anderson JC, Schultz PG (2001) Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. J Mol Biol 307:755–769
28. DeBenedictis EA, Carver GD, Chung CZ et al (2021) Multiplex suppression of four quadruplet codons via tRNA directed evolution. Nat Commun 12:1–13
29. Neumann H, Wang K, Davis L et al (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 464:441–444
30. Schmied WH, Tnimov Z, Uttamapinant C et al (2018) Controlling orthogonal ribosome subunit interactions enables evolution of new function. Nature 564:444–448
31. Malyshev DA, Dhami K, Lavergne T et al (2014) A semi-synthetic organism with an expanded genetic alphabet. Nature 509:385–388
32. Zhang Y, Ptacin JL, Fischer EC et al (2017) A semi-synthetic organism that stores and retrieves increased genetic information. Nature 551:644–647
33. Fischer EC, Hashimoto K, Zhang Y et al (2020) New codons for efficient production of unnatural proteins in a semisynthetic organism. Nat Chem Biol 16(5):570–576
34. Wan W, Tharp JM, Liu WR (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. Biochim Biophys Acta BBA Proteins Proteom 1844(6):1059–1070
35. Dumas A, Lercher L, Spicer CD et al (2015) Designing logical codon reassignment-expanding the chemistry in biology. Chem Sci 6:50–69
36. Italia JS, Addy PS, Wrobel CJJ et al (2017) An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat Chem Biol 13:446–450
37. Zhao H, Ding W, Zang J et al (2021) Directed-evolution of translation system for efficient unnatural amino acids incorporation and generalizable synthetic auxotroph construction. Nat Commun 12:1–12
38. Young TS, Ahmad I, Yin JA et al (2010) An enhanced system for unnatural amino acid mutagenesis in E. coli. J Mol Biol 395:361–374
39. Chatterjee A, Sun SB, Furman JL et al (2013) A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli. Biochemistry 52:1828–1837
40. Lajoie MJ, Rovner AJ, Goodman DB et al (2013) Genomically recoded organisms expand biological functions. Science 342(1979):357–360

41. Wang L, Schultz PG (2001) A general approach for the generation of orthogonal tRNAs. Chem Biol 8:883–890
42. Miller SM, Wang T, Liu DR (2020) Phage-assisted continuous and non-continuous evolution. Nat Protoc 15:4101–4127
43. Kunjapur AM, Stork DA, Kuru E et al (2018) Engineering posttranslational proofreading to discriminate nonstandard amino acids. Proc Natl Acad Sci USA 115:619–624
44. Chin JW, Santoro SW, Martin AB et al (2002) Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. J Am Chem Soc 124:9026–9027
45. Miyake-Stoner SJ, Refakis CA, Hammill JT et al (2010) Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. Biochemistry 49:1667–1677
46. Young DD, Young TS, Jahnz M et al (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. Biochemistry 50:1894–1900
47. Hohl A, Karan R, Akal A et al (2019) Engineering a polyspecific pyrrolysyl-tRNA synthetase by a high-throughput FACS screen. Sci Rep 9:1–9
48. Young DD, Jockush S, Turro NJ et al (2011) Synthetase polyspecificity as a tool to modulate protein function. Bioorg Med Chem Lett 21: 7502
49. Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812–3814
50. Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31:2745–2747
51. Gallagher RR, Li Z, Lewis AO et al (2014) Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. Nat Protoc 9:2301–2316
52. Wannier TM, Nyerges A, Kuchwara HM et al (2020) Improved bacterial recombineering by parallelized protein discovery. Proc Natl Acad Sci USA 117:13689–13698
53. Farasat I, Kushwaha M, Collens J et al (2014) Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Mol Syst Biol 10:731
54. Chen RH, Huang C-J, Newton BS et al (2009) Factors affecting endotoxin removal from recombinant therapeutic proteins by anion exchange chromatography. Protein Expr Purif 64:76–81
55. Wang L, Brock A, Herberich B et al (2001) Expanding the genetic code of Escherichia coli. Science 292(1979):498–500
56. Lee BS, Kim S, Ko BJ et al (2017) An efficient system for incorporation of unnatural amino acids in response to the four-base codon AGGA in Escherichia coli. Biochim Biophys Acta BBA Gen Subj 1861:3016–3023
57. Venkat S, Sturges J, Stahman A et al (2018) Genetically incorporating two distinct post-translational modifications into one protein simultaneously. ACS Synth Biol 7:689–695
58. Wan W, Huang Y, Wang Z et al (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. Angew Chem Int Ed 49: 3211–3214
59. Blight SK, Larue RC, Mahapatra A et al (2004) Direct charging of tRNACUA with pyrrolysine in vitro and in vivo. Nature 431:333–335
60. Mukai T, Yamaguchi A, Ohtake K et al (2015) Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli. Nucleic Acids Res 43:8111–8122
61. Ding W, Zhao H, Chen Y et al (2020) Chimeric design of pyrrolysyl-tRNA synthetase/tRNA pairs and canonical synthetase/tRNA pairs for genetic code expansion. Nat Commun 11:1–13
62. Kuznetsov G, Goodman DB, Filsinger GT et al (2017) Optimizing complex phenotypes through model-guided multiplex genome engineering. Genome Biol 18:1–12