Incorporation of non-standard amino acids into proteins: Challenges, recent achievements, and emerging applications

Xing Jin,[1] Oh-Jin Park,[1,2] and Seok Hoon Hong[1*]

[1]Department of Chemical and Biological Engineering, Illinois Institute of Technology,
Chicago, IL 60616, USA
[2]Department of Biological and Chemical Engineering, Yanbian University of Science and Technology,
Yanji, Jilin, People's Republic of China

*Corresponding author: shong26@iit.edu
Tel: (+1) 312 567 8950

Abstract

The natural genetic code only allows for 20 standard amino acids in protein translation, but genetic code reprogramming enables the incorporation of non-standard amino acids (NSAAs). Proteins containing NSAAs provide enhanced or novel properties and open diverse applications. With increased attention to the recent advancements in synthetic biology, various improved and novel methods have been developed to incorporate single and multiple distinct NSAAs into proteins. However, various challenges remain in regard to NSAA incorporation, such as low yield and misincorporation. In this review, we summarize the recent efforts to improve NSAA incorporation by utilizing orthogonal translational system optimization, cell-free protein synthesis, genomically recoded organisms, artificial codon boxes, quadruplet codons, and orthogonal ribosomes, before closing with a discussion of the emerging applications of NSAA incorporation.

Keywords non-standard amino acids, genetic code expansion, synthetic biology, orthogonal translational system

Introduction

20 In nature, the genetic code defines how 64 triplet codons translate into 20 standard amino acids. However, synthetic expansion of the genetic code enables the direct translational incorporation of non-standard amino acids (NSAAs) (Li and Liu 2014). To date, more than 200 distinct NSAAs have been incorporated into proteins (Vargas-Rodriguez et al. 2018). The incorporation of NSAAs provides enhanced or novel protein properties and thus enables diverse applications of this technology (Liu and 25 Schultz 2010; Kang *et al.* 2018). The incorporation of NSAAs helps us manipulate the physiochemical and biological properties of proteins, thereby enabling a range of promising applications (Liu and Schultz 2010), such as new approaches to the probing, imaging, and control of protein function and the precise engineering of therapeutics (Chin 2017). Although numerous previous studies have illustrated NSAA incorporation, many challenges remain in the practical application of this technology. Several seminal 30 works review advances in *in vitro* genetic code reprogramming (Hong *et al.* 2014a; Quast *et al.* 2015; Katoh et al. 2018), the *in vivo* incorporation of NSAAs in enzyme engineering (Ravikumar et al. 2015), the importance of tRNA design (Uhlenbeck and Schrader 2018), the natural incorporation of NSAAs and challenges of their misincorporation into recombinant proteins (Reitz et al. 2018), and the study of protein function (Neumann-Staubitz and Neumann 2016). In this review, we focus on the recent achievements 35 aimed at overcoming the challenges of NSAA incorporation and the emerging applications of proteins containing NSAAs.

Mechanisms of non-standard amino acid incorporation

40 In the natural protein translation process, an amino acid is first charged onto a transfer RNA (tRNA) by an aminoacyl-tRNA synthetase (aaRS) using a specific anti-codon. Then, the amino acid-charged tRNA is delivered by the elongation factor Tu (EF-Tu) to the ribosome, where translation occurs. However, expansion of the genetic code provides the possibility of incorporating NSAAs by assigning each NSAA to one of the redundant sense or stop codons (Forster *et al.* 2003) or to extra codons beyond 45 the 64 triplet codons (Katoh *et al.* 2018). Among the diversity of genetic code expansion approaches, amber suppression, in which an NSAA is assigned to the amber codon (UAG), is preferably used for NSAA incorporation because the amber codon is rarely used in organisms (*e.g.*, only 321 of 4505 stop codons in *E. coli* MG1655 (Lajoie et al. 2013; Keseler et al. 2017)), in contrast to the other two stop codons.

50 The site-specific incorporation of NSAAs uses an orthogonal translation system (OTS) similar to the natural translation process (**Fig. 1**). The OTS requires exploitation of the orthogonal aaRS/tRNA pair that

effectively charges the specific NSAA onto the tRNA with the reassigned codon. An ideal orthogonal aaRS/tRNA pair does not cross-react with the endogenous amino acids, aaRSs, or tRNAs in the host cell but can be recognized by the host ribosome (Dumas *et al.* 2015). To achieve site specificity, the orthogonal aaRS/tRNA pairs are generated from phylogenetically distinct organisms (Kim *et al.* 2013). For example, the tyrosyl(Tyr)-RS/$tRNA_{CUA}$ pair derived from *Methanocaldococcus jannaschii* is orthogonal to that from *Escherichia coli* and other bacteria, and pyrrolysyl(Pyl)-RS/$tRNA_{CUA}$ pairs from certain methanogens (*e.g.*, *Methanosarcina barkeri* and *Methanosarcina mazei*) are orthogonal in both prokaryotic and eukaryotic cells (Chin 2014). The orthogonal aaRS/tRNA pairs have been engineered to incorporate different types of NSAAs and improve the incorporation efficiency (Dumas *et al.* 2015).

Challenges and improvements of NSAA incorporation

Issues of poor NSAA incorporation efficiency, orthogonality between NSAAs and standard amino acids, and the ability to incorporate multiple distinct NSAAs must be addressed so that NSAA incorporation can be comparable to native amino acid incorporation. Recently, various approaches have been developed and further improved to address these challenges. These approaches include optimizing translation machineries, using *in vitro* protein synthesis, and applying genome engineering.

OTS optimization

The development of numerous OTSs promises the potential for advancements in the incorporation of NSAAs; nonetheless, the slow charging of NSAA to orthogonal tRNA, slow delivery by the elongation factor Tu (EF-Tu), and slow translation in the ribosome impede NSAA incorporation efficiency (O'Donoghue *et al*. 2013; Hong *et al.* 2014a). Previous studies improved NSAA incorporation efficiency by engineering orthogonal aaRS/tRNA pairs (Young and Schultz 2010; Chatterjee *et al.* 2012), EF-Tu (Park *et al.* 2011), and controlling the transcription and translation rates (Young and Schultz 2010; Chatterjee *et al.* 2013). However, these advances have been restricted to proteins containing a single or few NSAAs with low efficiency and accuracy. Further improvement of the NSAA incorporation efficiency is still required. Recently, Gruic-Sovulj group reported the importance of the aaRS and tRNA in post-transfer editing for high-fidelity translation of the amino acids (Dulic et al. 2018), suggesting that understanding of the binding and catalytic functions of aaRSs to substrates is important to promote orthogonal translation.

While previous studies have mostly focused on the optimization of components in the orthogonal translational system with the incorporation of one NSAA, increasing the number of NSAA incorporation

sites in a protein significantly decreases the incorporation efficiency by generating truncated immature proteins or the misincorporation of endogenous amino acids (Johnson *et al.* 2011). Gan *et al.* evolved aaRS and EF-Tu and tuned the expression of the evolved translation machinery components in a single vector, thereby enhancing the NSAA incorporation efficiency (Gan *et al*. 2017). These researchers found
90 that mutant aaRS improved activity by 8-fold, while engineered EF-Tu enhanced the production yield by 4-fold in single-site NSAA incorporation; however, the co-production of aaRS and EF-Tu reduced the production of superfolder green fluorescent protein (sfGFP) compared to that of aaRS only. Therefore, 27 plasmids containing diverse promotor combinations with orthogonal aaRS, orthogonal tRNA, and EF-Tu were constructed to optimize the NSAA incorporation efficiency. By combining the best mutants and
95 tuning the expression levels, the researchers achieved a 2-, 10-, and 5-fold increase in the incorporation of NSAA into 1, 3, and 5 UAG sites, respectively, compared to the progenitor enzymes (Gan *et al.* 2017). This study demonstrates that EF-Tu engineering (improving delivery) combined with aaRS/tRNA optimization (improving NSAA charging) improves the NSAA incorporation efficiency. Moreover, the authors suggested that coordinate optimization of translation machinery as a complex system (including
100 codons, tRNA, aaRS, EF-Tu and the ribosome) further improves the NSAA incorporation efficiency. Bryson *et al.* employed phage-assisted continuous evolution (PACE) selections to rapidly produce highly active and selective orthogonal aaRSs through hundreds of generations of evolution (Bryson *et al.* 2017). The PACE of a chimeric *Methanosarcina* spp. PylRS improved its enzymatic activity (k_{cat}/K_M^{tRNA}) by 45-fold compared to the parent enzyme, and transplantation of the evolved mutations into other PylRS-
105 derived synthetases increased yields of proteins containing NSAAs up to 9.7-fold. In addition, the PACE of a promiscuous *M. jannaschii* TyrRS variant significantly improved the selectivity for the site-specific incorporation of *p*-iodo-L-phenylalanine. These studies offer new aaRSs with the enhanced utility of OTSs.

110 ***Cell-free protein synthesis***

Studies of NSAA incorporation have predominantly been conducted using live cells. However, NSAA incorporation *in vivo* often causes problems of toxicity to the host cell and poor delivery of the NSAA inside the cells. Cell-free protein synthesis (CFPS) is a protein synthesis methodology that does
115 not require living cells. CFPS utilizes crude cell extracts to enable transcriptional and translational machinery for protein production (Swartz *et al.* 2004). The addition of energy sources, cofactors, buffers, salts, nucleotides, and amino acids further mimics cytoplasmic environments to enhance the CFPS (Jin and Hong 2018). To enable NSAA incorporation *in vitro*, NSAA, orthogonal tRNA, and orthogonal aaRS are also required (**Fig. 2A**). The toxicity associated with the overexpression of the orthogonal tRNA/aaRS

120 pair is not considered because no live cells are involved during the cell-free reaction (Hong *et al.* 2014a), and the openness of the reaction environment without cell walls and membranes enables the direct addition of NSAAs into the translation reaction. In addition to supplying OTSs in CFPS reactions (Noren *et al.* 1989), several other approaches have recently been applied to improve NSAA incorporation. One method is to co-express the orthogonal tRNA and the target gene simultaneously (Albayrak and Swartz
125 2013). A transzyme construct containing a hammer-head ribozyme that cleaves the 5'-end of tRNA upon transcription (Fechter *et al.* 1998) increased orthogonal tRNA supplement in the CFPS reaction. The use of optimized orthogonal tRNA enhances the recognition of tRNA by *E. coli* EF-Tu (Young *et al.* 2010) and further enhanced the NSAA incorporation efficiency compared to the original orthogonal tRNA. With this approach, 0.9 ~ 1.7 mg/mL of modified soluble GFP containing NSAA was produced with 50 ~ 88%
130 incorporation efficiency (Albayrak and Swartz 2013). Bundy and colleagues prepared the cell extracts to contain active orthogonal tRNA and aaRS by expressing the orthogonal components during the cell growth in the optimized fermentation condition prior to cell lysing. Modified CFPS systems using OTS-containing extract improved the yield of proteins with NSAA by as much as 150% compared to the standard CFPS when OTS are produced or supplied during the reaction (Smith *et al.* 2014). Moreover,
135 CFPS was developed as a rapid screening tool for NSAA-containing protein production. Schinn *et al.* investigated the efficiency of stop codon suppression at various positions within a protein by utilizing CFPS technology and screening 24 different T4-lysozyme mutants with randomly assigned and distinct NSAA incorporation sites (Schinn *et al.* 2017). The efficiency of NSAA incorporation was measured with green fluorescence by tagging the NSAA-containing T4-lysozyme with a GFP fragment that was self-
140 assembled with the rest of the GFP fragments to form active GFP (Cabantous *et al.* 2005). The CFPS approach enabled the rapid characterization of the positional effect of NSAA incorporation at different locations within the protein. Taken together, CFPS provides the rapid and simple production of modified proteins with NSAA with high incorporation efficiency and yield compared to *in vivo* incorporation.

145 ***Genomically recoded strain development***

The NSAA incorporation efficiency with amber codon suppression is inhibited by release factor 1 (RF1), which terminates protein translation by recognizing the amber codon, although the deletion of RF1 is lethal to the host cell (Voloshchuk and Montclare 2010; Hong *et al.* 2014a). However, a recent
150 advancement in genetic engineering enabled the construction of *E. coli* with an RF1 knockout but retained viability by fixing release factor 2 to be more efficient in terminating all UAA codons that are found in-frame after the UAG stop codon (Johnson *et al.* 2011), as well as by improving the efficiency of the UAG triplet decoding process (Ohtake *et al.* 2012). Lajoie *et al*. first constructed a genomically recoded *E. coli*

strain by reassigning all 321 UAG codons to UAA codons *via* multiplex automatable genome engineering
155 (MAGE) and thereby deleted RF1 without any growth defects (Lajoie *et al*. 2013). These researchers
demonstrated that the genomically recoded strain exhibits a greater NSAA incorporation efficiency
compared to the native strain.

Genomically recoded *E. coli* has become a foundational platform for improving NSAA incorporation
in vivo as well as *in vitro* (**Fig. 2B**). Recently, Amiram *et al*. evolved the orthogonal translation machinery
160 to enhance the NSAA incorporation efficiency and achieve multi-site NSAA incorporation by using
genetically recoded *E. coli* as a host (Amiram *et al*. 2015). The authors screened a library of orthogonal
aaRS variants to improve the efficiency of three sites of NSAA incorporation and found that one aaRS
variant enhanced GFP production by 25-fold compared to the progenitor aaRS. Furthermore, the
engineered orthogonal aaRS was able to successfully incorporate 30 NSAAs at one time into elastin-like
165 polypeptides at 50 mg/L production yield and over 95% accuracy. Hong *et al*. utilized partially recoded
RF1-deleted *E.coli* to prove the importance of removing RF1 for activating efficient and multiple site-
specific NSAA incorporations (Hong *et al*. 2014b). The RF1-deficient cell extract prepared from the RF1-
deficient strain improved the yield of modified GFP containing an NSAA p-propargyloxy-L-
phenylalanine by more than 250% compared to the RF1-present cell-free reaction, and sfGFP derived
170 from multiple NSAAs and containing two and five amber sites was easily achieved without RF1 (Hong *et
al*. 2014b). RF-deficient CFPS has been further advanced by eliminating negative effectors in the *E. coli*
cell extract to enhance the efficiency and accuracy of single and multiple NSAA incorporation strategies
(Hong *et al.* 2015; Martin *et al.* 2018). By disrupting the genes that encode negative effectors (*endA*, *gor*,
rne, and *mazF*) *via* MAGE (**Fig. 2B**) in the completely recoded *E. coli* genome (Martin *et al*. 2018), cell
175 extract activity was significantly increased, yielding sfGFP in concentrations that were 1,700 mg/L, over
4-fold greater than those of the progenitor strains. They further demonstrated that the incorporation of 40
p-acetyl-L-phenylalanine residues into an elastin-like polypeptide (ELP) results in an 85% ELP
production yield and 98% accuracy with the cell extract upon the deletion of these negative effectors.

Adaptive evolution also has been applied to convert a native amino acid into an NSAA. Hoesl *et al.*
180 demonstrated that a long-term cultivation experiment in defined synthetic media resulted in the evolution
of cells that were capable of surviving in the presence of NSAA L-β-(thieno[3,2-b]pyrrolyl)alanine
instead of tryptophan (Hoesl *et al.* 2015). The Shultz group recently reported that the functional
replacement of amino acid residues with NSAAs in proteins generated NSAA-dependent organisms
whose growth is strictly dependent on the presence of the NSAA, p-benzoyl-L-phenylalanine (Koh *et al.*
185 2017), and 3-methyl-L-histidine (Gan *et al.* 2018). Generating organisms that are dependent on the
presence of an NSAA is useful for the generation of live vaccines and the biological containment of
recombinant organisms (Rovner *et al.* 2015).

Artificial codon box division

190

NSAA incorporation with amber suppression has been widely used and studied, but this method can only incorporate one type of NSAA. To simultaneously incorporate various NSAAs into one protein, artificial codons have been generated (Link and Tirrell 2005; Krishnakumar and Ling 2014). In the natural genetic codon box, 18 of the standard amino acids, except for methionine and tryptophan, are 195 coded by multiple triplet codons, which means that there are duplicate codons that can be reassigned to NSAAs (**Fig. 2C**). Iwane *et al*. developed the artificial division of natural codon boxes to incorporate various NSAAs (Iwane *et al.* 2016). They replaced redundant $tRNA_{SNN}$ (N = A, C, G or U; S = G or C) with the NSAA-pre-charged $tRNA_{SNN}$. Using this artificial division of the genetic box, the authors successfully incorporated three distinct NSAAs into a 32-mer linear peptide and five distinct NSAAs into 200 a 14-mer macrocyclic peptide *in vitro*. More recently, Zhang *et al.* utilized an unnatural base pair (dNaM-dTPT3) to direct the incorporation of NSAA into proteins (Zhang *et al*. 2017). The study authors generated an *E. coli* strain that contains an unnatural base pair in its plasmid DNA. Upon the transcription of DNA containing dNaM and dTPT3 into mRNAs, sfGFP containing the unnatural codon AXC (X denotes NaM) was decoded by an unnatural tRNA containing the anticodon GYT (Y denotes TPT3) at the 205 ribosome. By adding another unnatural codon GXC and using an evolved orthogonal aaRS/tRNA pair, the authors demonstrated that two distinct NSAAs could be site-specifically incorporated into sfGFP. Rewriting a genetic code with artificial codon box divisions demonstrates the possibility of incorporating multiple distinct NSAAs into a single protein.

210 ### *Quadruplet codon and orthogonal ribosome*

In addition to the artificial division of the triplet codon box, a quadruplet codon was used to simultaneously incorporate three distinct NSAAs into streptavidin, although a low expression level was obtained due to the resulting truncated byproducts (Ohtsuki *et al.* 2005), which implies that the native 215 ribosome is an inefficient decoder of the quadruplet codon. Later, this method was improved by developing an orthogonal ribosome (Ribo-Q1) by assessing 11 saturation mutagenesis libraries in the 16S ribosomal RNA (Neumann *et al.* 2010) of the previously evolved ribosome that had been shown to increase the decoding of amber codons (Wang et al. 2007). Ribo-Q1 enhanced the incorporation efficiency and fidelity of NSAAs with the amber codon and a series of quadruplet codons when combined 220 with mutually orthogonal aaRS/tRNA pairs (**Fig. 2D**). This approach illustrates the possibility of incorporating more than 200 NSAAs into one protein *via* quadruplet codons. Recently, Orelle *et al.*

attempted to separate orthogonal translation from native translation by developing ribosomes with tethered subunits (Orelle *et al.* 2015). Covalently linked 16S and 23S ribosomal RNAs enabled a functional ribosome that was only able to translate orthogonal mRNA without affecting natural translation
225 by the native ribosome (**Fig. 2D**). Ribosome design and engineering provide the potential for the development of synthetic biomanufacturing machines (d'Aquino et al. 2018). Improving the NSAA incorporation efficiency and expanding the genetic code will be further advanced by the evolution of orthogonal ribosomes.

230 Emerging Applications of NSAA incorporation

NSAA incorporation enables the creation of novel functional proteins and biopolymers that have been either difficult or impossible to achieve due to the limited repertoire of 20 standard amino acids (Mukai *et al.* 2011). With advancements in NSAA incorporation, innovative emerging applications in medical and
235 industrial areas have recently been demonstrated.

Real-time tracking and imaging in vivo

Resident microbes in the gut microbiome closely interact with host cells and affect the human
240 physiology. Thus, developing an understanding of these interactions will result in new diagnostic, prognostic, and therapeutic capabilities (Shreiner *et al.* 2015). An ideal real-time tracking system of gut microbes *in vivo* should be noninvasive and provide spatiotemporal localization; unfortunately, the existing methods are either invasive or indirect. Recently, the Joshi group developed a real-time tracking method employing NSAA incorporation *in vivo* within the gastrointestinal tract (Praveschotinunt *et al.*
245 2018). The researchers incorporated p-azido-L-phenylalanine into a cell surface protein, CsgA, in probiotic *E. coli* using amber suppression and labeled the engineered strain with a Cy5 dye and dibenzocycloctyl *via* a copper-free click reaction. This approach showed a high fluorescence intensity within 24 h after oral administration of the dye and while using an *in vivo* imaging system that enabled the spatiotemporal monitoring of specific microbial populations in the gastrointestinal tract.
250 Installing artificial probes and reactive moieties directly onto proteins in the live cell is also an interesting objective that could be achieved through the incorporation of NSAAs. A robust fluorescence-based assay to evaluate the efficiency of orthogonal pairs in mammalian cells enabled the screening of tens to hundreds of orthogonal aaRS/tRNA variants (Serfling *et al.* 2018). This approach was used both to incorporate chemical probes into G-protein-coupled receptors for photo-crosslink mapping and to achieve
255 biocatalyst-free receptor labeling with a fluorescent dye *via* biorthogonal chemistry in the live cell.

Bioorthogonal reactions in living systems have been increasingly studied, and pretargeted imaging methods and therapies using reactive monoclonal antibodies are some of the most developed applications toward clinical testing (Devaraj 2018). Reactive NSAAs will contribute to further improvement and applications of such pretargeted approaches. In addition, the development of peptide- and protein-
260 nanoparticle conjugates *via* NSAA incorporation can be used for biomedical applications such as nano-drug delivery vehicles, imaging species, and active therapeutics (Spicer *et al.* 2018).

Biomolecular targeting

265 Because some NSAAs are able to bind or react with small molecules and metal ions, genetically encoded fluorescent probes for the detection of ions such as Cu^{2+} (Ayyadurai *et al.* 2011), Mn^{3+} (Liu *et al.* 2014), and peroxynitrite ($ONOO^-$) (Chen *et al.* 2013) have been studied. Recently, the Schultz group developed a genetically encoded fluorescent probe to monitor the protein sirtuin activities in living cells (Xuan *et al*. 2017). The misregulation of sirtuin activities leads to a number of diseases including cancer,
270 type 2 diabetes, and cardiovascular disease (Imai and Guarente 2014). NSAA N-ε-acetyl-L-lysine was incorporated by replacing lysine-85 in the enhanced GFP (EGFP), which resulted in a non-fluorescent labeler, but upon subsequent deacetylation by sirtuin, fluorescent EGFP was formed. This fluorescent probe could be used to monitor sirtuin activity in bacteria and in mammalian cells (Xuan *et al.* 2017). More recently, chemoselective reactive pyrrolysine anologues were site-specifically incorporated into
275 proteins using a CFPS platform allowing site-direct fluorescence labeling of target proteins with wide range of different chemoselective reactions (Gerrits et al. 2019).

NSAA incorporation also allows for the identification of protein-protein interactions (Wang 2017a; Nguyen *et al*. 2018). Functions of proteins previously considered to be predominately useless or a nuisance can be revealed *via* bioreactive NSAA incorporation (Wang 2017b). Yang *et al.* demonstrated
280 that NSAAs developed with a proximity-enabled reactive group, alkyl bromide, induces chemical cross-linking with the cysteine residues of the interacting protein, thereby enabling the identification of low-affinity protein binding, enzyme-substrate interaction, and endogenous protein interaction (Yang *et al.* 2017). Proximity-dependent cross-linking was further strengthened in combination with chemical cross-linking mass spectrometry, allowing researchers to explore the functions and structures of protein
285 complexes (Yang *et al.* 2018). New chemical cross-linker functions such as sulfate-fluoride exchange (Wang *et al.* 2018) and oxidation-induced generation (Shang et al. 2018) are being developed and tested for proximity-enhanced protein-protein interactions. The site-specific incorporation of NSAAs was also used to mimic post-translational modification that sometimes recruits interacting partners. Zheng *et al.* used dual NSAA mutagenesis and found that lysine-23 acetylation at histone 3 covalently captured its

290 interacting protein (Zheng *et al.* 2018). Ultimately, genetically encoded chemical cross-linkers that utilize bioactive NSAAs are crucial for identifying and understanding protein interactions.

Biocontainment

295 In nature, genetic elements can be transferred between certain single- and multicellular organisms *via* the process of horizontal gene transfer (HGT) (Kubyshkin *et al*. 2018), which increases the diversity of living organisms and allows them to survive in harsh environments (Aminov 2011). It is also possible that the occurrence of HGT from genetically engineered organisms to native organisms may cause unexpected threats to public safety or the environment (Moe-Behrens *et al.* 2013). Preventing the accidental release of 300 engineered organisms is of utmost concern and hence such studies must be strictly regulated and carried out only in closed industrial settings. However, to use new technology developed in this field in settings such as bioremediation efforts, healthcare applications, and agriculture would require intrinsic biocontainment approaches to block the distribution of engineered organisms. Alienating the genetic code from the current 20 amino acids *via* the incorporation of NSAAs could enable the biocontainment 305 mechanisms in engineered cells, effectively providing them with a "genetic firewall," *i.e.,* orthogonality, which might prevent HGT to natural systems (Kubyshkin and Budisa 2017). Mandell *et al.* computationally redesigned essential enzymes in the first genomically recoded organism derived from *E. coli* (Lajoie *et al.* 2013) to require an NSAA for proper translation, folding, and function so that the resulting cells cannot metabolically escape biocontainment mechanisms (Mandell *et al.* 2015). They 310 observed undetectable escape levels with double- and triple-enzyme synthetic auxotrophs during 14 days of monitoring. A similar NSAA incorporation approach that instead targeted different essential enzymes (Rovner *et al.* 2015) also demonstrated that NSAA incorporation could be the superior method for building an intrinsic safeguard into genetically engineered organisms, enabling the application of these in open systems. Both studies reported that synthetic auxotrophs requiring NSAAs impeded HGT and 315 mutagenesis for evolutionary escape from the biocontainment strategy.

Synthetic auxotrophy was also generated by incorporating 3-iodo-l-tyrosine into an antidote protein against a toxin, colicin E3 (Kato 2015). Upon the removal of 3-iodo-l-tyrosine, decay of the antidote thus activated colicin E3, which killed the host cells with an escaper frequency of 1.4 mutations per 10^5 cell divisions. This study demonstrated that the toxin-antidote approach with NSAA incorporation can be used 320 to establish a biocontainment system by only plasmid introduction without genome editing.

Biomaterials

The Swartz group was the first to demonstrate the feasibility of making polymers composed of
325 functional proteins covalently bonded to each other using NSAAs as linkers (Albayrak and Swartz 2014). Multiple copies of the NSAAs p-azido-L-phenylalanine and p-propargyloxy-l-phenylalanine were site-specifically incorporated into sfGFP by CFPS, and the direct polymerization of sfGFP was achieved by azide-alkyne cycloaddition between the two NSAA residues. The protein polymers retained more than 63% fluorescence after coupling. This demonstration suggests that functional biopolymer production can
330 be achieved by using simple click chemistry to insert NSAAs into the protein of interest.

The adhesive proteins of mussels exhibit incredible underwater adhesion capabilities that have been attracted significant attention due to their potential for application in the medical and industrial fields (Kord Forooshani and Lee 2017). Recently, the Budisa group demonstrated the production of 3,4-dihydroxyphenylalanine (DOPA)-rich adhesive proteins *via* genetic code expansion by using
335 computational design and genetic selection methods (Hauf *et al.* 2017). The study authors engineered aaRSs to enhance the incorporation of the photocaged NSAA ortho-nitrobenzyl DOPA. The NSAA was incorporated *in vivo* using a 5 site-specific strategy that elevated the spatiotemporal and tunable adhesion properties (ONB cleavage and generation of catechol moiety necessary for adhesive properties) when triggered by exposure to UV light. This strategy provides a direct means of producing recombinant
340 DOPA-based wet bioadhesives without post-translational modification.

In addition, the repertoire of NSAAs can be applied to improve fluorescent proteins that suffer from lower fluorescence quantum yield and inferior photo-stability (Wang et al. 2003). The presence of an aromatic amino acid at the chromophore triads of fluorescent proteins is essential. However, only four aromatic residues (histidine, phenylalanine, tryptophan, and tyrosine) are available among the 20 standard
345 amino acids. Therefore, standard amino acid replacement at the chromophore triad to construct GFP variants cannot cover the far-red spectral region (>600 nm) of fluorescence, limiting fluorescent microscopic approaches that require long-wavelength biological markers, *e.g.*, deep tissue imaging (Shcherbakova *et al.* 2012). The Budisa group introduced aromatic NSAAs in the replacement of tryptophan residues in the chromophore triad by using a selective pressure incorporation approach
350 (Baumann *et al.* 2018). An electron-donating 4-aminotryptophan incorporation resulted in a strong red-shift of the emission wavelength that was thermodynamically more stable than its predecessor. The NSAA approach enables the development of a tailored spectral diversity of fluorescent proteins that can be used in *in vivo* imaging of various biological samples.

355 ***Antimicrobial peptides***

Antimicrobial peptides (AMPs) are small oligopeptides (5 to 100 amino acids) (Bahar and Ren 2013) produced by living organisms for innate immune responses (Peters *et al.* 2010). AMPs exhibit broad antimicrobial activity to microbes including viruses, bacteria, protozoa, and fungi (Giuliani *et al.* 2007).
360 NSAA incorporation could be used to achieve chemical diversification of recombinant AMPs by improving their pH, protease resistance, solubility, oral availability, and half-life (Baumann *et al.* 2017). Specifically, ribosomally synthesized and post-translationally modified peptides (RiPPs) (Hudson and Mitchell 2018) can be significantly improved by the incorporation of NSAAs that mimic post-translational modification and broaden the chemical space (Baumann et al. 2017). Nisin, perhaps the most
365 thoroughly studied bacteriocin, is a broad spectrum lanthipeptide and a subclass of a large family of RiPPs. The Schultz group incorporated several NSAAs into nisin by expressing the *nisA* encoding structural peptide, *nisB* encoding dehydratase, and *nisC* encoding cyclase with the coexpression of OTS for amber suppression in *E. coli* and found that α-chloroacetamide-containing NSAA incorporation resulted in recombinant nisin variants with novel macrocyclic topologies (Zambaldo *et al.* 2017). The
370 Budisa group employed the selection pressure incorporation method to enable the residue-specific incorporation of six proline analogs into the antimicrobial peptide nisin (Nickling *et al.* 2018). These researchers also produced bioactive nisin variants containing Nε-Boc-L-lysine both in *Lactobacillus lactis* and *E. coli.* van der Donk and colleagues produced two analogues of NSAA-containing lantibiotics, lacticin 481 and nisin, using genomically recoded organisms *in vivo* (Kakkar *et al.* 2018). In addition, the
375 study authors developed the yeast surface and phage display of lanthipeptides to identify the variants with enhanced biological activities (Hetrick *et al.* 2018). Therefore, the employment of NSAA incorporation in RiPPs engineering can broaden the generation of lantibiotic variants with new biological activities. Along with the use of various algorithms, such as machine learning and evolutionary algorithms, to design effective antimicrobial peptides, the incorporation of NSAAs into AMPs will open a pathway for
380 identifying AMP variants and optimizing improved AMPs (Porto *et al.* 2017; Yoshida *et al.* 2018; Lee *et al.* 2018). It is possible that the same NSAA incorporation approach can be applied to the engineering of large antimicrobial toxins such as colicin and pyocin to produce chimeric antimicrobials (Kageyama et al. 1996) on advanced CFPS platforms (Jin et al. 2018). Novel antimicrobials containing NSAAs may be used to eradicate antibiotic-resistant bacteria which cause an urgent healthcare problem (Banin et al.
385 2017).

Biocatalysis

Enzyme engineering has been dependent on classical protein engineering techniques and the more
390 advanced directed evolution methods. The incorporation of NSAAs into enzymes can offer new chemical

functionalities to improve the various aspects of enzymatic catalysis. Standard protein engineering approaches and unnatural mutagenesis are expected to complement each other to generate efficient biocatalysts that can be used in various applications (Agostini *et al.* 2017). The Schultz group reported on a catalytically-improved TEM-1 β-lactamase variant following p-acrylamido-phenylalanine incorporation
395 *via* the replacement of valine-216, which had never been generated with the conventional 20 canonical amino acids. The employment of p-acrylamido-phenylalanine at the 216 position led to conformational changes in key active site residues, lowering the free energy of activation of the substrate (Xiao *et al.* 2015). The incorporation of the N_{δ}-methyl histidine ligand into an engineered ascorbate peroxidase resulted in a catalytically higher turnover number compared to the wild type enzyme (Green et al. 2016).
400 In addition, the incorporation of several NSAAs enabled improved activity of *Thermoanaerobacter thermohydrosulfuricus* lipase (TTL) in organic solvents or other hostile environments necessary for industrial transformations (Acevedo-Rocha *et al.* 2013). The Schultz group recently reported that the substitution of phenylalanine-22 by *p*-benzoyl-phenylalanine enhanced the thermal stability of *E. coli* homoserine O-succinyltransferase (Li *et al.* 2018). As shown in these examples, the engineering of
405 biocatalysts can be envisioned *via* expansion of the genetic code, which, in theory, could provide an unlimited set of NSAAs with new chemical functionalities. With the development of more rapid screening processes utilizing high-throughput screening, mutational scanning, and machine learning, NSAA incorporation will allow an unprecedented sequence space for novel biocatalysts (Badenhorst and Bornscheuer 2018).
410

Industrial applications of NSAAs

Genetic code expansion with NSAAs has emerged in the pharmaceutical industry with great promise in applications ranging from drug discovery and clinical trials to the manufacture of biopharmaceuticals
415 (Kang *et al.* 2018). Molecular engineering of a protein to introduce the conjugation site at a defined position is very important for biological activity, stability, and pharmacokinetics. Site-selective conjugation such as PEGylation can be achieved by the incorporation of NSAAs. However, protein expression using stop codon reassignment can result in leaky expression and low yields of protein due to low overall suppression efficiency. Efforts such as using cell-free protein synthesis have been made to
420 address limitations otherwise encountered by the specific incorporation of NSAAs (Schinn *et al.* 2017; Martin *et al.* 2018).

Several biotech companies are utilizing NSAAs or genetic code expansion to produce therapeutic proteins. Ambrx (La Jolla, CA, USA) is developing NSAA-containing ADCs, αHER2 ADC (ARX788) and PEG-FGF21 (ARX618), which are currently in phase 1 and phase 2 clinical trials, respectively

425 (Huang and Liu 2018). To incorporate the NSAAs, Ambrx developed a mammalian cell-based production platform called EuCODE™ (Tian *et al.* 2014). Significant improvements in the mammalian cell expression platform were achieved with a titer up to 1.0 ~ 1.5 g/L of the desired protein in a fed-batch 2,000 L bioreactor (Kang *et al.* 2018). Abzena (Cambridge, UK) uses site-specific conjugation for the PEGylation of interferon, a naturally occurring cytokine, to maximize the retention of its biological 430 activity (Peciak et al. 2019). Synthorx (La Jolla, CA, USA) is developing Synthorin IL-2 cytokines (THOR-707) to deliver sustained tumor-killing activity without compromising safety liability *via* protein modification using their proprietary expanded genetic alphabet platform based on an unnatural base pair (Dien *et al.* 2018). These modified cytokines provide protein therapeutics that can be used to tune receptor pharmacology and extend receptor half-lives.

435 Along with therapeutic proteins, the pharmacokinetic properties of lanthipeptides undergoing clinical studies were reviewed (Baumann et al. 2017; Ongey et al. 2017; Gomes et al. 2018; Ongey et al. 2018). Rational modifications and incorporation of NSAAs into AMPs may further expand the chemical space and provide new diversified functional characteristics. In collaboration with Intrexon, Oragenics (Florida, USA) has developed a bioengineered and novel lantibiotic called OG716 to treat *C. difficile* infection 440 (Kers et al. 2018). Novacta Biosystems (Welwyn Garden City, UK) completed a phase I clinical trial of an actagardine lantibiotic derivative NVB302 for *C. difficile* infection in healthy volunteers (Ongey et al. 2017).

Conclusion

445

NSAA incorporation is becoming more attractive for a range of biotechnology uses following recent developments of the platform technology. Molecular imaging, protein interactions, biocontainment, biomaterials, therapeutic proteins, and biocatalysts with novel functional and structural properties are available *via* NSAA incorporation. We speculate that the biological synthesis of modified proteins 450 containing NSAAs will be economically feasible for large-scale production, thereby circumventing biological barriers in the production of novel materials in living systems. Continued effort in the development of new methodologies that combine OTS optimization, biological platform improvement, and high-throughput screening/selection methods is necessary to advance the efficiency and fidelity of NSAA incorporation techniques. More applications will emerge as the field grows.

455

Funding This study was funded by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (R15AI130988).

Compliance with ethical standards

460

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare that they have no conflict of interest.

465 **References**

Acevedo-Rocha CG, Hoesl MG, Nehring S, Royter M, Wolschner C, Wiltschi B, Antranikian G, Budisa N (2013) Non-canonical amino acids as a useful synthetic biological tool for lipase-catalysed reactions in hostile environments. Catal Sci Technol 3:1198–1201 . doi: 10.1039/c3cy20712a

Agostini F, Völler J-S, Koksch B, Acevedo-Rocha CG, Kubyshkin V, Budisa N (2017) Biocatalysis with
470 unnatural amino acids: enzymology meets xenobiology. Angew Chemie Int Ed 56:9680–9703 . doi: 10.1002/anie.201610129

Albayrak C, Swartz JR (2014) Direct polymerization of proteins. ACS Synth Biol 3:353–362 . doi: 10.1021/sb400116x

Albayrak C, Swartz JR (2013) Cell-free co-production of an orthogonal transfer RNA activates efficient
475 site-specific non-natural amino acid incorporation. Nucleic Acids Res 41:5949–5963 . doi: 10.1093/nar/gkt226

Aminov RI (2011) Horizontal gene exchange in environmental microbiota. Front Microbiol 2:158 . doi: 10.3389/fmicb.2011.00158

Amiram M, Haimovich AD, Fan C, Wang Y-S, Aerni H-R, Ntai I, Moonan DW, Ma NJ, Rovner AJ,
480 Hong SH, Kelleher NL, Goodman AL, Jewett MC, Söll D, Rinehart J, Isaacs FJ (2015) Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. Nat Biotechnol 33:1272–1279 . doi: 10.1038/nbt.3372

Ayyadurai N, Saravanan Prabhu N, Deepankumar K, Lee S-G, Jeong H-H, Lee C-S, Yun H (2011) Development of a selective, sensitive, and reversible biosensor by the genetic incorporation of a
485 metal-binding site into green fluorescent protein. Angew Chemie Int Ed 50:6534–6537 . doi: 10.1002/anie.201008289

Badenhorst CPS, Bornscheuer UT (2018) Getting momentum: from biocatalysis to advanced synthetic biology. Trends Biochem Sci 43:180–198 . doi: 10.1016/J.TIBS.2018.01.003

Bahar AA, Ren D (2013) Antimicrobial peptides. Pharmaceuticals 6:1543–1575 . doi:
490 10.3390/ph6121543

Banin E, Hughes D, Kuipers OP (2017) Editorial: bacterial pathogens, antibiotics and antibiotic resistance. FEMS Microbiol Rev 41:450–452 . doi: 10.1093/femsre/fux016

Baumann T, Nickling JH, Bartholomae M, Buivydas A, Kuipers OP, Budisa N (2017) Prospects of *in vivo* incorporation of non-canonical amino acids for the chemical diversification of antimicrobial
495 peptides. Front Microbiol 8:124 . doi: 10.3389/fmicb.2017.00124

Baumann T, Schmitt F-J, Pelzer A, Spiering VJ, Freiherr Von Sass GJ, Friedrich T, Budisa N, Baumann C:, Schmitt T, Spiering A, Von Sass F, Friedrich GJ (2018) Engineering “golden” fluorescence by selective pressure incorporation of non-canonical amino acids and protein analysis by mass spectrometry and fluorescence. J Vis Exp 134:57017 . doi: 10.3791/57017

500 Bryson D, Fan C, Guo L, Miller C, Söll D, Liu D (2017) Continuous directed evolution of aminoacyl-tRNA synthetases. Nat Chem Biol 13:1253–1260

Cabantous S, Terwilliger TC, Waldo GS (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat Biotechnol 23:102–107 . doi: 10.1038/nbt1044

505 Chatterjee A, Sun SB, Furman JL, Xiao H, Schultz PG (2013) A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. Biochemistry 52:1828–1837 . doi: 10.1021/bi4000244

Chatterjee A, Xiao H, Schultz PG (2012) Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*. Proc Natl Acad Sci
510 U S A 109:14841–14846 . doi: 10.1073/pnas.1212454109

Chen Z, Ren W, Wright QE, Ai H (2013) Genetically encoded fluorescent probe for the selective detection of peroxynitrite. J Am Chem Soc 135:14940–14943 . doi: 10.1021/ja408011q

Chin JW (2017) Expanding and reprogramming the genetic code. Nature 550:53–60 . doi: 10.1038/nature24031

Chin JW (2014) Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem 83:379–408 . doi: 10.1146/annurev-biochem-060713-035737
d'Aquino AE, Kim DS, Jewett MC (2018) Engineered ribosomes for basic science and synthetic biology. Annu Rev Chem Biomol Eng 9:311–340 . doi: 10.1146/annurev-chembioeng-060817-084129
Devaraj NK (2018) The future of bioorthogonal chemistry. ACS Cent Sci 4:952–959 . doi: 10.1021/acscentsci.8b00251
Dien VT, Morris SE, Karadeema RJ, Romesberg FE (2018) Expansion of the genetic code via expansion of the genetic alphabet. Curr Opin Chem Biol 46:196–202 . doi: 10.1016/J.CBPA.2018.08.009
Dulic M, Cvetesic N, Zivkovic I, Palencia A, Cusack S, Bertosa B, Gruic-Sovulj I (2018) Kinetic origin of substrate specificity in post-transfer editing by leucyl-tRNA synthetase. J Mol Biol 430:1–16 . doi: 10.1016/J.JMB.2017.10.024
Dumas A, Lercher L, Spicer CD, Davis BG (2015) Designing logical codon reassignment – expanding the chemistry in biology. Chem Sci 6:50–69 . doi: 10.1039/C4SC01534G
Fechter P, Rudinger J, Giegé R, Théobald-Dietrich A (1998) Ribozyme processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: production and activity. FEBS Lett 436:99–103 . doi: 10.1016/S0014-5793(98)01096-5
Forster AC, Tan Z, Nalam MNL, Lin H, Qu H, Cornish VW, Blacklow SC (2003) Programming peptidomimetic syntheses by translating genetic codes designed *de novo*. Proc Natl Acad Sci U S A 100:6353–6357 . doi: 10.1073/pnas.1132122100
Gan F, Liu R, Wang F, Schultz PG (2018) Functional replacement of histidine in proteins to generate noncanonical amino acid dependent organisms. J Am Chem Soc 140:3829–3832 . doi: 10.1021/jacs.7b13452
Gan R, Perez JG, Carlson ED, Ntai I, Isaacs FJ, Kelleher NL, Jewett MC (2017) Translation system engineering in *Escherichia coli* enhances non-canonical amino acid incorporation into proteins. Biotechnol Bioeng 114:1074–1086 . doi: 10.1002/bit.26239
Gerrits M, Budisa N, Merk H (2019) Site-specific chemoselective pyrrolysine analogues incorporation using the cell-free protein synthesis system. ACS Synth Biol online . doi: 10.1021/acssynbio.8b00421
Giuliani A, Pirri G, Nicoletto S (2007) Antimicrobial peptides: an overview of a promising class of therapeutics. Cent Eur J Biol 2:1–33 . doi: 10.2478/s11535-007-0010-5
Gomes B, Augusto MT, Felício MR, Hollmann A, Franco OL, Gonçalves S, Santos NC (2018) Designing improved active peptides for therapeutic approaches against infectious diseases. Biotechnol Adv 36:415–429 . doi: 10.1016/J.BIOTECHADV.2018.01.004
Green AP, Hayashi T, Mittl PRE, Hilvert D (2016) A chemically programmed proximal ligand enhances the catalytic properties of a heme enzyme. J Am Chem Soc 138:11344–11352 . doi: 10.1021/jacs.6b07029
Hauf M, Richter F, Schneider T, Faidt T, Martins BM, Baumann T, Durkin P, Dobbek H, Jacobs K, Möglich A, Budisa N (2017) Photoactivatable mussel-based underwater adhesive proteins by an expanded genetic code. ChemBioChem 18:1819–1823 . doi: 10.1002/cbic.201700327
Hetrick KJ, Walker MC, Van Der Donk WA (2018) Development and application of yeast and phage display of diverse lanthipeptides. ACS Cent Sci 4:458–467 . doi: 10.1021/acscentsci.7b00581
Hoesl MG, Oehm S, Durkin P, Darmon E, Peil L, Aerni H-R, Rappsilber J, Rinehart J, Leach D, Söll D, Budisa N (2015) Chemical evolution of a bacterial proteome. Angew Chemie Int Ed 54:10030–10034 . doi: 10.1002/anie.201502868
Hong SH, Kwon Y-C, Jewett MC (2014a) Non-standard amino acid incorporation into proteins using *Escherichia coli* cell-free protein synthesis. Front Chem 2:34 . doi: 10.3389/fchem.2014.00034
Hong SH, Kwon Y-C, Martin RW, Des Soye BJ, de Paz AM, Swonger KN, Ntai I, Kelleher NL, Jewett MC (2015) Improving cell-free protein synthesis through genome engineering of *Escherichia coli* lacking release factor 1. ChemBioChem 16:844–853 . doi: 10.1002/cbic.201402708
Hong SH, Ntai I, Haimovich AD, Kelleher NL, Isaacs FJ, Jewett MC (2014b) Cell-free protein synthesis from a release factor 1 deficient *Escherichia coli* activates efficient and multiple site-specific non-

standard amino acid incorporation. ACS Synth Biol 3:398–409 . doi: 10.1021/sb400140t
Huang Y, Liu T (2018) Therapeutic applications of genetic code expansion. Synth Syst Biotechnol 3:150–158 . doi: 10.1016/j.synbio.2018.09.003
Hudson GA, Mitchell DA (2018) RiPP antibiotics: biosynthesis and engineering potential. Curr Opin
570 Microbiol 45:61–69 . doi: 10.1016/J.MIB.2018.02.010
Imai S, Guarente L (2014) NAD+ and sirtuins in aging and disease. Trends Cell Biol 24:464–471 . doi: 10.1016/j.tcb.2014.04.002
Iwane Y, Hitomi A, Murakami H, Katoh T, Goto Y, Suga H (2016) Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. Nat Chem 8:317–325 .
575 doi: 10.1038/nchem.2446
Jin X, Hong SH (2018) Cell-free protein synthesis for producing 'difficult-to-express' proteins. Biochem Eng J 138:156–164 . doi: 10.1016/j.bej.2018.07.013
Jin X, Kightlinger W, Kwon Y-C, Hong SH (2018) Rapid production and characterization of antimicrobial colicins using *Escherichia coli*-based cell-free protein synthesis. Synth Biol 3:ysy004 .
580 doi: 10.1093/synbio/ysy004
Johnson DBF, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, Xiang Z, Ecker JR, Briggs SP, Wang L (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 7:779–786 . doi: 10.1038/nchembio.657
Kageyama M, Kobayashi M, Sano Y, Masaki H (1996) Construction and characterization of pyocin-
585 colicin chimeric proteins. J Bacteriol 178:103–110
Kakkar N, Perez JG, Liu WR, Jewett MC, van der Donk WA (2018) Incorporation of nonproteinogenic amino acids in class I and II lantibiotics. ACS Chem Biol 13:951–957 . doi: 10.1021/acschembio.7b01024
Kang M, Lu Y, Chen S, Tian F (2018) Harnessing the power of an expanded genetic code toward next-
590 generation biopharmaceuticals. Curr Opin Chem Biol 46:123–129 . doi: 10.1016/j.cbpa.2018.07.018
Kato Y (2015) An engineered bacterium auxotrophic for an unnatural amino acid: a novel biological containment system. PeerJ 3:e1247 . doi: 10.7717/peerj.1247
Katoh T, Passioura T, Suga H (2018) Advances in *in vitro* genetic code reprogramming in 2014–2017. Synth Biol 3:ysy008 . doi: 10.1093/synbio/ysy008
595 Kers JA, DeFusco AW, Park JH, Xu J, Pulse ME, Weiss WJ, Handfield M (2018) OG716: Designing a fit-for-purpose lantibiotic for the treatment of *Clostridium difficile* infections. PLoS One 13:e0197467 . doi: 10.1371/journal.pone.0197467
Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, Latendresse M, Muñiz-Rascado L, Ong Q, Paley S,
600 Peralta-Gil M, Subhraveti P, Velázquez-Ramírez DA, Weaver D, Collado-Vides J, Paulsen I, Karp PD (2017) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res 45:D543–D550 . doi: 10.1093/nar/gkw1003
Kim CH, Axup JY, Schultz PG (2013) Protein conjugation with genetically encoded unnatural amino acids. Curr Opin Chem Biol 17:412–419 . doi: http://dx.doi.org/10.1016/j.cbpa.2013.04.017
605 Koh M, Nasertorabi F, Han GW, Stevens RC, Schultz PG (2017) Generation of an orthogonal protein–protein interface with a noncanonical amino acid. J Am Chem Soc 139:5728–5731 . doi: 10.1021/jacs.7b02273
Kord Forooshani P, Lee BP (2017) Recent approaches in designing bioadhesive materials inspired by mussel adhesive protein. J Polym Sci A Polym Chem 55:9–33 . doi: 10.1002/pola.28368
610 Krishnakumar R, Ling J (2014) Experimental challenges of sense codon reassignment: an innovative approach to genetic code expansion. FEBS Lett 588:383–388 . doi: 10.1016/j.febslet.2013.11.039
Kubyshkin V, Acevedo-Rocha CG, Budisa N (2018) On universal coding events in protein biogenesis. Biosystems 164:16–25 . doi: 10.1016/J.BIOSYSTEMS.2017.10.004
Kubyshkin V, Budisa N (2017) Synthetic alienation of microbial organisms by using genetic code
615 engineering: why and how? Biotechnol J 12:1600097 . doi: 10.1002/biot.201600097
Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH,

Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ (2013) Genomically recoded organisms expand biological functions. Science 342:357–360 . doi: 10.1126/science.1241459
620 Lee EY, Wong GCL, Ferguson AL (2018) Machine learning-enabled discovery and design of membrane-active peptides. Bioorg Med Chem 26:2708–2718 . doi: 10.1016/J.BMC.2017.07.012
Li JC, Liu T, Wang Y, Mehta AP, Schultz PG (2018) Enhancing protein stability with genetically encoded noncanonical amino acids. J Am Chem Soc 140:15997–16000 . doi: 10.1021/jacs.8b07157
Li X, Liu CC (2014) Biological applications of expanded genetic codes. ChemBioChem 15:2335–2341 .
625 doi: 10.1002/cbic.201402159
Link AJ, Tirrell DA (2005) Reassignment of sense codons *in vivo*. Methods 36:291–298 . doi: 10.1016/j.ymeth.2005.04.005
Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. Annu Rev Biochem 79:413–444 . doi: doi:10.1146/annurev.biochem.052308.105824
630 Liu X, Jiang L, Li J, Wang L, Yu Y, Zhou Q, Lv X, Gong W, Lu Y, Wang J (2014) Significant expansion of fluorescent protein sensing ability through the genetic incorporation of superior photo-induced electron-transfer quenchers. J Am Chem Soc 136:13094–13097 . doi: 10.1021/ja505219r
Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, Gregg CJ, Stoddard BL, Church GM (2015) Biocontainment of genetically modified organisms by synthetic protein design.
635 Nature 518:55–60
Martin RW, Soye BJ Des, Kwon Y-C, Kay J, Davis RG, Thomas PM, Majewska NI, Chen CX, Marcum RD, Weiss MG, Stoddart AE, Amiram M, Charna AKR, Patel JR, Isaacs FJ, Kelleher NL, Hong SH, Jewett MC (2018) Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. Nat Commun 9:1203 . doi: 10.1038/s41467-018-03469-
640 5
Moe-Behrens GHG, Davis R, Haynes KA (2013) Preparing synthetic biology for the world. Front Microbiol 4:5 . doi: 10.3389/fmicb.2013.00005
Mukai T, Yanagisawa T, Ohtake K, Wakamori M, Adachi J, Hino N, Sato A, Kobayashi T, Hayashi A, Shirouzu M, Umehara T, Yokoyama S, Sakamoto K (2011) Genetic-code evolution for protein
645 synthesis with non-natural amino acids. Biochem Biophys Res Commun 411:757–761 . doi: 10.1016/j.bbrc.2011.07.020
Neumann-Staubitz P, Neumann H (2016) The use of unnatural amino acids to study and engineer protein function. Curr Opin Struct Biol 38:119–128 . doi: 10.1016/j.sbi.2016.06.006
Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW (2010) Encoding multiple unnatural amino acids
650 via evolution of a quadruplet-decoding ribosome. Nature 464:441–444 . doi: 10.1038/nature08817
Nguyen T-A, Cigler M, Lang K (2018) Expanding the genetic code to study protein-protein interactions. Angew Chemie Int Ed 57:14350–14361 . doi: 10.1002/anie.201805869
Nickling JH, Baumann T, Schmitt F-J, Bartholomae M, Kuipers OP, Friedrich T, Budisa N (2018) Antimicrobial peptides produced by selective pressure incorporation of non-canonical amino acids. J
655 Vis Exp 135:e57551 . doi: 10.3791/57551
Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. Science 244:182–188
O'Donoghue P, Ling J, Wang Y-S, Söll D (2013) Upgrading protein synthesis for synthetic biology. Nat Chem Biol 9:594–598 . doi: 10.1038/nchembio.1339
660 Ohtake K, Sato A, Mukai T, Hino N, Yokoyama S, Sakamoto K (2012) Efficient decoding of the UAG triplet as a full-fledged sense codon enhances the growth of a *prfA*-deficient strain of *Escherichia coli*. J Bacteriol 194:2606–2613 . doi: 10.1128/jb.00195-12
Ohtsuki T, Manabe T, Sisido M (2005) Multiple incorporation of non-natural amino acids into a single protein using tRNAs with non-standard structures. FEBS Lett 579:6769–6774 . doi:
665 10.1016/j.febslet.2005.11.010
Ongey EL, Pflugmacher S, Neubauer P (2018) Bioinspired designs, molecular premise and tools for evaluating the ecological importance of antimicrobial peptides. Pharmaceuticals 11:68 . doi:

10.3390/ph11030068
Ongey EL, Yassi H, Pflugmacher S, Neubauer P (2017) Pharmacological and pharmacokinetic properties of lanthipeptides undergoing clinical studies. Biotechnol Lett 39:473–482 . doi: 10.1007/s10529-016-2279-9
Orelle C, Carlson ED, Szal T, Florin T, Jewett MC, Mankin AS (2015) Protein synthesis by ribosomes with tethered subunits. Nature 524:119–124 . doi: 10.1038/nature14862
Park H-S, Hohn MJ, Umehara T, Guo L-T, Osborne EM, Benner J, Noren CJ, Rinehart J, Söll D (2011) Expanding the genetic code of *Escherichia coli* with phosphoserine. Science 333:1151–1154 . doi: 10.1126/science.1207203
Peciak K, Laurine E, Tommasi R, Choi J, Brocchini S (2019) Site-selective protein conjugation at histidine. Chem Sci 10:427–439 . doi: 10.1039/C8SC03355B
Peters BM, Shirtliff ME, Jabra-Rizk MA (2010) Antimicrobial peptides: primeval molecules or future drugs? PLoS Pathog 6:e1001067 . doi: 10.1371/journal.ppat.1001067
Porto WF, Pires AS, Franco OL (2017) Computational tools for exploring sequence databases as a resource for antimicrobial peptides. Biotechnol Adv 35:337–349 . doi: 10.1016/J.BIOTECHADV.2017.02.001
Praveschotinunt P, Courchesne NMD, den Hartog I, Lu C, Kim JJ, Nguyen PQ, Joshi NS (2018) Tracking of engineered bacteria *in vivo* using nonstandard amino acid incorporation. ACS Synth Biol 7:1640–1650 . doi: 10.1021/acssynbio.8b00135
Quast RB, Mrusek D, Hoffmeister C, Sonnabend A, Kubick S (2015) Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis. FEBS Lett 589:1703–1712 . doi: 10.1016/j.febslet.2015.04.041
Ravikumar Y, Nadarajan SP, Yoo TH, Lee C, Yun H (2015) Unnatural amino acid mutagenesis-based enzyme engineering. Trends Biotechnol 33:462–470 . doi: 10.1016/j.tibtech.2015.05.002
Reitz C, Fan Q, Neubauer P (2018) Synthesis of non-canonical branched-chain amino acids in *Escherichia coli* and approaches to avoid their incorporation into recombinant proteins. Curr Opin Biotechnol 53:248–253 . doi: 10.1016/J.COPBIO.2018.05.003
Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW, Gassaway BM, Amiram M, Patel JR, Gallagher RR, Rinehart J, Isaacs FJ (2015) Recoded organisms engineered to depend on synthetic amino acids. Nature 518:89–93
Schinn S-M, Bradley W, Groesbeck A, Wu JC, Broadbent A, Bundy BC (2017) Rapid *in vitro* screening for the location-dependent effects of unnatural amino acids on protein expression and activity. Biotechnol Bioeng 114:2412–2417 . doi: 10.1002/bit.26305
Serfling R, Seidel L, Böttke T, Coin I (2018) Optimizing the genetic incorporation of chemical probes into GPCRs for photo-crosslinking mapping and bioorthogonal chemistry in live mammalian cells. J Vis Exp 134:e57069 . doi: 10.3791/57069
Shang X, Chen Y, Wang N, Niu W, Guo J (2018) Oxidation-induced generation of a mild electrophile for proximity-enhanced protein–protein crosslinking. Chem Commun 54:4172–4175 . doi: 10.1039/C8CC01639A
Shcherbakova DM, Subach OM, Verkhusha V V. (2012) Red fluorescent proteins: advanced imaging applications and future design. Angew Chemie Int Ed 51:10724–10738 . doi: 10.1002/anie.201200408
Shreiner AB, Kao JY, Young VB (2015) The gut microbiome in health and in disease. Curr Opin Gastroenterol 31:69–75 . doi: 10.1097/MOG.0000000000000139
Smith MT, Hawes AK, Shrestha P, Rainsdon JM, Wu JC, Bundy BC (2014) Alternative fermentation conditions for improved *Escherichia coli*-based cell-free protein synthesis for proteins requiring supplemental components for proper synthesis. Process Biochem 49:217–222 . doi: http://dx.doi.org/10.1016/j.procbio.2013.10.012
Spicer CD, Jumeaux C, Gupta B, Stevens MM (2018) Peptide and protein nanoparticle conjugates: versatile platforms for biomedical applications. Chem Soc Rev 47:3574–3620 . doi: 10.1039/C7CS00877E

Swartz JR, Jewett MC, Woodrow K a (2004) Cell-free protein synthesis with prokaryotic combined transcription-translation. Methods Mol Biol 267:169–182 . doi: 10.1385/1-59259-774-2:169
720
Tian F, Lu Y, Manibusan A, Sellers A, Tran H, Sun Y, Phuong T, Barnett R, Hehli B, Song F, DeGuzman MJ, Ensari S, Pinkstaff JK, Sullivan LM, Biroc SL, Cho H, Schultz PG, DiJoseph J, Dougher M, Ma D, Dushin R, Leal M, Tchistiakova L, Feyfant E, Gerber H-P, Sapra P (2014) A general approach to site-specific antibody drug conjugates. Proc Natl Acad Sci U S A 111:1766–1771 . doi: 10.1073/pnas.1321237111
725
Uhlenbeck OC, Schrader JM (2018) Evolutionary tuning impacts the design of bacterial tRNAs for the incorporation of unnatural amino acids by ribosomes. Curr Opin Chem Biol 46:138–145 . doi: 10.1016/j.cbpa.2018.07.016
Vargas-Rodriguez O, Sevostyanova A, Söll D, Crnković A (2018) Upgrading aminoacyl-tRNA synthetases for genetic code expansion. Curr Opin Chem Biol 46:115–122 . doi: 10.1016/J.CBPA.2018.07.014
730
Voloshchuk N, Montclare JK (2010) Incorporation of unnatural amino acids for synthetic biology. Mol Biosyst 6:65–80
Wang K, Neumann H, Peak-Chew SY, Chin JW (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol 25:770–777 . doi: http://www.nature.com/nbt/journal/v25/n7/suppinfo/nbt1314_S1.html
735
Wang L (2017a) Engineering the genetic code in cells and animals:biological considerations and impacts. Acc Chem Res 50:2767–2775 . doi: 10.1007/s10995-015-1800-4.Alcohol
Wang L (2017b) Genetically encoding new bioreactivity. N Biotechnol 38:16–25 . doi: 10.1016/j.nbt.2016.10.003
740
Wang L, Xie J, Deniz AA, Schultz PG (2003) Unnatural amino acid mutagenesis of green fluorescent protein. J Org Chem 68:174–176 . doi: 10.1021/JO026570U
Wang N, Yang B, Fu C, Zhu H, Zheng F, Kobayashi T, Liu J, Li S, Ma C, Wang PG, Wang Q, Wang L (2018) Genetically encoding fluorosulfate-L-tyrosine to react with lysine, histidine, and tyrosine via SuFEx in proteins *in vivo*. J Am Chem Soc 140:4995–4999 . doi: 10.1021/jacs.8b01087
745
Xiao H, Nasertorabi F, Choi S, Han GW, Reed SA, Stevens RC, Schultz PG (2015) Exploring the potential impact of an expanded genetic code on protein function. Proc Natl Acad Sci U S A 112:6961–6966 . doi: 10.1073/pnas.1507741112
Xuan W, Yao A, Schultz PG (2017) Genetically encoded fluorescent probe for detecting sirtuins in living cells. J Am Chem Soc 139:12350–12353 . doi: 10.1021/jacs.7b05725
750
Yang B, Tang S, Ma C, Li S-T, Shao G-C, Dang B, DeGrado WF, Dong M-Q, Wang PG, Ding S, Wang L (2017) Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions. Nat Commun 8:2240 . doi: 10.1038/s41467-017-02409-z
Yang B, Wu H, Schnier PD, Liu Y, Liu J, Wang N, DeGrado WF, Wang L (2018) Proximity-enhanced SuFEx chemical cross-linker for specific and multitargeting cross-linking mass spectrometry. Proc Natl Acad Sci U S A 115:11162–11167 . doi: 10.1073/PNAS.1813574115
755
Yoshida M, Hinkley T, Tsuda S, Abul-Haija YM, McBurney RT, Kulikov V, Mathieson JS, Galiñanes Reyes S, Castro MD, Cronin L (2018) Using evolutionary algorithms and machine learning to explore sequence space for the discovery of antimicrobial peptides. Chem 4:533–543 . doi: 10.1016/J.CHEMPR.2018.01.005
760
Young TS, Ahmad I, Yin JA, Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. J Mol Biol 395:361–374 . doi: 10.1016/j.jmb.2009.10.030
Young TS, Schultz PG (2010) Beyond the canonical 20 amino acids: expanding the genetic lexicon. J Biol Chem 285:11039–11044 . doi: 10.1074/jbc.R109.091306
765
Zambaldo C, Luo X, Mehta AP, Schultz PG (2017) Recombinant macrocyclic lanthipeptides incorporating non-canonical amino acids. J Am Chem Soc 139:11646–11649 . doi: 10.1021/jacs.7b04159
Zhang Y, Ptacin JL, Fischer EC, Aerni HR, Caffaro CE, San Jose K, Feldman AW, Turner CR, Romesberg FE (2017) A semi-synthetic organism that stores and retrieves increased genetic

770 information. Nature 551:644–647 . doi: 10.1038/nature24659

Zheng Y, Gilgenast MJ, Hauc S, Chatterjee A (2018) Capturing post-translational modification-triggered protein–protein interactions using dual noncanonical amino acid mutagenesis. ACS Chem Biol 13:1137–1141 . doi: 10.1021/acschembio.8b00021

775

Figure Legends

Figure 1. Scheme of non-standard amino acid incorporation using an orthogonal translation system via amber suppression. Orthogonal aminoacyl tRNA synthetase, o-aaRS; non-standard amino acid, NSAA; orthogonal tRNA, o-tRNA; elongation factor Tu, EF-Tu. The anti-amber codon sequence on o-tRNA is CUA.

Figure 2. A) NSAA incorporation using a cell-free protein synthesis reaction. Cell extracts containing transcription/translation machinery, co-factors, energy sources and other components are added to the reaction to mimic the bacterial environment. The addition of the o-aaRS/o-tRNA pair, NSAA and plasmid DNA template containing amber sites enables NSAA incorporation. **B) Genomically recoded organism evolution via multiplex automated genome engineering (MAGE).** Single-strand DNA oligos are added to the MAGE cycle to mutate multiple genomic locations simultaneously. **C) Redundant sense codon reassignment.** Codons for the standard amino acid (SAA) are artificially reprogramed for NSAA. N denotes A, U, G or C. **D) Orthogonal ribosome and quadruplet codon.** The orthogonal ribosome is developed by recognizing the orthogonal ribosome binding site (o-RBS) and tethering the 50S subunit and orthogonal 30S subunit. The orthogonal ribosome translates orthogonal mRNA containing o-RBS and incorporates NSAAs via quadruplet codon and amber codon suppression.

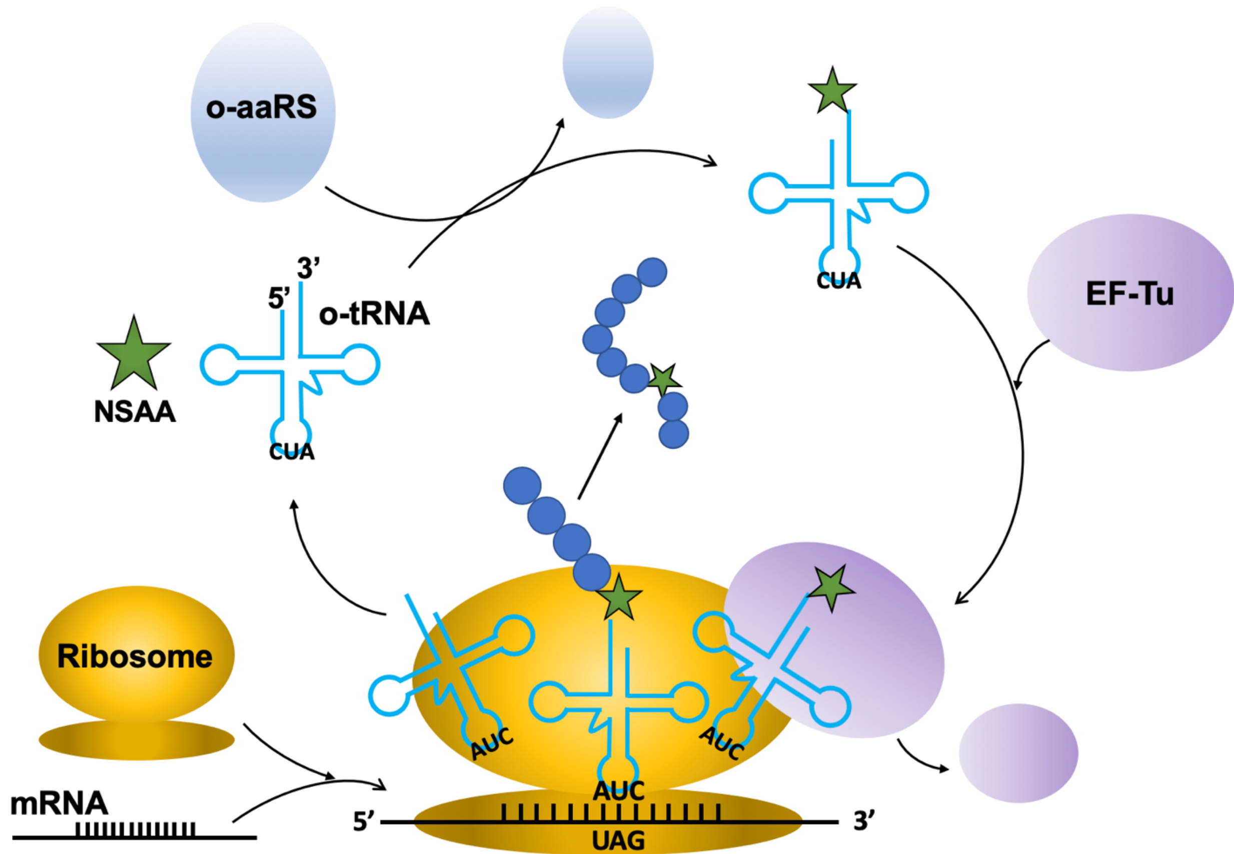
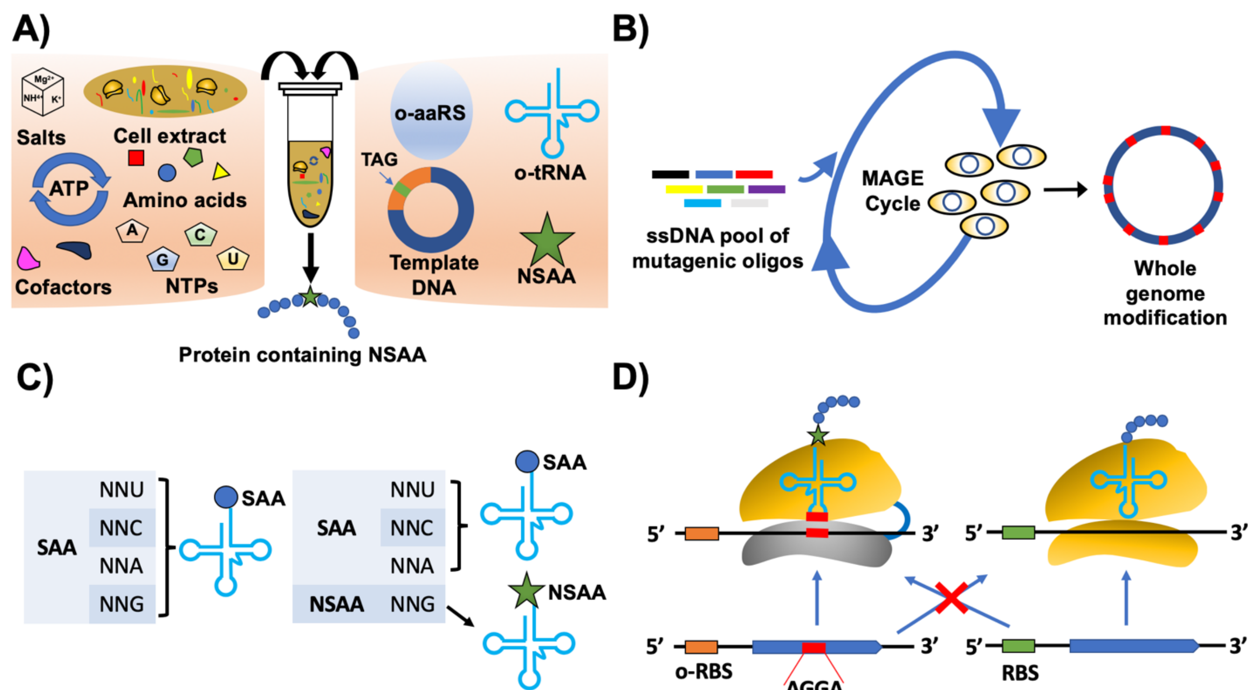


Figure 1



800

Figure 2