Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Upgrading aminoacyl-tRNA synthetases for genetic code expansion

Oscar Vargas-Rodriguez[1], Anastasia Sevostyanova[1], Dieter Söll[1,2] and Ana Crnković[1]

Synthesis of proteins with non-canonical amino acids via genetic code expansion is at the forefront of synthetic biology. Progress in this field has enabled site-specific incorporation of over 200 chemically and structurally diverse amino acids into proteins in an increasing number of organisms. This has been facilitated by our ability to repurpose aminoacyl-tRNA synthetases to attach non-canonical amino acids to engineered tRNAs. Current efforts in the field focus on overcoming existing limitations to the simultaneous incorporation of multiple non-canonical amino acids or amino acids that differ from the L-α-amino acid structure (e.g. D-amino acid or β-amino acid). Here, we summarize the progress and challenges in developing more selective and efficient aminoacyl-tRNA synthetases for genetic code expansion.

Addresses
[1] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA
[2] Department of Chemistry, Yale University, New Haven, CT 06520, USA

Corresponding authors: Vargas-Rodriguez, Oscar (oscar.vargas@yale.edu), Crnković, Ana (ana.crnkovic@yale.edu)

Current Opinion in Chemical Biology 2018, **46C**:115–122

This review comes from a themed issue on **Synthetic biology**

Edited by **Michael Ledbetter** and **Floyd E Romesberg**

https://doi.org/10.1016/j.cbpa.2018.07.011

Introduction

All living organisms synthesize proteins using the same set of twenty canonical amino acids (AAs) according to the rules of the genetic code. Over the past two decades, efforts to change these rules have resulted in the co-translational incorporation of numerous structurally diverse non-canonical AAs (ncAAs) into proteins by codon reassignment. Genetic code expansion (GCE) is now possible in a wide range of model organisms, which has facilitated investigation of protein structure and function in various biological processes, as well as development of therapeutics and the biocontainment of genetically modified organisms.

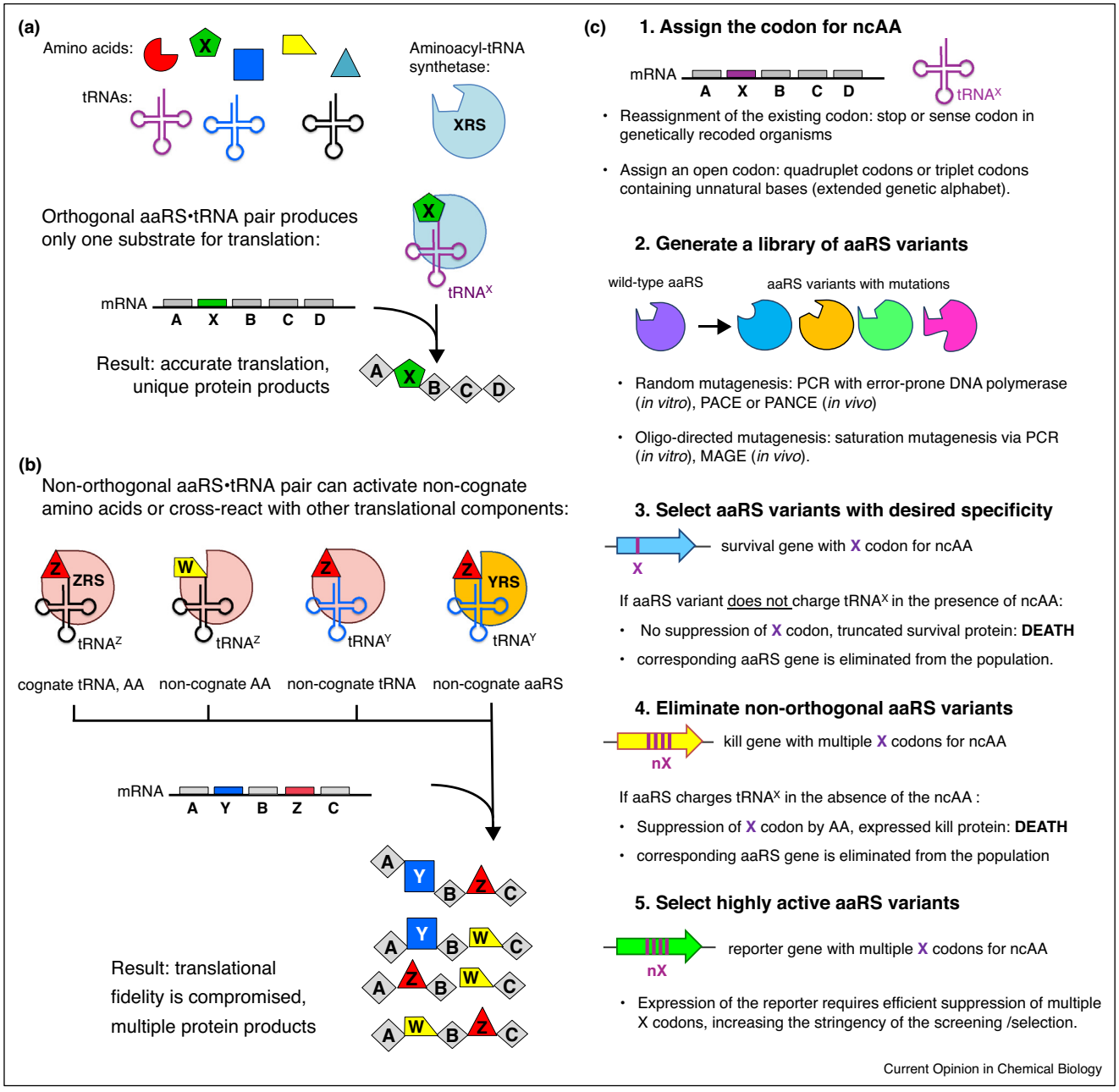
Expanding the genetic code entails altering two fundamental steps in translation: synthesis of aminoacyl-tRNAs (AA-tRNAs), and codon-anticodon pairing in the ribosome [1••]. This is achieved by repurposing an aminoacyl-tRNA synthetase (aaRS) to ligate the desired ncAA to a dedicated tRNA capable of inserting the ncAA at a defined codon (Figure 1a). For correct *in vivo* incorporation of ncAAs, engineered aaRS•tRNA pairs should not cross-react with the host aaRSs and tRNAs (*i.e.* they should be orthogonal), a requirement often met by relying on aaRS•tRNA pairs from organisms belonging to different domains of life.

Although over 200 ncAAs have been incorporated into proteins, engineered orthogonal aaRSs (o-aaRSs) usually have lower catalytic activity than their parental aaRSs, yet they maintain their AA polyspecificity (Figure 1b). These characteristics have hampered the development of more sophisticated applications (*e.g.* mass production of ncAA-containing proteins, multi-site incorporation of an ncAA, simultaneous incorporation of various ncAAs into a protein, or sense codon reassignment). Here we review fundamental aspects of aaRS specificity and efficiency and discuss recent advances in the development and discovery of superior engineered aaRSs that can contribute to overcoming the current limitations of GCE.

Fundamentals of aaRS substrate specificity

Each canonical AA has its cognate aaRS. This ensures accurate genetic code interpretation by specific pairing of AAs with tRNAs with the correct anticodon (Figure 1a). aaRSs select their cognate tRNAs through a defined set of nucleotides known as identity elements [2]. The majority of these elements are located in the anticodon loop and the acceptor stem of the tRNA, but the specific positions are unique to each tRNA for a particular AA [2]. Specific interactions with the anticodon are characteristically absent in some aaRSs, such as pyrrolysyl-tRNA synthetase (PylRS), seryl-tRNA synthetase (SerRS), or leucyl-tRNA synthetase (LeuRS), making their corresponding tRNAs well-suited for codon reassignment (Figure 1c) [2,3]. tRNA identity elements are not fully conserved throughout evolution for many aaRS•tRNA pairs, which — in some cases — prevents cross-reaction between aaRSs and tRNAs from a different species. The phylogenetic diversity of tRNA identity elements for an AA underlies the principle of orthogonality that enables GCE *in vivo*.

Figure 1



Accurate protein synthesis and conventional methodology to alter aaRS specificity. **(a)** Each aaRS (XRS) must selectively attach a cognate AA (denoted as X) to a dedicated tRNA with the correct anticodon to form aminoacyl-tRNA (X-tRNA^X). Correct codon-anticodon matching in the ribosome ensures accurate incorporation of X at a defined position in a protein sequence. **(b)** Polyspecific aaRSs recognize more than one amino acid (depicted by YRS and ZRS) or more than one tRNA (ZRS). These errors in aminoacylation result in expression of a gene into proteins with varied primary structures. **(c)** General steps for directed evolution of aaRSs.

Because AAs provide a less diverse set of structural and chemical features to help ensure accurate substrate selection, many aaRSs occasionally produce noncognate AA-tRNAs (Figure 1b). Despite detrimental consequences associated with mistranslation [4], organisms can tolerate errors in their proteomes [5] (Figure 1b). This is commonly exploited by using wild-type methionyl-tRNA synthetase to randomly incorporate selenomethionine or azidohomoalanine for X-ray crystallography applications [6,7] or *in vivo* protein labeling [8••]. However, site-specific incorporation requires development of an aaRS•tRNA pair dedicated to the particular ncAA.

AaRS•tRNA pairs for GCE applications

Over two-thirds of all ncAAs incorporated to date are mediated by the wild-type or engineered *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (TyrRS)•$tRNA^{Tyr}$ pair or the *Methanosarcina* PylRS•$tRNA^{Pyl}$ pair (Table 1). These pairs are now standard tools for GCE applications because of their robust orthogonality in *Escherichia coli* and the tolerance of both aaRSs to anticodon mutations in their tRNAs. These features simplify the engineering of PylRS and TyrRS for a desired ncAA by *in vivo* selection methods (Figure 1c) and facilitate the targeting of different codons for reassignment (Table 1). The *M. jannaschii* TyrRS•$tRNA^{Tyr}$ pair is used in bacteria such as *E. coli* [1••], *Salmonella* [9], *Streptomyces* [10], and *Mycobacterium* [11] (Table 1); yet this pair is not suitable for eukaryotic applications because identity elements overlap with eukaryotic TyrRS•$tRNA^{Tyr}$. However, the *E. coli* TyrRS•$tRNA^{Tyr}$ pair has been adapted for GCE in diverse eukaryotic species [12]; although, the repertoire of ncAA substrates is smaller (Table 1). In contrast to the TyrRS system, the *Methanosarcina* PylRS•$tRNA^{Pyl}$ pairs are orthogonal to a much wider range of hosts, including higher eukaryotes, due to the distinctive structure of $tRNA^{Pyl}$ and its unique mode of recognition by PylRS [3,13•]. Several other o-aaRS•tRNA pairs have been developed for GCE applications, but poor orthogonality and/or limited ncAA specificity restrict their use to only a few hosts (Table 1).

Efficiency of o-aaRS•tRNA pairs

The lower catalytic activity of evolved o-aaRSs results in modest concentrations of ncAA-tRNAs. This impedes successful incorporation of ncAAs into multiple positions within one protein, and prevents efficient competition with endogenous AA-tRNAs for reassignment of sense codons. High-quality libraries of diverse aaRS variants are paramount to obtaining highly active and ncAA-specific aaRSs (Figure 1c). Oligo-directed libraries depend on the availability of high-resolution crystal structures of aaRSs [14], as well as docking studies to identify active site residues that may contact the ncAA [15,16]. In practice, the size of site saturated libraries that can be subjected to comprehensive analysis with standard lab resources is limited to 10^9, which corresponds to 6 residues

Table 1

Orthogonal aaRS•tRNA pairs applied for genetic code expansion

	Source	Orthogonal hosts	ncAA substrates[a,b]	Reassigned codons[a]	Mutual orthogonality
TyrRS	*Methanocaldococcus jannaschii*	Bacteria[a]	>50 Tyr analogs	UAG, UGA, AGGA, UAGA	PylRS, SepRS
TyrRS	*Escherichia coli*	Yeasts[a], *Caenorhabditis elegans* [70], *Danio rerio*, *Mus musculus* [71], mammalian cell lines[a]	mY, pPpa, pMpa, AzF, AcF, BnzylF, pIF	UAG	N.D.
PylRS	*Methanosarcina barkeri or mazei*	Bacteria[a], *Mus musculus* [72], *Danio rerio* [71], *Saccharomyces cerevisiae* [73,74], *Arabidopsis thaliana* [75], *Drosophila melanogaster* [74], mammalian cell lines[a]	>100 diverse AA	UAA, UAG, UGA, AGU	MjTyrRS, EcLeuRS, EcTrpRS
LeuRS	*Escherichia coli*	mammalian cell lines [76], *Saccharomyces cerevisiae* [77], *Caenorhabditis elegans* [70]	DanAla, 2-aminocaprylic acid, mY, nbC, dMnbS	UAG	PylRS
TrpRS	*Escherichia coli*	Human cell lines [78]	5PrW, 5HTP, 5AzW, 5AmW	UAG	N.D.
TrpRS	*Saccharomyces cerevisiae*	*Escherichia coli* [79,80]	NapAla, 1MeW, BtAla, MeW	UAG	N.D.
SepRS	*Methanocaldococcus jannaschii*	*Escherichia coli* [25]	Sep, pThr	UAG	PylRS, MjTyrRS
LysRS	*Pyrococcus horikoshii*	*Escherichia coli* [48]	Homoglutamine	AGGA	MjTyrRS
PheRS	*Saccharomyces cerevisiae*	*Escherichia coli* [77,81]	pFF, 2Nal, pBrF	UAG, UUU, UUG	N.D.

[a] For a full list see reference [1••].

[b] For a comprehensive list see [1••,34•]. N.D., not determined. MjTyrRS, TyrRS from *M. jannaschii*; EcLeuRS, LeuRS from *E. coli*; EcTrpRS, TrpRS from *E. coli.* mY, *p*-methyl-Tyr; AcF, *p*-acetyl-Phe; BnzylF, *p*-benzoyl-Phe; AzF, *p*-azido-Phe; pIF, *p*-iodo-Phe; pPpa, propargyloxy-Phe; pMpa, methoxy-Phe; DanAla, 2-amino-3-(5-(dimethylamino) naphthalene-1-sulfonamido) propanoic acid; nbC, *O*-nitrobenzyl-cysteine; dMnbS, 4,5-dimethoxy-2-nitrobenzyl-L-serine; 5PrW, 5-(2-Propyn-1-yloxy)-L-tryptophan; 5HTP, 5-hydroxy-L-tryptophan; 5AzW, 5-azido-L-tryptophan; 5AmW, 5-amino-L-tryptophan; Sep, *O*-phospho-L-serine; pThr, *O*-phospho-L-threonine; NapAla, 3-(1-naphthyl)-L-alanine; 1MeW, 1-methyl-L-tryptophan; BtAla, 3-benzothienyl-L-alanine; 6MeW, 6-methyl-L-tryptophan; pFF, *p*-fluoro-Phe; 2Nal, 2-naphthylalanine; pBrF, *p*-bromo-Phe.

Box 1 Cellular factors that influence the artificial evolution of orthogonal aaRSs

- Expression levels of the o-aaRS•tRNA pair

Fine-tuning the expression levels of the o-aaRS•tRNA pairs is essential as overexpression can be detrimental for cell sustainability and/or increase cross-reaction with the host aaRSs and tRNAs. Conversely, low expression of the orthogonal pair may lead to a false negative result (too low expression of the selection reporter).

- Intracellular concentration of the ncAA

The intracellular concentration of many ncAAs is limited by host cell permeability, which increases the stringency of the selection. To overcome this, ncAAs can be delivered as dipeptides [66] or, alternatively, can be synthesized *in situ* synthesis through metabolic engineering [33••].

- Orthogonality of the ncAA

Many ncAAs may be substrates of host aaRSs, which can lead to proteome-wide incorporation of the ncAA [28]. Moreover, ncAAs may interfere with essential metabolic pathways or be metabolized into a different substrate [43].

- Translation factors and quality control mechanisms

Several AA-tRNA deacylases have been recently shown to hydrolyze ncAA-tRNAs [67,68]. The delivery of ncAA-o-tRNA to the ribosome may be impaired by weak binding with elongation factor Tu [25]. Additionally, incorporation of the ncAA into the nascent protein may be limited by the ribosome [45]. Both of these issues may require further engineering of the elongation factor and the ribosome.

- Sensitivity of the selection reporter

Most selection reporters do not depend on the identity of the incorporated AA. Thus, canonical AA mischarging by o-aaRS variants [69] and suppressor tRNAs that are not fully orthogonal obscure results. The latter may abolish the possibility of negative selection [33••].

randomized using redundant NNK primers [16]. Proficient o-aaRS variants are identified by selection or screening techniques, typically using reporter genes that confer antibiotic resistance or encode fluorescent proteins (Figure 1c). However, robust and stringent phenotypes that unequivocally signal the incorporation of the ncAA are rare. Furthermore, most phenotype-based selections are sensitive to other factors (e.g. the protein expression level or relative cellular solubility of o-aaRS mutants) that can obstruct the selection of genuinely specific variants (Box 1). Despite the limitations imposed by conventional methods of mutagenesis and selection [17••], many ncAA-charging aaRS variants have been evolved and ncAA-containing proteins produced (Table 1). Nonetheless, innovative methods have been recently applied to the evolution of o-aaRSs that enabled engineering of more active and selective aaRSs.

Multiplex automated genome engineering (MAGE) facilitates the creation of combinatorial genetic diversity within a cell population by simultaneously mutagenizing different loci [18] (Figure 2a). With MAGE, a large library of chromosomally integrated *M. jannaschii* TyrRS variants was created with up to 12 mutational changes in the active site and anticodon binding domains [19••]. Genomic integration of the TyrRS variant genes increased selection stringency and prevented gene dosage fluctuations linked to plasmid-based expression. TyrRS variants generated by MAGE were capable of mediating specific ncAA incorporation in response to 30 UAG codons of a reporter gene after a single selection step and two rounds of cell sorting [19••].

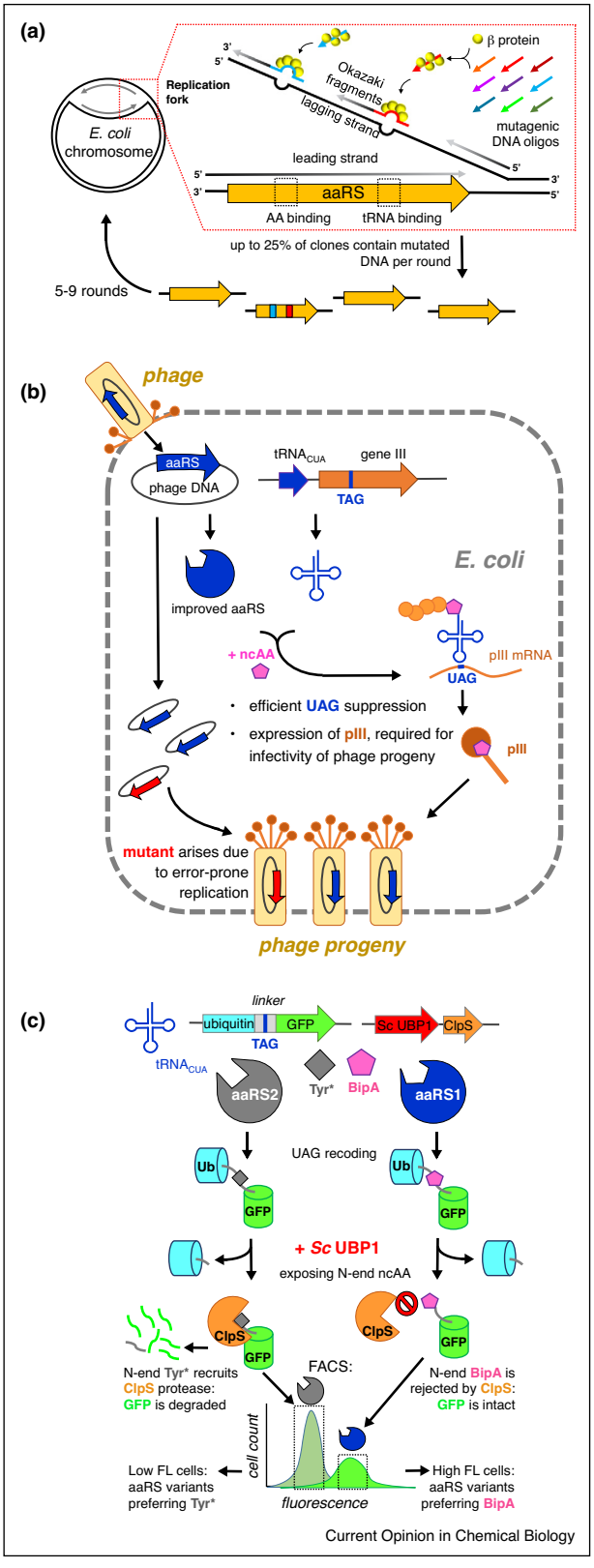
Phage-assisted continuous evolution (PACE) has emerged as powerful technology that links bacteriophage propagation with the evolution of a single gene expressed in *E. coli* [20] (Figure 1c). PACE employs a plasmid-based mutagenesis system that enables creation of random libraries with a defined mutational frequency [21]. Thus, by coupling mutational and selection cycles, PACE allows ncAA-specific aaRSs to emerge after only 48 hours or highly active variants after 20 days [22••] (Figure 2b). The latter was demonstrated by the evolution of a chimeric PylRS variant with 45-fold higher activity toward Boc-lysine [22••]. Similar results were obtained using phage-assisted noncontinuous evolution (PANCE) [13•]. This adaptation of PACE relies on serial flask transfers instead of a continuous flow machine although highly active variants require more time to emerge [13•]. Importantly, PACE and PANCE do not rely on prior structural or functional knowledge, like oligo-based mutagenesis methods do. Thus, while rational design assisted by computational methods [15] may help identify proficient aaRS variants harboring mutations outside the AA binding site [23•], approaches for random mutagenesis *in vivo* may reveal otherwise unpredictable effects of residues governing o-aaRS activity.

tRNA engineering can also improve the efficiency of an o-aaRS•tRNA pair by optimizing orthogonality of the tRNA, its binding to the o-aaRS or elongation factor Tu, or the decoding strength of the targeted codon [24–29] (Box 1). Also, the role of heterologous tRNA post-transcriptional modifications is emerging as a considerable factor in improving o-tRNA proficiency [30–32]. In cases when poor tRNA orthogonality impedes negative selection (Figure 1c), parallel positive selection schemes may be developed [33••]. With the advances in deep sequencing techniques, the most abundant members of the resulting population can be identified as potential gain-of-function variants.

Polyspecificity of o-aaRS•tRNA pairs

As described above, many aaRSs are polyspecific (i.e. they can charge multiple ncAAs), which can be exploited to incorporate ncAAs without altering the active site of the enzyme. Wild-type PylRS has a remarkably large AA binding site that accommodates >20 lysine derivatives [34•] (Table 1). This inherent polyspecificity is also displayed by engineered o-aaRSs. For example, one engineered PylRS mutant was shown to acylate at least

Figure 2



Advanced methodology for aaRS evolution. **(a)** Multiplex-automated genome engineering (MAGE) [18,19••] uses a modified λ-Red allelic replacement system to introduce changes in a targeted chromosomal gene (e.g. orthogonal aaRS) by integrating mutagenic oligos. Short regions flanking the randomized sequence of the oligo direct β protein to introduce the primer into the replication fork. After two rounds of duplication, only one of the DNA copies from the depicted replication fork will incorporate mutagenic oligos, resulting in the production of no more than 25% of recombinant clones for every round of MAGE. The desired level of diversity is achieved by repeating MAGE cycles, increasing the size of the oligo library, and the number of targeted sites. **(b)** Phage-assisted continuous evolution (PACE) [21,22••] employs a bacteriophage that lacks the gene for protein III (pIII), which is required for infectivity. A copy of the pIII gene, with at least one TAG codon, is encoded in a plasmid in *E. coli*. Active aaRS mutants aminoacylate the suppressor tRNA leading to pIII expression and propagation of bacteriophage. **(c)** Post-translational proofreading (PTP) [39•] exploits the N-end rule for protein degradation in *E. coli*. Green fluorescent protein (GFP) is fused to ubiquitin (Ub), separated by a TAG stop codon. Expression of the full-length reporter is the result of aminoacylation of a suppressor tRNA by orthogonal aaRS variants. Ubiquitin cleavase protein 1 (UBP1) from *S. cerevisiae* cleaves Ub from the reporter, exposing the N-terminus residue of GFP. ClpS surveys the N-terminus of GFP and applies the N-end rule to facilitate degradation of GFP if an undesired AA is incorporated. In contrast, if the desired ncAA is installed at the N-terminus, GFP accumulates and cells with high fluorescence emission are sorted and collected. The system is adaptable as ClpS can be engineered to discriminate against specific ncAAs.

15 different phenylalanine (Phe) analogs [35]. The ncAA polyspecificity of both wild-type and evolved aaRSs may be caused by the lack of exposure to non-cognate substrates in the course of their evolution. In some cases, the ability of o-aaRS to use structurally similar ncAAs can be advantageous [36–38]. In other cases it can hamper unambiguous site-specific incorporation of multiple ncAAs into a single protein (see below).

Using PACE with a round of negative selection, a *M. jannaschii* TyrRS variant that acylates Phe, *p*-nitro-Phe, and iodo-Phe was evolved to increase its selectivity for iodo-Phe >23 fold [22••]. An alternative approach for screening more selective aaRSs was recently devised (Figure 2c) [39•]. This method hijacks a natural protein degradation pathway in *E. coli* to screen for *M. jannaschii* TyrRS variants capable of selectively charging the ncAA biphenylalanine (BipA) and discriminating against other Tyr analogs that were substrates of the progenitor TyrRS. By supplying undesired ncAA substrates during the rounds of negative screening, a TyrRS variant emerged with a much narrower substrate selectivity for BipA. Surprisingly, similar to the PACE-evolved PylRS variants with improved activity, the resulting mutations in *M. jannaschii* TyrRS were located outside of the active site region.

Acylation of amino acids with non-canonical backbones

Recent characterization and engineering of translation apparatus components have shown that genetic incorporation of ncAAs with structures beyond the L-α structure is possible *in vitro* [40•,41]. These efforts have relied on

the synthesis of ncAA-tRNAs using alternative *in vitro* aminoacylation techniques such as flexizyme [42]. Although wild-type aaRSs were known to acylate L-β-AAs and D-α-AAs [43,44] and can facilitate the incorporation L-β-Phe *in vivo* [45], the recombinant protein yield was low. The progress in evolving aaRSs for these AAs will depend on engineered ribosomes that effectively catalyze peptide bond formation between AAs with altered backbones [45,46].

Codon reassignment and development of mutually orthogonal o-aaRS•tRNA pairs

One of the current challenges in GCE is the incorporation of multiple different ncAAs into a single protein. This task faces two difficulties: the need for mutually orthogonal o-aaRS•tRNA pairs (Table 1, [47–49]) and emancipation of more codons that can be reassigned with ncAAs. While UAG-codon and rare sense codon recoded strains [50–53] improve the incorporation of individual ncAAs in response to reassigned codons, additional efforts are needed to make these strains amenable to multiple ncAA installment. ncAA incorporation in response to quadruplet codons capitalizes on the ability of natural translational machinery to tolerate frameshift suppression [54] but requires ribosome adaptation [55] and proficient o-tRNA variants [27]. Emerging techniques to rapidly liberate sense codons [56,57] highlight the need for deeper understanding of viable codon replacements [58] due to the inherent influence of codon usage on various cellular processes, including translation [59].

Organisms with 6-letter genetic alphabets provide an alternative to genomically recoded organisms (Figure 1c) [60]. tRNAs containing unnatural bases within the anticodons can decode cognate unnatural codons and direct site-specific incorporation of ncAAs [61••]. Importantly, the anticodon-recognizing aaRSs can tolerate these unnatural bases within the anticodon [61••,62]. It remains unknown if tRNAs with unnatural bases can achieve full orthogonality.

Outlook

Discovery of new orthogonal aaRS•tRNA pairs through bioinformatics analyses combined with innovative techniques for directed evolution allows selection of o-aaRSs variants with high catalytic efficiency and specificity. This will advance the application of orthogonal translation systems in cutting-edge research, such as biocontainment of synthetic organisms with rewired genomes [53,57,63] or expanded genetic alphabets [61••], production of avirulent, yet fully infectious viruses for effective vaccines [64], development of stable delivery vehicles for gene therapy [65], or a host of still unknown applications.

Conflict of interest statement

Nothing declared.

Acknowledgements

We thank Kyle Hoffman, Hui Si Kwok, Anna Merkuryev, Takahito Mukai and Noah Reynolds for critical reading of the manuscript. Research in the Söll laboratory is supported by grants from the US National Institutes of Health (NIH) (R35GM122560), the US National Science Foundation (CHE-1740549) and from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the Department of Energy (DE-FG02-98ER20311) to D.S.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Mukai T, Lajoie MJ, Englert M, Söll D: **Rewriting the genetic code**. *Annu Rev Microbiol* 2017, **71**:557-577.
This recent review discusses all aspects necessary for full reassigment of a codon with a ncAA and comprehensively annotates the different milestones in the GCE field with a list of all aaRS•tRNA pairs developed to date.

2. Giegé R, Sissler M, Florentz C: **Universal rules and idiosyncratic features in tRNA identity**. *Nucleic Acids Res* 1998, **26**:5017-5035.

3. Nozawa K, O'Donoghue P, Gundllapalli S, Araiso Y, Ishitani R, Umehara T, Söll D, Nureki O: **Pyrrolysyl-tRNA synthetase-tRNA^{Pyl} structure reveals the molecular basis of orthogonality**. *Nature* 2009, **457**:1163-1167.

4. Bullwinkle T, Lazazzera B, Ibba M: **Quality control and infiltration of translation by amino acids outside of the genetic code**. *Annu Rev Genet* 2014, **48**:149-166.

5. Ruan B, Palioura S, Sabina J, Marvin-Guy L, Kochhar S, Larossa RA, Söll D: **Quality control despite mistranslation caused by an ambiguous genetic code**. *Proc Natl Acad Sci U S A* 2008, **105**:16502-16507.

6. Cowie DB, Cohen GN: **Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur**. *Biochim Biophys Acta* 1957, **26**:252-261.

7. Hendrickson WA, Horton JR, LeMaster DM: **Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure**. *EMBO J* 1990, **9**:1665-1672.

8. •• Dieterich DC, Hodas JJ, Gouzer G, Shadrin IY, Ngo JT, Triller A, Tirrell DA, Schuman EM: ***In situ* visualization and dynamics of newly synthesized proteins in rat hippocampal neurons**. *Nat Neurosci* 2010, **13**:897-905.
In this study a pulse-chase application of two ncAAs allowed visualization of newly synthesized neuronal proteins*in situ*.

9. Gan Q, Lehman BP, Bobik TA, Fan C: **Expanding the genetic code of *Salmonella* with non-canonical amino acids**. *Sci Rep* 2016, **6**:39920.

10. He J, Van Treeck B, Nguyen HB, Melançon CE 3rd: **Development of an unnatural amino acid incorporation system in the actinobacterial natural product producer *Streptomyces venezuelae* ATCC 15439**. *ACS Synth Biol* 2016, **5**:125-132.

11. Wang F, Robbins S, Guo JT, Shen WJ, Schultz PG: **Genetic incorporation of unnatural amino acids into proteins in *Mycobacterium tuberculosis***. *PLoS ONE* 2010, **5**.

12. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG: **An expanded eukaryotic genetic code**. *Science* 2003, **301**:964-967.

13. • Suzuki T, Miller C, Guo LT, Ho JML, Bryson DI, Wang YS, Liu DR, Söll D: **Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase**. *Nat Chem Biol* 2017, **13**:1261-1266.
In this study, the crystal structure of a previously uncharacterized N-terminal domain of PylRS in complex with tRNA^{Pyl} reveals the basis of orthogonality of this system in a broad range of organisms. A simplified version of PACE was also developed.

14. Yanagisawa T, Umehara T, Sakamoto K, Yokoyama S: **Expanded genetic code technologies for incorporating modified lysine at multiple sites**. *Chembiochem* 2014, **15**:2181-2187.

15. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D: ***De novo* enzyme design using Rosetta3**. *PLoS ONE* 2011, **6**:e19230.

16. Zeymer C, Hilvert D: **Directed evolution of protein catalysts**. *Annu Rev Biochem* 2018.

17. •• Liu DR, Schultz PG: **Progress toward the evolution of an organism with an expanded genetic code**. *Proc Natl Acad Sci U S A* 1999, **96**:4780-4785.
A pioneering study that established reporter genes that confer antibiotic resistance as the basis for*in vivo* selections. Methods described in this paper became the standard for aaRS evolution.

18. Haimovich AD, Muir P, Isaacs FJ: **Genomes by design**. *Nat Rev Genet* 2015, **16**:501-516.

19. •• Amiram M, Haimovich AD, Fan C, Wang YS, Aerni HR, Ntai I, Moonan DW, Ma NJ, Rovner AJ, Hong SH *et al.*: **Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids**. *Nat Biotechnol* 2015, **33**:1272-1279.
This paper describes the first application of the MAGE technology to accelerate the evolution of the M. jannaschii TyrRS•tRNA pair.

20. Packer MS, Liu DR: **Methods for the directed evolution of proteins**. *Nat Rev Genet* 2015, **16**:379-394.

21. Badran AH, Liu DR: **Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra**. *Nat Commun* 2015, **6**:8425.

22. •• Bryson DI, Fan C, Guo LT, Miller C, Söll D, Liu DR: **Continuous directed evolution of aminoacyl-tRNA synthetases**. *Nat Chem Biol* 2017, **13**:1253-1260.
This study reports the evolution of a PylRS variant with improved catalytic activity using PACE.

23. • Hauf M, Richter F, Schneider T, Faidt T, Martins BM, Baumann T, Durkin P, Dobbek H, Jacobs K, Möglich A *et al.*: **Photoactivatable mussel-based underwater adhesive proteins by an expanded genetic code**. *Chembiochem* 2017, **18**:1819-1823.
The authors used an*in silico* approach (Rosetta3) to create a library of *M. jannaschii* TyrRS with focused diversity at 17 positions incluiding first- and second-shell residues in the active site.

24. Thyer R, Robotham SA, Brodbelt JS, Ellington AD: **Evolving $tRNA^{Sec}$ for efficient canonical incorporation of selenocysteine**. *J Am Chem Soc* 2015, **137**:46-49.

25. Park HS, Hohn MJ, Umehara T, Guo LT, Osborne EM, Benner J, Noren CJ, Rinehart J, Söll D: **Expanding the genetic code of *Escherichia coli* with phosphoserine**. *Science* 2011, **333**:1151-1154.

26. Rogerson DT, Sachdeva A, Wang K, Haq T, Kazlauskaite A, Hancock SM, Huguenin-Dezot N, Muqit MM, Fry AM, Bayliss R *et al.*: **Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog**. *Nat Chem Biol* 2015, **11**:496-503.

27. Wang K, Sachdeva A, Cox DJ, Wilf NM, Lang K, Wallace S, Mehl RA, Chin JW: **Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET**. *Nat Chem* 2014, **6**:393-403.

28. Fan C, Ho JML, Chirathivat N, Söll D, Wang YS: **Exploring the substrate range of wild-type aminoacyl-tRNA synthetases**. *Chembiochem* 2014, **15**:1805-1809.

29. Haruna K, Alkazemi MH, Liu Y, Söll D, Englert M: **Engineering the elongation factor Tu for efficient selenoprotein synthesis**. *Nucleic Acids Res* 2014, **42**:9976-9983.

30. Biddle W, Schmitt MA, Fisk JD: **Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment**. *Nucleic Acids Res* 2016, **44**:10042-10050.

31. Crnković A, Vargas-Rodriguez O, Merkuryev A, Söll D: **Effects of heterologous tRNA modifications on the production of proteins containing noncanonical amino acids**. *Bioengineering (Basel)* 2018, **5**.

32. Baldridge KC, Jora M, Maranhao AC, Quick MM, Addepalli B, Brodbelt JS, Ellington AD, Limbach PA, Contreras LM: **Directed evolution of heterologous tRNAs leads to reduced dependence on post-transcriptional modifications**. *ACS Synth Biol* 2018 http://dx.doi.org/10.1021/acssynbio.7b00421.

33. •• Zhang MS, Brunner SF, Huguenin-Dezot N, Liang AD, Schmied WH, Rogerson DT, Chin JW: **Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing**. *Nat Methods* 2017, **14**:729-736.
In this study, the authors describe a method to rapidly discover aaRS variants with selectivity for the ncAA phosphothreonine. The method uses deep sequencing to screen populations of aaRS variants obtained via parallel positive selection in the presence and absence of phosphothreonine. The method circumvents the need for the negative selection step. Additionally, phosphothreonine was synthesized in situ by introducing the*Samonella enterica* PduX protein in *E. coli*, which converts threonine to phosphothreonine.

34. • Dumas A, Lercher L, Spicer CD, Davis BG: **Designing logical codon reassignment — Expanding the chemistry in biology**. *Chem Sci* 2015, **6**:50-69.
This review provides a comprehensive list of diverse ncAAs with their the corresponding o-aaRS•tRNA pairs used for protein synthesis via GCE.

35. Wang YS, Fang X, Chen HY, Wu B, Wang ZU, Hilty C, Liu WR: **Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant**. *ACS Chem Biol* 2013, **8**:405-415.

36. Peng T, Hang HC: **Site-specific bioorthogonal labeling for fluorescence imaging of intracellular proteins in living cells**. *J Am Chem Soc* 2016, **138**:14423-14433.

37. Bi X, Pasunooti KK, Lescar J, Liu CF: **Thiazolidine-masked α-oxo aldehyde functionality for peptide and protein modification**. *Bioconjug Chem* 2017, **28**:325-329.

38. Joiner CM, Breen ME, Clayton J, Mapp AK: **A bifunctional amino acid enables both covalent chemical capture and isolation of *in vivo* protein-protein interactions**. *Chembiochem* 2017, **18**:181-184.

39. • Kunjapur AM, Stork DA, Kuru E, Vargas-Rodriguez O, Landon M, Söll D, Church GM: **Engineering posttranslational proofreading to discriminate nonstandard amino acids**. *Proc Natl Acad Sci U S A* 2018, **115**:619-624.
This paper reports the engineering of the N-end rule of protein degradation to rapidly screen o-aaRS•tRNA variants with improved specificity toward a ncAA.

40. • Fujino T, Goto Y, Suga H, Murakami H: **Ribosomal synthesis of peptides with multiple β-amino acids**. *J Am Chem Soc* 2016, **138**:1962-1969.
This study demonstrates that wild-type ribosomes can incorporate multiple β-amino acids*in vitro*.

41. Katoh T, Tajima K, Suga H: **Consecutive elongation of D-amino acids in translation**. *Cell Chem Biol* 2017, **24**:46-54.

42. Murakami H, Ohta A, Ashigai H, Suga H: **A highly flexible tRNA acylation method for non-natural polypeptide synthesis**. *Nat Methods* 2006, **3**:357-359.

43. Hartman MC, Josephson K, Szostak JW: **Enzymatic aminoacylation of tRNA with unnatural amino acids**. *Proc Natl Acad Sci U S A* 2006, **103**:4356-4361.

44. Kobayashi T, Yanagisawa T, Sakamoto K, Yokoyama S: **Recognition of non-α-amino substrates by pyrrolysyl-tRNA synthetase**. *J Mol Biol* 2009, **385**:1352-1360.

45. Melo Czekster C, Robertson WE, Walker AS, Söll D, Schepartz A: ***In vivo* biosynthesis of a β-amino acid-containing protein**. *J Am Chem Soc* 2016, **138**:5194-5197.

46. Maini R, Nguyen DT, Chen SX, Dedkova LM, Chowdhury SR, Alcala-Torano R, Hecht SM: **Incorporation of β-amino acids into dihydrofolate reductase by ribosomes having modifications in the peptidyltransferase center**. *Bioorg Med Chem* 2013, **21**:1088-1096.

47. Tharp JM, Liu WR: **Using amber and ochre nonsense codons to code two different noncanonical amino acids in one protein gene**. *Methods Mol Biol* 2018, **1728**:147-154.

48. Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG: **An expanded genetic code with a functional**

quadruplet codon. *Proc Natl Acad Sci U S A* 2004, **101**:7566-7571.

49. Zheng Y, Mukherjee R, Chin MA, Igo P, Gilgenast MJ, Chatterjee A: **Expanding the scope of single- and double-noncanonical amino acid mutagenesis in mammalian cells using orthogonal polyspecific leucyl-tRNA synthetases**. *Biochemistry* 2018, **57**:441-445.

50. Mukai T, Hoshi H, Ohtake K, Takahashi M, Yamaguchi A, Hayashi A, Yokoyama S, Sakamoto K: **Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon**. *Sci Rep* 2015, **5**.

51. Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA *et al.*: **Genomically recoded organisms expand biological functions**. *Science* 2013, **342**:357-360.

52. Mukai T, Yamaguchi A, Ohtake K, Takahashi M, Hayashi A, Iraha F, Kira S, Yanagisawa T, Yokoyama S, Hoshi H *et al.*: **Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli***. *Nucleic Acids Res* 2015, **43**:8111-8122.

53. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CL, Chandrasegaran S, Cai Y *et al.*: **Design of a synthetic yeast genome**. *Science* 2017, **355**:1040-1044.

54. Roth JR: **Frameshift mutations**. *Annu Rev Genet* 1974, **8**:319-346.

55. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW: **Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome**. *Nature* 2010, **464**:441-444.

56. Wang K, Fredens J, Brunner SF, Kim SH, Chia T, Chin JW: **Defining synonymous codon compression schemes by genome recoding**. *Nature* 2016, **539**:59-64.

57. Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M *et al.*: **Design, synthesis, and testing toward a 57-codon genome**. *Science* 2016, **353**:819-822.

58. Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, Liang L, Wang Z, Zeitoun R, Alexander WG *et al.*: **Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering**. *Nat Biotechnol* 2017, **35**:48-55.

59. Chaney JL, Clark PL: **Roles for synonymous codon usage in protein biogenesis**. *Annu Rev Biophys* 2015, **44**:143-166.

60. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Correa IR Jr, Romesberg FE: **A semi-synthetic organism with an expanded genetic alphabet**. *Nature* 2014, **509**:385-388.

61. •• Zhang Y, Ptacin JL, Fischer EC, Aerni HR, Caffaro CE, San Jose K, Feldman AW, Turner CR, Romesberg FE: **A semi-synthetic organism that stores and retrieves increased genetic information**. *Nature* 2017, **551**:644-647.

This paper shows that tRNAs with unnatural bases in the anticodon can specifically decode codons in mRNAs containing unnatural bases in a semi-synthetic organism with an expanded genetic alphabet.

62. Fukunaga R, Harada Y, Hirao I, Yokoyama S: **Phosphoserine aminoacylation of tRNA bearing an unnatural base anticodon**. *Biochem Biophys Res Commun* 2008, **372**:480-485.

63. Gan F, Liu R, Wang F, Schultz PG: **Functional replacement of histidine in proteins to generate noncanonical amino acid dependent organisms**. *J Am Chem Soc* 2018, **140**:3829-3832.

64. Si LL, Xu H, Zhou XY, Zhang ZW, Tian ZY, Wang Y, Wu YM, Zhang B, Niu ZL, Zhang CL *et al.*: **Generation of influenza A viruses as live but replication-incompetent virus vaccines**. *Science* 2016, **354**:1170-1173.

65. Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, Bale JB, Ke S, Lenz GH, Yehdego A *et al.*: **Evolution of a designed protein assembly encapsulating its own RNA genome**. *Nature* 2017, **552**:415-420.

66. Luo X, Fu G, Wang RE, Zhu X, Zambaldo C, Liu R, Liu T, Lyu X, Du J, Xuan W *et al.*: **Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria**. *Nat Chem Biol* 2017, **13**:845-849.

67. McMurry JL, Chang MCY: **Fluorothreonyl-tRNA deacylase prevents mistranslation in the organofluorine producer *Streptomyces cattleya***. *Proc Natl Acad Sci U S A* 2017, **114**:11920-11925.

68. Bacusmo JM, Kuzmishin AB, Cantara WA, Goto Y, Suga H, Musier-Forsyth K: **Quality control by trans-editing factor prevents global mistranslation of non-protein amino acid α-aminobutyrate**. *RNA Biol* 2017:1-10.

69. Ko JH, Wang YS, Nakamura A, Guo LT, Söll D, Umehara T: **Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function**. *FEBS Lett* 2013, **587**:3243-3248.

70. Parrish AR, She X, Xiang Z, Coin I, Shen Z, Briggs SP, Dillin A, Wang L: **Expanding the genetic code of *Caenorhabditis elegans* using bacterial aminoacyl-tRNA synthetase/tRNA pairs**. *ACS Chem Biol* 2012, **7**:1292-1302.

71. Chen Y, Ma J, Lu W, Tian M, Thauvin M, Yuan C, Volovitch M, Wang Q, Holst J, Liu M *et al.*: **Heritable expansion of the genetic code in mouse and zebrafish**. *Cell Res* 2017, **27**:294-297.

72. Han S, Yang A, Lee S, Lee HW, Park CB, Park HS: **Expanding the genetic code of *Mus musculus***. *Nat Commun* 2017, **8**:14568.

73. Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S: **Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases**. *Biochem Biophys Res Commun* 2008, **371**:818-822.

74. Hancock SM, Uprety R, Deiters A, Chin JW: **Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair**. *J Am Chem Soc* 2010, **132**:14819-14824.

75. Li F, Zhang H, Sun Y, Pan Y, Zhou J, Wang J: **Expanding the genetic code for photoclick chemistry in *E. coli*, mammalian cells, and *A. thaliana***. *Angew Chem Int Ed Engl* 2013, **52**:9700-9704.

76. Wang W, Takimoto JK, Louie GV, Baiga TJ, Noel JP, Lee KF, Slesinger PA, Wang L: **Genetically encoding unnatural amino acids for cellular and neuronal studies**. *Nat Neurosci* 2007, **10**:1063-1072.

77. Wang Q, Wang L: **New methods enabling efficient incorporation of unnatural amino acids in yeast**. *J Am Chem Soc* 2008, **130**:6066-6067.

78. Italia JS, Addy PS, Wrobel CJ, Crawford LA, Lajoie MJ, Zheng Y, Chatterjee A: **An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes**. *Nat Chem Biol* 2017, **13**:446-450.

79. Chatterjee A, Xiao H, Yang PY, Soundararajan G, Schultz PG: **A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in *E. coli***. *Angew Chem Int Ed Engl* 2013, **52**:5106-5109.

80. Hughes RA, Ellington AD: **Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA**. *Nucleic Acids Res* 2010, **38**:6813-6830.

81. Kwon I, Kirshenbaum K, Tirrell DA: **Breaking the degeneracy of the genetic code**. *J Am Chem Soc* 2003, **125**:7512-7513.