



Cell-Free Approach for Non-canonical Amino Acids Incorporation Into Polypeptides

Zhenling Cui*, Wayne A. Johnston and Kirill Alexandrov*

Synthetic Biology Laboratory, School of Biology and Environmental Science, Queensland University of Technology, Brisbane, QLD, Australia

OPEN ACCESS

Edited by:

Chenguang Fan,
University of Arkansas, United States

Reviewed by:

Noah Reynolds,
University of Illinois at Springfield,
United States

Patrick O'Donoghue,
University of Western
Ontario, Canada

*Correspondence:

Zhenling Cui
zhenling.cui@qut.edu.au
Kirill Alexandrov
kirill.alexandrov@qut.edu.au

Specialty section:

This article was submitted to
Synthetic Biology,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 19 June 2020

Accepted: 07 August 2020

Published: 28 September 2020

Citation:

Cui Z, Johnston WA and Alexandrov K
(2020) Cell-Free Approach for
Non-canonical Amino Acids
Incorporation Into Polypeptides.
Front. Bioeng. Biotechnol. 8:1031.
doi: 10.3389/fbioe.2020.01031

Synthetic biology holds promise to revolutionize the life sciences and biomedicine via expansion of macromolecular diversity outside the natural chemical space. Use of non-canonical amino acids (ncAAs) via codon reassignment has found diverse applications in protein structure and interaction analysis, introduction of post-translational modifications, production of constrained peptides, antibody-drug conjugates, and novel enzymes. However, simultaneously encoding multiple ncAAs *in vivo* requires complex engineering and is sometimes restricted by the cell's poor uptake of ncAAs. In contrast the open nature of cell-free protein synthesis systems offers much greater freedom for manipulation and repurposing of the biosynthetic machinery by controlling the level and identity of translational components and reagents, and allows simultaneous incorporation of multiple ncAAs with non-canonical side chains and even backbones (N-methyl, D-, β -amino acids, α -hydroxy acids etc.). This review focuses on the two most used *Escherichia coli*-based cell-free protein synthesis systems; cell extract- and PURE-based systems. The former is a biological mixture with >500 proteins, while the latter consists of 38 individually purified biomolecules. We delineate compositions of these two systems and discuss their respective advantages and applications. Also, we dissect the translational components required for ncAA incorporation and compile lists of ncAAs that can be incorporated into polypeptides via different acylation approaches. We highlight the recent progress in using unnatural nucleobase pairs to increase the repertoire of orthogonal codons, as well as using tRNA-specific ribozymes for *in situ* acylation. We summarize advances in engineering of translational machinery such as tRNAs, aminoacyl-tRNA synthetases, elongation factors, and ribosomes to achieve efficient incorporation of structurally challenging ncAAs. We note that, many engineered components of biosynthetic machinery are developed for the use *in vivo* but are equally applicable to the *in vitro* systems. These are included in the review to provide a comprehensive overview for ncAA incorporation and offer new insights for the future development in cell-free systems. Finally, we highlight the exciting progress in the genomic engineering, resulting in *E. coli* strains free of amber and some redundant sense codons. These strains can be used for preparation of cell extracts offering multiple reassignment options.

Keywords: genetic code expansion, genetic code reprogramming, cell-free protein synthesis, non-canonical amino acids, amber suppression, codon reassignment

1. INTRODUCTION

Proteins are the central functional constituents in all living organisms that are formed by ribosomal biosynthesis using genetically determined sequences of the 20 canonical amino acids with the rare exception of two recently discovered proteinogenic amino acids selenocysteine (Sec) (Bock et al., 1991) and pyrrolysine (Pyl) (Srinivasan et al., 2002). Genetic encoding of non-proteinogenic or non-canonical amino acids (ncAAs) in living organisms poses many challenges due to the requirement for additional transport and biosynthetic components. Using multiple rounds of negative and positive screening, the Schultz group pioneered engineering orthogonal tRNAs and aminoacyl-tRNA synthetases (aaRSs) for incorporation of tyrosine analogs into proteins *in vivo* (Wang et al., 2001). This pioneering work utilizing amber codon suppression opened up a new field of genetic code expansion (Wang et al., 2006b; Liu and Schultz, 2010). Subsequent efforts resulted in demonstration of ribosomal site specific incorporation of >200 chemical entities into proteins endowing them with novel physical, chemical, and biological properties (Neumann, 2012; Dumas et al., 2015; Xiao and Schultz, 2016).

While the focus of the initial efforts has been on utilization of the self-replicating nature of living organisms to produce polypeptides with ncAAs, recently cell-free systems have become progressively more utilized for this purpose. As a testimony to the practical value of this approach, two successful biotechnology companies, Sutro Biopharma Inc. and PeptiDream Inc., utilize ncAA incorporation into polypeptides using the *E. coli* cell-free protein synthesis systems (CFPS). At present, these companies are estimated to have approximately \$220M and \$4.9B market value, respectively.

In this review, we first compare the crude (S30) and fully reconstituted (PURE) *E. coli* cell-free protein synthesis systems (Table 1) and discuss their strengths and weaknesses. We then delineate the strategies and components required for the template-directed ncAA incorporation, including creation of vacant codons and synthesis of acylated tRNAs. The recent invention of unnatural nucleic acid base pairs can be used as orthogonal codons to dramatically expand the genetic code *in vivo* and *in vitro*. We summarize advances in engineering of the translational machinery (tRNAs, aaRSs, elongation factors, and ribosomes). Finally, we highlight the exciting progress in the genome engineering field that led to *E. coli* strains free of amber codons and some redundant sense codons and their use for preparation of more efficient CFPS.

2. TYPES OF *Escherichia coli* CFPS

2.1. Crude Extract CFPS

The gene transcription and translation can take place in crude *E. coli* extract supplemented with rNTPs, amino acids, energy sources, and RNA polymerases. The first report of an *E. coli* crude cell-free translation system dates back to the 1960s when it was used to decipher the genetic code (Nirenberg and Matthaei, 1961; Nirenberg and Leder, 1964). Since then, many variants of the crude CFPS have been developed (Dopp B. J. L. et al.,

2019) and used for elucidation of many biochemical processes, prototyping of metabolic and gene expression circuits including the glycosylation pathway, and for high yield therapeutic biomanufacturing (Carlson et al., 2012; Guarino and DeLisa, 2012; Quast et al., 2015; Jaroentomeechai et al., 2018; Kightlinger et al., 2019; Silverman et al., 2019).

Preparation of the most widely used S30 cell extract for CFPS is simple and inexpensive. It involves cell disruption followed by a run-off incubation procedure that frees the ribosomes and degrades the mRNAs, several high-speed centrifugations that removes non-essential proteins. Approximately 500–1,000 proteins were identified in *E. coli* S30 lysate (A19 and BL21 Rosetta2) (Foshag et al., 2018; Garenne et al., 2019). These represent 20–40% of the *E. coli* proteome (Iwasaki et al., 2010; Schmidt et al., 2016). The extract also contains residual amounts of membrane proteins including the respiratory chain involving several subunits of the ATP-synthase and NADH-quinone oxidoreductase. These proteins enable oxidative phosphorylation in the cell free system for energy regeneration under the condition known as “Cytomim” (Jewett and Swartz, 2004; Zawada et al., 2011; Cai et al., 2015).

The Cytomim system closely mimics *E. coli* cellular metabolism and integrates complex metabolic networks to provide new sources for ATP production (Jewett and Swartz, 2004; Jewett et al., 2008). This system has been effectively scaled up from 10 μ l to 100 L and used for production of disulfide bond containing cytokine proteins, achieving 0.7 mg/ml yield in a long (>10 h) batch reaction (Zawada et al., 2011). Better understanding of and control over the chemical reactions in the cell extracts dramatically enhanced the productivity and scalability of protein synthesis, thereby transforming the crude CFPS systems into useful platforms for high yield protein production and therapeutic biomanufacturing (Carlson et al., 2012). Other energy generation systems include the PANOX-SP (acronym for PEP, Amino acids, NAD, Oxalic acid, Spermidine, and Putrescine) (Jewett and Swartz, 2004; Albayrak and Swartz, 2013; Martin et al., 2018) and 3-phosphoglyceric acid/Maltose dual system (Caschera and Noireaux, 2014; Garamella et al., 2016). They also take advantage of the *E. coli* cellular metabolism to perform ATP generation for extended periods achieving model protein production yields in excess of 1 mg/ml in batch mode.

However, the S30 system contains other factors, such as nucleases, RNases, and proteases, that negatively affect the translation productivity (Foshag et al., 2018; Garenne et al., 2019). The presence of endo- and exo- nucleases causes rapid decay of linear PCR templates. In efforts to avoid time-consuming and costly plasmid amplification and purification, linear templates were utilized to prime the translation reaction using S30 lysate from dedicated *E. coli* strains including BL21 star (DE3) (Ahn et al., 2005; Shrestha et al., 2014), SL119 (Lesley et al., 1991), NMR5 (A19 Δ recCBD::red-kan Δ endA) (Michel-Reydellet et al., 2005). These strains are deficient in endonuclease RNase E, exonuclease V, endonuclease I, and exonuclease V, respectively. Other strategies to protect linear DNA include utilization of Lambda Gam protein (Sun et al., 2014), short auxiliary double-stranded DNAs encoding χ sites that preferentially bound to

TABLE 1 | Comparison of the crude and PURE *E. coli* cell-free protein synthesis systems.

	Crude CFPS	PURE CFPS
Developed	Year 1964	Year 2001
Cost	<\$0.05/reaction	\$10/reaction
Composition	Crude extract (>500 proteins) + necessary factors	36 purified proteins+ tRNAs + ribosomes + necessary factors
Energy source	Versatile energy sources including variants that take advantage of cellular metabolism	Creatine phosphate/creatine kinase ATP regeneration system
Flexibility	Partial control of translational machinery	Full control of translational machinery
Linear template	Unstable, requires special strains, or lysate treatment	Stable
Peptide synthesis	Unstable, requires special strains, or lysate treatment	Stable
Applications	Small scale protein expression Large scale protein biomanufacturing Biological process prototyping	Peptide expression for mRNA display Isotopically labeled peptides (MS standard) Translation factor characterization
Compatibility with selection systems	Ribosome display Microbead display (<i>in vitro</i> compartmentation)	mRNA display cDNA display
Amber suppression	Frequently used	Seldom used
Initiation reassignment	Seldom used Limited amino acid structures	Commonly used A wide range of ncAAs (Figure 5)
Elongator codon reassignment	Less used	Commonly used (residue-specific manner) A wide range of ncAAs (Figure 6)

RecBCD recombination machinery (Marshall et al., 2017), steric protection of the linear DNAs with 250–500 bp protective sequences (Sun et al., 2014) and circularization of the linear fragments (Dopp J. L. et al., 2019). The presence of RNases in the cell extract causes rapid RNA degradation and prevents its use for mRNA display (Liu et al., 2000). However, S30 extract from the MRE600 strain, which is deficient in ribonuclease I, is compatible for ribosome display presumably due to at least partial RNA protection by the ribosome (Hanes and Pluckthun, 1997; Yonezawa et al., 2003; Dreier and Pluckthun, 2018). Further the commercial S30 based CFPS kit has been integrated with *in vitro* bead display for binder selection (Huang et al., 2013).

Utilization of crude CFPS for ncAA incorporation, coupled with chemoenzymatic acylated tRNAs for amber or frameshift suppression, dates back nearly 40 years to the 1980s (Noren et al., 1989). This approach suffers from low protein yield due to the single turnover of the pre-charged, non-regeneratable ncAA-tRNAs and has been largely replaced by the subsequently developed orthogonal tRNA and aminoacyl-tRNA synthetase pairs (o-tRNA/aaRSs). These are defined as orthogonal translation systems (OTSs) and have been widely interfaced with the crude CFPS for high yield ncAA-protein production mainly via amber suppression (Goerke and Swartz, 2009; Hong et al., 2014a). Recent advances in using this approach for ncAA incorporation include production of antibody drug conjugates (Zimmerman et al., 2014) and phosphoproteins with single or multiple phosphoserines (Oza et al., 2015); site-specific PEGylation (Wilding et al., 2018); incorporation of ncAAs via reassigned sense codons (Cui et al., 2017, 2018) and the use of cell-extract from genome recoded bacteria with high amber suppression efficiency (Martin et al., 2018; Des Soye et al., 2019).

2.2. PURE CFPS

PURE (Protein synthesis Using purified Recombinant Elements) system represents a key breakthrough in cell free synthetic biology. The system was first reported in 2001 and was a result of a thorough understanding of the mechanism of prokaryotic transcription and translation, along with advancements in the isolation of functional translational machinery (Shimizu et al., 2001; Shimizu and Ueda, 2010). The system is composed of 36 individually purified protein factors combined with purified ribosomes, total tRNA mixture, and the necessary small chemical compounds and ions. The protein factors include the T7 RNA polymerase for transcription, a full set of 20 aminoacyl-tRNA synthetases for continuous tRNA aminoacylation, 10 translation factors for initiation, elongation, and termination (initiation factors IF1, IF2, IF3, elongation factors EF-Tu, EF-Ts, and EF-G, release factors RF1, RF2, RF3, and a ribosome recycling factor RRF), as well as 5 enzymes including methionyl-tRNA transformylase, creatine kinase, myokinase, nucleoside diphosphate kinase, and pyrophosphatase. The creatine phosphate: creatine kinase system is commonly used for ATP generation in PURE system and may be responsible for its relatively short productive phase.

PURE contains negligible amounts of factors that negatively affect protein translation. Low levels of nuclease, RNase, and protease activities allow the use of linear DNA or mRNA as templates for efficient polypeptide translation. Protein folding can be improved by addition of appropriate molecular chaperons as well as by adjusting the redox potential of the reaction (Shimizu et al., 2005). This system has been used to dissect the functional mechanism of different components of protein biosynthetic machinery such as ribosome rescue factors (Shimizu, 2012) and protein chaperones (Niwa et al., 2012). It

has also been used to produce membrane proteins (Kuruma and Ueda, 2015) as well as isotope-labeled peptides as mass spectrometry standards (Narumi et al., 2016, 2018).

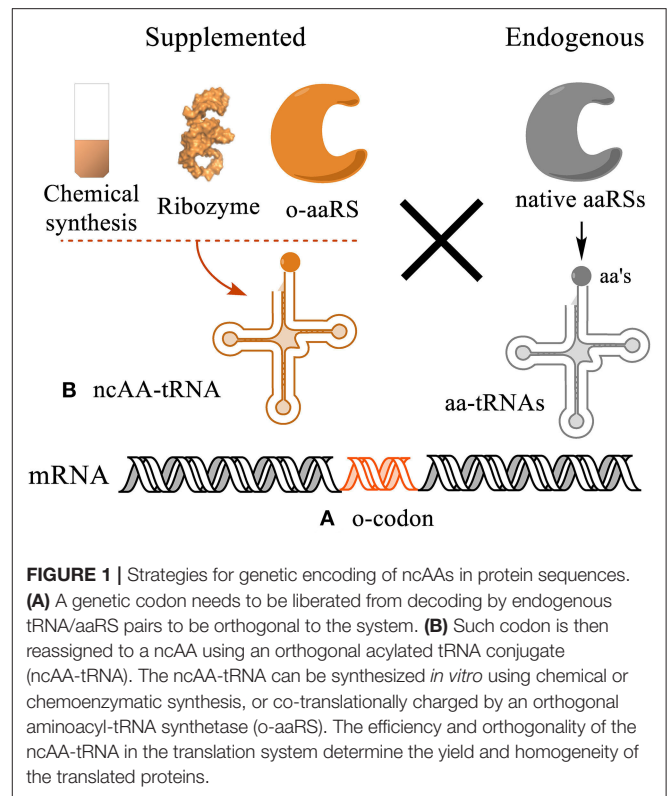
PURE system provides superior control over individual components that can be manipulated to achieve the desired translational outcomes. Majority of the work using PURE has focused on reprogramming sense codons by omitting the corresponding amino acids and aaRSs. Development of Flexizyme (*Flexible tRNA acylation Ribozyme*) transformed the PURE system into a very powerful platform for production of non-standard peptides. In particular it enables the production of peptides cyclized in various ways including backbone cyclisation, backbone-side chain cyclisation, as well as producing bicyclic and tricyclic peptides (Ito et al., 2013; Passioura and Suga, 2014; Bashiruddin et al., 2015; Yin et al., 2019). More than 300 different ncAAs including N-methyl and D-amino acids were incorporated into such peptides endowing them with drug-like properties (Figures 5, 6).

Integration of the PURE system with Flexizyme tRNA aminoacylation technology and mRNA display has resulted in a powerful *in vitro* selection platform for constrained peptides (Passioura and Suga, 2017; Huang et al., 2019). The platform can generate macrocyclic peptide libraries with diversity approaching 10^{13} . Screening of such libraries consistently yields a range of highly selective and potent protein binders capable of stabilizing protein conformations and disrupting protein: protein interactions. Such peptides can be developed as research agents, affinity ligands, vaccines as well as pharmaceutical drugs (Passioura, 2019). Recent examples include HiP-8, a macrocyclic thioether peptide consisting of 12 amino acids, that selectively recognizes active hepatocyte growth factor and can be used in potential diagnosis and treatment of cancers (Sakai et al., 2019). Another example is ub4ix that tightly and specifically binds to K48-linked Ubiquitin chains and protects them from deubiquitinating enzymes and degradation (Nawatha et al., 2019). Additionally, this macrocyclic peptide can enter cells, inhibit growth, and induce programmed cell death, opening new opportunities for therapeutic intervention (Nawatha et al., 2019).

2.3. Comparison of the S30 and PURE Systems

A comparison of the S30 and PURE systems is provided in Table 1. Crude CFPS is generally more economical and efficient in terms of protein production. One study demonstrated that the S30 system generated 5 times more proteins than PURE (190 and 41 $\mu\text{g/ml}$ of *p*-propargyloxy-L-phenylalanine-sfGFP), and is much more cost-effective (<\$0.05/reaction and \$10/reaction) (Hong et al., 2014b). Recently reported variants of the S30 system produced ncAA-containing proteins at mg/ml quantities with >95% fidelity using o-tRNA/aaRS (Martin et al., 2018). However, peptides are generally unstable in S30 cell extract, which requires additional treatment to inhibit protease activity and improve product stability (Cui et al., 2015). In contrast, PURE is well-suited for peptide production due to the lack of contaminating proteases (Pardee et al., 2016).

Currently, genetic code expansion or reprogramming in crude CFPS is far less advanced than that in the PURE CFPS system as well as in the *in vivo* system. Only a limited number of



ncAAs has been explored (Quast et al., 2015). Manipulation of the translational machinery, the tRNAs, aaRSs, and ribosomes, in crude CFPS is less straightforward as in the PURE system. However, it is possible to separate the native tRNA pool and the ribosome from S30 cell extract. Ablation/inactivation of a specific tRNA can liberate the corresponding sense codon(s) for coding ncAAs (Cui et al., 2018). The cell extract centrifuged at 150,000 g (S150) can be separated from the ribosomes and used for *in vitro* construction of modified ribosomes with specialized activity (Jewett et al., 2013).

3. GENETIC ENCODING OF ncAAs

The genetic code determines translation of 64 nucleotide triplets into 20 canonical amino acids. To incorporate ncAAs into protein sequences, a genetic codon needs to be reassigned to a ncAA by an orthogonal acylated tRNA (Figure 1). This process includes two key elements: (1) Creation of orthogonal codons decoded only by orthogonal acylated tRNAs but not by any endogenous acylated tRNAs. (2) Synthesis of orthogonal ncAA-tRNA conjugates that specifically recognize the orthogonal codon, but not the other codons. This can be carried out both *in vivo* and *in vitro* in two fashions: site specific and residue specific.

3.1. Creation of Vacant Codons for Reassignment

The first prerequisite for ncAA incorporation is creation of an orthogonal codon that is not decoded by the endogenous tRNA/aaRS pair(s). This could be natural triplets (stop

TABLE 2 | Choice and Method for creating Orthogonal Codons.

Potential o-codons	Number	Orthogonal	Orthogonality creation	Competition	Reassignment mode
Stop codons	3	Yes	—	RF1	Site-specific
Sense codons	61	No	Deletion of a specific aa ± aaRS	aa-tRNA(s)	Residue-specific
		No	Deletion of the cognate tRNA(s)	aa-tRNA(s)	Site-specific
Quadruplets	256	Yes	—	aa-tRNA(s) that recognize the 1st three bases as triplets	Site-specific
Triplets with novel base(s)	variable	Yes	—	Depending on the orthogonality of new bases	Site-specific

and sense codons), unnatural triplets, as well as four-base-codons i.e., Quadruplets (Table 2). Incorporation of ncAAs via reassignment of stop or quadruplet codons results in site-specific incorporation. Sense codon reassignment can be performed in both site- and residue-specific fashion depending on the orthogonality generation approach (Table 2). Site-specific sense codon reassignment means the reassignment of a subset of synonymous sense codons (mostly one or two) to a ncAA while the rest of synonymous codons are retained for the natural amino acid, leading to genetic code expansion. Residue-specific approach indicates the reassignment of all the synonymous codons of a natural amino acid to a ncAA (genetic code reprogramming). This will be discussed later in detail.

3.1.1. Stop Codon and Nonsense Suppression

Nonsense codons (*amber*, *opal*, and *ochre*) that do not have decoding tRNA/aaRS pairs, are naturally orthogonal to both *pro*- and *eukaryotic* protein synthesis systems (living organisms and CFPSs). Protein translation generally terminates at these codons through the action of release factors (Figure 2). Amber codon is the least used stop codon in the *E. coli* genome (frequency of 7–9%) and is the most used codon for site-specific ncAA incorporation. Many approaches have been developed to increase the *amber* suppression efficiency by silencing or removing competing RF1 from the system (Martin et al., 2018; Adachi et al., 2019). *Ochre* codon that faces competition from both RF1 and RF2 is less frequently used for genetic code expansion (Wan et al., 2010; Odoi et al., 2013). Background suppression of *opal* codon is generally high (10–25% of the wild type protein activity), leading to contamination of the modified proteins with mis-incorporated Trp (O'Donoghue et al., 2012; Odoi et al., 2013). This is likely to occur due to a near-cognate suppression of UGA codon by tRNA^{Trp}CCA through wobble base pairing at the 3rd codon position.

Nonsense suppression generally limits the number of incorporable ncAAs types to two, as one stop codon needs to be reserved for translational termination. However, a recent publication demonstrated the simultaneous utilization of all the three nonsense codons to encode three distinct ncAAs (Italia et al., 2019).

3.1.2. Sense Codons and Sense Codon Reassignment

The key obstacle in the reassignment of sense codons is the presence of aminoacyl-tRNAs (aa-tRNAs) that compete with the introduced ncAA-tRNA conjugates (Figure 2). Depending on the approach used to prevent the interference of native acylated tRNAs, sense codon could be reassigned either in residue- or site-specific fashion.

In principle eliminating certain amino acid from the translation reaction can free all its synonymous codons leading to reassignment of all triplets encoding for that amino acid. However, in reality it is very difficult to completely remove residual amount of that amino acid from the system. A notable exception is the AUG start codon which could be faithfully assigned to a ncAA by simply omitting Met in the translation reaction. In the PURE system, simultaneous exclusion of both the amino acid and the corresponding aaRS is effectively used to liberate the codons (Goto et al., 2011). All the synonymous codons for the same amino acid are reassigned to a ncAA. This approach is referred to as “genetic code reprogramming” since the genetic code is recoded from one canonical amino acid to a ncAA in residue-specific manner (Ohta et al., 2007).

In the crude CFPS, S30 lysate prepared from an auxotroph strain or treated by gel filtration was used to deplete certain amino acid from the system (Torizawa et al., 2004; Singh-Blom et al., 2014). By taking advantage of the promiscuity of certain aaRS, this approach was adapted for *in vitro* synthesis of streptavidin protein with tryptophan analogs (Singh-Blom et al., 2014). However, drastic change of protein composition in residue-specific manner from one canonical amino acid to a ncAA is sometimes detrimental to protein functionality and limits this approach to applications such probing proteomic changes in specific cell types (Saleh et al., 2019).

In contrast, to achieve site-specific ncAA incorporation using sense codon reassignment the degeneracy of the genetic code should be broken to liberate at least one codon for ncAA incorporation while maintaining the decoding of other synonymous codon(s) to the native amino acid. In the *E. coli* genome, more than 40 tRNAs decode the 61 sense codons to 20 amino acids (Figure 2) (Cui et al.,

aa. Codon	tRNA	aa. Codon	tRNA	aa. Codon	tRNA	aa. Codon	tRNA
F UUU	← gaa	S UCU	← gga	Y UAU	← g*ua	C UGU	← gca
F UUC	← u*aa	S UCC	← cga	Y UAC	← RF1/2	C UGC	← cca
L UUA	← c*aa	S UCA	← ggg	Ochre UAA	← RF1	Opal UGA	← RF2
L UUG	← gag	S UCG	← u*gg	Amber UAG	← ccg	w UGG	← a*cg
L CUU	← cag	P CCU	← gcu	H CAU	← g*ug	R CGU	← gcu
L CUC	← u*ag	P CCC	← cgu	H CAC	← u*uu	R CGC	← ccu
L CUA	← ggu	P CCA	← gcc	Q CAA	← g*uc	R CGA	← gcc
L CUG	← cgu	P CCG	← g*uc	Q CAG	← u*cc	R CGG	← ccc
I AUU	← gau	T ACU	← gac	N AAU	← u*ac	S AGU	← gac
I AUC	← c*ag	T ACC	← gcc	N AAC	← u*uc	S AGC	← gcc
I AUA	← c*au	T ACA	← g*uc	K AAA	← u*cc	R AGA	← g*uc
M AUG	← c*au	T ACG	← g*uc	K AAG	← u*cc	R AGG	← c*uu
V GUU	← gac	A GCU	← g*uc	D GAU	← u*cc	G GGU	← g*uc
V GUC	← u*ac	A GCC	← g*uc	D GAC	← u*cc	G GGC	← g*uc
V GUA	← g*uc	A GCA	← g*uc	E GAA	← u*cc	G GGA	← g*uc
V GUG	← g*uc	A GCG	← g*uc	E GAG	← u*cc	G GGG	← g*uc

FIGURE 2 | The genetic code and its decoding tRNA isoacceptors of *E. coli*. The genetic codons are shown in upper case while their decoding tRNAs are shown in lower case representing by their respective anticodons. The * indicates the modification on the nucleotide at 34 position of the native tRNAs (N_{34}). The three stop codons (red) are read by release factors to signal the translation termination. The 6- and 4-fold degenerate amino acids that can be used for sense codon reassignment are highlighted in blue and green, respectively. The Met AUG codon (in bold) acts as initiator and elongator via decoding by fMet-tRNA^{Met}CAU and Met-tRNA^{Met}CAU, respectively. These two acylated tRNAs are delivered to the ribosomal P site by IF2 and to A site by EF-Tu for translation initiation and elongation, respectively.

2015). The decoding preference provides a valuable guide for identifying “orthogonal” vs. “native” codon pairs from the synonymous codons of a particular amino acid. Such pairs can either be created from the codons of different families of 6-fold-degenerate amino acids or from those derived from the unsplit codon family boxes with restricted wobble recognitions. The cognate tRNA(s) should be inactivated or excluded from the translation system to create orthogonal codons.

The Suga group replaced the native tRNA mixture with 32 synthetic tRNAs in the PURE system (Iwane et al., 2016). This resulted in genetic code expansion to ncAAs while maintaining 20 canonical amino acids. They demonstrated the expansion of the amino acid repertoire from 20 to 23 by artificially dividing the upsplitted codon boxes of Arg, Gly, and Val where the G- and C-ending codons in the same codon boxes were translated to 2 amino acids (1 canonical amino acid +1 ncAA) using wild type synthetic tRNASNN/aaRS pair and precharged ncAA-tRNA^{AsnE2}SNN (S=G or C), respectively. The drawback of this approach is in its complexity, and in the effort required to synthesize and purify the system’s individual components including synthetic tRNAs.

In an alternative approach our group created S30 extract chromatographically depleted of all tRNAs. The translational activity of such cell extract is dependent on the supplementation with a semi-synthetic tRNA complement (Cui et al., 2015). Similar to Suga’s method, the requirement to synthesize the semi-synthetic tRNA complement complicates broad utilization of the method. We circumvented this problem by utilizing DNA-hybridization chromatography to deplete the native tRNA mixture of certain tRNA isoacceptors thereby creating vacant codons (Cui et al., 2017). In a further development of this approach we bypassed chromatography steps by sequestering specific tRNA(s) in lysate using methylated antisense oligonucleotides (Cui et al., 2018). These form essentially irreversible complex with the desired tRNA(s) sequestering them from the translational machinery. Using this approach, we demonstrated reassignment of AGC/U and AGG codon to ncAAs while retaining UCN and CGN for Ser and Arg, respectively. This approach is simple and scalable as it involves only liquid handling. Furthermore, it is applicable to cell-free systems of any origin. We demonstrated this by applying the approach to a *Leishmania* derived eukaryotic cell-free expression system (Cui et al., 2018).

3.1.3. Quadruplet Codons and Frameshift Suppression

Another approach for site-specific ncAA incorporation is frameshift suppression. This approach was applied to crude CFPS in conjunction with chemoenzymatic synthesized ncAA-tRNAs (Quast et al., 2015). Theoretically quadruplet codons could provide 256 blank codons. However, encoding a ncAA via four-base codons is far less efficient than the use of nonsense suppression, whereby the yield of a target protein is lowered by misreading of the quadruplet as a triplet. The most commonly used quadruplets *in vivo* comprise rare triplets plus a fourth nucleotide, such as AGGN (Neumann et al., 2010; O'Donoghue et al., 2012). Use of rare codons is preferred to reduce the competition with endogenous acylated tRNA(s). Modified ribosomes were used to enhance the quadruplet's suppression efficiency (Neumann et al., 2010).

3.1.4. New Base Pairing

Unnatural base pairs (UBPs) hold the promise to dramatically increase the information storage of DNA, RNA, and therein encoded protein sequences. However, retrieval of the genetic information stored in the DNA molecules requires a series of biochemical steps ranging from faithful DNA replication, transcription of the UBPs in the context of the DNA templates to RNAs (mRNAs and tRNAs), and finally efficient aminoacylation of tRNAs harboring the modified anticodons that are capable of decoding the triplets with UBPs in the ribosomal center. Ensuring that all these steps proceed with sufficient affinities and fidelity requires a large engineering effort.

Many UBP pairs have been reported (Malyshev and Romesberg, 2015). These include isoG (keto): isoC pair which mimics the natural G:C pair forming 3 hydrogen bonding (Figure 3). The ribosomal decoding of the unnatural codons (*iosCAG*) using a tRNA harboring its complementary anticodon (*CUisoG*) is effective in mediating ncAA incorporation (Bain et al., 1992). However, isoG could isomerize and form hydrogen bonding with T, which diminishes its orthogonality causing elimination after several rounds of DNA replications.

UBPs based on hydrophobic interactions such as dNaM-dTPT3 in DNA molecules have been used to build semisynthetic organism (SSO) (Zhang Y. et al., 2017; Fischer et al., 2020). DNA templates with these UBPs can be transcribed into mRNA and tRNA molecules with the respective unnatural nucleotides containing TPT3 and NaM (Figure 3). These novel tRNAs could decode the unnatural codons to ncAAs using o-aaRSs. Theoretically, one extra UBP pair could expand the genetic code from 64 to 216 triplets. However, some UBP-containing codons suffer with poor *in vivo* retention and require further optimization (Dien et al., 2018). Fischer et al. recently systematically analyzed a range of unnatural codons containing this UBP pair and identified at least nine codons capable of mediating protein production with little or no detectable contamination with native proteins (Fischer et al., 2020). Impressively, they demonstrated that at least 3 of these unnatural codons were mutually orthogonal and could be simultaneously used to decode three distinct amino acids including two ncAAs. This study showcased the first 67-codon organism with UBPs

demonstrating a possibility of creating stable SSO with chemical metabolism significantly different from native organisms. Of note is that the lack of interbase hydrogen bonding in these UBPs restricts their usage to the 2nd or 3rd position of the codons. Poor decoding of the unnatural triplets with a UBP at the 1st position is possibly due to their failure to adopt an appropriate structure or to engage the type I A-minor interaction, which is crucial for selecting correct Watson-Crick like geometry at the ribosomal decoding center.

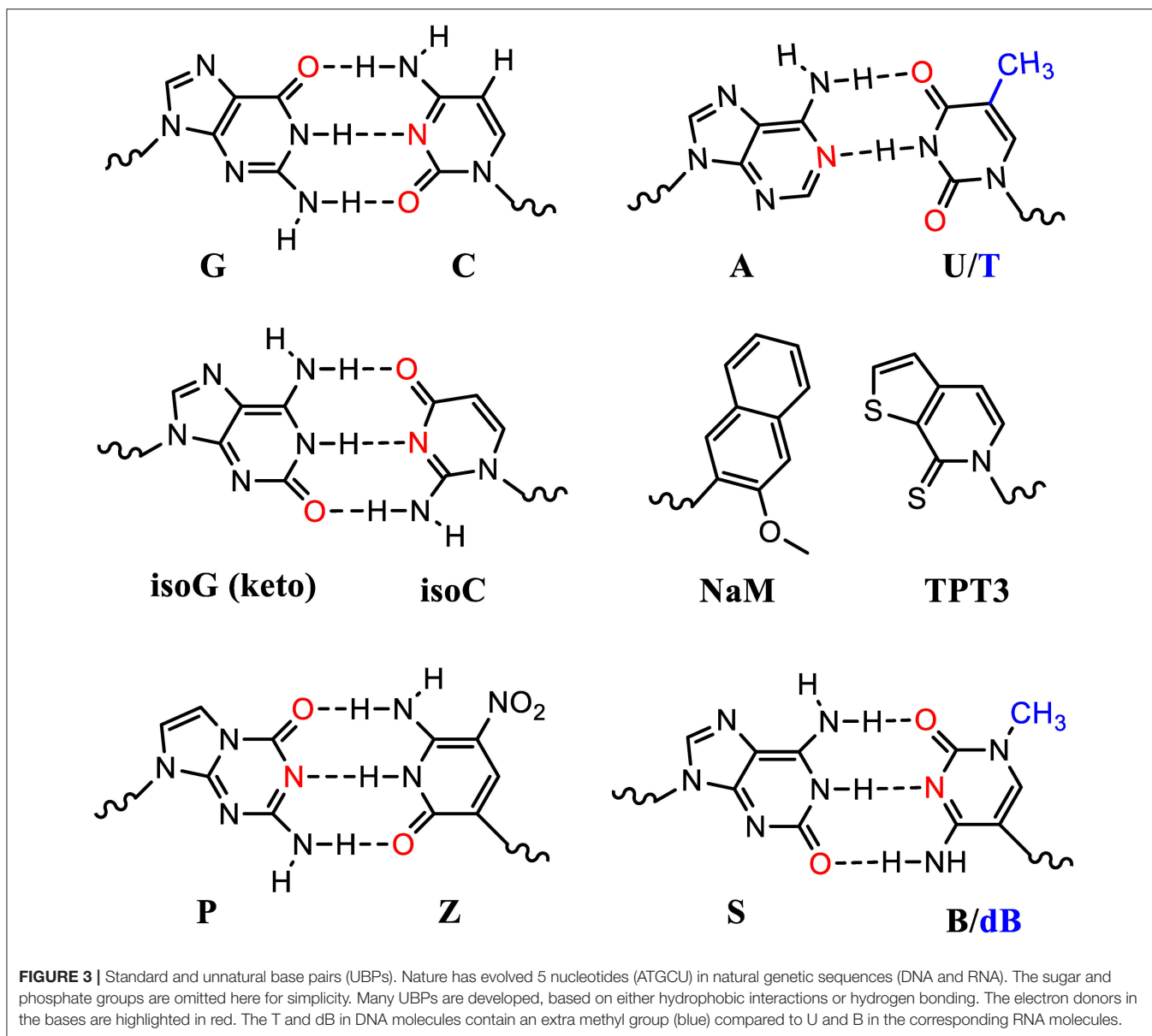
Another two nucleotide analog pairs interacting via hydrogen bonding (P:Z, B:S) were recently reported increasing the number of units in the genetic alphabet from 4 to 8 (A, T, G, and C, purine analogs P and B, and pyrimidine analogs Z and S) (Hoshika et al., 2019). The P:Z and S:B (or S:dB in RNA) pairs are promising UBPs that mimic the natural nucleotides and are compatible with DNA replication and RNA transcription (Figure 3). With increased information density and storage capacity, this 8-letter DNA/RNA genetic system holds the promise of dramatically increasing diversity of building blocks for protein synthesis. However, the efficiency and orthogonality of these UBPs compared to the natural ones in terms of codon decoding in the ribosomal center still need to be evaluated.

3.2. Orthogonal ncAA-tRNAs

Acylation of an orthogonal tRNA with the desired ncAA to generate ncAA-tRNA conjugate is another key step in genetic code expansion and reprogramming. Numerous synthetic routes have been developed, including the aforementioned chemoenzymatic tRNA acylation (Hecht et al., 1978), enzymatic acylation with orthogonal aaRSs (Wang et al., 2006b) or ribozymes such as Flexizyme (Murakami et al., 2006), as well as some other less used approaches which take advantage of the promiscuity of native aaRSs (Hartman et al., 2006, 2007; Iqbal et al., 2018) and post-aminoacylation modifications (Merryman and Green, 2004; Gubbens et al., 2010). Besides the aaRS route, the majority of the acylation methods generate ncAA-tRNA conjugates *in vitro* and are supplemented into the translation reaction where they support single turnover translation thereby limiting the yield of modified polypeptides.

Chemoenzymatic and Flexizyme acylation are among the most used *in vitro* tRNA acylation methods, enabling the incorporation of chemical functionalities that are quite distinct from native amino acids and therefore cannot be conjugated via aaRS route. The former involves enzymatic ligation of chemically prepared acylated dinucleotides to truncated tRNAs lacking the 3'-CA dinucleotide (Hecht et al., 1978). This, well-established, method is theoretically applicable to any substrates but requires multiple steps of chemical synthesis and suffers from low yield. The Flexizyme represents an important advancement in tRNA acylation allowing incorporation of a wide range of ncAAs into polypeptides.

A recent exciting work on directed evolution of ribozymes identified the so-called T-boxzyme, which enabled *in situ* aminoacylation of tRNA^{Gly}GCC and one-pot synthesis of peptides with N-terminal biotin group (Ishida et al., 2020). Unlike the Flexizyme, this T-boxzyme is based on the T-box riboswitch of *Bacillus subtilis* glyQS and thereby is tRNA^{Gly}



specific. Although co-translational charging of ncAA via T-boxzyme was showcased, the efficiency of peptide synthesis was several times lower than that using supplemented ncAA-tRNA conjugate. Further optimization is expected to turn this technology into a useful tool.

3.2.1. Flexizyme Approach

3.2.1.1. Flexizyme and its substrates

Flexizyme was created to mimic the function of aaRSs but using RNA as catalyst (Morimoto et al., 2011). A Flexizyme family includes dinitro-flexizyme (dFx) (Murakami et al., 2006), enhanced flexizyme (eFx) (Murakami et al., 2006), and amino-flexizyme (aFx) (Niwa et al., 2009). These ~45 nucleotide long RNA molecules can acylate tRNAs using a wide range of substrates with appropriated leaving groups (**Figure 4A**).

Acylation efficiency varies depending on the structure of the acyl-donor substrate, the leaving group, and the type of flexizyme (Goto et al., 2011).

Recently, a large panel of substrates including Phe analogs, benzoic acid derivatives, heteroaromatic amino acids, and aliphatic acids were synthesized to study the Fx acceptance rules (Lee J. et al., 2019). Reaction rates of eFx toward its aromatic substrates generally correlate with the electronic character of the substrate benzol ring, increasing for electron-poor substrates while decreasing for electron-rich substrates. The effective substrates could form either T-stacked or parallel stacked interactions with eFx. The aFx prefers aliphatic substrates with straight chain, while the substrates with increased steric bulk decrease aFx-catalyzed acylation activity.

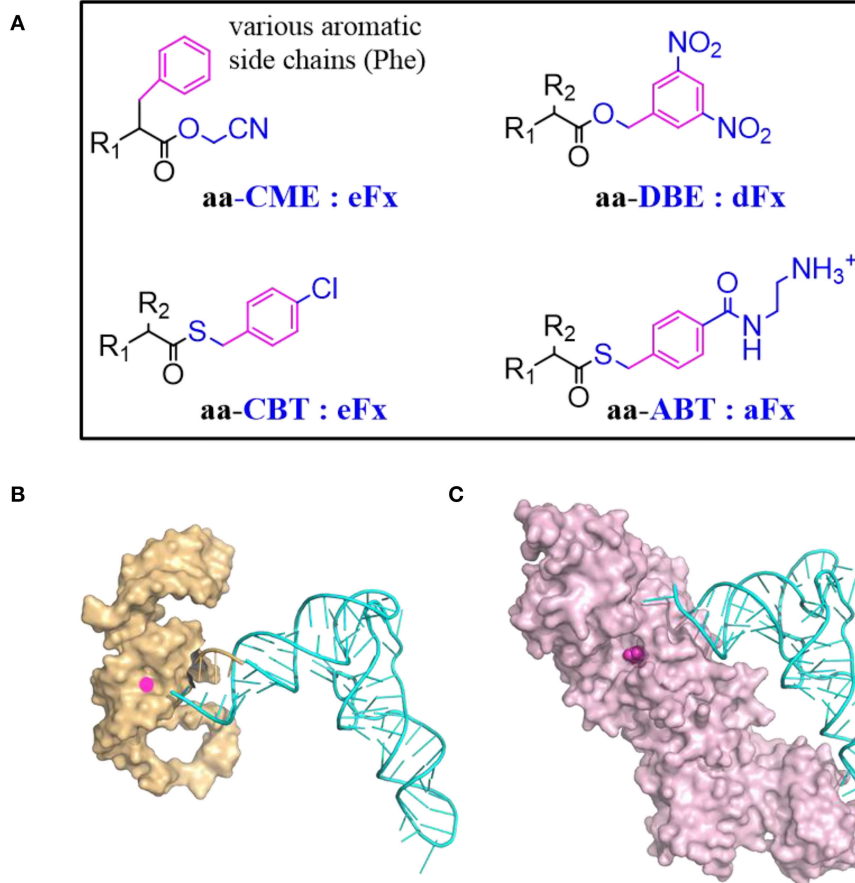


FIGURE 4 | tRNA acylation via Flexizyme and o-aaRS approach. **(A)** The choice of amino acid activation group and the charging Flexizyme (Fx). The Fx can charge a range of substrates including amino acids, hydroxyl acids, N-acyl amino acids etc. $R_1 = \text{NH}_2, \text{OH}, \text{RCONH}$, etc; $R_2 =$ various side chains. The acid substrates are colored in black while the activation or leaving group is colored in blue. The aromatic moiety in the substrate that recognized by Fx is highlighted in magenta. Aromatic amino acids are activated by cyanomethyl ester (CME) and charged on tRNAs via eFx while non-aromatic amino acids are activated by 5-dinitrobenzyl ester (DBE) and charged on tRNAs via dFx. For acyl-donor substrates with α -N-acyl, β -branched, or bulky side chains, the best choice is eFx paired with the 4-chlorobenzyloxy thioester (CBT) leaving group. If the activated acids with above leaving groups are of poor solubility, pairing with aFx and 4-[(2-aminoethyl) carbamoyl] benzyl (ABT) thioester leaving group might be a better choice. **(B)** A representative structure of Fx and tRNA complex. Fx (PDB: 3CUN) recognizes the tRNA through its 3'-terminal region (5'-DCCA-3' where D means A/G/U) therefore it accepts most of tRNAs regardless of their body sequences. The "GGU" end of Flexizyme is shown as cartoon in orange to highlight its complementary interaction with the "CCA" end of tRNA (Cyan). **(C)** A representative structure of tRNA and aaRS complex (Mj)tRNA^{Tyr}/TyrRS, PDB: 1J1U). The amino acid binding pocket is filled with tyrosine (magenta). Most of aaRSs recognize tRNAs through the two distal extremes involving the anticodon sequence and the N₇₃ discriminator base. The substrate binding is malleable but limited to analogs of canonical amino acids.

3.2.1.2. tRNAs engineering for Flexizyme mediated ncAA incorporation

The Flexizyme theoretically can recognize any tRNA with 3'-D₇₃C₇₄C₇₅A₇₆, regardless of their body sequences (**Figure 4B**). An engineered tRNA^{fMetE}, derived from *E. coli* initiator tRNA with a single mutation (C1G), is commonly used for reassignment of the initiation codon (Goto et al., 2011). This tRNA harbors elements that interact with many translational factors (IF2 and IF3) to initiate the translation (Rasmussen et al., 2009). The C1G mutation increases T7 transcription efficiency and thereby the yield of this synthetic tRNA.

Engineering elongator tRNAs is a common practice to enhance ncAA incorporation efficiency in CFPS, and several generations of o-tRNAs have designed. *E. coli* tRNA^{AsnE2}NNN

maintains orthogonality to all the 20 aaRSs and is effective in incorporating a single copy of many functional chemistries but shows little or no yield for consecutive incorporation of structurally challenging ncAAs, such as D-amino acids (Ohta et al., 2007; Goto et al., 2009, 2011). The engineered tRNA^{GluE2} displays strong EF-Tu binding and was used to achieve incorporation of consecutive D-Ser amino acids (Katoh et al., 2017a). The rationally designed tRNA^{ProE2} (**Figure 8E**) has the optimal T-stem motif from tRNA^{GluE2} for enhanced EF-Tu binding and is capable of recruiting EF-P to facilitate synthesis of peptides with challenging substrates. The tRNA^{ProE2} is effective in mediating ribosomal synthesis of polypeptides with not only consecutive D-amino acids (D-Phe, D-Ser, D-Ala, and/or D-Cys) (Katoh et al., 2017a) but also β -amino acids (Katoh and Suga,

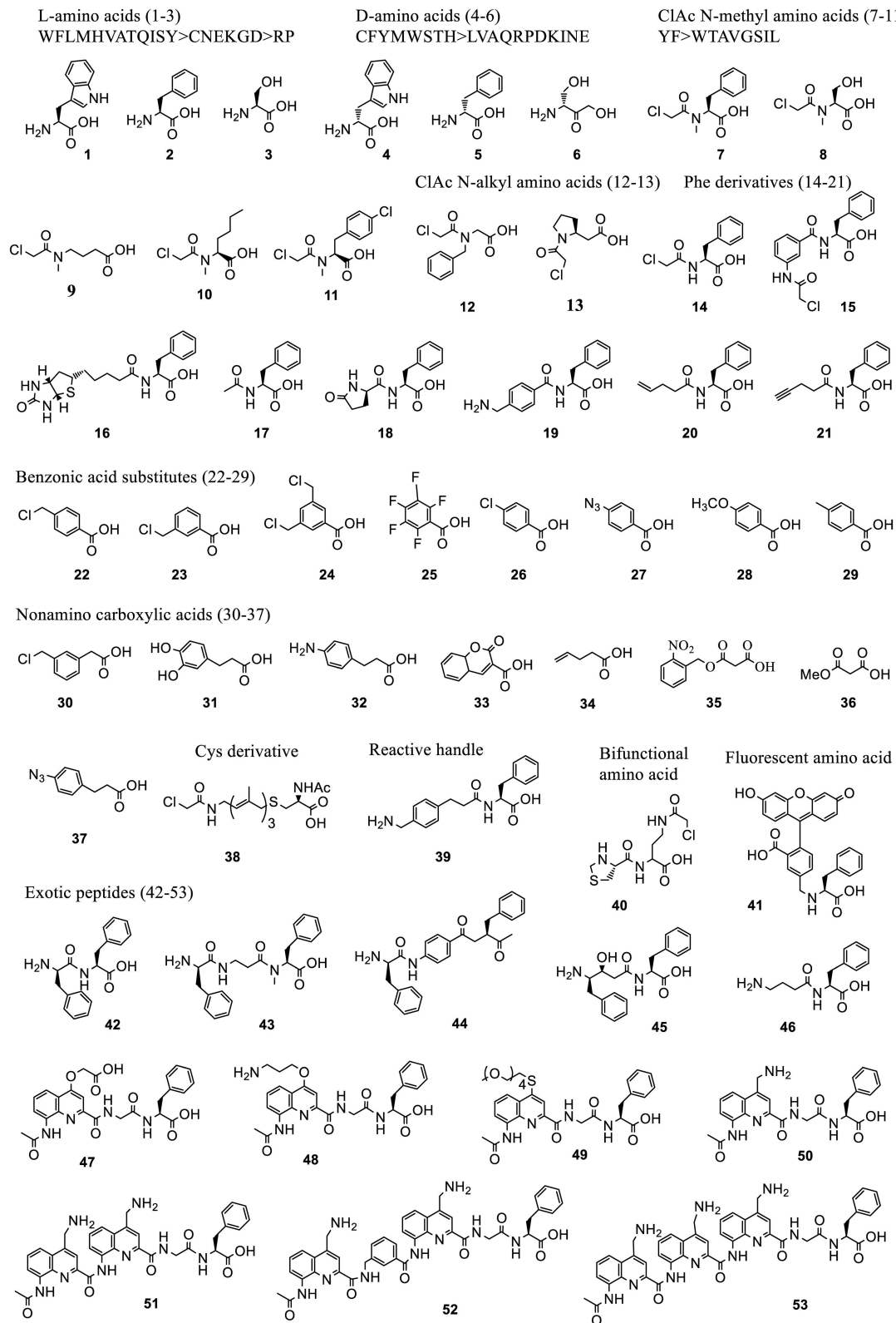
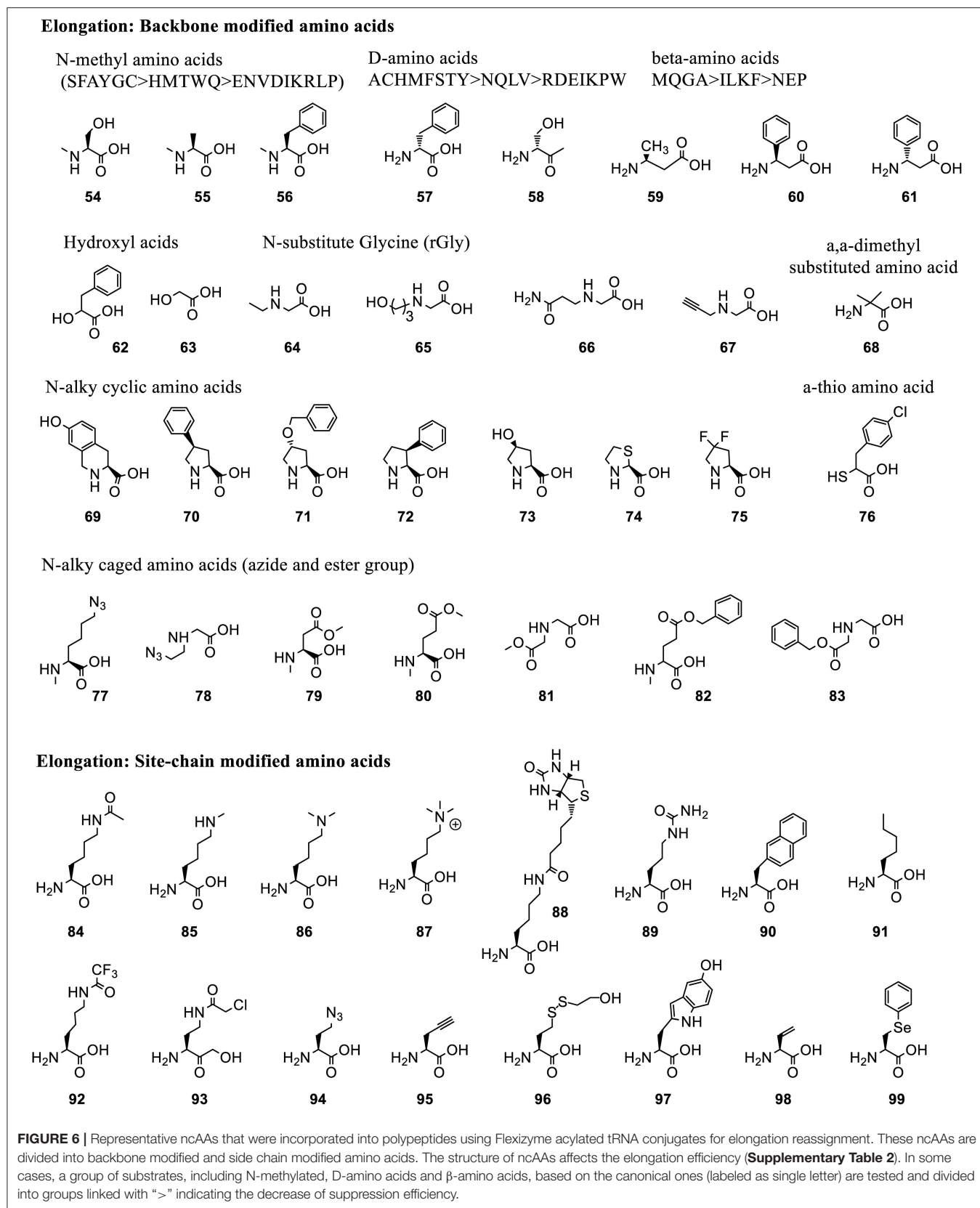


FIGURE 5 | Representative acid substrates that were incorporated into polypeptides using Flexizyme acylated tRNA conjugates for reassignment of the initiation codon. Most of the tested substrates except the L-amino acids (1-3) are backbone modified. The incorporation efficiency of these substrates vary (**Supplementary Table 1**). In some cases, a group of substrates, including 20 L-, 19 D-amino acids, 10 ClAc N-methyl amino acids, based on the canonical aa's (labeled as single letters) were tested and divided into groups linked with ">" indicating the decrease of incorporation efficiency.



2018) and 2-aminoisobutyric acid (**Figure 6**, compound **57-58**, **59-61**, and **68**) (Katoh et al., 2017b). The design of tRNA^{ProE2} is inspired by the recent breakthrough in discovery of the recognition elements in tRNA for recruiting EF-P for enhanced peptide bond formation (Katoh et al., 2016). The determinant of EF-P recognition relies on the D-loop structure (9-nt loop size) closed by the stable 4 base pairs in the D-stem (**Figure 8E**).

Noticeably, changing the codon-anticodon base pairing or the whole progenitor tRNAs might affect ncAA incorporation. For instance, precharged tRNA^{AsnE2}CUA with D-Cys or D-Met is not effective for amber suppression (Goto et al., 2008a) while the same D-amino acid-tRNA but with GGA anticodon is capable for reassigning UCC (Ser) codon (Fujino et al., 2013). Compared to tRNA^{AsnE2}GGA that affords single-incorporation of 12 D-amino acids, an unmodified tRNA^{Gly} supports single incorporation of 17 out of 18 tested D-amino acids into a polypeptides (Achenbach et al., 2015). This might be due to the stronger EF-Tu binding affinity of tRNA^{Gly} (Asahara and Uhlenbeck, 2002). The peptide yield mediated via precharged-tRNA^{GluE2} is several times higher than that of pre-charged ncAA-tRNA^{AsnE2} (Terasaka et al., 2014).

3.2.1.3. Flexizyme enabled ncAAs incorporation

Hundreds of acid substrates, including α -amino acids (D-, N-acyl, N-alkyl), β -amino acids, γ -amino acids, α -hydroxyl acids, α -mercapto acids, and thio acids, have been charged on tRNAs using Flexizyme and successfully installed into polypeptide sequences (**Figures 5**, **6**). Here we review backbone- and side chain-modified ncAAs used in initiation and elongation reassignment.

3.2.1.3.1. Initiation reassignment. Using Flexizyme synthesized ncAA-tRNA^{fMet}CAU in a Met-free PURE system, peptides with a range of ncAAs of exotic structures at N-term were synthesized with varying efficiency. A summary of these ncAAs and their incorporation efficiency are provided in **Supplementary Table 1**. Some of the representative structures are shown in **Figure 5**.

Briefly, these ncAAs include L-amino acids (**1**, **2**, **3**) (Goto et al., 2008b), D-amino acids (**4**, **5**, **6**) (Goto et al., 2008a), chloroacetyl N ^{α} -alkyl amino acids (**7-13**) (Kawakami et al., 2016), Phe derivatives containing chloroacetyl- (**14-15**), biotinyl- (**16**), N-acetyl- (**17**), pyroglutamate (**18**), and other diverse groups (**19-21**) (Goto et al., 2008b; Kawakami et al., 2013a), benzoic acids derivatives (**22-29**) and other carboxylic acids (**30-37**) (Kawakami et al., 2016; Ad et al., 2019; Lee J. et al., 2019) which include aramid precursors and malonyl (1,3-dicarbonyl) substrates to generate a diverse set of aramid-peptide and polyketidepeptide hybrid molecules. Reassignment of the initiator codon has also been achieved to N-acetyl-S-12-(ClAc)farnesyl-D-Cys (**38**) (Torikai and Suga, 2014) mimicking the polyene moiety in the natural biologically active products, phenylalanine derivative with benzylamine group (**39**) for oxidative coupling, a thiazolidine-protected cysteine with 2-chloroacetoamide side chain (Thz-Cab) (**40**) for backbone cyclization (Takatsuji et al., 2019), bulky fluorescent amino acids for peptide detection and quantitation (**41**) (Terasaka et al., 2014), exotic peptides of non-standard structures containing

monomers, such as D-, β -, N-methyl, aminobenzoic acids (**42-44**), γ -amino acids (**45-46**) (Goto and Suga, 2009; Ohshiro et al., 2011), as well as aromatic foldamers comprising of quinoline (Q) and pyridine (P) amino acids (**47-53**) (Rogers et al., 2018; Tsiamantas et al., 2019).

The initiator tRNAs charged with ncAAs containing N ^{α} -acyl groups generally afford higher peptide expression levels (**Supplementary Table 1**). They circumvent the formylation step by methionyl-tRNA formyltransferase (MTF) and bind to IF2 for delivery to the P-site of the ribosome. L-amino acid initiated peptides are generally formylated at their N terminus by MTF (Goto et al., 2008b) while the D-amino acid initiator is not (Goto et al., 2008a) possibly due to the steric clash of these amino acids in MTF catalytic center. The pre-acetylation of D-amino acids could enhance the translational efficiency.

Initiation reassignment to N-chloroacetyl-Phe/Trp/Tyr either in L- or D- stereochemistry integrated with downstream Cys is the most used method for generating a macrocyclic peptide library for mRNA display (Passioura and Suga, 2017; Huang et al., 2019). The N-ClAc group reacts with Cys *in situ* after translation and leads the formation of head-side chain thioether. A recently developed multiple-step strategy demonstrated the first example of backbone cyclization compatible with mRNA display (Takatsuji et al., 2019). This takes advantage of two ncAAs, the backbone modified acid (**40**) and a-thio acid (**76**), with a downstream Cys. The puromycin molecule at the C-termini of the peptide is maintained after backbone cyclization preserving the genotype and phenotype linkage.

3.2.1.3.2. Elongation reassignment to backbone modified residues. Many backbone-modified residues are compatible with ribosomal elongation, such as N-methyl amino acids (**54-56**) (Kawakami et al., 2008), D-amino acids (**57-58**) (Fujino et al., 2013), β -amino acids (**59-60**), D- β -amino acids (**61**) (Fujino et al., 2016), α -hydroxyl acids (**62-63**) (Ohta et al., 2007), N-substituted Gly (rGly, **64-67**) (Kawakami et al., 2008), α , α -dimethyl substituted amino acids (**68**) (Katoh et al., 2017b), cyclic N-alkyl amino acids (**69-75**) (Kawakami et al., 2013a), α -Thio acids (**76**) (Takatsuji et al., 2019), and N-alkyl amino acids with caged side-chains (**77-83**) (Kawakami et al., 2014) (**Figure 6**). The incorporation efficiency is affected by the properties of the backbone as well as of the side chains (**Supplementary Table 2**). Generally, substrates with bulky and charged side chains are unfavorable for elongation.

Polyesters with up to 12 α -hydroxyl acids including phenyllactic acid (**62**), non-aromatic α -hydroxy (**63**) and their derivatives can be synthesized (Ohta et al., 2007). Up to 10 N-methyl amino acids with aromatic or non-charged and non-bulky side chains could be incorporated into the peptide at around 10–20% suppression efficiency (Kawakami et al., 2008). N-substituents of rGly bearing non-branched alkyl chain, or with no bulky group near the amino group and uncharged functional group are well-incorporated in peptoids synthesis (Kawakami et al., 2008). The expression level of peptoids decreases with the increasing number of rGly in the sequence. This is less efficient than N-methyl-peptidyl elongation with the same number of monomers. D- and β -amino acids are also compatible with

the ribosomal synthesis when combined with appropriately engineered tRNAs (Fujino et al., 2013, 2016). N-alkyl amino acids with charged side chains, such as amine and carboxyl group, are poor substrates for ribosomal synthesis. This issue is circumvented by masking the charge of the side chains with precursor amino acids, azide and ester (77-83), followed by their chemical or enzymatic conversion back to the original group (Kawakami et al., 2014).

3.2.1.3.3. Elongation reassignment to side chain modified residues. Many ncAAs with side chain modifications are incorporated into peptide via Flexizyme-mediated approach (Figure 6). Lysine analogs including ϵ -N-acylation, ϵ -N-methylation (mono-, di-, or trimethylation) (84-87) (Kang et al., 2008) were used to study the role of post-translational modifications in protein-protein interaction and cell signaling. A lysine derivative with biotin group (88) was used as an affinity handle. The ncAAs with varying hydrophobicity and properties (89-91) (Passioura et al., 2018) were used to build libraries with expanded diversity for peptide selection (Kawakami et al., 2013b). A warhead ncAA, ϵ -N-trifluoroacetyl-Lysine (92), in peptide sequence was used to achieve selective inhibition of the de-acylation activity of Human Deacetylase SIRT2 (Morimoto et al., 2012).

Many side chain modified ncAAs are employed to form intermolecular bonds, such as N^{γ} -(2-chloroacetyl)- α , γ -diaminobutylic acid (Cab, 93) that reacts with Cys (Sako et al., 2008a), γ -azidohomoalanine (Aha, 94) that reacts with propargylglycine (95) (Sako et al., 2008b), benzylamine in F^{bza} (39) that reacts with 5-hydroxytryptophan (W^{OH} , 97) in the presence of potassium ferricyanide $K_3Fe(CN)_6$ (Yamagishi et al., 2009). Other ncAAs are used to indirectly form intermolecular interactions or mimic the natural bioactive products, such as backbone cyclization of linear peptides bearing a C-terminal Cys-Pro-Glycolic acid (63) sequence with an N-terminal amino group (Kawakami et al., 2009) or with γ -amino acids (Ohshiro et al., 2011). Synthesis of methylsulfonium-containing cyclic peptides takes advantage of vinylglycine (98) (Goto et al., 2009). Synthesis of alkylamides uses Aha (94) or 2-mercaptoethanol-masking homocysteine (96), and phenyllactic acid (62) (Nakajima et al., 2009), while synthesis of thiopeptide scaffolds utilizes Se-phenylselenocysteine (99) (Fleming et al., 2019).

3.2.1.3.4. Comparison of structural acceptance in translation initiation and elongation. The presence of an N-acyl group in ncAA positively affects its initiation but not elongation. The higher structural tolerance of ncAAs as initiators rather than elongators might be due to the relaxed recognition pattern in the ribosomal P-site. The larger substrate binding pocket in P site allows installation of not only bulky amino acids but also exotic peptides with variable chemistries in the monomer, such as N-methyl, D-, D-methyl, γ -amino acids, and foldamers (Figure 5). These substrates can not serve as elongators as they do not fit in the ribosomal A-site where precise positioning of ncAA-tRNA is critical for nucleophilic attack of the α -amino group of elongator aa-tRNA on the carbonyl group of the initiator or peptidyl-tRNAs. However, elongation reassignment can mediate multiple

ncAAs incorporation while the use of the initiator codon results in a single incorporation.

3.2.2. O-tRNA/aaRS Approach

Engineering o-tRNA/aaRS pairs, which do not recognize any endogenous amino acids and tRNAs but specifically charge o-tRNAs with ncAAs, has revolutionized the genetic code expansion field (Wang et al., 2006b). Although the incorporable chemistries are generally limited to amino acid side-chain analogs, this approach requires fewer chemical manipulation steps, is applicable to both *in vivo* and *in vitro* translation systems and enables co-translational charging with multiple turnovers and higher protein yields. While majority of the o-tRNA/aaRS pairs were developed in living organisms using multiple rounds of negative and positive selections, some of them have been expressed in *E. coli* strains that were used to generate cell extracts (Oza et al., 2015) or supplemented as purified components into the CFPS (Cui et al., 2018) for producing ncAA(s)-proteins. Here we provide an overview of available OTSs developed to use in *E. coli*. Although some of them, at present, were only investigated in cells, they are useful elements as ncAA encoding toolbox for more diverse applications of CFPS.

3.2.2.1. O-tRNA/aaRS pairs

Two o-tRNA/aaRS pairs, *Methanocaldococcus jannaschii* (Mj) tRNA^{Tyr}/TyrRS and *Methanosarcina* tRNA^{Pyl}/PylRS, that have minimal or no cross-aminoacylation in *E. coli* have become the most commonly used starting points for directed evolution aimed at creation of specificity to ncAAs (Wang et al., 2006b; Liu and Schultz, 2010; Dumas et al., 2015). Mutants from these two enzymes account for two thirds of the ~200 ncAAs that were successfully incorporated into proteins *in vivo* (Vargas-Rodriguez et al., 2018). MjTyrRS mutants are active toward >50 ncAAs containing either β - or γ -aromatic side chains with functional groups while PylRS mutants are responsible for >100 ncAAs including lysine derivatives containing aliphatic side chains and functional groups, as well as phenylalanine derivatives with aromatic side chains. Noticeably, the recognition profiles of these enzymes overlap on certain phenylalanine analogs. The challenges, methods and future perspectives for directed evolution of o-aaRSs were reviewed elsewhere (Crnkovic et al., 2019) including a powerful evolution strategy using phage-assisted continuous evolution (PACE) (Bryson et al., 2017).

Alternative OTS has been developed for direct encoding of phosphoserine (Sep), which was enlightened by the discovery of an RNA dependent cysteine biosynthesis pathway in archaea (Sauerwald et al., 2005). It takes advantage of a dedicated tRNA^{Sep} (derived from MjtRNA^{Cys}) and an optimized o-phosphoseryl-tRNA synthetase (SepRS) (Park et al., 2011) to generate the Sep-tRNA^{Sep}. The Sep-tRNA^{Sep} can be delivered to amber codon using an engineered EF-Tu albeit with low efficiency. Directed evolution of the SepRS and/or tRNA^{Sep} led to new variants with enhanced Sep incorporation efficiency and was used to produce homogeneously modified proteins in amounts sufficient for biological function analysis (Lee et al., 2013; Rogerson et al., 2015). This orthogonal pair was also evolved for genetic encoding of phosphothreonine (Zhang M. S. et al., 2017).

The engineered tRNA^{UTU}/(SerRS, SelA) pair enables site-specific incorporation of selenocysteine (Sec) via amber suppression. This approach bypasses the need for the Sec-dedicated elongation factor SelB and the conserved Sec-insertion sequence element (SECIS) on mRNA, which are required in natural translation of *opal* codon to Sec. The chimera tRNA^{UTU}, based on tRNA^{Ser} and tRNA^{Sec}, was subjected to a two-step conversion: firstly is aminoacylated by SerRS and then converted to Sec-tRNA^{UTU} by selenocysteine synthase (SelA) (Aldag et al., 2013). The Sec was then delivered to *amber* codon using EF-Tu, instead of SelB. Further efforts generated more effective tRNAs, such as tRNA^{UTUX} (Miller et al., 2015) and tRNA^{UTUT6} (A59C mutation) (Fan et al., 2018), that afford production of proteins with high Sec/Ser ratios (>80%). An alternative effort focused on a newly discovered allo-tRNA family which has unusual acceptor branches acting as efficient serine acceptors. Engineering of the allo-tRNA and SelA from *Aeromonas salmonicida*, as well as the *E. coli* selenium metabolism improved the cellular expression yield and purity (>80%) of recombinant human glutathione peroxidase (Mukai et al., 2018).

3.2.2.2. aaRS and tRNA engineering

The aaRS based aminoacylation approach poses many restrictions on tRNA engineering, as many elements of the latter are involved in the interaction with the former (Figure 4C). Most of the reported o-tRNA engineering focuses on variations of EF-Tu interacting nucleotides. For instance, the directed evolution and/or rational design of MjtRNA^{Tyr} and tRNA^{Pyl} in the T- and acceptor-stem yield optimized tRNAs with better ncAA incorporation efficiency (Guo et al., 2009; Fan et al., 2015).

MjTyrRS is lacking most of the non-conserved domain that binds to the anticodon loop of its cognate tRNA^{Tyr} (Steer and Schimmel, 1999; Kobayashi et al., 2003). As a result it has some promiscuity toward tRNAs anticodon allowing it to be used not only for amber suppression, but also reassignment of *opal*, *ochre*, sense codons (AGA/U) (Wang and Tsao, 2016; Vargas-Rodriguez et al., 2018) as well as unnatural codons (ANaMC) (Fischer et al., 2020). Surprisingly, an engineered initiator tRNA^{fMet} harboring one mutation in the acceptor stem (A72G) and two mutations in the anticodon nucleotides (A35U, U36A) is a substrate of the MjTyrRS for ncAA acylation (Tharp et al., 2019). Despite its low efficiency, this is the first example of o-tRNA/aaRS being used for initiation reassignment. The ncAA incorporation efficiency on AUG start codon is enhanced by deletion of tRNA^{fMet} gene in the *E. coli* genome (Tharp et al., 2019).

The tRNA^{Pyl}/PylRS is orthogonal to both *pro*- and *eukaryotic* systems including living organisms as well as CFPSs (Nozawa et al., 2009). The anticodon of tRNA^{Pyl} is not involved in PylRS recognition making this system suitable for genetic code expansion technique, not only for nonsense codons but also for sense, quadruplet codons (Cui et al., 2018; Vargas-Rodriguez et al., 2018; Oller-Salvia and Chin, 2019) as well as unnatural triplets (Fischer et al., 2020). The commonly used PylRS variants originate from *Methanosarcina barkeri* (Mb) and *Methanosarcina mazei* (Mm), comprising N-terminal tRNA binding domains and C-terminal catalytic domains. These enzymes have higher activity than their bacterial counterparts

from *Desulfitobacterium hafniense* (Dh) whereby DhPylSn and DhPylSc are expressed separately and gain functionality after assembly *in vivo*. Recently, genome data mining identified a new class of PylRS lacking the N-terminal domain, such as the one from *Methanomethylophilus alvus* (Ma), that is both efficient and orthogonal in *E. coli* (Willis and Chin, 2018). The anticodon stem loop of MatRNA^{Pyl} features a nucleotide bulge. By optimizing the variable loop, an efficient mutant, MatRNA^{Pyl(6)}, was obtained that is only accepted by MaPylRS but not MmPylRS resulting in an orthogonal tRNA/aaRS pair (Willis and Chin, 2018).

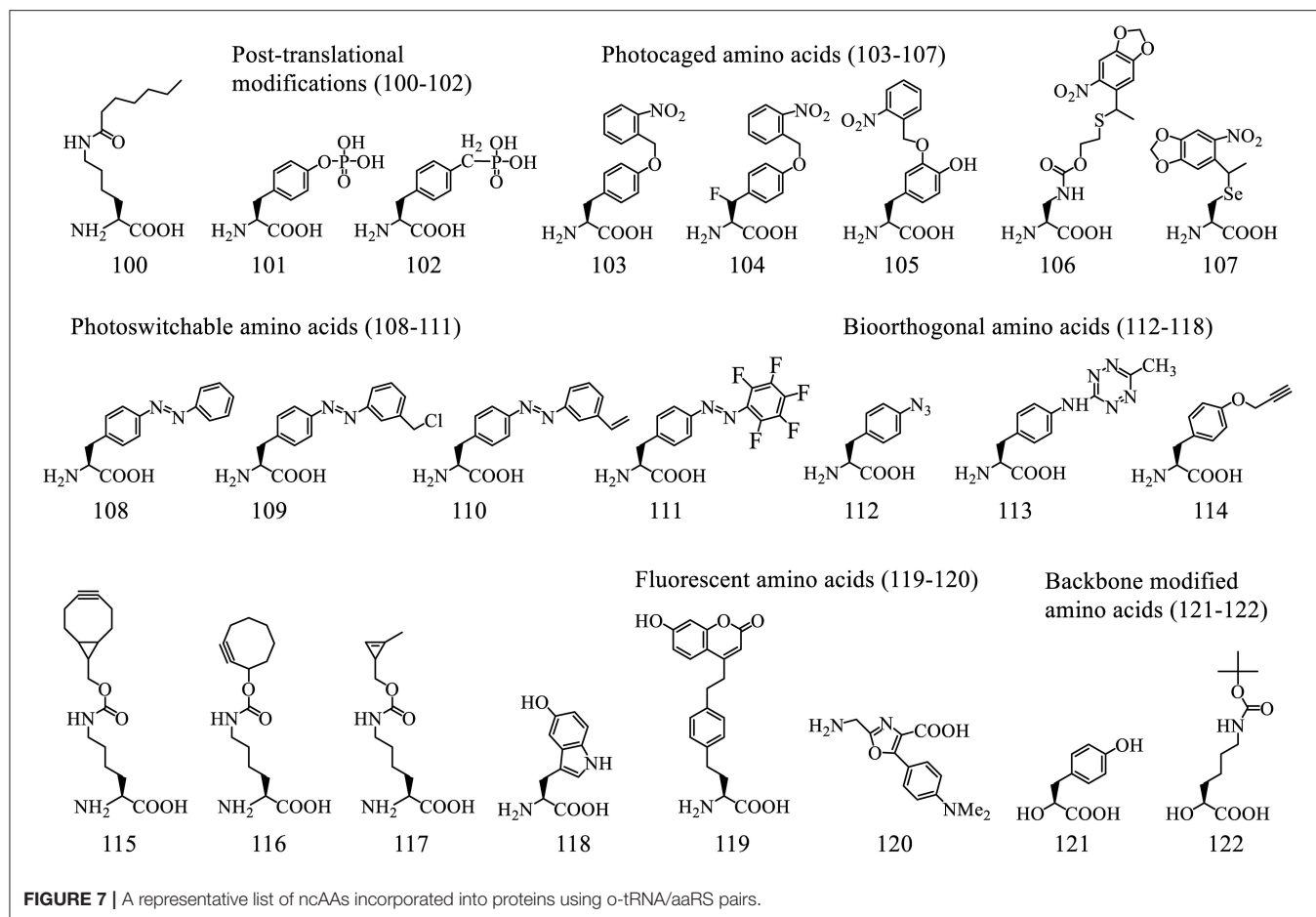
3.2.2.3. Diversity of genetically encoded ncAAs mediated by o-aaRSs

The amino acid binding pockets of aaRSs have been subjected to directed evolution to accommodate a range of ncAAs. These ncAAs contain useful functional groups such as post-translational modifications, photoreactive handles, bioorthogonal reactive groups as well as fluorescent groups. A representative list and some recent additions to these ncAAs are shown in Figure 7. Reports of aaRS-mediated incorporation of ncAAs including the ones used as probes for NMR, IR, and crystallographic analysis, used for enhancing the enzymatic activity, enantioselectivity, ancillary function, creating novel metal-binding sites, or making catalytic residues, are reviewed elsewhere (Dumas et al., 2015; Xiao and Schultz, 2016; Yu et al., 2018; Drienovská and Roelfes, 2020).

3.2.2.3.1. *Post-translational modifications (PTMs)*. Post-translational modifications (PTMs), such as Lysine acetylation (Neumann et al., 2009; Lammers et al., 2010), methylation (Nguyen et al., 2009), ubiquitination (Virdee et al., 2010, 2011), can be incorporated into protein sequences either directly or indirectly to facilitate their characterization in various biological process. PTMs could alter the properties of protein/peptide-based therapeutics. A recent example shows the incorporation of ϵ -N-heptanoyl-l-lysine (HepoK, **100**) that enhances GLP1 peptide binding to human albumin and confers a more potent and long-lasting ability to decrease blood glucose levels (Fu et al., 2019).

A common PTM phosphotyrosine (pTyr, **101**) is a difficult target for direct genetic decoding due to the negative charge of its side chain that causes poor interaction with EF-Tu. Its incorporation was achieved using engineered EF-Tu (Figure 8C) in a phosphatase knockout strain (Fan et al., 2016). Although the yield of pTyr-containing sfGFP is only about 5% that of wild type (wt) sfGFP, 20 mg/L is sufficient for biological studies of tyrosine phosphorylation. Alternative methods circumvent this issue by genetically encoding a non-hydrolyzable analog of pTyr (**102**) (Luo et al., 2017). Although feasible, it sometimes could not fully represent the pTyr due to the minor structural difference.

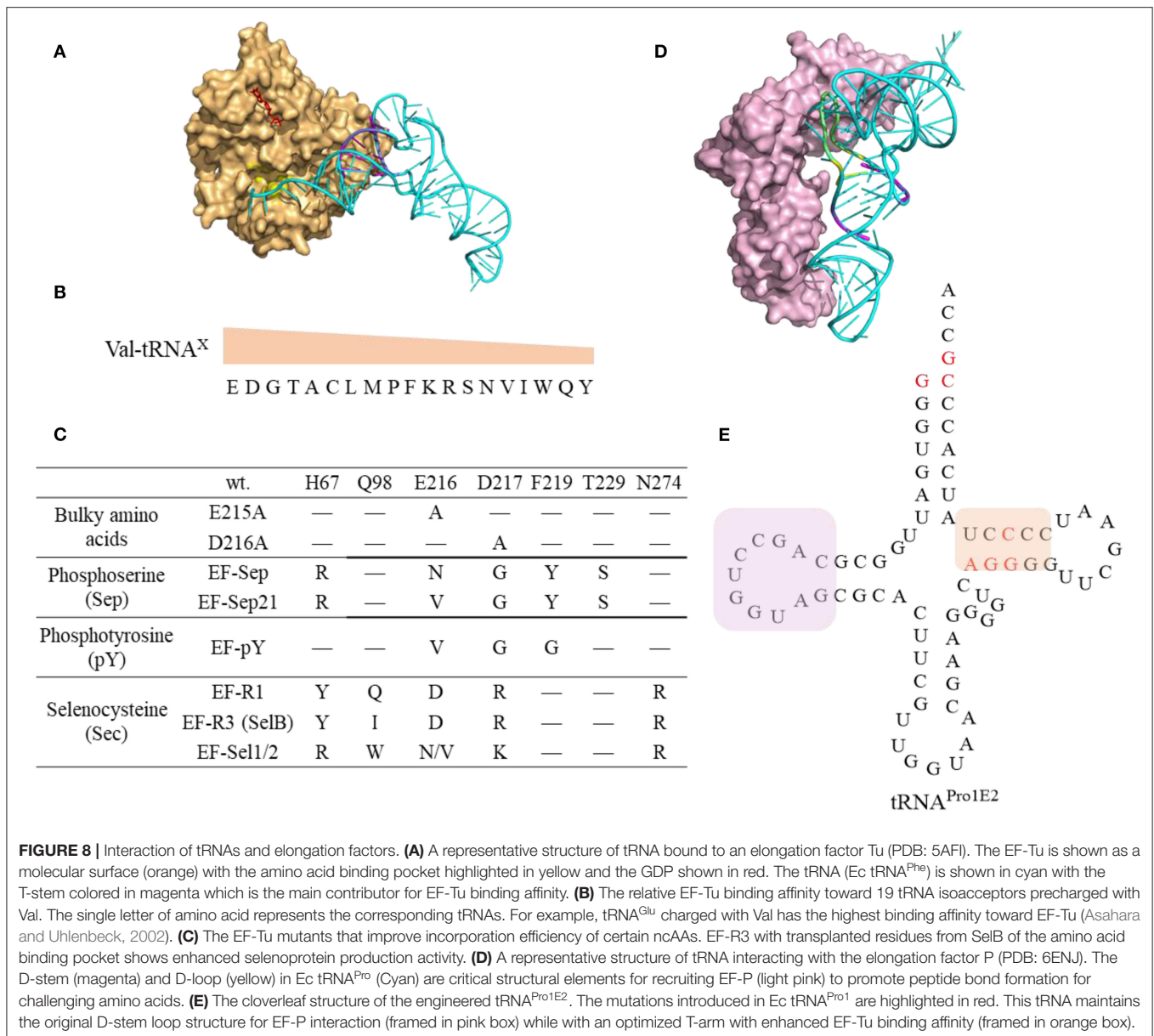
3.2.2.3.2. *Photo-responsive ncAAs*. Photocaged ncAAs can be used to regulate protein functions by exposure to light, which makes them useful cell-biological tools. For instance *o*-nitrobenzyl-Tyr (**103**) was used to construct light-activatable nanobodies (Jedlitzke et al., 2020) and antibody fragment (Bridge et al., 2019), *o*-nitrobenzyl-4-hydroxyphenylalanine



(**105**) was used for generating a photo-controlled variant of wet adhesive protein (Hauf et al., 2017), a caged diaminopropionic acid (**106**) was used for catalytic mechanism characterization of valinomycin synthetase (Huguenin-Dezot et al., 2019), a photocaged Sec, 4,5-dimethoxy-2-nitrobenzyl-Sec (**107**), was used for production of selenoproteins as well as for site specific alkylation (Welegedara et al., 2018).

Photocaged crosslinkers such as *o*-2-nitrobenzyl- β -fluorotyrosine (FnbY, **104**) were used to cross-link proteins to identify their interacting partners *in vivo* (Liu et al., 2019). Photoswitchable biomolecules, such as phenylalanine-4'-azobenzene (**108**), harboring photoisomerizable functionalities that could induce reversible changes in protein conformation and functionality, offer minimal invasive approach for precise spatiotemporal control. Photoswitchable click amino acids (PSCaa) have an additional click functional group such as benzyl chloride (**109**, trans), alkene (**110**, trans), keto, and pentafluoro PSCaa (**111**, trans), were used to generate a covalent protein bridge by reacting with a nearby cysteine residue through proximity-enabled bioreactivity (Hoppmann et al., 2014, 2015; Klippenstein et al., 2017). The resultant azobenzene bridge isomerized in response to light, thereby changing the conformation of the protein and its activity.

3.2.2.3.3. Bioorthogonal handles. Bioorthogonal handles are useful for installation of desired chemical groups, such as small affinity tags, fluorophores, or linkers for protein conjugation (Lee K. J. et al., 2019). The ncAAs with groups such as ketones, azides (**112**), tetrazine (**113**), terminal alkynes (**114**), strained alkynes and alkenes (**115-117**), remain inert in the endogenous biological environment and react with complementary probes. The most popular reaction schemes are copper-catalyzed azide alkyne cycloaddition (CuAAC), photoclick cycloaddition, strain-promoted azide alkyne cycloaddition (SPAAC), and inverse electron-demand Diels-Alder cycloadditions (IEDDA) (Lee K. J. et al., 2019). The IEDDA reactions between electron-poor tetrazines and strained alkyne/alkene moieties, such as bicyclo[6.1.0]nonyne (**115**) (Lang et al., 2012a), trans-cyclooctene (**116**) (Nikic et al., 2014), cyclopropene (**117**) (Elliott et al., 2014), and norbornene (Lang et al., 2012b), are rapid under physiological conditions. A recently developed chemoselective rapid azo-coupling reaction (CRACR) between an aromatic diazonium ion and 5-hydroxyl-Trp (**118**) (Addy et al., 2017) is orthogonal to both SPAAC and IEDDA chemistries. These chemistries were used for installing multiple ncAAs into polypeptides site-specific fashion (Italia et al., 2019).



3.2.2.3.4. Fluorescent amino acids. Direct incorporation of fluorescent ncAAs minimizes the functional interference with target proteins, compared to other labelling methods. However, these ncAAs, such as coumarin-derived amino acid (**119**), suffer from low quantum yield (Wang et al., 2006a; Lee K. J. et al., 2019). A recent report demonstrated genetic encoding in cells with a bright fluorescent oxazole amino acid (**120**) that does not contain either an α -amine moiety or any asymmetric center. This was achieved by using a modified ribosome albeit with low efficiency (Chen et al., 2019).

3.2.2.3.5. Backbone modified ncAAs. Unlike the Flexizyme-mediated approach, only a very limited number of backbone-modified amino acids have been installed via o-tRNA/aaRS pairs. These include MjTyrRS for α -hydroxy-L-phenyllactic acid (**121**)

(Guo et al., 2008), PylRS for α -hydroxyl acid analog of Boc-lysine (Boc-LysOH, **122**) (Kobayashi et al., 2009) as well as a fluorescent oxazole amino acid (**120**) (Chen et al., 2019). All these ncAAs were incorporated into proteins with very low efficiency (<10%), which might be caused by poor aminoacylation, inefficient EF-Tu interactions, and slow ester/amide bond formation in the ribosomal decoding center.

3.2.2.4. Simultaneous incorporation of multiple ncAAs

Creating mutual orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs is necessary for faithful installation of multiple ncAAs. The tRNA^{Pyl}/PylRS pairs are orthogonal to the MjtRNA^{Tyr}/TyrRS and have been used for orthogonal dual incorporation of structurally distinct substrates. For instance, a cyclopropene derivative of Lys (CypK, **117**) and

p-propargyloxy-phenylalanine (PrpF, **114**) are genetically encoded by MbPylRS-PylT_{UACU} and MjPrpFRS-tRNA^{Tyr}CUA, respectively, and were used in phage display for selecting polypeptides with expanded chemistries (Oller-Salvia and Chin, 2019). The Sep OTS is orthogonal to Pyl OTS and their combination was used for simultaneous incorporation of phosphoserine (tRNA^{Sep}/SepRS/EF^{Sep}) and acetylsine (tRNA^{Pyl}/AcKRS) (Venkat et al., 2018).

Evolution of substrate selectivity of two polyspecific MjTyrRSs, pCNFRS that recognizes >18 ncAAs and AcKRS, generated variants that are highly selective for *p*-azidophenylalanine (**112**) and *m*-iodo-L-phenylalanine, respectively (Kwok et al., 2019). However, using two MjTyrRS derived mutants for simultaneous incorporation of two ncAAs is still problematic due to the lack of orthogonality of the tRNAs. In contrast, tRNA^{Pyl}/PylRS pairs of different origins were evolved to be orthogonal to each other and were used for dual site-specific ncAAs incorporation via amber and frameshift suppression using MaPylRS(CbzKRS)/tRNA^{Pyl(6)} and MmPylRS(CypKRS)/tRNA^{Pyl} combination (Willis and Chin, 2018). The active sites of these two PylRS enzymes are specific for two ncAAs, N^ε-benzyloxycarbonyl-lysine (CbzK) and CypK (**117**) (Willis and Chin, 2018).

Simultaneous incorporation of three ncAAs (**112**, **117**, **118**) was achieved by combining MjtRNA^{Tyr}/TyrRS, MbtRNA^{Pyl}/PylRS and EctRNA^{Trp}/TrpRS pairs in an *E. coli* ATMW1 strain whereby the endogenous EctRNA^{Trp}/TrpRS pair was functionally replaced by its yeast counterpart (Italia et al., 2019). The resulting sfGFP protein contains the most types of genetically encoded ncAAs so far. Although impressive, this strategy relies on reassignment of all three stop codons and one of them is read either as an ncAA or as termination signal. This complicates the procedure and requires aforementioned special design for protein expression.

3.3. Engineering Other Elements of Translational Machinery for Enhanced ncAA Incorporation

Not all the proteinogenic amino acids are incorporated into proteins using identical mechanisms. Nature has evolved dedicated elongation factors EF-P and SelB for efficient delivery of consecutive Pro (Doerfel et al., 2013) and selenocysteine (Aldag et al., 2013), respectively. Other modified amino acids, such as phosphoserine, phosphotyrosine as well as many other modifications, are post-translationally introduced into protein sequences. Modification of translational machinery is expected to provide solutions to limitations in incorporating structurally complex ncAAs.

3.3.1. Elongation Factors

The amino acids and tRNA sequences are fine-tuned to ensure appropriate EF-Tu binding affinity and kinetics that sufficient to form ternary complex and yet being able to release EF-Tu for codon decoding (Schrader et al., 2011) (**Figure 8A**). The thermodynamic contributions of the esterified amino acid and the tRNA to the overall binding affinity are independent and compensate each other (LaRiviere et al., 2001). The binding

affinity of tRNA^{Glu} and tRNA^{Asp} to the EF-Tu are among the highest indicating the relatively little contribution from the negatively charged Glu and Asp (Asahara and Uhlenbeck, 2002) (**Figure 8B**). It is possible to fine tune not only the sequence of the orthogonal tRNA, but also the EF-Tu to achieve the best incorporation efficiency of challenging ncAAs.

EF-Tu has been engineered to display enhanced binding affinity toward challenging ncAAs (**Figure 8C**). For instance, incorporation of bulky ncAAs, 1-pyrenylalanine, DL-2-anthraquinonylalanine, L-2-pyrenylalanine, which are poorly incorporated into streptavidin via the wt EF-Tu, was achieved by introducing mutations in EF-Tu (E215A or D216A) to expand its amino acid binding pocket leading to incorporation efficiency at 8–25% in PURE system (Doi et al., 2007). An engineered EF-Tu (EF-Sep) containing five mutations in the amino acid binding pocket is capable of mediating incorporation of negatively charged phosphoserine into *amber* codon, albeit with low efficiency (1–25 μg of MEK1 proteins with single or double phosphoserine were produced per liter of *E. coli* culture) (Park et al., 2011). Further optimized SepRS and EF-Sep (EF-Sep21) afforded production of significant amounts (3 mg per L culture) of recombinant full length phosphohistone H3 (Lee et al., 2013). However, subsequent work on systematic evolution of the anticodon stem loop of tRNA^{Sep}, and the anticodon recognition region of SepRS, identified new pairs of tRNA^{Sep}/SepRS that have enhanced Sep incorporation efficiency without optimization of EF-Sep interaction (Rogerson et al., 2015). Although many engineered OTSs for Sep incorporation are generated and investigated in the *in vivo* system, these elements could be transferred to the cell-free system. For instance, the cell free extract derived from *E. coli* strains expressing the 1st generation of EF-Sep, SepRS, and tRNA^{Sep} enables production of up to a milligram of phosphorylated human MEK1 kinase (Oza et al., 2015). The produced protein contained some non-phosphorylated species. Using more efficient OTSs might be beneficial for increasing protein purity.

EF-Tu could be engineered to support increased selenocysteine (Sec) protein production either by its directed evolution (EF-Sel) or by rational transplanting of the residues from elongation factor SelB to EF-Tu, in order to generate variants with positively charged amino acid binding pockets (EF-R1, EF-R3) (Haruna et al., 2014). The EF-Sel mutants selected through directed evolution is beneficial for the modified protein yields but not as good as EF-R1/R3 variants, which may represent a compromise between selenoprotein toxicity and yield. Cellular fitness puts restrictions on the range of engineering of the orthogonal translation systems. The efficiency of these *in vivo* engineered EF-Tus is yet to be characterized for *in vitro* protein synthesis.

EF-P could stimulate the formation of the first peptide bond and accelerate synthesis of proteins containing consecutive prolines by preventing ribosomal stalling and promoting peptide bond formation (Blaha et al., 2009; Doerfel et al., 2013). The discovery of recognition elements in tRNA^{Pro} for EF-P recruitment (Katoh et al., 2016) inspired the design of tRNA^{ProE2} (**Figure 8E**) as aforementioned. EF-P recognizes both the tRNA^{Pro} (**Figure 8D**) as well as the peptidyl-Pro residue

in the P site to accelerate Pro-aa bond formation, where aa is a poor A-site substrate such as Pro, Gly, or other secondary amino acids. Experimental data shows EF-P is not only effective in alleviating the ribosome stalling on the consecutive Pro, but also on D-amino acids (Katoh et al., 2017a), β -amino acids (Katoh and Suga, 2018) and other structurally challenging ncAAs (Katoh et al., 2017b), at optimal concentration of 5–10 μ M. Higher concentrations are detrimental possibly due to the longer residence of EF-P in the vicinity of the ribosomal E site thus inhibiting the translocation of deacylated tRNAs from ribosomal P site to E site.

3.3.2. Ribosome Engineering

Ribosome engineering continues pushing the limits of this biocatalyst. Recent efforts have focused in several areas: enhancing ribosome ability to polymerize non-native monomers, generating orthogonal ribosomes that function in parallel to the wt systems and hence could be subjected to directed evolution, and creating double genetic codes of the same mRNA templates to increase the information encoding density (Figure 9).

Point mutations in the peptidyl transferase center (PTC) of 23S rRNA enhanced D-amino acid incorporation efficiency (Dedkova et al., 2003, 2006). Compared to wt ribosome, the S30 extract prepared from the cells containing both the wt and a mutant (A4, Figure 9B) has increased yield of proteins harboring D-Phe or D-Met from 3–5% to 12–23%. Noticeably, alteration of a single nucleotide in the key position of the ribosomal PTC can be lethal to the cells, while multiple mutations can rescue activity by maintaining general PTC architecture (Thompson et al., 2001). Due to the opposite chirality of the C $^{\alpha}$ -atoms, the α -amino group of the D-amino acid is positioned further away from the peptidyl-tRNA carbonyl group carbon in the wt ribosome P-site, thus resulting in non-optimal nucleophilic attack and blocking proton shuttling during the peptide bond formation (Melnikov et al., 2019). The A4 mutant has four-nucleotide changes at position 2447–2450 and is expected to lead to a larger A-site cleft that is beneficial for D-amino acid incorporation.

The reorganization of ribosomal PTC architecture through mutagenesis of two regions in 23S rRNA, instead of simple mutations in one region, is effective in decoding other structurally challenging ncAAs (Figure 9B). Two of the modified ribosomes (040329 and 0403x4) were able to suppress UAG codon in *E. coli* producing β -Ala modified proteins (Dedkova et al., 2012) with suppression efficiency of \sim 10–15% compared to 4% of wt ribosome. The same research group succeeded in incorporation of various β -amino acids, but all with <20% efficiency (Maini et al., 2013). Directed evolution also identified ribosomes that are capable of accepting dipeptides, dipeptide analogs, oxazole fluorescent amino acid (010326R6/010328R4) (Maini et al., 2015; Chen et al., 2019) as well as phosphorylated tyrosine (030449 and 040412 mutants) (Chen et al., 2017), as substrates.

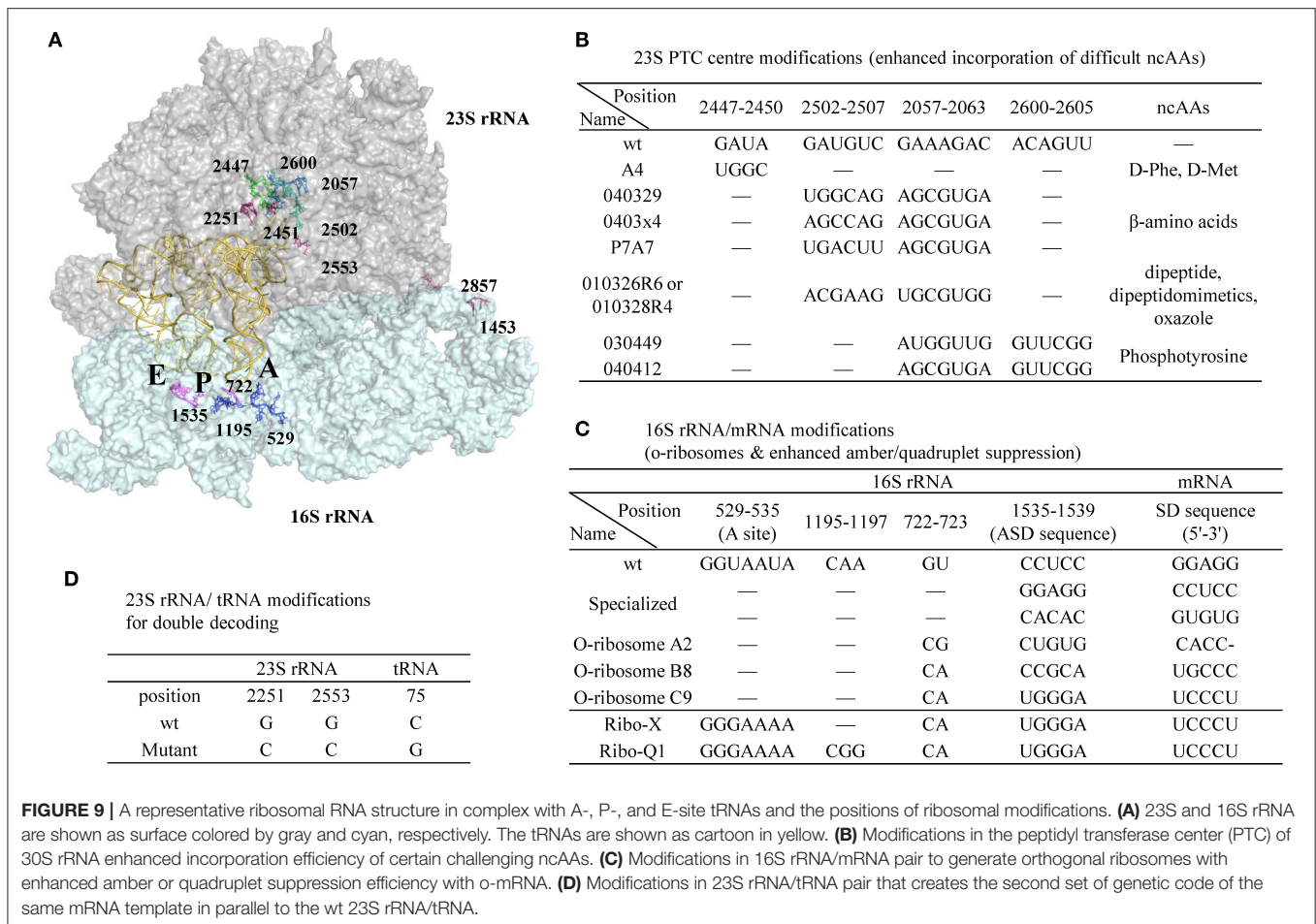
Another direction in ribosome engineering is the generation of orthogonal ribosomes that operate as independent translation systems translating target proteins while the wt ribosomes continue to synthesize the genome-encoded proteins to ensure cell viability (Figure 9C). This is achieved by co-modifying the

anti-Shine Dalgarno in the 16S rRNA as well as the Shine Dalgarno sequence in the engineered mRNAs to enable gene-specific translation (Hui and de Boer, 1987; Rackham and Chin, 2005a). Combination of the three o-ribosome (A2, B8, C9)/o-mRNA pairs to control downstream translation of variable protein fragments enables creation of combinatorial logic gates in the living cells, and opens the door for generating cellular computers controlled by the biomolecules (Rackham and Chin, 2005b).

Further optimization of the orthogonal ribosome A-site located at the 530 loop of the 16S rRNA was performed to reduce its interaction with RF1 thereby enhancing suppression efficiency (Wang et al., 2007). The evolved o-ribosome termed Ribo-X, with two-nucleotide mutation in the loop region, demonstrated 60% and 20% suppression efficiency toward an o-mRNA that harbored one and two amber codons, respectively, to *p*-benzoyl-L-phenylalanine. Much larger libraries covering 127 residues within 12Å of the tRNA bound in 16S rRNA decoding center were generated, and the evolved orthogonal ribosome Ribo-Q1 could efficiently decode both quadruplet (AGGA) and amber codon (Neumann et al., 2010). This modified ribosome has been recently used to incorporate CypK (117) and PrpF (114) at two sites of ScFv followed by subsequent conjugation with two distinct fluorophores (Oller-Salvia and Chin, 2019).

In order to restrict the exchange of the large subunits between pools of the native and orthogonal small subunits, the later versions of orthogonal ribosomes (Ribo-T) were generated where the small and large subunits were tethered together via helix 44 of the 16S rRNA and helix 101 of the 23S rRNA using either polyadenine linkers or RNA hinges from self-splicing introns [16S(1-1453)-linker-23S(2858-2902-1-2857)-linker-16S(1454-1542)] (Fried et al., 2015; Orelle et al., 2015; Schmied et al., 2018) (Figure 9A, reconnecting position at 1453 in 16S rRNA and 2857 in 23S rRNA). These new ribosomes enable selection of mutants 30S subunit capable of re-programming cellular logic (Rackham and Chin, 2005b) and enabling new decoding properties (Wang et al., 2007; Neumann et al., 2010). However, the altered arrangement of rRNA sequences with a permuted 23S rRNA inserted into 16S rRNA reduced their ability to rapidly assemble into a macromolecular machine. The 2nd generation of Ribo-T was evolved by selecting new RNA tethers from a more diversified library that connect the 16S and 23S rRNA (Carlson et al., 2019). Ribosome profiling of Ribo-T cells shows no defects in translation elongation but reveals somewhat higher occupancy by Ribo-T of the start codons and to a lesser extent stop codons, suggesting that subunit tethering mildly affects the initiation and termination stages of translation.

Two concomitantly functioning orthogonal translation machineries can be achieved not only by manipulating the aforementioned mRNA/16S rRNA pair, but also by engineering the tRNA/30S rRNA recognition patterns (Terasaka et al., 2014). Two highly conserved residues G2251G2252 at P-site of the 23S rRNA interact with C75C74 (tRNA) via Watson-Crick base pairing while another nucleotide G2553 in A-site base pairs C75 of the incoming tRNA. Engineering these conserved nucleotides in the peptidyl transferase center in 23S rRNA



as well as the ones in tRNA 3'-CCA-end demonstrated that the compensatory restoration of these base pairs recovers the translational activity. Two mutant ribosome-tRNA pairs were found to be orthogonal to the wt pair. One of these pairs (23S rRNA-G2251CG2553C/tRNA-C75G) demonstrated good activity and functioned together with the wt pair in the translation reaction (**Figure 9D**). They acted orthogonally and translated the same mRNA sequence into two distinct peptide sequences (Terasaka et al., 2014).

4. GENOME WIDE CODON ENGINEERING

A genomically recoded *E. coli* strain (C321.ΔA), where all the 321 UAG amber stop codons were replaced by synonymous UAA codons, permitted the deletion of RF1 and more efficient reassignment of UAG codons (Lajoie et al., 2013). This strain has served as a synthesis platform for modified proteins such as phosphoproteins (Pirman et al., 2015), as well as a platform for enzyme evolution (Amiram et al., 2015). In the latter case, chromosomally integrated aaRS variants based on this strain have been evolved to enable multi-site ncAAs incorporation. Despite the lack of competition from the RF1 and superior ability for amber suppression, this strain grows slower and sometimes

produces less modified proteins than BL21(DE3) (George et al., 2016). An unexpected codon skipping phenomenon was observed in this strain when producing Sep-ubiquitin proteins that missing one amino acid residue, the ncAA, with the +3 frameshifting (George et al., 2016). The mechanism behind this phenomenon is unexplored.

The C321.ΔA strain can serve as a chassis strain for crude CFPS. The cell extract from this strain integrated with an improved Sep-OTS components enable production of phosphoserine-protein at mg per ml yield (Oza et al., 2015). However, the standard cell extract from this strain supports the production of wt proteins with about 3 times lower efficiency than that of the standard BL21(DE3) strain (Martin et al., 2018). Inactivation of negative effectors in the host strain using multiplexed automated genome engineering (MAGE) generated a new strain, C321.ΔA.759 (*endA-*, *gor-*, *rne-*, *mazF-*). The derived cell extract from this strain has improved productivity for both wt and modified proteins (Martin et al., 2018). With an increasing number of UAG codon present in the sfGFP template, an exponential decrease of protein yield was observed for the cell extract derived from BL21Star (DE3), leading to no detectable active protein for sfGFP-5UAG. The yield from release factor 1 (RF1) deficient MCJ.559 strain in which only a small set of essential genes were recoded (Hong et al., 2014b),

is roughly 2 times less compared to the one from C321.ΔA.759. Incorporation of 8 and 9 consecutive ncAAs is also achieved with the full-length protein ratio at 75 and 60%, respectively. Keeping a proper surface area to volume ratio, the batch translation reaction could be scaled up to 17-fold from 15 to 255 μ l without loss of productivity (Martin et al., 2018).

Genomic engineering of *E. coli* strains free of sense codons was also attempted. The first ambitious project to build a 57-codon *E. coli* was reported in 2016, which aimed to replace 7 codons (R-AGG, R-AGA, S-AGC, S-AGT, S-TCG, S-TCA) with their synonymous codons (Ostrov et al., 2016). However, it turned out that replacement of 13 out of 123 AGR codons to the synonymous codon CUG was detrimental to the cells, requiring diversification to viable codon alternatives (Napolitano et al., 2016). It is not surprising that drastic change in the genome structure might cause cell fitness issues.

Impressively, recent effort in the Chin lab demonstrated successful deletion of 3 of the 64 codons (amber-TAG, S-TCG, S-TCA) from *E. coli* genome and their replacement with synonymous codons, creating the 61-codon bacteria (Syn61) (Fredens et al., 2019). Due to the absence of sense codons TCG and TCA, deletion of the genes of their cognate tRNAs (*serU* and *SerT*) from the genome is non-lethal. Co-translational incorporation of ncAAs through TCG codon reassignment was achieved using tRNA^{Pyl} CGA/PylRS pair in Syn61, but not in the non-recoded MDS42 strain. This is expected that this pair of OTS leads to wrongly synthesized proteome which is toxic to the standard cells. The Syn61 strain enables site-specific sense codon reassignment without scrambling the proteome and represents a new synthetic biology platform for synthesis of polymers with enhanced and novel activities.

5. SUMMARY AND PERSPECTIVES

We reviewed technical elements for ncAA incorporation ranging from creation of vacant codons, synthesis of ncAA-tRNAs, to translational machinery engineering. These strategies and elements are generally applicable in both types of *E. coli* CFPS, i.e., crude and PURE CFPS. However, due to their distinct compositions, cost and ease to manipulate the translational components, these two *in vitro* translation platforms have advantages and disadvantages for different applications (Table 1).

PURE is the preferred system for peptide synthesis. Impressive work on the genetic reprogramming of the sense codons using this system integrated with Flexizyme tRNA acylation system reveals the extraordinary flexibility of ribosomal synthesis (Figures 5, 6). This affords production of peptides with high diversity and forms an extraordinary screening platform when integrated with mRNA display. However, due to its low productivity and high cost, characterization of the selected non-standard peptide hits relies on chemical synthesis adding an extra layer of complexity. With advances in expression productivity, S30 CFPS might be a promising platform that could combine selection and small-scale production of peptides with ncAAs. This however requires development of approaches for increasing the stability of mRNA and the product peptides.

Substrate promiscuity of the ribosomes is much higher than that of aaRSs. Therefore, the ability to synthesize the ncAA-tRNA largely determines the diversity of novel chemistries that can be incorporated into the polypeptides sequences. Flexizyme represents one solution to this problem and is a very versatile *in vitro* acylation tool that dominates the *in vitro* acylation landscape. The OTSs (o-tRNA/aaRS pairs) are capable of co-translational aminoacylation and have superior capacity to support protein synthesis both *in vitro* and *in vivo* achieving relatively high yields. The Flexizyme can recognize practically any tRNAs with a 3'-DCCA- end, while the aaRS requires many nucleotide identity elements for efficient aminoacylation thereby complicating tRNA engineering. The recently developed T-boxzyme, comprising a tRNA recognition and a catalytic domain, showcased its capacity of co-translational acylation of tRNA^{Gly}GCC achieving one-pot *in vitro* synthesis of unnatural peptides (Ishida et al., 2020). Although in its infancy, this T-boxzyme holds promise to transform the tRNA acylation technology and to dramatically expand the structural diversity and increase the productivity of ncAAs-proteins, not only *in vitro* but also in cells.

Reassignment of initiation codon supported by Flexizyme-mediated acylation affords incorporation of more exotic substrate structures such as peptides and foldamers. This is due to the intrinsically large flexibility of the ribosomal initiation P-site. These substrates are too bulky to fit in the ribosomal A-site for elongation. This is consistent with the fact that the dipeptide incorporation via elongation reassignment requires modified ribosomes (Maini et al., 2015). In contrast, initiation reassignment via OTS is far less sophisticated that only one recent example demonstrated its feasibility using an o-aaRS and an engineered initiator tRNA^{fMet} (Tharp et al., 2019). This work demonstrates unexpected versatility of aaRSs. As the initiation accepts a more versatile structures than elongation, it is exciting to see how far the OTSs can reach toward the diversity of the ncAAs.

Poor ribosomal synthesis of proteins with ncAAs might be due to the inefficient binding of ncAA-tRNAs to the EF-Tu and/or compatibility issue in the ribosomal decoding center. Engineering of the translational components is an effective approach to increase the incorporation efficiency of structurally challenging ncAAs. Transfer RNA engineering is beneficial to D-, β -, and many other ncAAs by enhancing EF-Tu binding affinity and/or recruiting of EF-P to facilitate the peptide bond formation. Modified EF-Tu variants have increased capacity for bulky amino acids, negatively charged amino acids including Sep, phosphotyrosine, selenocysteine, but not for D-amino acids. Ribosomes with modified PTCs show beneficial for D-, β -amino acids, dipeptide, fluorescent amino acids, as well as phosphotyrosine. Ribosomes with altered A-site could enhance the *amber* and quadruplet suppression by reducing their interaction to RF1.

Noticeably, the effect of tRNA engineering is specific toward the ncAAs while mutations introduced in EF-Tu and ribosomes cause global effect not only for ncAAs but also for the other standard amino acids in the polypeptide sequences. For instance, the EF-Tu mutants with enhanced D-amino acid

incorporation efficiency actually only retain 20–80% activity of wt EF-Tu for canonical amino acids (Doi et al., 2007). Therefore, the benefit brought by the mutated EF-Tu and ribosomes should be balanced when generating proteins containing both canonical amino acids and ncAAs. On the other hand, it seems promising to engineer EF-Tu or ribosome specialized for production of a single polymer type, such as polyesters, peptoids etc. However, current ribosome engineering is generally confined in the *in vivo* system that is governed by cell fitness, ncAA permeability and stability, that set significant limitations. From this perspective *in vitro* translation systems appear as much better platforms for selection of ribosomes capable of synthesizing non-peptide polymers (Hammerling et al., 2019).

Nature has inspired many elegant strategies for enhancing ncAA incorporation. For instance, the direct genetic encoding of phosphotyrosine and phosphothreonine was advanced by the discovery of the RNA dependent cysteine biosynthesis pathway. Consecutive incorporation of D- and β -amino acids were dramatically improved by deciphering of the identity elements in tRNA for recruiting EF-P elongation factor. Discovery of the Δ NPylRS facilitates the evolution of mutually orthogonal tRNA/PylRS pairs. Nature has also evolved strategies to diversify the ribosomal produced polypeptides through post-translational modifications that confer them with a range of added structural and functional features, such as introducing diverse and multiple α -keto- β -amino acids into proteins (Morinaka et al., 2018). Therefore, exploring the diversity of translational components and post-translation modifications from the three kingdoms of life provide sources of further increase in polypeptide chemical diversity.

Using genomically recoded bacteria free of amber codons, RF1 coding genes, as well as several negative effectors, for generating cell free extract enables high yields of proteins containing up to

40 ncAAs with no observable truncated products (Martin et al., 2018). The new genomically recoded organisms free of amber and two Ser codons (Fredens et al., 2019) can serve as a chassis strain for the development of highly efficient CFPS systems capable of site specific incorporation of two or more distinct ncAAs into a single protein or sequence-defined polymers. Utilization of non-natural codons with orthogonal nucleotides beyond the AT(U)GC bases holds the potential to significantly increase the information storage capacity of genes and mRNAs thereby transforming the field of synthetic biology (Zhang Y. et al., 2017; Fischer et al., 2020).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported in part by funding from Australian Research Council grant LP150100689 to KA, National Health and Medical Research Council APP1037320 to KA, and CSIRO-QUT Synthetic Biology alliance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2020.01031/full#supplementary-material>

REFERENCES

- Achenbach, J., Jahnz, M., Bethge, L., Paal, K., Jung, M., Schuster, M., et al. (2015). Outwitting EF-Tu and the ribosome: translation with d-amino acids. *Nucleic Acids Res.* 43, 5687–5698. doi: 10.1093/nar/gkv566
- Ad, O., Hoffman, K. S., Cairns, A. G., Featherston, A. L., Miller, S. J., Soll, D., et al. (2019). Translation of diverse aramid- and 1,3-Dicarbonyl-peptides by wild type ribosomes *in vitro*. *ACS Cent. Sci.* 5, 1289–1294. doi: 10.1021/acscentsci.9b00460
- Adachi, J., Katsura, K., Seki, E., Takemoto, C., Shirouzu, M., Terada, T., et al. (2019). Cell-free protein synthesis using S30 extracts from *Escherichia coli* RFzero strains for efficient incorporation of non-natural amino acids into proteins. *Int. J. Mol. Sci.* 20:492. doi: 10.3390/ijms20030492
- Addy, P. S., Erickson, S. B., Italia, J. S., and Chatterjee, A. (2017). A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for unclickable bioconjugation. *J. Am. Chem. Soc.* 139, 11670–11673. doi: 10.1021/jacs.7b05125
- Ahn, J. H., Chu, H. S., Kim, T. W., Oh, I. S., Choi, C. Y., Hahn, G. H., et al. (2005). Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochem. Biophys. Res. Commun.* 338, 1346–1352. doi: 10.1016/j.bbrc.2005.10.094
- Albayrak, C., and Swartz, J. R. (2013). Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res.* 41, 5949–5963. doi: 10.1093/nar/gkt226
- Aldag, C., Brocker, M. J., Hohn, M. J., Prat, L., Hammond, G., Plummer, A., et al. (2013). Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. *Angew. Chem. Int. Ed Engl.* 52, 1441–1445. doi: 10.1002/anie.201207567
- Amiram, M., Haimovich, A. D., Fan, C., Wang, Y. S., Aerni, H. R., Ntai, I., et al. (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
- Asahara, H., and Uhlenbeck, O. C. (2002). The tRNA specificity of *Thermus thermophilus* EF-Tu. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3499–3504. doi: 10.1073/pnas.052028599
- Bain, J. D., Switzer, C., Chamberlin, A. R., and Benner, S. A. (1992). Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature* 356, 537–539. doi: 10.1038/356537a0
- Bashiruddin, N. K., Nagano, M., and Suga, H. (2015). Synthesis of fused tricyclic peptides using a reprogrammed translation system and chemical modification. *Bioorg. Chem.* 61, 45–50. doi: 10.1016/j.bioorg.2015.06.002
- Blaha, G., Stanley, R. E., and Steitz, T. A. (2009). Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. *Science* 325, 966–970. doi: 10.1126/science.1175800
- Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sowers, G., Veprek, B., et al. (1991). Selenocysteine: the 21st amino acid. *Mol. Microbiol.* 5, 515–520. doi: 10.1111/j.1365-2958.1991.tb00722.x

- Bridge, T., Shaikh, S. A., Thomas, P., Botta, J., McCormick, P. J., and Sachdeva, A. (2019). Site-specific encoding of photoactivity in antibodies enables light-mediated antibody-antigen binding on live cells. *Angew. Chem. Int. Ed Engl.* 58, 17986–17993. doi: 10.1002/anie.201908655
- Bryson, D. L., Fan, C., Guo, L. T., Miller, C., Soll, D., and Liu, D. R. (2017). Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* 13, 1253–1260. doi: 10.1038/nchembio.2474
- Cai, Q., Hanson, J. A., Steiner, A. R., Tran, C., Masikat, M. R., Chen, R., et al. (2015). A simplified and robust protocol for immunoglobulin expression in *Escherichia coli* cell-free protein synthesis systems. *Biotechnol. Prog.* 31, 823–831. doi: 10.1002/btpr.2082
- Carlson, E. D., d'Aquino, A. E., Kim, D. S., Fulk, E. M., Hoang, K., Szal, T., et al. (2019). Engineered ribosomes with tethered subunits for expanding biological function. *Nat. Commun.* 10:3920. doi: 10.1038/s41467-019-11427-y
- Carlson, E. D., Gan, R., Hodgman, C. E., and Jewett, M. C. (2012). Cell-free protein synthesis: applications come of age. *Biotechnol. Adv.* 30, 1185–1194. doi: 10.1016/j.biotechadv.2011.09.016
- Caschera, F., and Noireaux, V. (2014). Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription-translation system. *Biochimie* 99, 162–168. doi: 10.1016/j.biochi.2013.11.025
- Chen, S., Ji, X., Gao, M., Dedkova, L. M., and Hecht, S. M. (2019). *In cellulo* synthesis of proteins containing a fluorescent oxazole amino acid. *J. Am. Chem. Soc.* 141, 5597–5601. doi: 10.1021/jacs.8b12767
- Chen, S., Maini, R., Bai, X., Nangreave, R. C., Dedkova, L. M., and Hecht, S. M. (2017). Incorporation of phosphorylated tyrosine into proteins: *in vitro* translation and study of phosphorylated ikappab-alpha and its interaction with NF-kappaB. *J. Am. Chem. Soc.* 139, 14098–14108. doi: 10.1021/jacs.7b05168
- Crnkovic, A., Vargas-Rodriguez, O., and Soll, D. (2019). Plasticity and constraints of tRNA aminoacylation define directed evolution of Aminoacyl-tRNA synthetases. *Int. J. Mol. Sci.* 20:2294. doi: 10.3390/ijms20092294
- Cui, Z., Mureev, S., Polinkovsky, M. E., Tnimov, Z., Guo, Z., Durek, T., et al. (2017). Combining sense and nonsense codon reassignment for site-selective protein modification with unnatural amino acids. *ACS Synth. Biol.* 6, 535–544. doi: 10.1021/acssynbio.6b00245
- Cui, Z., Stein, V., Tnimov, Z., Mureev, S., and Alexandrov, K. (2015). Semisynthetic tRNA complement mediates *in vitro* protein synthesis. *J. Am. Chem. Soc.* 137, 4404–4413. doi: 10.1021/ja5131963
- Cui, Z., Wu, Y., Mureev, S., and Alexandrov, K. (2018). Oligonucleotide-mediated tRNA sequestration enables one-pot sense codon reassignment *in vitro*. *Nucleic Acids Res.* 46, 6387–6400. doi: 10.1093/nar/gky365
- Dedkova, L. M., Fahmi, N. E., Golovine, S. Y., and Hecht, S. M. (2003). Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc.* 125, 6616–6617. doi: 10.1021/ja035141q
- Dedkova, L. M., Fahmi, N. E., Golovine, S. Y., and Hecht, S. M. (2006). Construction of modified ribosomes for incorporation of D-amino acids into proteins. *Biochemistry* 45, 15541–15551. doi: 10.1021/bi060986a
- Dedkova, L. M., Fahmi, N. E., Paul, R., del Rosario, M., Zhang, L., Chen, S., et al. (2012). Beta-Puromycin selection of modified ribosomes for *in vitro* incorporation of beta-amino acids. *Biochemistry* 51, 401–415. doi: 10.1021/bi2016124
- Des Soye, B. J., Gerbasi, V. R., Thomas, P. M., Kelleher, N. L., and Jewett, M. C. (2019). A highly productive, one-pot cell-free protein synthesis platform based on genomically recoded *Escherichia coli*. *Cell. Chem. Biol.* 26, 1743–1754 e9. doi: 10.1016/j.chembiol.2019.10.008
- Dien, V. T., Holcomb, M., Feldman, A. W., Fischer, E. C., Dwyer, T. J., and Unersberg, F. E. (2018). Progress toward a semi-synthetic organism with an unrestricted expanded genetic alphabet. *J. Am. Chem. Soc.* 140, 16115–16123. doi: 10.1021/jacs.8b08416
- Doerfel, L. K., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H., and Rodnina, M. V. (2013). EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* 339, 85–88. doi: 10.1126/science.1229017
- Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T., and Sisido, M. (2007). Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *J. Am. Chem. Soc.* 129, 14458–14462. doi: 10.1021/ja075557u
- Dopp, B. J. L., Tamiev, D. D., and Reuel, N. F. (2019). Cell-free supplement mixtures: elucidating the history and biochemical utility of additives used to support *in vitro* protein synthesis in *E. coli* extract. *Biotechnol. Adv.* 37, 246–258. doi: 10.1016/j.biotechadv.2018.12.006
- Dopp, J. L., Jo, Y. R., and Reuel, N. F. (2019). Methods to reduce variability in *E. Coli*-based cell-free protein expression experiments. *Synth. Syst. Biotechnol.* 4, 204–211. doi: 10.1016/j.synbio.2019.10.003
- Dreier, B., and Pluckthun, A. (2018). Rapid Selection of High-affinity antibody scFv fragments using ribosome display. *Methods Mol. Biol.* 1827, 235–268. doi: 10.1007/978-1-4939-8648-4_13
- Drienovská, I., and Roelfes, G. (2020). Expanding the enzyme universe with genetically encoded unnatural amino acids. *Nat. Catal.* 3, 193–202. doi: 10.1038/s41929-019-0410-8
- Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015). Designing logical codon reassignment - expanding the chemistry in biology. *Chem. Sci.* 6, 50–69. doi: 10.1039/C4SC01534G
- Elliott, T. S., Townsley, F. M., Bianco, A., Ernst, R. J., Sachdeva, A., Elsasser, S. J., et al. (2014). Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat. Biotechnol.* 32, 465–472. doi: 10.1038/nbt.2860
- Fan, C., Ip, K., and Soll, D. (2016). Expanding the genetic code of *Escherichia coli* with phosphotyrosine. *FEBS Lett.* 590, 3040–3047. doi: 10.1002/1873-3468.12333
- Fan, C., Xiong, H., Reynolds, N. M., and Soll, D. (2015). Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.* 43:e156. doi: 10.1093/nar/gkv800
- Fan, Z., Song, J., Guan, T., Lv, X., and Wei, J. (2018). Efficient expression of glutathione peroxidase with chimeric tRNA in amber-less *Escherichia coli*. *ACS Synth. Biol.* 7, 249–257. doi: 10.1021/acssynbio.7b00290
- Fischer, E. C., Hashimoto, K., Zhang, Y., Feldman, A. W., Dien, V. T., Karadeema, R. J., et al. (2020). New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nat. Chem. Biol.* 16, 570–576. doi: 10.1038/s41589-020-0507-z
- Fleming, S. R., Bartges, T. E., Vinogradov, A. A., Kirkpatrick, C. L., Goto, Y., Suga, H., et al. (2019). Flexizyme-enabled benchtop biosynthesis of thiopeptides. *J. Am. Chem. Soc.* 141, 758–762. doi: 10.1021/jacs.8b11521
- Foshag, D., Henrich, E., Hiller, E., Schafer, M., Kerger, C., Burger-Kentscher, A., et al. (2018). The *E. coli* S30 lysate proteome: a prototype for cell-free protein production. *N. Biotechnol.* 40, 245–260. doi: 10.1016/j.nbt.2017.09.005
- Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., et al. (2019). Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 569, 514–518. doi: 10.1038/s41586-019-1192-5
- Fried, S. D., Schmied, W. H., Uttamapinant, C., and Chin, J. W. (2015). Ribosome subunit stapling for orthogonal translation in *E. coli*. *Angew. Chem. Weinheim Bergstr Ger.* 127, 12982–12985. doi: 10.1002/ange.201506311
- Fu, C., Chen, Q., Zheng, F., Yang, L., Li, H., Zhao, Q., et al. (2019). Genetically Encoding a lipidated amino acid for extension of protein half-life *in vivo*. *Angew. Chem. Int. Ed Engl.* 58, 1392–1396. doi: 10.1002/anie.201811837
- Fujino, T., Goto, Y., Suga, H., and Murakami, H. (2013). Reevaluation of the D-amino acid compatibility with the elongation event in translation. *J. Am. Chem. Soc.* 135, 1830–1837. doi: 10.1021/ja309570x
- Fujino, T., Goto, Y., Suga, H., and Murakami, H. (2016). Ribosomal synthesis of peptides with multiple beta-amino acids. *J. Am. Chem. Soc.* 138, 1962–1969. doi: 10.1021/jacs.5b12482
- Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016). The all *E. coli* TX-TL Toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth. Biol.* 5, 344–355. doi: 10.1021/acssynbio.5b00296
- Garenne, D., Beisel, C. L., and Noireaux, V. (2019). Characterization of the all-*E. coli* transcription-translation system myTXTL by mass spectrometry. *Rapid. Commun. Mass Spectrom.* 33, 1036–1048. doi: 10.1002/rcm.8438
- George, S., Aguirre, J. D., Spratt, D. E., Bi, Y., Jeffery, M., Shaw, G. S., et al. (2016). Generation of phospho-ubiquitin variants by orthogonal translation reveals codon skipping. *FEBS Lett.* 590, 1530–1542. doi: 10.1002/1873-3468.12182
- Goerke, A. R., and Swartz, J. R. (2009). High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnol. Bioeng.* 102, 400–416. doi: 10.1002/bit.22070
- Goto, Y., Iwasaki, K., Torikai, K., Murakami, H., and Suga, H. (2009). Ribosomal synthesis of dehydrobutyrine- and methyllanthionine-containing peptides. *Chem. Commun.* 21, 3419–3421. doi: 10.1039/b904314d
- Goto, Y., Katoh, T., and Suga, H. (2011). Flexizymes for genetic code reprogramming. *Nat. Protoc.* 6, 779–790. doi: 10.1038/nprot.2011.331

- Goto, Y., Murakami, H., and Suga, H. (2008a). Initiating translation with D-amino acids. *RNA* 14, 1390–1398. doi: 10.1261/rna.1020708
- Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., and Suga, H. (2008b). Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* 3, 120–129. doi: 10.1021/cb700233t
- Goto, Y., and Suga, H. (2009). Translation initiation with initiator tRNA charged with exotic peptides. *J. Am. Chem. Soc.* 131, 5040–5041. doi: 10.1021/ja900597d
- Guarino, C., and DeLisa, M. P. (2012). A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology* 22, 596–601. doi: 10.1093/glycob/cwr151
- Gubbens, J., Kim, S. J., Yang, Z. Y., Johnson, A. E., and Skach, W. R. (2010). *In vitro* incorporation of nonnatural amino acids into protein using tRNA(Cys)-derived opal, ochre, and amber suppressor tRNAs. *RNA* 16, 1660–1672. doi: 10.1261/rna.2024810
- Guo, J., Melancon, C. E. 3rd, Lee, H. S., Groff, D., and Schultz, P. G. (2009). Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem. Int. Ed Engl.* 48, 9148–9151. doi: 10.1002/anie.200904035
- Guo, J., Wang, J., Anderson, J. C., and Schultz, P. G. (2008). Addition of an alpha-hydroxy acid to the genetic code of bacteria. *Angew. Chem. Int. Ed Engl.* 47, 722–725. doi: 10.1002/anie.200704074
- Hammerling, M. J., Kruger, A., and Jewett, M. C. (2019). Strategies for *in vitro* engineering of the translation machinery. *Nucleic Acids Res.* 48, 1068–1083. doi: 10.1093/nar/gkz1011
- Hanes, J., and Pluckthun, A. (1997). *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.* 94, 4937–4942. doi: 10.1073/pnas.94.10.4937
- Hartman, M. C., Josephson, K., Lin, C. W., and Szostak, J. W. (2007). An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS ONE* 2:e972. doi: 10.1371/journal.pone.0000972
- Hartman, M. C., Josephson, K., and Szostak, J. W. (2006). Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4356–4361. doi: 10.1073/pnas.0509219103
- Haruna, K., Alkazemi, M. H., Liu, Y., Soll, D., and Englert, M. (2014). Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic Acids Res.* 42, 9976–9983. doi: 10.1093/nar/gku691
- Hauf, M., Richter, F., Schneider, T., Faidt, T., Martins, B. M., Baumann, T., et al. (2017). photoactivatable mussel-based underwater adhesive proteins by an expanded genetic code. *ChemBiochem* 18, 1819–1823. doi: 10.1002/cbic.201700327
- Hecht, S. M., Alford, B. L., Kuroda, Y., and Kitano, S. (1978). “Chemical aminoacylation” of tRNAs. *J. Biol. Chem.* 253, 4517–4520.
- Hong, S. H., Kwon, Y. C., and Jewett, M. C. (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, S. H., Ntai, I., Haimovich, A. D., Kelleher, N. L., Isaacs, F. J., and Jewett, M. C. (2014b). Cell-free protein synthesis from a release factor 1 deficient *Escherichia coli* activates efficient and multiple site-specific nonstandard amino acid incorporation. *ACS Synth. Biol.* 3, 398–409. doi: 10.1021/sb400140t
- Hoppmann, C., Lacey, V. K., Louie, G. V., Wei, J., Noel, J. P., and Wang, L. (2014). Genetically encoding photoswitchable click amino acids in *Escherichia coli* and mammalian cells. *Angew Chem Int Edit* 53, 3932–3936. doi: 10.1002/anie.201400001
- Hoppmann, C., Maslennikov, I., Choe, S., and Wang, L. (2015). *In situ* Formation of an Azo Bridge on Proteins Controllable by Visible Light. *J. Am. Chem. Soc.* 137, 11218–11221. doi: 10.1021/jacs.5b06234
- Hoshika, S., Leal, N. A., Kim, M. J., Kim, M. S., Karalkar, N. B., Kim, H. J., et al. (2019). Hachimoji DNA and RNA: a genetic system with eight building blocks. *Science* 363, 884–887. doi: 10.1126/science.aat0971
- Huang, L. C., Pan, X., Yang, H., and Wan, L. K. (2013). Linking genotype to phenotype on beads: high throughput selection of peptides with biological function. *Sci. Rep.* 3:3030. doi: 10.1038/srep03030
- Huang, Y. C., Wiedmann, M. M., and Suga, H. (2019). RNA display methods for the discovery of bioactive macrocycles. *Chem. Rev.* 119, 10360–10391. doi: 10.1021/acs.chemrev.8b00430
- Huguenin-Desot, N., Alonzo, D. A., Heberlig, G. W., Mahesh, M., Nguyen, D. P., Dornan, M. H., et al. (2019). Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* 565, 112–117. doi: 10.1038/s41586-018-0781-z
- Hui, A., and de Boer, H. A. (1987). Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 84, 4762–4766. doi: 10.1073/pnas.84.14.4762
- Iqbal, E. S., Dods, K. K., and Hartman, M. C. T. (2018). Ribosomal incorporation of backbone modified amino acids via an editing-deficient aminoacyl-tRNA synthetase. *Org. Biomol. Chem.* 16, 1073–1078. doi: 10.1039/C7OB02931D
- Ishida, S., Terasaka, N., Katoh, T., and Suga, H. (2020). An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. *Nat. Chem. Biol.* 16, 702–709. doi: 10.1038/s41589-020-0500-6
- Italia, J. S., Addy, P. S., Erickson, S. B., Peeler, J. C., Weerapana, E., and Chatterjee, A. (2019). Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites. *J. Am. Chem. Soc.* 141, 6204–6212. doi: 10.1021/jacs.8b12954
- Ito, K., Passioura, T., and Suga, H. (2013). technologies for the synthesis of mRNA-encoding libraries and discovery of bioactive natural product-inspired non-traditional macrocyclic peptides. *Molecules* 18, 3502–3528. doi: 10.3390/molecules18033502
- Iwane, Y., Hitomi, A., Murakami, H., Katoh, T., Goto, Y., and Suga, H. (2016). Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nat. Chem.* 8, 317–325. doi: 10.1038/nchem.2446
- Iwasaki, M., Miwa, S., Ikegami, T., Tomita, M., Tanaka, N., and Ishihama, Y. (2010). One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale. *Anal. Chem.* 82, 2616–2620. doi: 10.1021/ac100343q
- Jaroentomeechai, T., Stark, J. C., Natarajan, A., Glasscock, C. J., Yates, L. E., Hsu, K. J., et al. (2018). Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat. Commun.* 9:2686. doi: 10.1038/s41467-018-05110-x
- Jedlitzke, B., Yilmaz, Z., Dörner, W., and Mootz, H. D. (2020). Photobodies: light-activatable single-domain antibody fragments. *Angew. Chem. Int. Ed Engl.* 59, 1506–1510. doi: 10.1002/anie.201912286
- Jewett, M. C., Calhoun, K. A., Voloshin, A., Wu, J. J., and Swartz, J. R. (2008). An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* 4:220. doi: 10.1038/msb.2008.57
- Jewett, M. C., Fritz, B. R., Timmerman, L. E., and Church, G. M. (2013). *In vitro* integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol. Syst. Biol.* 9:678. doi: 10.1038/msb.2013.31
- Jewett, M. C., and Swartz, J. R. (2004). Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* 86, 19–26. doi: 10.1002/bit.20026
- Kang, T. J., Yuzawa, S., and Suga, H. (2008). Expression of histone H3 tails with combinatorial lysine modifications under the reprogrammed genetic code for the investigation on epigenetic markers. *Chem. Biol.* 15, 1166–1174. doi: 10.1016/j.chembiol.2008.09.014
- Katoh, T., Iwane, Y., and Suga, H. (2017b). Logical engineering of D-arm and T-stem of tRNA that enhances D-amino acid incorporation. *Nucleic Acids Res.* 45, 12601–12610. doi: 10.1093/nar/gkx1129
- Katoh, T., and Suga, H. (2018). Ribosomal incorporation of consecutive beta-amino acids. *J. Am. Chem. Soc.* 140, 12159–12167. doi: 10.1021/jacs.8b07247
- Katoh, T., Tajima, K., and Suga, H. (2017a). Consecutive elongation of D-Amino acids in translation. *Cell Chem. Biol.* 24, 46–54. doi: 10.1016/j.chembiol.2016.11.012
- Katoh, T., Wohlgenuth, I., Nagano, M., Rodnina, M. V., and Suga, H. (2016). Essential structural elements in tRNA(Pro) for EF-P-mediated alleviation of translation stalling. *Nat. Commun.* 7:11657. doi: 10.1038/ncomms11657
- Kawakami, T., Ishizawa, T., Fujino, T., Reid, P. C., Suga, H., and Murakami, H. (2013b). *In vitro* selection of multiple libraries created by genetic code reprogramming to discover macrocyclic peptides that antagonize VEGFR2 activity in living cells. *ACS Chem. Biol.* 8, 1205–1214. doi: 10.1021/cb300697h
- Kawakami, T., Ishizawa, T., and Murakami, H. (2013a). Extensive reprogramming of the genetic code for genetically encoded synthesis of highly N-alkylated polycyclic peptidomimetics. *J. Am. Chem. Soc.* 135, 12297–12304. doi: 10.1021/ja405044k

- Kawakami, T., Murakami, H., and Suga, H. (2008). Messenger RNA-Programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol.* 15, 32–42. doi: 10.1016/j.chembiol.2007.12.008
- Kawakami, T., Ogawa, K., Hatta, T., Goshima, N., and Natsume, T. (2016). Directed evolution of a cyclized peptid-peptide chimera against a cell-free expressed protein and proteomic profiling of the interacting proteins to create a protein-protein interaction inhibitor. *ACS Chem. Biol.* 11, 1569–1577. doi: 10.1021/acscchembio.5b01014
- Kawakami, T., Ohta, A., Ohuchi, M., Ashigai, H., Murakami, H., and Suga, H. (2009). Diverse backbone-cyclized peptides via codon reprogramming. *Nat. Chem. Biol.* 5, 888–890. doi: 10.1038/nchembio.259
- Kawakami, T., Sasaki, T., Reid, P. C., and Murakami, H. (2014). Incorporation of electrically charged N-alkyl amino acids into ribosomally synthesized peptides via post-translational conversion. *Chem. Sci.* 5, 887–893. doi: 10.1039/c3sc52744a
- Kightlinger, W., Duncker, K. E., Ramesh, A., Thames, A. H., Natarajan, A., Stark, J. C., et al. (2019). A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nat. Commun.* 10:5404. doi: 10.1038/s41467-019-12024-9
- Klippenstein, V., Hoppmann, C., Ye, S., Wang, L., and Paoletti, P. (2017). Optocontrol of glutamate receptor activity by single side-chain photoisomerization. *Elife* 6:e25808. doi: 10.7554/eLife.25808.028
- Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tkalco, M., Cusack, S., et al. (2003). Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* 10, 425–432. doi: 10.1038/nsb934
- Kobayashi, T., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2009). Recognition of non-alpha-amino substrates by pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* 385, 1352–1360. doi: 10.1016/j.jmb.2008.11.059
- Kuruma, Y., and Ueda, T. (2015). The PURE system for the cell-free synthesis of membrane proteins. *Nat. Protoc.* 10, 1328–1344. doi: 10.1038/nprot.2015.082
- Kwok, H. S., Vargas-Rodriguez, O., Melnikov, S. V., and Soll, D. (2019). Engineered Aminoacyl-tRNA synthetases with improved selectivity toward noncanonical amino acids. *ACS Chem. Biol.* 14, 603–612. doi: 10.1021/acscchembio.9b00088
- Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., et al. (2013). Genomically recoded organisms expand biological functions. *Science* 342, 357–360. doi: 10.1126/science.1241459
- Lammers, M., Neumann, H., Chin, J. W., and James, L. C. (2010). Acetylation regulates Cyclophilin A catalysis, immunosuppression and HIV isomerization. *Nat. Chem. Biol.* 6, 331–337. doi: 10.1038/nchembio.342
- Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A., and Chin, J. W. (2012b). Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat. Chem.* 4, 298–304. doi: 10.1038/nchem.1250
- Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D. J., Blackman, M. L., et al. (2012a). Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling *in vitro* and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. *J. Am. Chem. Soc.* 134, 10317–10320. doi: 10.1021/ja302832g
- LaRiviere, F. J., Wolfson, A. D., and Uhlenbeck, O. C. (2001). Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* 294, 165–168. doi: 10.1126/science.1064242
- Lee, J., Schwieter, K. E., Watkins, A. M., Kim, D. S., Yu, H., Schwarz, K. J., et al. (2019). Expanding the limits of the second genetic code with ribozymes. *Nat. Commun.* 10:5097. doi: 10.1038/s41467-019-12916-w
- Lee, K. J., Kang, D., and Park, H. S. (2019). Site-specific labeling of proteins using unnatural amino acids. *Mol. Cells* 42, 386–396. doi: 10.14348/molcells.2019.0078
- Lee, S., Oh, S., Yang, A., Kim, J., Soll, D., Lee, D., et al. (2013). A facile strategy for selective incorporation of phosphoserine into histones. *Angew. Chem. Int. Ed Engl.* 52, 5771–5775. doi: 10.1002/anie.201300531
- Lesley, S. A., Brow, M. A., and Burgess, R. R. (1991). Use of *in vitro* protein synthesis from polymerase chain reaction-generated templates to study interaction of *Escherichia coli* transcription factors with core RNA polymerase and for epitope mapping of monoclonal antibodies. *J. Biol. Chem.* 266, 2632–2638.
- Liu, C. C., and Schultz, P. G. (2010). Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 79, 413–444. doi: 10.1146/annurev.biochem.052308.105824
- Liu, J., Li, S. S., Aslam, N. A., Zheng, F., Yang, B., Cheng, R. J., et al. (2019). Genetically encoding photocaged quinone methide to multitarget protein residues covalently *in vivo*. *J. Am. Chem. Soc.* 141, 9458–9462. doi: 10.1021/jacs.9b01738
- Liu, R., Barrick, J. E., Szostak, J. W., and Roberts, R. W. (2000). Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. *Meth. Enzymol.* 318, 268–293. doi: 10.1016/S0076-6879(00)18058-9
- Luo, X., Fu, G., Wang, R. E., Zhu, X., Zambaldo, C., Liu, R., et al. (2017). Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* 13, 845–849. doi: 10.1038/nchembio.2405
- Maini, R., Dedkova, L. M., Paul, R., Madathil, M. M., Chowdhury, S. R., Chen, S., et al. (2015). Ribosome-mediated incorporation of dipeptides and dipeptide analogues into proteins *in vitro*. *J. Am. Chem. Soc.* 137, 11206–11209. doi: 10.1021/jacs.5b03135
- Maini, R., Nguyen, D. T., Chen, S., Dedkova, L. M., Chowdhury, S. R., Alcalá-Torano, R., et al. (2013). Incorporation of beta-amino acids into dihydrofolate reductase by ribosomes having modifications in the peptidyltransferase center. *Bioorg. Med. Chem.* 21, 1088–1096. doi: 10.1016/j.bmc.2013.01.002
- Malyshev, D. A., and Romesberg, F. E. (2015). The expanded genetic alphabet. *Angew. Chem. Int. Ed Engl.* 54, 11930–11944. doi: 10.1002/anie.201502890
- Marshall, R., Maxwell, C. S., Collins, S. P., Beisel, C. L., and Noireaux, V. (2017). Short DNA containing chi sites enhances DNA stability and gene expression in *E-coli* cell-free transcription-translation systems. *Biotechnol. Bioeng.* 114, 2137–2141. doi: 10.1002/bit.26333
- Martin, R. W., Des Soye, B. J., Kwon, Y. C., Kay, J., Davis, R. G., Thomas, P. M., et al. (2018). Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nat. Commun.* 9:1203. doi: 10.1038/s41467-018-03469-5
- Melnikov, S. V., Khabibullina, N. F., Mairhofer, E., Vargas-Rodriguez, O., Reynolds, N. M., Micura, R., et al. (2019). Mechanistic insights into the slow peptide bond formation with D-amino acids in the ribosomal active site. *Nucleic Acids Res.* 47, 2089–2100. doi: 10.1093/nar/gky1211
- Merryman, C., and Green, R. (2004). Transformation of aminoacyl tRNAs for the *in vitro* selection of “drug-like” molecules. *Chem. Biol.* 11, 575–582. doi: 10.1016/j.chembiol.2004.03.009
- Michel-Reydellet, N., Woodrow, K., and Swartz, J. (2005). Increasing PCR fragment stability and protein yields in a cell-free system with genetically modified *Escherichia coli* extracts. *J. Mol. Microbiol. Biotechnol.* 9, 26–34. doi: 10.1159/000088143
- Miller, C., Brocker, M. J., Prat, L., Ip, K., Chirathivat, N., Feiock, A., et al. (2015). A synthetic tRNA for EF-Tu mediated selenocysteine incorporation *in vivo* and *in vitro*. *FEBS Lett.* 589, 2194–2199. doi: 10.1016/j.febslet.2015.06.039
- Morimoto, J., Hayashi, Y., Iwasaki, K., and Suga, H. (2011). Flexizymes: their evolutionary history and the origin of catalytic function. *Acc. Chem. Res.* 44, 1359–1368. doi: 10.1021/ar2000953
- Morimoto, J., Hayashi, Y., and Suga, H. (2012). Discovery of macrocyclic peptides armed with a mechanism-based warhead: isoform-selective inhibition of human deacetylase SIRT2. *Angew. Chem. Int. Ed Engl.* 51, 3423–3427. doi: 10.1002/anie.201108118
- Morinaka, B. I., Lakis, E., Verest, M., Helf, M. J., Scalvenzi, T., Vagstad, A. L., et al. (2018). Natural noncanonical protein splicing yields products with diverse beta-amino acid residues. *Science* 359, 779–782. doi: 10.1126/science.aao0157
- Mukai, T., Sevostyanova, A., Suzuki, T., Fu, X., and Soll, D. (2018). A facile method for producing selenocysteine-containing proteins. *Angew. Chem. Int. Ed Engl.* 57, 7215–7219. doi: 10.1002/anie.201713215
- Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods.* 3, 357–359. doi: 10.1038/nmeth877
- Nakajima, E., Goto, Y., Sako, Y., Murakami, H., and Suga, H. (2009). Ribosomal synthesis of peptides with C-terminal lactams, thiolactones, and alkylamides. *ChemBiochem* 10, 1186–1192. doi: 10.1002/cbic.200900058
- Napolitano, M. G., Landon, M., Gregg, C. J., Lajoie, M. J., Govindarajan, L., Mosberg, J. A., et al. (2016). Emergent rules for codon choice elucidated by editing rare arginine codons in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 113, E5588–E5597. doi: 10.1073/pnas.1605856113
- Narumi, R., Masuda, K., Tomonaga, T., Adachi, J., Ueda, H. R., and Shimizu, Y. (2018). Cell-free synthesis of stable isotope-labeled internal standards for targeted quantitative proteomics. *Synth. Syst. Biotechnol.* 3, 97–104. doi: 10.1016/j.synbio.2018.02.004

- Narumi, R., Shimizu, Y., Ukai-Tadenuma, M., Ode, K. L., Kanda, G. N., Shinohara, Y., et al. (2016). Mass spectrometry-based absolute quantification reveals rhythmic variation of mouse circadian clock proteins. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3461–E3467. doi: 10.1073/pnas.1603799113
- Nawatha, M., Rogers, J. M., Bonn, S. M., Livneh, I., Lemma, B., Mali, S. M., et al. (2019). De novo macrocyclic peptides that specifically modulate Lys48-linked ubiquitin chains. *Nat. Chem.* 11, 644–652. doi: 10.1038/s41557-019-0278-x
- Neumann, H. (2012). Rewiring translation - genetic code expansion and its applications. *FEBS Lett.* 586, 2057–2064. doi: 10.1016/j.febslet.2012.02.002
- Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., et al. (2009). A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell* 36, 153–163. doi: 10.1016/j.molcel.2009.07.027
- Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464, 441–444. doi: 10.1038/nature08817
- Nguyen, D. P., Alai, M. M. G., Kapadnis, P. B., Neumann, H., and Chin, J. W. (2009). Genetically encoding N-epsilon-Methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.* 131, 14194–14195. doi: 10.1021/ja906603s
- Nikic, I., Plass, T., Schraidt, O., Szymanski, J., Briggs, J. A. G., Schultz, C., et al. (2014). Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew. Chem. Int. Edit.* 53, 2245–2249. doi: 10.1002/anie.201309847
- Nirenberg, M., and Leder, P. (1964). Rna codewords and protein synthesis. the effect of trinucleotides upon the binding of srna to ribosomes. *Science* 145, 1399–1407. doi: 10.1126/science.145.3639.1399
- Nirenberg, M., and Matthaei, J. H. (1961). Dependence of cell-free protein synthesis in *E coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 47:1588–1602. doi: 10.1073/pnas.47.10.1588
- Niwa, N., Yamagishi, Y., Murakami, H., and Suga, H. (2009). A flexizyme that selectively charges amino acids activated by a water-friendly leaving group. *Bioorg. Med. Chem. Lett.* 19, 3892–3894. doi: 10.1016/j.bmcl.2009.03.114
- Niwa, T., Kanamori, T., Ueda, T., and Taguchi, H. (2012). Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8937–8942. doi: 10.1073/pnas.1201380109
- Noren, C. J., Anthonycahill, S. J., Griffith, M. C., and Schultz, P. G. (1989). A general-method for site-specific incorporation of unnatural amino-acids into proteins. *Science* 244, 182–188. doi: 10.1126/science.2649980
- Nozawa, K., O'Donoghue, P., Gundllapalli, S., Araiso, Y., Ishitani, R., Umehara, T., et al. (2009). Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* 457, 1163–1167. doi: 10.1038/nature07611
- Odoi, K. A., Huang, Y., Rezenom, Y. H., and Liu, W. R. (2013). Nonsense and sense suppression abilities of original and derivative Methanosarcina mazei pyrrolysyl-tRNA synthetase-tRNA(Pyl) pairs in the *Escherichia coli* BL21(DE3) cell strain. *PLoS ONE* 8:e57035. doi: 10.1371/journal.pone.0057035
- O'Donoghue, P., Prat, L., Heinemann, I. U., Ling, J., Odoi, K., Liu, W. R., et al. (2012). Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA^{Pyl} for genetic code expansion. *FEBS Lett.* 586, 3931–3937. doi: 10.1016/j.febslet.2012.09.033
- Ohshiro, Y., Nakajima, E., Goto, Y., Fuse, S., Takahashi, T., Doi, T., et al. (2011). Ribosomal synthesis of backbone-macrocyclic peptides containing gamma-amino acids. *ChemBiochem* 12, 1183–1187. doi: 10.1002/cbic.201100104
- Ohta, A., Murakami, H., Higashimura, E., and Suga, H. (2007). Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol.* 14, 1315–1322. doi: 10.1016/j.chembiol.2007.10.015
- Oller-Salvia, B., and Chin, J. W. (2019). efficient phage display with multiple distinct non-canonical amino acids using orthogonal ribosome-mediated genetic code expansion. *Angew. Chem. Int. Ed Engl.* 58, 10844–10848. doi: 10.1002/anie.201902658
- Orelle, C., Carlson, E. D., Szal, T., Florin, T., Jewett, M. C., and Mankin, A. S. (2015). Protein synthesis by ribosomes with tethered subunits. *Nature* 524, 119–124. doi: 10.1038/nature14862
- Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., et al. (2016). Design, synthesis, and testing toward a 57-codon genome. *Science* 353, 819–822. doi: 10.1126/science.aaf3639
- Oza, J. P., Aerni, H. R., Pirman, N. L., Barber, K. W., ter Haar, C. M., Rogulina, S., et al. (2015). Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nat. Commun.* 6:8168. doi: 10.1038/ncomms9168
- Pardee, K., Slomovic, S., Nguyen, P. Q., Lee, J. W., Donghia, N., Burrill, D., et al. (2016). Portable, on-demand biomolecular manufacturing. *Cell* 167, 248–259 e12. doi: 10.1016/j.cell.2016.09.013
- Park, H. S., Hohn, M. J., Umehara, T., Guo, L. T., Osborne, E. M., Benner, J., et al. (2011). Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* 333, 1151–1154. doi: 10.1126/science.1207203
- Passioura, T. (2019). The road ahead for the development of macrocyclic peptide ligands. *Biochemistry.* 59, 139–145. doi: 10.1021/acs.biochem.9b00802
- Passioura, T., Liu, W., Dunkelmann, D., Higuchi, T., and Suga, H. (2018). Display selection of exotic macrocyclic peptides expressed under a radically reprogrammed 23 amino acid genetic code. *J. Am. Chem. Soc.* 140, 11551–11555. doi: 10.1021/jacs.8b03367
- Passioura, T., and Suga, H. (2014). Reprogramming the genetic code *in vitro*. *Trends Biochem. Sci.* 39, 400–408. doi: 10.1016/j.tibs.2014.07.005
- Passioura, T., and Suga, H. (2017). A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun.* 53, 1931–1940. doi: 10.1039/C6CC06951G
- Pirman, N. L., Barber, K. W., Aerni, H. R., Ma, N. J., Haimovich, A. D., Rogulina, S., et al. (2015). A flexible codon in genomically recoded *Escherichia coli* permits programmable protein phosphorylation. *Nat. Commun.* 6:8130. doi: 10.1038/ncomms9130
- Quast, R. B., Mrusek, D., Hoffmeister, C., Sonnabend, A., and 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
- Rackham, O., and Chin, J. W. (2005a). A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol.* 1, 159–166. doi: 10.1038/nchembio719
- Rackham, O., and Chin, J. W. (2005b). Cellular logic with orthogonal ribosomes. *J. Am. Chem. Soc.* 127, 17584–17585. doi: 10.1021/ja055338d
- Rasmussen, L. C. V., Laursen, B. S., Mortensen, K. K., and Sperling-Petersen, H. U. (2009). “Initiator tRNAs in bacteria and eukaryotes,” in *Encyclopedia of Life Sciences (ELS)*. Chichester: John Wiley & Sons, Ltd. doi: 10.1002/9780470015902.a0000543.pub2
- Rogers, J. M., Kwon, S., Dawson, S. J., Mandal, P. K., Suga, H., and Huc, I. (2018). Ribosomal synthesis and folding of peptide-helical aromatic foldamer hybrids. *Nat. Chem.* 10, 405–412. doi: 10.1038/s41557-018-0007-x
- Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaitė, A., Hancock, S. M., et al. (2015). Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* 11, 496–503. doi: 10.1038/nchembio.1823
- Sakai, K., Passioura, T., Sato, H., Ito, K., Furuhashi, H., Umitsu, M., et al. (2019). Macrocyclic peptide-based inhibition and imaging of hepatocyte growth factor. *Nat. Chem. Biol.* 15, 598–606. doi: 10.1038/s41589-019-0285-7
- Sako, Y., Goto, Y., Murakami, H., and Suga, H. (2008a). Ribosomal synthesis of peptidase-resistant peptides closed by a nonreducible inter-side-chain bond. *ACS Chem. Biol.* 3, 241–249. doi: 10.1021/cb800010p
- Sako, Y., Morimoto, J., Murakami, H., and Suga, H. (2008b). Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions. *J. Am. Chem. Soc.* 130, 7232–7234. doi: 10.1021/ja800953c
- Saleh, A. M., Wilding, K. M., Calve, S., Bundy, B. C., and Kinzer-Ursem, T. L. (2019). Non-canonical amino acid labeling in proteomics and biotechnology. *J. Biol. Eng.* 13:43. doi: 10.1186/s13036-019-0166-3
- Sauerwald, A., Zhu, W., Major, T. A., Roy, H., Palioura, S., Jahn, D., et al. (2005). RNA-dependent cysteine biosynthesis in archaea. *Science* 307, 1969–1972. doi: 10.1126/science.1108329
- Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrne, E., Volkmer, B., Callipo, L., et al. (2016). The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* 34, 104–110. doi: 10.1038/nbt.3418
- Schmied, W. H., Tnimov, Z., Uttamapinant, C., Rae, C. D., Fried, S. D., and Chin, J. W. (2018). Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* 564, 444–448. doi: 10.1038/s41586-018-0773-z
- Schrader, J. M., Chapman, S. J., and Uhlenbeck, O. C. (2011). Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5215–5220. doi: 10.1073/pnas.1102128108
- Shimizu, Y. (2012). ArfA recruits RF2 into stalled ribosomes. *J. Mol. Biol.* 423, 624–631. doi: 10.1016/j.jmb.2012.08.007
- Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* 19, 751–755. doi: 10.1038/90802

- Shimizu, Y., Kanamori, T., and Ueda, T. (2005). Protein synthesis by pure translation systems. *Methods* 36, 299–304. doi: 10.1016/j.ymeth.2005.04.006
- Shimizu, Y., and Ueda, T. (2010). PURE technology. *Cell-Free Protein Produc.* 607, 11–21. doi: 10.1007/978-1-60327-331-2_2
- Shrestha, P., Smith, M. T., and Bundy, B. C. (2014). Cell-free unnatural amino acid incorporation with alternative energy systems and linear expression templates. *N. Biotechnol.* 31, 28–34. doi: 10.1016/j.nbt.2013.09.002
- Silverman, A. D., Karim, A. S., and Jewett, M. C. (2019). Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* 21, 151–170. doi: 10.1038/s41576-019-0186-3
- Singh-Blom, A., Hughes, R. A., and Ellington, A. D. (2014). An amino acid depleted cell-free protein synthesis system for the incorporation of non-canonical amino acid analogs into proteins. *J. Biotechnol.* 178, 12–22. doi: 10.1016/j.jbiotec.2014.02.009
- Srinivasan, G., James, C. M., and Krzycki, J. A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* 296, 1459–1462. doi: 10.1126/science.1069588
- Steer, B. A., and Schimmel, P. (1999). Major anticodon-binding region missing from an archaeobacterial tRNA synthetase. *J. Biol. Chem.* 274, 35601–35606. doi: 10.1074/jbc.274.50.35601
- Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. (2014). Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* 3, 387–397. doi: 10.1021/sb400131a
- Takatsuji, R., Shinbara, K., Katoh, T., Goto, Y., Passioura, T., Yajima, R., et al. (2019). Ribosomal synthesis of backbone-cyclic peptides compatible with *in vitro* display. *J. Am. Chem. Soc.* 141, 2279–2287. doi: 10.1021/jacs.8b05327
- Terasaka, N., Hayashi, G., Katoh, T., and Suga, H. (2014). An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. *Nat. Chem. Biol.* 10, 555–557. doi: 10.1038/nchembio.1549
- Tharp, J. M., Ad, O., Amikura, K., Ward, F. R., Garcia, E. M., Cate, J. H. D., et al. (2019). Initiation of protein synthesis with non-canonical amino acids *in vivo*. *Angew. Chem. Int. Ed Engl.* 59, 3122–3126. doi: 10.1002/anie.201914671
- Thompson, J., Kim, D. F., O'Connor, M., Lieberman, K. R., Bayfield, M. A., Gregory, S. T., et al. (2001). Analysis of mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proc. Natl. Acad. Sci. U.S.A.* 98, 9002–9007. doi: 10.1073/pnas.151257098
- Torikai, K., and Suga, H. (2014). Ribosomal synthesis of an amphoterin-B inspired macrocycle. *J. Am. Chem. Soc.* 136, 17359–17361. doi: 10.1021/ja508648s
- Torizawa, T., Shimizu, M., Taoka, M., Miyano, H., and Kainosho, M. (2004). Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol. *J. Biomol. NMR* 30, 311–325. doi: 10.1007/s10858-004-3534-2
- Tsiamantas, C., Kwon, S., Douat, C., Huc, I., and Suga, H. (2019). Optimizing aromatic oligoamide foldamer side-chains for ribosomal translation initiation. *Chem. Commun.* 55, 7366–7369. doi: 10.1039/C9CC03547H
- Vargas-Rodriguez, O., Sevostyanova, A., Soll, D., and Crnkovic, 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
- Venkat, S., Sturges, J., Stahman, A., Gregory, C., Gan, Q., and Fan, C. (2018). Genetically incorporating two distinct post-translational modifications into one protein simultaneously. *ACS Synth. Biol.* 7, 689–695. doi: 10.1021/acssynbio.7b00408
- Virdee, S., Kapadnis, P. B., Elliott, T., Lang, K., Madrzak, J., Nguyen, D. P., et al. (2011). Traceless and site-specific ubiquitination of recombinant proteins. *J. Am. Chem. Soc.* 133, 10708–10711. doi: 10.1021/ja202799r
- Virdee, S., Ye, Y., Nguyen, D. P., Komander, D., and Chin, J. W. (2010). Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nat. Chem. Biol.* 6, 750–757. doi: 10.1038/nchembio.426
- Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P. J., Russell, D. H., et al. (2010). A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angew. Chem. Int. Ed Engl.* 49, 3211–3214. doi: 10.1002/anie.201000465
- Wang, J., Xie, J., and Schultz, P. G. (2006a). A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.* 128, 8738–8739. doi: 10.1021/ja062666k
- Wang, K., Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2007). Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* 25, 770–777. doi: 10.1038/nbt1314
- Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001). Expanding the genetic code of *Escherichia coli*. *Science* 292, 498–500. doi: 10.1126/science.1060077
- Wang, L., Xie, J., and Schultz, P. G. (2006b). Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* 35, 225–249. doi: 10.1146/annurev.biophys.35.101105.121507
- Wang, Y., and Tsao, M. L. (2016). Reassigning sense codon AGA to encode noncanonical amino acids in *Escherichia coli*. *Chembiochem* 17, 2234–2239. doi: 10.1002/cbic.201600448
- Welegedara, A. P., Adams, L. A., Huber, T., Graham, B., and Otting, G. (2018). Site-specific incorporation of selenocysteine by genetic encoding as a photocaged unnatural amino acid. *Bioconjug. Chem.* 29, 2257–2264. doi: 10.1021/acs.bioconjchem.8b00254
- Wilding, K. M., Smith, A. K., Wilkerson, J. W., Bush, D. B., Knotts, T. A., and Bundy, B. C. (2018). The locational impact of site-specific PEGylation: streamlined screening with cell-free protein expression and coarse-grain simulation. *ACS Synth. Biol.* 7, 510–521. doi: 10.1021/acssynbio.7b00316
- Willis, J. C. W., and Chin, J. W. (2018). Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem.* 10, 831–837. doi: 10.1038/s41557-018-0052-5
- Xiao, H., and Schultz, P. G. (2016). At the interface of chemical and biological synthesis: an expanded genetic code. *Cold Spring Harb. Perspect. Biol.* 8:a023945. doi: 10.1101/cshperspect.a023945
- Yamagishi, Y., Ashigai, H., Goto, Y., Murakami, H., and Suga, H. (2009). Ribosomal synthesis of cyclic peptides with a fluorogenic oxidative coupling reaction. *Chembiochem* 10, 1469–1472. doi: 10.1002/cbic.200900021
- Yin, Y., Fei, Q., Liu, W., Li, Z., Suga, H., and Wu, C. (2019). Chemical and ribosomal synthesis of topologically controlled bicyclic and tricyclic peptide scaffolds primed by selenoether formation. *Angew. Chem. Int. Ed Engl.* 58, 4880–4885. doi: 10.1002/anie.201813827
- Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., and Yanagawa, H. (2003). DNA display for *in vitro* selection of diverse peptide libraries. *Nucleic Acids Res.* 31:e118. doi: 10.1093/nar/gng119
- Yu, Y., Hu, C., Xia, L., and Wang, J. Y. (2018). Artificial metalloenzyme design with unnatural amino acids and non-native cofactors. *ACS Catal.* 8, 1851–1863. doi: 10.1021/acscatal.7b03754
- Zawada, J. F., Yin, G., Steiner, A. R., Yang, J., Naresh, A., Roy, S. M., et al. (2011). Microscale to manufacturing scale-up of cell-free cytokine production—a new approach for shortening protein production development timelines. *Biotechnol. Bioeng.* 108, 1570–1578. doi: 10.1002/bit.23103
- Zhang, M. S., Brunner, S. F., Huguenin-Dezot, N., Liang, A. D., Schmed, W. H., Rogerson, D. T., et al. (2017). Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat. Methods* 14, 729–736. doi: 10.1038/nmeth.4302
- Zhang, Y., Ptacin, J. L., Fischer, E. C., Aerni, H. R., Caffaro, C. E., San Jose, K., et al. (2017). A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* 551, 644–647. doi: 10.1038/nature24659
- Zimmerman, E. S., Heibeck, T. H., Gill, A., Li, X., Murray, C. J., Madlansacay, M. R., et al. (2014). Production of site-specific antibody-drug conjugates using optimized non-natural amino acids in a cell-free expression system. *Bioconjug. Chem.* 25, 351–361. doi: 10.1021/bc400490z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cui, Johnston and Alexandrov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.