



# Genetic Code Engineering by Natural and Unnatural Base Pair Systems for the Site-Specific Incorporation of Non-Standard Amino Acids Into Proteins

Michiko Kimoto\* and Ichiro Hirao\*

Institute of Bioengineering and Bioimaging (IBB), Agency for Science, Technology and Research (A\*STAR), Singapore, Singapore

## OPEN ACCESS

### Edited by:

Nediljko Budisa,  
University of Manitoba, Canada

### Reviewed by:

Nikolaj Georg Koch,  
Technical University of Berlin,  
Germany  
Lluís Ribas De Pouplana,  
Institute for Research in Biomedicine,  
Spain

### \*Correspondence:

Michiko Kimoto  
michiko@ibb.a-star.edu.sg  
Ichiro Hirao  
ichiro@ibb.a-star.edu.sg

### Specialty section:

This article was submitted to  
Protein Biochemistry for Basic and  
Applied Sciences,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 10 January 2022

**Accepted:** 25 April 2022

**Published:** 24 May 2022

### Citation:

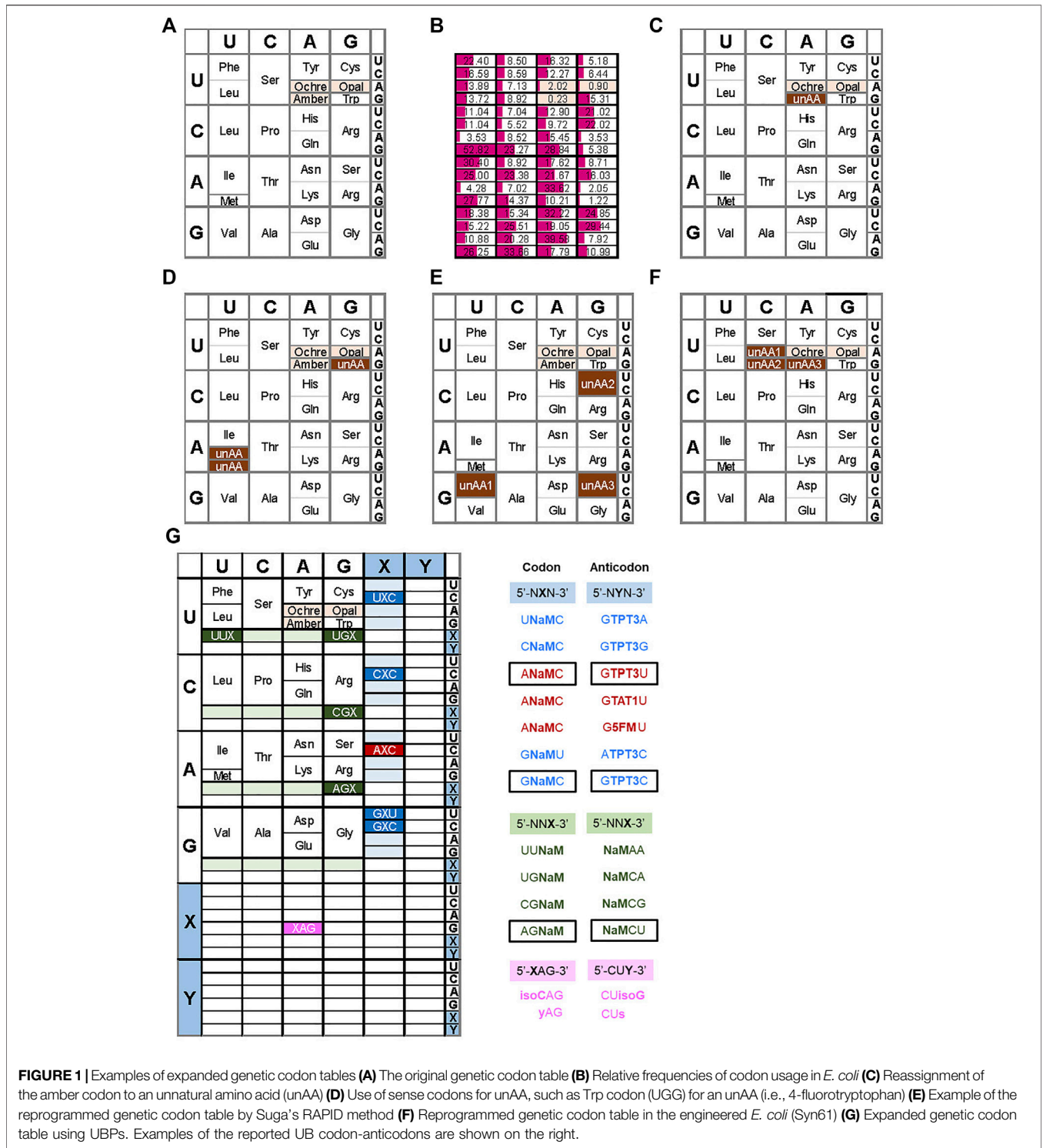
Kimoto M and Hirao I (2022) Genetic  
Code Engineering by Natural and  
Unnatural Base Pair Systems for the  
Site-Specific Incorporation of Non-  
Standard Amino Acids Into Proteins.  
Front. Mol. Biosci. 9:851646.  
doi: 10.3389/fmolb.2022.851646

Amino acid sequences of proteins are encoded in nucleic acids composed of four letters, A, G, C, and T(U). However, this four-letter alphabet coding system limits further functionalities of proteins by the twenty letters of amino acids. If we expand the genetic code or develop alternative codes, we could create novel biological systems and biotechnologies by the site-specific incorporation of non-standard amino acids (or unnatural amino acids, unAAs) into proteins. To this end, new codons and their complementary anticodons are required for unAAs. In this review, we introduce the current status of methods to incorporate new amino acids into proteins by *in vitro* and *in vivo* translation systems, by focusing on the creation of new codon-anticodon interactions, including unnatural base pair systems for genetic alphabet expansion.

**Keywords:** translation, genetic alphabet expansion, unnatural amino acid, genetic code expansion, unnatural base pair

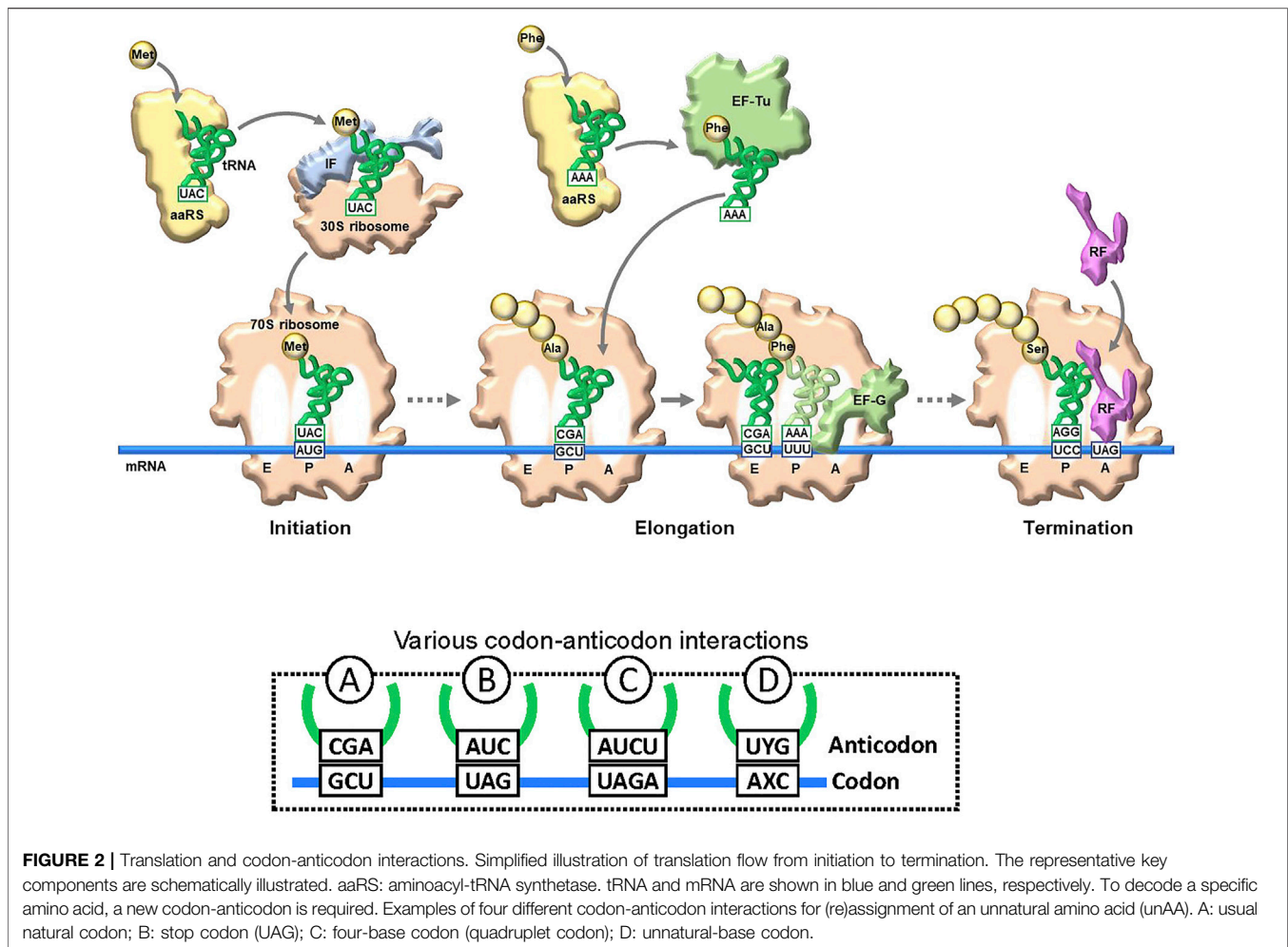
## INTRODUCTION

The genetic code on earth is ruled by the combinations of three consecutive base sequences as codons corresponding to each amino acid to construct proteins. Sixty-four codons composed of four letters, A, G, C, and T(U), are assigned to the twenty letters of standard amino acids and the three termination signals (stop codons) in translation (**Figure 1A**). Living organisms maintain the integrity of nucleic acids and proteins within the constraints of the four natural bases and twenty standard amino acids, respectively, by the evolutionary equilibrium between precise information flow through the cognate base pairings, A–T and G–C, and mutations through non-cognate mispairings. However, the limited chemical and biological diversity of these canonical components restricts further improvement toward the development of increased functionalities of nucleic acids and proteins and their biosystems. In fact, living organisms use a wide variety of modified nucleotides and non-standard amino acids (Ambrogelly et al., 2007). For example, D-amino acids, instead of the standard L-amino acids, often appear in peptides and proteins (Heck et al., 1994; Kreil, 1994; Kreil, 1997). Modified nucleotides produced by posttranscriptional modifications of tRNAs increase the efficiency and fidelity of the near cognate codon-anticodon interactions (Agris et al., 2007; Vare et al., 2017; Koh and Sarin, 2018). Therefore, artificially introducing unnatural bases (UBs) and unnatural amino acids (unAAs) into nucleic acids and proteins could increase their functionalities by expanding the genetic alphabet, and thus lead to the creation of newly engineered organisms.



The artificial incorporation of unAAs into proteins through the natural base pair (NBP) or unnatural base pair (UBP) systems requires an orthogonal “bypassing” system for the pre-existing genetic information flow in the central dogma: replication, transcription, and translation. Living organisms have evolved mechanisms to avoid the misincorporation (non-cognate) events

of UBs and unAAs and remove these extra components as errors during nucleic acid and protein biosynthesis. Accordingly, in nature, most of the unnatural components in biopolymers are introduced by post-biosynthesis modifications or other biosynthetic mechanisms. To circumvent the proofreading systems of living organisms, researchers have created several



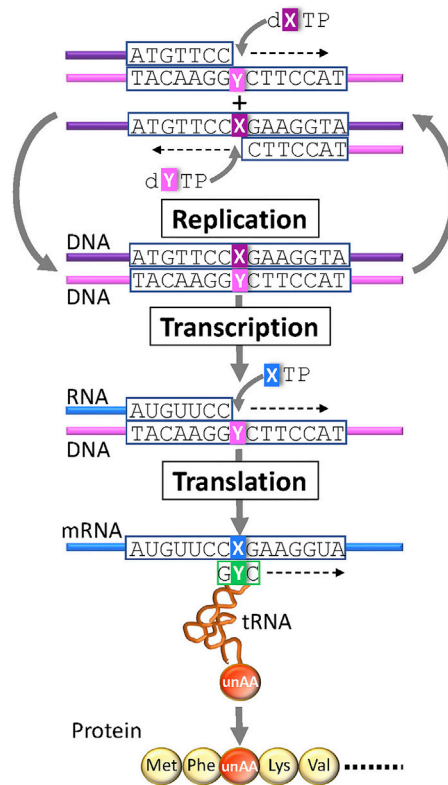
“bypassing” schemes by modifying the genetic information flow systems, including the codon table.

Genetic code engineering based on the NBP system has a long research history. In translation, there are several checkpoints for unAA incorporation into proteins (**Figure 2**). A specific aminoacyl-tRNA synthetase (aaRS) is required for esterifying the unAA to a specific tRNA to generate the unAA-tRNA. Namely, an orthogonal pair of an unAA and its aaRS must be created. The unAA-tRNA should be recognized by elongation factor Tu (EF-Tu), which binds specifically to the aminoacyl-tRNA. Ribosomes must catalyze protein synthesis by incorporating the unAA, using the unAA-tRNA as a substrate. Over the past few decades, several methods to bypass these checkpoints have been developed, for genetic code expansion systems using the existing NBP system, such as the use of stop codons (Type B in **Figure 2** and **Figure 1C**), four-base codon-anticodon interactions (Type C in **Figure 2**), and sense codon reprogramming (**Figures 1D–F**), for unAA incorporation into proteins. Importantly, these NBP methods are also employed in UBP systems, and these NBP and UBP systems could potentially be complementary to each other

for the further advancement of novel translation systems involving unAA incorporation.

In the last quarter century, several UBPs that function as a third base pair in replication, transcription, and/or translation have been developed (**Figure 3**) (Benner et al., 2016; Kimoto and Hirao, 2020; Manandhar et al., 2021). DNAs containing UBPs are amplified and transcribed to RNA by polymerases. UBPs also create novel codon-anticodon interactions involving new letters, enabling the site-specific incorporation of unAAs into proteins by ribosome-mediated translation. Additional UBPs could largely expand the existing codon table and theoretically make 152 additional new codons [ $216 (= 6 \times 6 \times 6) - 64 (= 4 \times 4 \times 4)$ ] in a six-letter UB system for multiple unAA incorporations (**Figure 1G** and type D in **Figure 2**).

In this review, we will introduce the methods to create new codon-anticodon interactions for the site-specific incorporation of unAAs into proteins, using NBP and UBP systems. Basic methods for employing the NBP system to expand the codon table will be briefly mentioned. For details of the related topics, including unAA-aminoacylations of tRNAs and eukaryotic translation systems, refer to these reviews (Young and Schultz,



**FIGURE 3** | Genetic alphabet expansion using an unnatural base pair (UBP) system for genetic code expansion. A third base pair (X–Y) that functions in replication, transcription, and translation, together with the natural A–T(U) and G–C base pairs, enables the site-specific incorporation of unnatural X and Y nucleotides and unnatural amino acids (unAAs) into nucleic acids and proteins.

2010; Budisa, 2013; Lajoie et al., 2016; Agostini et al., 2017; Mukai et al., 2017; Kubyshkin et al., 2018; Melnikov and Soll, 2019; Chung et al., 2020; De la Torre and Chin, 2021). Since the UBP systems are relatively new, we will describe them in detail. Finally, we will also discuss the combination of the UBP and NBP systems and future perspectives.

## GENETIC CODE EXPANSION USING THE NBP SYSTEM

First, we will briefly introduce the genetic code expansion using the NBP system, which includes the use of stop codons, four-base codon-anticodon interactions, and sense codon reprogramming in prokaryotic systems.

**Use of stop codons:** The most common method for site-specific unAA incorporation is the use of stop codons. As a specific case, archaea and eukaryotes also use stop codons for the incorporation of selenocysteine and pyrrolysine into proteins, using suppressor tRNAs. There are three stop codons, amber (UAG), ochre (UAA), and opal (UGA). Among them, amber is the most popular codon for unAA incorporation (Figure 1C and type B in Figure 2), because it has the lowest frequency as a stop codon in *Escherichia coli*. The stop codon usage in *E. coli* K12 is 7% UAG, 64% UAA, and 29% UGA (Figure 1B) (Openwetware, 2012).

Interestingly, the amber codon usage is also low in other organisms, but it is especially low in *E. coli* (Belin and Puigbo, 2022).

There are two methods for the unAA-aminoacylation of the suppressor tRNA with the CUA anticodon corresponding to the UAG amber codon. One is enzymatic ligation between a suppressor tRNA without the 3'-CA sequence and a chemically synthesized dinucleotide, pCA, which is aminoacylated with an unAA at the 3'-terminus (Heckler et al., 1984). This method has been used in *in vitro* translation systems for site-specific unAA incorporation into proteins (Bain et al., 1989; Noren et al., 1989). The other method is the use of a specific tRNA and aaRS pair, which can be employed in both *in vitro* and *in vivo* translation systems. Some of the tRNA-aaRS pairs are specific in each archaeon, prokaryote, and eukaryote, and tRNA-aaRS engineering studies revealed that they can potentially be used as orthogonal pairs in archaea or eukaryotes and assigned for unAAs in bacterial translation systems for unAA incorporation (Wang et al., 2001; Melnikov and Soll, 2019). For example, a tyrosyl-tRNA<sub>CUA</sub> variant and its aaRS from *Methanocaldococcus jannaschii* and a pyrrolysyl-tRNA<sub>CUA</sub> variant and its aaRS from *Methanosarcina barkeri* have been used as representative orthogonal pairs for unAA-tRNAs in *E. coli* and eukaryotic translation systems (Steer and Schimmel, 1999; Wang et al., 2001; Wang and Schultz, 2001; Chin et al., 2002a; Nguyen et al., 2009; Syed et al., 2019).

The issue with using stop codons is that the suppressor tRNA competes with release factors (see the left side in **Figure 7A**). When the ribosome reaches the stop codon on the mRNA, one of the release factors (RFs) binds to the A site for the ribosome dissociation from the mRNA (**Figure 2**). For example, RF1 recognizes UAG and competitively inhibits the binding of unAA-tRNA<sub>CUA</sub> to UAG in mRNA, reducing the efficiency of unAA-incorporation into proteins.

To address this issue, Ueda's team developed the PURE system (Protein synthesis Using Recombinant Element system), in which ribosomes, tRNAs, and translation factors isolated from *E. coli* are mixed for *in vitro* translation (Shimizu et al., 2001). By removing RF1 from the recombinant system, the UAG codon becomes free to encode an unAA, which increases the translation efficiency. Only UGA and UAA, which are recognized by RF2, are employed as the stop codons in the system.

Another method for *in vivo* translation was developed by removing RF1 from living organisms. Sakamoto's team created an organism (RF1-deficient *E. coli* strain, RF<sub>ZERO</sub>) by replacing seven essential UAG amber codons (Mukai et al., 2010), and then further improved the strategy, by replacing 95 of the 273 UAG codons in *E. coli* with UAA or UGA stop codons (Mukai et al., 2015a). In the strain, UAG codons are used for the site-specific incorporation of unAAs. Isaacs' team replaced all of the UAG codons at 321 positions in the *E. coli* genome with UAA, and constructed a genomically recoded organism (GRO), the C321.ΔA strain (Lajoie et al., 2013a; Lajoie et al., 2013b). Interestingly, in addition to efficient unAA incorporation into proteins, the GRO exhibited increased resistance to bacteriophage T7 infection.

**Four-base codon-anticodon interactions:** As a codon alternative, Sisido's team developed a four-base codon system, instead of the natural three-base codon system (type C in **Figure 2**) (Hohsaka et al., 1996; Murakami et al., 1998). In the system, unAA-tRNAs contain four-base anticodons corresponding to the four-base codons in mRNA, and the four-base codon-anticodon interactions function in ribosome-mediated translation. The problem is that the existing tRNA<sub>XYZ</sub> competes with the four-base anticodon tRNA<sub>XYZW</sub> and *vice versa*. To address this issue, they first chose AGG as the four-base codon, because the AGG codon for arginine is the least used codon in *E. coli* (AGG: 2%, AGA: 4%, CGG: 10%, CGA: 6%, CGU: 38%, and CGC: 40% for arginine) (**Figure 1B**). In addition, they embedded a stop codon, such as UAA, in the following frame-shifted position (i.e., AGGUCGU-AAU) (see the left side in **Figure 7B**). If the AGGU codon was undesirably used by tRNA<sub>CCU</sub>, then the translation would pause at the stop codon (i.e., AGGUCG-UAAU).

In subsequent experiments, they found that GGGU exhibits the most efficient translation efficiency among the four-base codon contexts (Hohsaka and Sisido, 2002). Using two four-base codons, AGGU and CGGG, they succeeded in the site-specific incorporations of two unAAs into streptavidin (Hohsaka et al., 1999). Hohsaka's team achieved the site-specific labeling of proteins by using dye-conjugated amino acids as unAAs by the four-base codon system (Abe et al., 2010). Schultz's team comprehensively examined the translation efficiency of the

four-codon system and found the best four-codon contexts, AGGA, UAGA, CCCU, and CUAG (Magliery et al., 2001).

Toward *in vivo* translation systems combining the four-base codon system and the amber codon suppression, improved orthogonal pairs of aaRSs and tRNAs with four-base anticodons were developed (Magliery et al., 2001; Chatterjee et al., 2012). Chin's team evolved a ribosome (ribo-Q1) that efficiently decodes a series of four-base and amber codons to increase the multiple incorporations of unAAs in *E. coli*. Using this ribo-Q1 system, including AGGA and UAG codons and their orthogonal tRNA-aaRS pairs, they performed the site-specific incorporation of two clickable unAA pairs, azide- and alkyne-containing amino acids, allowing for the cyclization of the generated proteins (Neumann et al., 2010). These amino acids are encoded by only one codon (Met: AUG; Trp: UGG), facilitating the replacement of these sense codons with unAAs.

**Reprogramming sense codons:** Historically, the reassignment of sense codons was first reported for unAA incorporations, in which auxotrophic bacteria were starved for one natural amino acid and supplemented with an unAA. Cowie and Cohen replaced methionine with selenomethionine, using an *E. coli* methionine auxotroph (Cowie and Cohen, 1957). Wong reported a variant of tryptophan-auxotrophic *Bacillus subtilis* using 4-fluorotryptophan, instead of tryptophan, which was created by gradually decreasing tryptophan and increasing 4-fluorotryptophan in the culture medium (Wong, 1983). Tryptophan is coded with only one codon (UGG) (**Figure 1D**).

As in the use of the amber codon, rare codons in a synonymous codon family for the same amino acids are useful for unAA assignment as a 21st amino acid. As mentioned above, the rare AGG arginine codon in *E. coli* was used as an unAA codon (Zeng et al., 2014). Sakamoto's team developed an *E. coli* system to incorporate L-homoarginine into proteins, using an engineered pair of an archaeal pyrrolysyl-tRNA synthetase and tRNA<sub>CCU</sub><sup>Pyl</sup> for the unAA (Mukai et al., 2015b). Furthermore, they replaced AGG codons in essential genes with other synonymous Arg codons for an efficient *in vivo* unAA translation system in *E. coli*.

Another rare codon, AUA, a sense codon (**Figure 1B**), has also been suggested for unAA incorporation (Bohlke and Budisa, 2014). The AUA codon in *E. coli* is recognized by tRNA<sup>Leu</sup> with a modified LAU anticodon (L: lysidine (2-lysyl-cytidine)), enabling L to pair with A in the codon (Suzuki and Miyauchi, 2010). Since this modification is catalyzed by TilS (Soma et al., 2003), the unmodified tRNA with CAU does not recognize the AUA codon, and thus could be used for an unAA in a TilS-depleted *E. coli* strain (**Figure 1D**).

The AUG codon is also a candidate for the sense codon reassignment for unAA incorporation (De Simone et al., 2016) (**Figure 1D**). The methionine codon AUG is used in two tRNAs: initiator tRNA<sup>Met</sup> for the initiation codon and elongator tRNA<sup>Met</sup> for internal AUG codons. By eliminating the elongator tRNA<sup>Met</sup> from a methionine auxotrophic *E. coli* strain, the introduction of a heterologous MetRS-tRNA<sup>Met</sup> pair from the archaeon *Sulfolobus acidocaldarius* to the system allows the incorporation of methionine analogs into proteins. The initiator tRNA<sup>Met</sup> could also be used for unAA incorporation at the N-terminal position of proteins.

In *in vitro* translation systems to reassign sense codons, the PURE system is useful to specifically remove the endogenous tRNA and aaRS for each unAA. A representative method for multiple unAA incorporations is the FIT (Flexible *In-vitro* Translation) system developed by Suga's team (Goto and Suga, 2009; Torikai and Suga, 2014). For example, they assigned GUU/C, CGU/C, and GGU/C codons to three different unAAs for the preparation of 23-letter proteins (Figure 1E) (Iwane et al., 2016). Their system also used ribozymes called Flexizymes for unAA-aminoacylation of tRNAs (Ohuchi et al., 2007; Passioura et al., 2014). Flexizymes were generated by an *in vitro* selection method using RNA libraries and an activated unAAs (Lee et al., 2000; Saito and Suga, 2001; Passioura and Suga, 2014). By applying the FIT system to ribosome display methods, they developed a platform system (RaPID, Random non-standard Peptide Integrated Discovery) to generate functional cyclic peptides from peptide libraries containing several unAAs (Passioura et al., 2014). Recently, they established a system for multiple incorporations of  $\beta$ -amino acids into peptides (Katoh et al., 2020), in which the tRNAs were engineered by modifying the T-stem and D-arm to increase the binding affinity to EF-Tu (Iwane et al., 2021). In this system, they used AUU/C, CAU/C, and UGU/C codons for  $\beta$ -amino acids, as well as AUG for unAAs, to promote the cyclization of the generated peptides. To generate stabilized inhibitor peptides, they recently reported a successful screening using a random peptide library with aromatic cyclic  $\beta^{2,3}$  amino acids, prepared by ribosomal incorporation (Katoh and Suga, 2022).

In the area of codon reprogramming, Chin's team established another GRO system by the total synthesis of the *E. coli* genome with defined synonymous codon compression (Wang et al., 2016; Fredens et al., 2019). They designed and synthesized the 4-Mb *E. coli* genome, in which 18,214 codons for two serine codons (TCA and TCG) and the TAG amber codon in all of the genes were replaced with AGC, AGT, and TAA, respectively. Furthermore, the genes encoding tRNA<sup>Ser</sup><sub>UGA</sub>, tRNA<sup>Ser</sup><sub>CGA</sub>, and RFI were also removed from the genome. Therefore, the synthesized *E. coli* (Syn61) uses 61 codons, and the three vacant codons can be used for three unAA codons (Figure 1F).

## GENETIC CODE EXPANSION USING UBP SYSTEMS

**Development of UBP systems *in vitro*:** In 1962, Alexander Rich proposed the potential use of a UBP, isoguanine (isoG) and isocytosine (isoC) (Figure 4) with different hydrogen-bonding patterns from those of G–C, for a new codon-anticodon system (Rich, 1962). Even though the codon table was still being deciphered at that time, he imagined that two-base genetic codons, instead of three-base codons, using six-letter genetic alphabets could cover the 20 standard amino acids ( $20 < 6 \times 6$ ). Over 2 decades later, in the late 1980s, Benner's team designed several UBPs with alternative hydrogen bonding patterns, including the isoG–isoC pair, and chemically synthesized these UB units. Their biological results opened a new world in which the UBPs could be used for replication and transcription, with

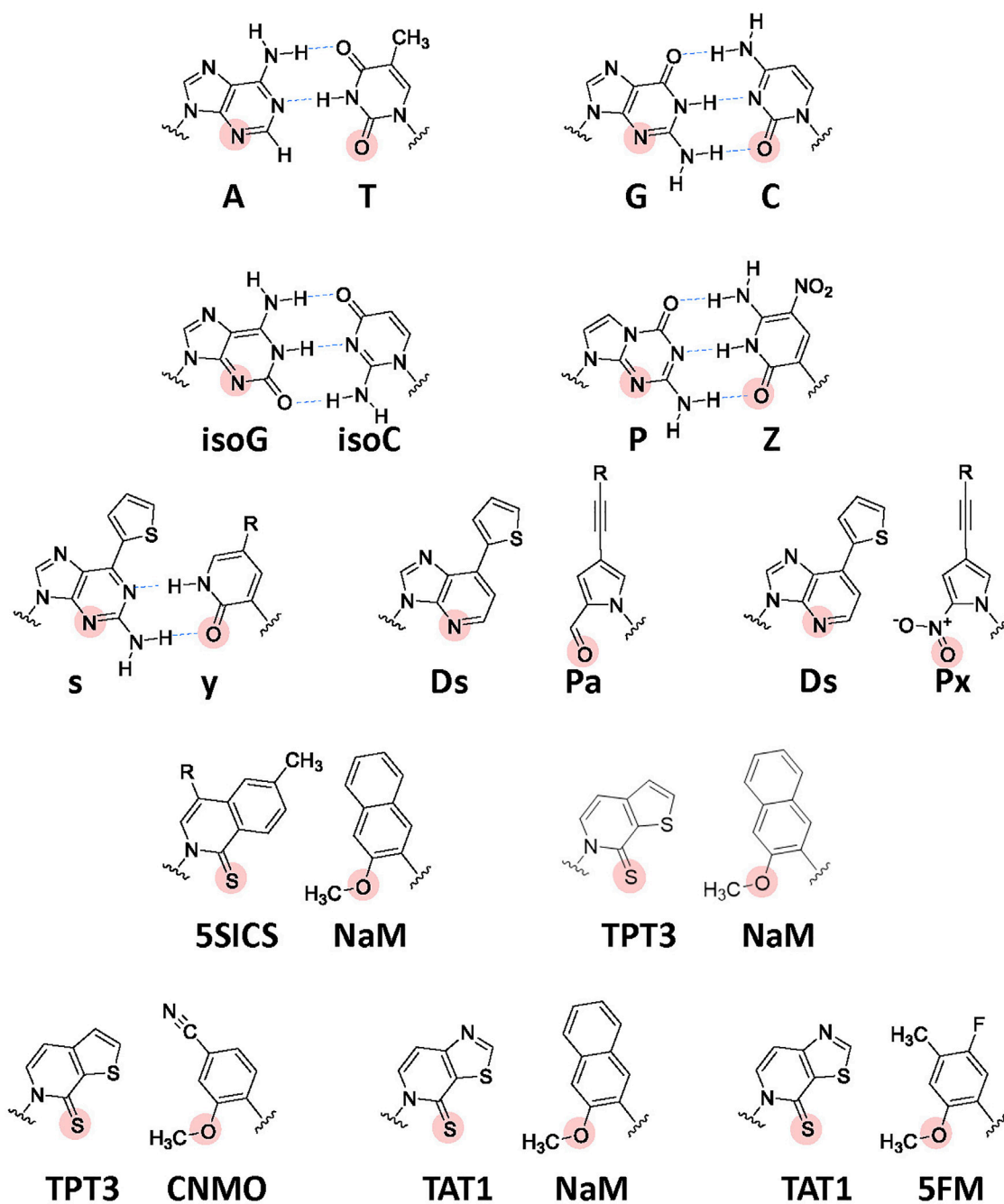
orthogonal base pairing to the two natural base pairs (Switzer et al., 1989; Piccirilli et al., 1990). In 1992, they reported an *in vitro* translation system for the site-specific incorporation of 3-iodotyrosine into a peptide, using chemically synthesized mRNA with an (isoC)AG codon and tRNA with a CU(isoG) anticodon (Figure 1G) (Bain et al., 1992). Their efforts toward further UBP development and optimization led to the replicable and transcribable P–Z pair, with higher fidelity than those of the isoG–isoC pair (Figure 4) (Yang et al., 2010; Yang et al., 2011).

In the late 1990s, Romesberg's and Hirao's teams also started to develop UBPs, based on different concepts, toward practical applications to increase the functionalities of nucleic acids and proteins beyond the canonical four-letter biological system. More than 20 years on, several representative UBPs have become applicable in replication, transcription, or translation *in vitro* and/or *in vivo*, including the s–y and Ds–Pa/Ds–Px pairs from Hirao's team and the NaM–5SICS/NaM–TPT3/CNMO–TPT3/NaM–TAT1 pairs from Romesberg's team (Figure 4) (Hirao et al., 2002; Hirao and Kimoto, 2012; Malyshev and Romesberg, 2015; Kimoto and Hirao, 2020; Manandhar et al., 2021).

Hirao's team developed the s–y pair, which functions as a third base pair in transcription (Figure 4). The bulky thienyl group in the s base eliminates its pairing with the natural bases, and the y substrate (yTP) is site-specifically incorporated into RNA opposite s in the DNA template by T7 RNA polymerase (Fujiwara et al., 2001; Hirao et al., 2002). The s–y pair was applied to an *in vitro* transcription-translation system for the incorporation of unAAs into a specific position of the 185-aa Ras protein. Using an 863-mer DNA template containing s and 3-chlorotyrosyl-tRNA<sub>CU<sub>s</sub></sub> (ClTyr-tRNA<sub>CU<sub>s</sub></sub>), they coupled the T7 transcription with *in vitro* translation using an *E. coli* cell-free system. The LC-MS analysis of the obtained protein confirmed that the yAG codon in the transcribed *ras* mRNA was decoded by the CU<sub>s</sub> anticodon of ClTyr-tRNA<sub>CU<sub>s</sub></sub> (Figure 1G). Although the yAG codon was decoded by the native Lys-tRNA<sub>UUU</sub> and Gln-tRNA<sub>CUG</sub> in the absence of ClTyr-tRNA<sub>CU<sub>s</sub></sub>, the undesired misincorporation was competitively suppressed by the predominant ClTyr incorporation by ClTyr-tRNA<sub>CU<sub>s</sub></sub> (Hirao et al., 2002). Even though the yAG codon is closely related to the UAG amber codon, the translation experiments without ClTyr-tRNA<sub>CU<sub>s</sub></sub> revealed that the replacement of one of the bases in the termination codons with an unnatural base bypasses the competition with release factors (see Figure 7A).

In the translation system, tRNA<sub>CU<sub>s</sub></sub> was prepared by ligation of the 5'-half fragment derived from the native *Saccharomyces cerevisiae* tRNA<sup>Tyr</sup> with the chemically synthesized 3'-half fragment containing CU<sub>s</sub> (Ohtsuki et al., 1996). The aminoacylation of tRNA<sub>CU<sub>s</sub></sub> with ClTyr was performed by *S. cerevisiae* tyrosyl-tRNA synthetase, which does not recognize the third anticodon position (Chow and Rajbhandary, 1993; Tsunoda et al., 2007), the s position, and aminoacylates *S. cerevisiae* tRNA<sup>Tyr</sup> with tyrosine and tyrosine analogs, such as 3-halotyrosine and DOPA, under specific conditions in the presence of 20% dimethyl sulfoxide and 0.25% Tween-20. In addition, the *S. cerevisiae* tRNA<sup>Tyr</sup> is not aminoacylated by *E. coli* tRNA synthetase.

Hirao's team subsequently developed the hydrophobic Ds–Pa/Px pairs, which exhibit high fidelity in replication and



**FIGURE 4** | Chemical structures of the natural Watson-Crick base pairs and a variety of UBPs developed to expand the genetic alphabet. Hydrogen-bonding interactions between the cognate base pairs are shown by blue arrows. The important residues (hydrogen acceptors) recognized by DNA and RNA polymerases are indicated by solid circles. R: further modification is available by attaching a variety of functional groups via linkers.

transcription, by removing the hydrogen-bonding interactions between pairing bases (Hirao et al., 2006; Hirao et al., 2007; Kimoto et al., 2009; Yamashige et al., 2012). The Ds-Px pair was applied to high-affinity DNA aptamer generation, thus demonstrating how unnatural components greatly increase nucleic acid functionalities (Kimoto et al., 2013; Matsunaga et al., 2017; Futami et al., 2019; Matsunaga et al., 2021).

**Development of UBP systems *in vivo*:** Romesberg's team also developed a series of hydrophobic UBPs, such as NaM-5SICS and NaM-TPT3 (Figure 4), with high fidelity in replication and transcription (Malyshev et al., 2009; Seo et al., 2009; Seo et al., 2011; Malyshev et al., 2012; Lavergne et al., 2013). In 2014, using the NaM-5SICS/TPT3 pairs, Romesberg's team successfully created an engineered *E. coli* strain (Figure 5) (Semi-synthetic

organism, SSO), which replicates six-letter DNA, using their NaM–5SCIS/TPT3 pairs (Malyshev et al., 2014). To supply the UB substrates for the living cells, they employed media supplemented with UB-nucleoside triphosphates, dNaMTP and d5SCISTP. To facilitate the import of sufficient amounts of the UB substrates within the cells, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (*PtNTT2*) was expressed in *E. coli* [C41 (DE3) strain]. They prepared a plasmid DNA containing the NaM–TPT3 pair through PCR amplification and transformed it into the engineered *E. coli* expressing *PtNTT2*. An analysis of the cultured cells revealed the successful replication of the six-letter plasmid DNA with reasonable retention of the NaM–5SCIS pair. These results also demonstrated that the UBP was not extensively rejected as a foreign component by the repair system.

They further improved this first generation SSO by switching from the original *E. coli* strain to the BL21 (DE3) strain, which is more suitable for protein expression. The *PtNTT2* expression was also optimized because the extremely high expression using T7 RNA polymerase inhibited cell growth. They modified *PtNTT2* expression by 1) removing the N-terminal signal peptide (65 aa) of *PtNTT2*, which is toxic to cell growth, 2) using the optimized codon usage, 3) choosing the best RNA polymerase II promoter sequence, and 4) encoding the truncated *PtNTT2* gene within a *lacZYA* locus in the genome, to avoid expression plasmid copy number variations. The resultant second-generation SSO, called the YZ3 strain, greatly increased the retention rates of their UBPs in replicated DNA within various sequence contexts (Zhang et al., 2017a).

In 2017, their team reported successful *in vivo* transcription and translation using the YZ3 strain, to decode the AXC or GXC codon (X = NaM) by the corresponding GYU or GYC anticodon (Y = TPT3), thus enabling the site-specific incorporation of unAAs into a recombinant superfolder green fluorescent protein (sfGFP) (Figures 1G, 4) (Zhang et al., 2017b). The YZ3 cells carrying a specific aaRS expression plasmid were additionally transformed by the plasmid DNA containing the NaM–TPT3 pairs, encoding mRNA (X) and tRNA (Y). The induced T7 RNA polymerase expression in the SSO allowed successful T7 transcription of UB-containing mRNA and tRNA, and finally yielded the GFP with an unAA at the AXC or GXC codon position, decoded by unAA-tRNA<sub>GYU</sub> or unAA-tRNA<sub>GYC</sub>.

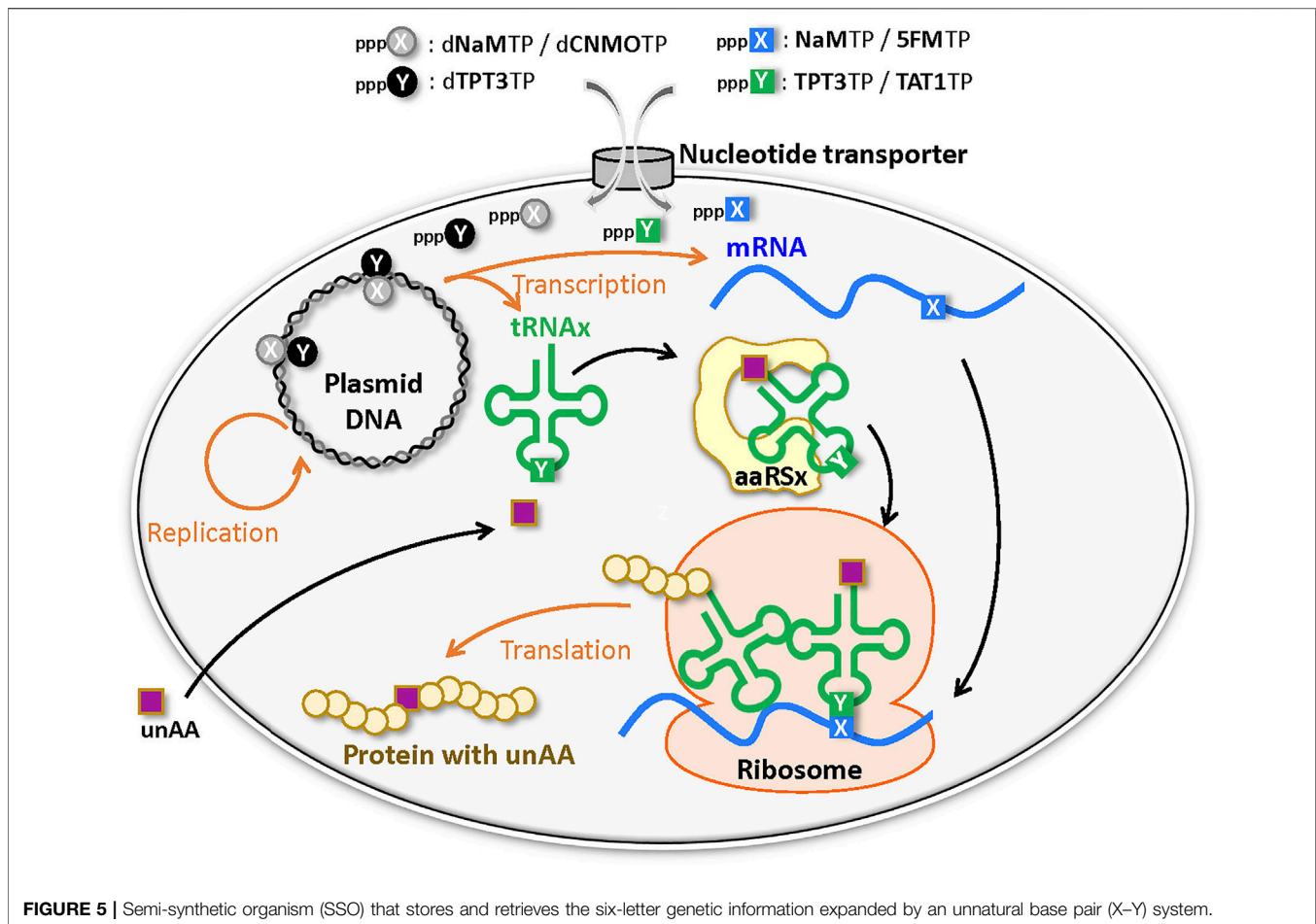
To evaluate the decoding of the new codon-anticodon interactions involving the NaM–TPT3 pair, they first investigated the incorporation of serine into GFP at position 151 (the TAC codon was replaced by the unnatural codon AXC) through *E. coli* tRNA<sup>Ser</sup> (*serT*), where the anticodon was replaced by the unnatural codon GYT. This system can eliminate complicated situations related to unAAs because *E. coli* serine aaRS does not recognize the anticodon for tRNA amino acylation (Shimizu et al., 1992). The efficient full-length sfGFP production with  $98.5 \pm 0.7\%$  incorporation of serine at position 151 was confirmed in the cells transformed with the plasmid encoding both sfGFP(AXC)<sup>151</sup> and tRNA<sup>Ser</sup><sub>GYT</sub>, cultured in the presence of deoxy- and ribo-nucleoside triphosphates (dNaMTP, dTPT3TP, NaMTP, and TPT3TP).

After the validation of the unnatural codon-anticodon interactions, they focused on unAA N<sup>6</sup>-[(2-propynyl)oxy] carbonyl]-L-lysine (PrK) incorporation in sfGFP(AXC)<sup>151</sup> or sfGFP(GXC)<sup>151</sup>, utilizing a pair of the *Methanosarcina mazei* tRNA<sub>GYU</sub><sup>Pyl</sup> or RNA<sub>GYC</sub><sup>Pyl</sup> and the *Methanosarcina barkei* pyrrolysyl-tRNA synthetase (PylRS) (Nguyen et al., 2009; Chatterjee et al., 2013). The PylRS was encoded in a separate plasmid with expression controlled by IPTG induction. For another unAA p-azido-phenylalanine (*pAzF*) in sfGFP(AXC)<sup>151</sup>, they utilized an evolved *Methanococcus jannaschii* TyrRS/tRNA<sup>Tyr</sup> pair (i.e. *pAzFRS.tRNA*<sub>GYU</sub><sup>pAzF</sup>) (Chin et al., 2002b).

By assessing the UBP retention in plasmid DNAs, the incorporation efficiencies of unAAs, and the cell growth, Romesberg's team has further optimized the second-generation SSO. First, they examined the *in vivo* replication mechanisms for UBPs and found that the elimination of RecA and the release of DNA Pol II from SOS repression increased the UBP retention. This study resulted in the third generation SSO with an error-avoidance mechanism, called the ML2 strain [BL21 (DE3) *lacZYA::PtNTT2(66-575) ΔrecA polB<sup>++</sup>*] (Ledbetter et al., 2018). Next, they explored a variety of UB substrate analogs for DNA replication and RNA transcription *in vivo*. They identified the CNMO–TPT3 pair, which is superior to the NaM–TPT3 pair for efficient *in vivo* DNA replication, and the 5FM–TPT3 and NaM–TAT1 pairs, which are better for the efficient production of GFP with an unAA. The optimized SSO with the dCNMOTP–dTPT3TP/NaMTP–TAT1TP system efficiently produced the GFP with three proximal unAAs, using the AXC-GYU codon-anticodon interaction (Feldman et al., 2019).

Using the ML2 strain, Romesberg's team further identified new codons for efficient production of proteins with unAAs. In 2021, they reported that twenty UB codons are available: seven of the NXN-NYN codon-anticodon interactions (X = NaM, Y = TPT3; including UXC-GYA, CXC-GYG, AXC-GYU, GXU-AYC, GXC-GYC in clonal SSOs) and thirteen of the NNX-XNN interactions (X = NaM; including UUX-XAA, UGX-XGA, CGX-XCG, AGX-XCU in clonal SSOs) (Figure 1G) (Fischer et al., 2020; Romesberg, 2021). Interestingly, only NaM, and not TPT3, is acceptable for the codon in the second and third positions, and the third position should be the self NaM–NaM pair, rather than the hetero NaM–TPT3 pair. In addition, they confirmed that at least three of the codon-anticodon interactions, AXC-GYU, GXC-GYC, and AGX-XCU, are orthogonal to each other, allowing for simultaneous decoding in the SSO (Fischer et al., 2020). By measuring the transcription fidelity *in vivo*, they found that the decoding at the ribosome is more sensitive than transcription (Zhou et al., 2020). This might be because the variable codon performance is the total output of the sequence-dependent translation efficiency, although the recognition of the codon-anticodon interaction might differ in eukaryotic cells (Zhou et al., 2019). They are currently exploring the recognition of the NaM–TPT3 pair by the multi-subunit *E. coli* RNA polymerase II, as compared to the single-subunit T7 RNA polymerase, to create next generation SSOs (Hashimoto et al., 2021; Oh et al., 2021).



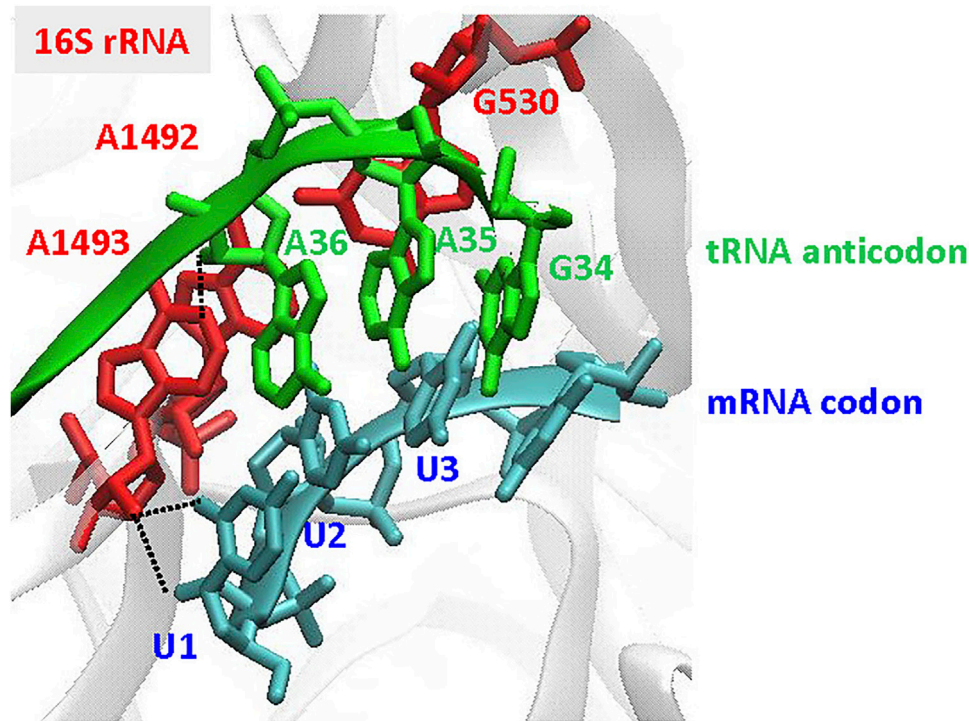


**FIGURE 5 |** Semi-synthetic organism (SSO) that stores and retrieves the six-letter genetic information expanded by an unnatural base pair (X–Y) system.

**Recognition of the codon-anticodon interaction in the bacterial ribosome:** For accurate discrimination between cognate and near- or non-cognate aa-tRNAs, the three highly conserved G530, A1492, and A1493 bases in 16S rRNA are prerequisite (Figure 6). To select the cognate tRNA through the minor groove interactions of the codon-anticodon interaction, the two adenine bases are flipped out from the internal loop of helix 44 of 16S rRNA in the 30S ribosomal subunit (Ogle et al., 2001; Ramakrishnan, 2002; Còchella et al., 2007). At the first position of the codon-anticodon, A1493 forms a type I A-minor motif interaction both the O2' and N3 of A1493 are located in the minor groove of the first position, maximizing the number of hydrogen bonds that can be formed (Nissen et al., 2001; Ogle et al., 2001). At the second position, A1492 and G530 are tightly packed in the minor groove of the codon-anticodon (a type II A-minor motif interaction) (Nissen et al., 2001; Ogle et al., 2001), but do not directly interact with the base moieties. Thus, the second position would accommodate small structural differences in the base pair (Fischer et al., 2020). The third position has more open space and is less monitored by the ribosome.

Recent *in vitro* translation studies using RNA nucleobase derivatives in the mRNA demonstrated that the hydrogen-bonding interaction between the N1 of purines and the N3 of

pyrimidines is sufficient for decoding at the first or second position. At the third “wobble” position, an adequate stacking force, not limited to the hydrogen-bonding interactions, could be essential (Hoernes et al., 2018). The UB studies clearly demonstrated that the acceptance of the hydrogen-bonded isoC-isoG and y-s pairs at the first codon-anticodon position might be reasonable, since their UB pairing structures effectively mimic the natural Watson-Crick base pairing geometry. In contrast, the non-hydrogen-bonded NaM-TPT3 pair at the first position might adopt a cross-strand intercalated structure (Manandhar et al., 2021), as found in the free DNA duplex form. The UB is quite different from the Watson-Crick like structure found in the polymerase active site (Betz et al., 2012; Betz et al., 2013), which might prevent recognition as a cognate base pair. Interestingly, although the previous UB study suggested that at least the hydrogen-bonding interaction between the N1 of purines and the N3 of pyrimidines at the second position is required for the decoding, the NaM-TPT3 pair is well accepted as cognate. Together with the acceptance of the self NaM-NaM pair at the third position, these results indicate that complementary packing and hydrophobic forces can “bypass” the requirement for precise decoding at the second and third positions (Hoernes et al., 2018). However, the reason why only NaM, and not TPT3, is acceptable for mRNA remains unclear.



**FIGURE 6** | Recognition of the codon-anticodon interaction at the ribosome. The illustration is focused on the decoding site of the 30S subunit, showing the A-site codon (UUU in blue) and the tRNA anticodon (GAA in green), using the coordinates in PDB: 1IBM (Ogle et al., 2001). The important nucleotides, G530, A1492, and A1493, in 16S rRNA are indicated in red. Type I A-minor motif interactions, found in A1493 at the first codon-anticodon position (U1 A36), are indicated by black dotted lines.

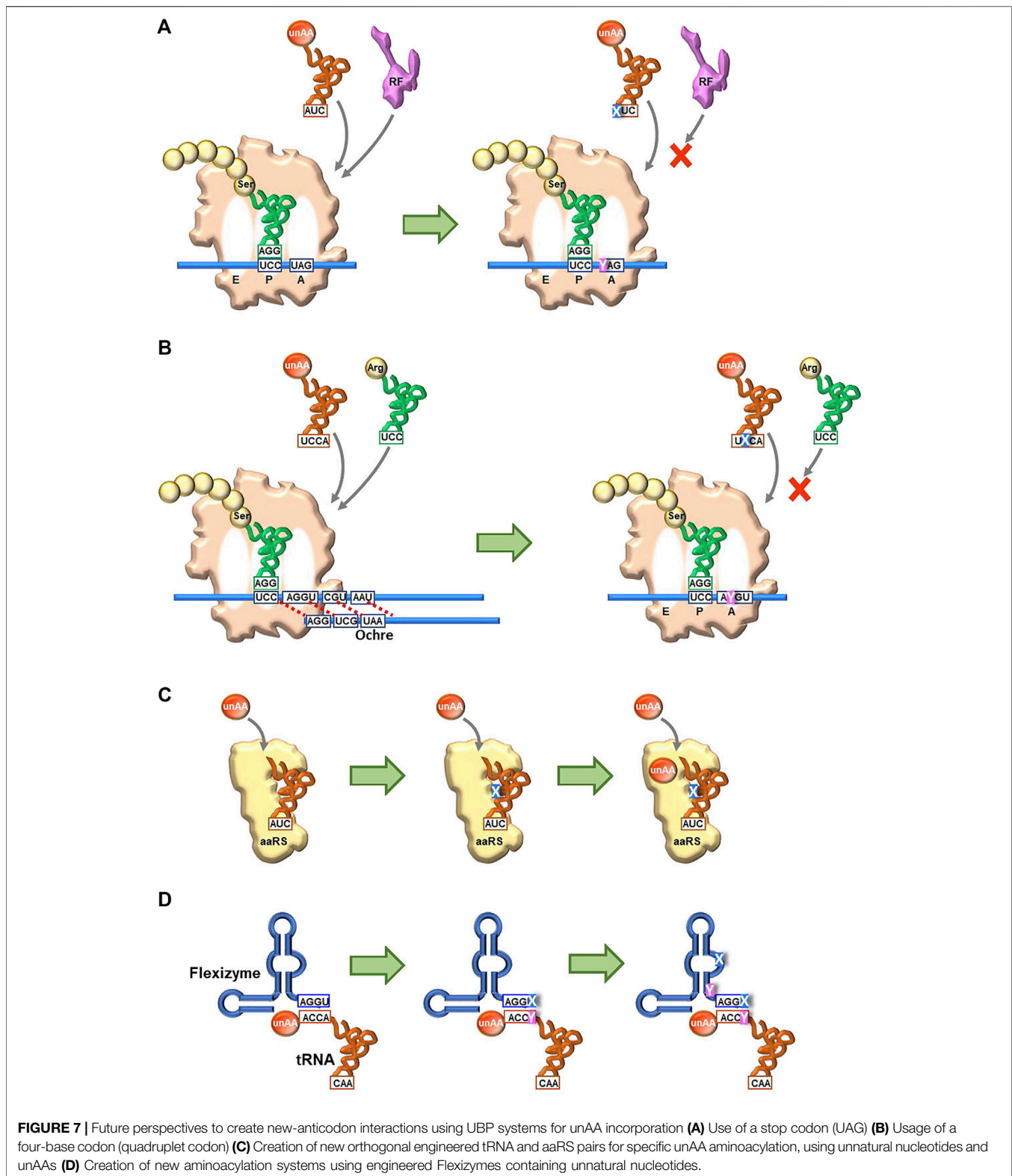
The discrimination mechanisms of cognate and non-cognate pairs by DNA and RNA polymerases may be similar to those of the decoding process. Both polymerases and the ribosome 30S subunit undergo a structural rearrangement from an open to a closed form in the cognate pairing, through interactions with the minor groove of the Watson-Crick base pairing (Ogle et al., 2002). Further detailed translation studies using other UBPs might elucidate the unknown mechanisms and driving forces by which the RNA-based decoding system precisely discriminates the cognate and non-cognate pairing, commonly and/or differently from those in polymerases.

## THERAPEUTIC APPLICATIONS OF ENGINEERED PROTEINS BY GENETIC CODE EXPANSION

These translation systems by genetic code expansion have facilitated the rational design and optimization of proteins suitable for therapeutic applications, by improving the biological functions and pharmacokinetics of biologics in a manner resembling the pursuit of small-molecule therapeutics. Several macrocyclic peptides containing unAAs, including those generated by the ribosomal translation system, are now undergoing clinical tests (Vinogradov et al., 2019). Currently, several engineered proteins generated by these

technologies are in pre-clinical and clinical trials as protein therapeutics. Representatives are, but not limited to, PEGylated interleukin-2 (SAR44425, THOR-707) (Manandhar et al., 2021), PEGylated fibroblast growth factor 21 (BMS-986036, pegbelfermin), Anti-HER2 antibody-drug conjugate (a site-specific Herceptin-monomethyl auristatin D (MMAD) conjugate, ARX788) (Sun et al., 2014), and anti-CD3 Folate Bi-Specific (Sun et al., 2014).

The recombinant human cytokine interleukin-2 (rhIL-2, or aldesleukin) was originally approved as a drug for immune oncology targeting renal cell carcinoma (Klapper et al., 2008; Krieg et al., 2010). However, rhIL-2/aldesleukin therapy, targeting the stimulation of tumor immune responses through CD8<sup>+</sup> effector T and natural-killer cells, which express the IL-2 receptor beta and gamma subunit complex (IL-2 Rβγ), has been limited due to the short half-life and off-target effects resulting from its interaction with the IL-2 receptor alpha subunit (IL-2 Rα). Scientists at Synthorx (founded by Romesberg in 2014, acquired by Sanofi in 2019) used Romesberg's SSO (YZ3 strain) to successfully identify a suitable PEGylated position (P65) in IL-2 from 10 candidates (K35, R38, T41, F42, K43, Y45, F62, P65, E68, and V69) and addressed the above two issues. These analyses resulted in the development of THOR-707, the IL-2 compound with an unAA at position 65, followed by further modification with a 30 kDa



mPEG, which retained the binding ability to IL-2 R $\beta$  but lacked that to the undesired IL-2 R $\alpha$  $\beta$ , and showed an extended half-life (Manandhar et al., 2021; Ptacin et al., 2021; Romesberg, 2021).

THOR-707 is currently in a phase I/II study, not only as a monotherapeutic, but also in combination with a checkpoint inhibitor (pembrolizumab or cemiplimab) (Romesberg, 2021).

## CONCLUSION AND FUTURE PERSPECTIVE

Continuous and comprehensive research on genetic alphabet rearrangement and expansion technologies has largely improved the unAA incorporation fidelity and efficiency and created new organisms. Even hydrophobic UBPs without any clear hydrogen-bond interactions between pairing bases can function as new letters of DNA and RNA, for information storage and retrieval in *in vivo* systems (SSOs). Recent breakthroughs in UBP development as a third base pair have created novel genetic alphabet systems of DNA and RNA, providing the expanded codon table in translation, which can bypass the checkpoints of the native translation system. The new UB-codons related to the stop codons, such as isoCAG, yAG, and UGNaM, predominantly interact with their UB-anticodons and prevent the interaction with RF (Figure 7A).

The UBP systems could improve the current NBP genetic code expansion systems. For example, the introduction of UBs into four-base codon systems might prevent the competition with the native tRNAs with three-base anticodons (Figure 7B). As shown in Romesberg's results, embedding the UB in the middle of a codon (for example, AXC and GXC) would maximize the UB's discrimination capabilities in decoding. A novel pair of tRNA and aaRS for unAAs could be created by introducing UBs into tRNAs and unAAs into aaRSs (Figure 7C) (Young and Schultz, 2018), as UB-containing nucleic acid aptamers significantly increase the affinities and specificities with target proteins (Kimoto et al., 2013). Such novel pairs would enhance the simultaneous incorporation of different multiple unAAs, as well as the known tRNA/aaRS pairs (Chin et al., 2003; Brustad et al., 2008; Tanrikulu et al., 2009; Italia et al., 2017; Melnikov and Soll, 2019; Ding et al., 2020). Flexizymes could also improve the efficiency and specificity of tRNA aminoacylation by introducing UBs, although the UBP applications to ribozymes have not yet been reported. The current Flexizymes recognize tRNAs by the interaction between the terminal GGU sequence of Flexizyme and the terminal ACCA sequence of tRNAs, and thus aminoacylate tRNAs non-specifically (Figure 7D). One possible improvement would be the introduction of a UBP (X-Y) to the terminal position (GGX) of Flexizyme and to the discriminator base (YCC) in tRNAs. Further research for the introduction of UBs (UBPs) and unAAs into biopolymers (DNA, RNA, and protein) would yield not only fruitful findings in translation mechanisms but also novel protein therapeutics, empowered by the cooperative fusion of chemistry and biology. Furthermore, the UBP-unAA systems have the potential to create novel organisms with increased functionalities, such as enhanced productivity of useful materials and highly sensitive sensors for detection. Thus, the combination of the NBP and UBP systems could further expand

the capability of genetic code engineering for multiple unAA incorporations.

Currently, only the UBPs developed by Romesberg's team have been demonstrated in the *in vivo* system. The potentials of other UBPs, such as Z-P and Ds-Px, for *in vivo* systems are still unknown. The fidelity and toxicity of UBPs and UB materials and the limitation of the number of UB-codons available in SSOs are important issues. Romesberg's team demonstrated that the use of a Cas9-based editing system allowed the increased retention (fidelity in replication) of their UBPs in living cells (Zhang et al., 2017a), but the additionally expressed sgRNAs might interfere with efficient translation. Although *in vitro* studies revealed that the replication fidelities of some UBPs are more than 99.8% per duplication, there is still room for further improvement of the specificity and stability of UBPs to reduce the mutation rates in replication, transcription, and translation. The toxicity of continuously supplementing unnatural base substrates as a third base pair for long term cultures and the possible increase in mutations have not been fully elucidated. The current UBP translation systems have mainly been studied in prokaryotic systems. In the future, UBP studies will be expanded to eukaryotic systems (Zhou et al., 2019) and provide further information and possibilities.

UBP research and its applications have only just begun. Nevertheless, the improvements of UBP systems have opened the door to novel biotechnologies, as described here. Replication with non-hydrogen bonded UBPs in the *E. coli* genome has also become an achievable target (Ledbetter et al., 2018). However, as compared to the NBP system, the utility of the current UBP systems is still limited due to the comparatively lower fidelity and efficiency with some sequence biases, which increase the mutation rates and change the evolutionary equilibrium of the system and SSOs. As for the codon-anticodon interactions, non-hydrogen bonded UBPs are less stable, but can be used in translation within the limits of the codon usage. Extensive studies using UBPs and modified nucleotides, including the further development of UBPs and the combination with NBP systems, might reveal the unknown secrets of the current life system.

## AUTHOR CONTRIBUTIONS

Both MK and IH wrote the manuscript and figures by introducing their ideas.

## FUNDING

This work was supported by the Institute of Bioengineering and Bioimaging (Biomedical Research Council, Agency for Science, Technology and Research, Singapore).

## REFERENCES

- Abe, R., Shiraga, K., Ebisu, S., Takagi, H., and Hohsaka, T. (2010). Incorporation of Fluorescent Non-natural Amino Acids into N-Terminal Tag of Proteins in Cell-free Translation and its Dependence on Position and Neighboring Codons. *J. Biosci. Bioeng.* 110, 32–38. doi:10.1016/j.jbiosc.2010.01.003
- Agostini, F., Völler, J. S., Kokschi, B., Acevedo-Rocha, C. G., Kubyskhin, V., and Budisa, N. (2017). Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chem. Int. Ed.* 56, 9680–9703. doi:10.1002/anie.201610129

- Agris, P. F., Vendeix, F. A. P., and Graham, W. D. (2007). tRNA's Wobble Decoding of the Genome: 40 Years of Modification. *J. Mol. Biol.* 366, 1–13. doi:10.1016/j.jmb.2006.11.046
- Ambrogelly, A., Palioura, S., and Söll, D. (2007). Natural Expansion of the Genetic Code. *Nat. Chem. Biol.* 3, 29–35. doi:10.1038/nchembio847
- Bain, J. D., Diala, E. S., Glabe, C. G., Dix, T. A., and Chamberlin, A. R. (1989). Biosynthetic Site-specific Incorporation of a Non-natural Amino Acid into a Polypeptide. *J. Am. Chem. Soc.* 111, 8013–8014. doi:10.1021/ja00202a052
- Bain, J. D., Switzer, C., Chamberlin, 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
- Belin, D., and Puigbò, P. (2022). Why Is the UAG (Amber) Stop Codon Almost Absent in Highly Expressed Bacterial Genes? *Life (Basel)* 12, 431. doi:10.3390/life12030431
- Benner, S. A., Karalkar, N. B., Hoshika, S., Laos, R., Shaw, R. W., Matsuura, M., et al. (2016). Alternative Watson-Crick Synthetic Genetic Systems. *Cold Spring Harb. Perspect. Biol.* 8, a023770. doi:10.1101/cshperspect.a023770
- Betz, K., Malyshev, D. A., Lavergne, T., Welte, W., Diederichs, K., Dwyer, T. J., et al. (2012). KlenTaq Polymerase Replicates Unnatural Base Pairs by Inducing a Watson-Crick Geometry. *Nat. Chem. Biol.* 8, 612–614. doi:10.1038/nchembio.966
- Betz, K., Malyshev, D. A., Lavergne, T., Welte, W., Diederichs, K., Romesberg, F. E., et al. (2013). Structural Insights into DNA Replication without Hydrogen Bonds. *J. Am. Chem. Soc.* 135, 18637–18643. doi:10.1021/ja409609j
- Bohlke, N., and Budisa, N. (2014). Sense Codon Emancipation for Proteome-wide Incorporation of Noncanonical Amino Acids: Rare Isoleucine Codon AUA as a Target for Genetic Code Expansion. *FEMS Microbiol. Lett.* 351, 133–144. doi:10.1111/1574-6968.12371
- Brustad, E., Bushey, M. L., Brock, A., Chittuluru, J., and Schultz, P. G. (2008). A Promiscuous Aminoacyl-tRNA Synthetase that Incorporates Cysteine, Methionine, and Alanine Homologs into Proteins. *Bioorg. Med. Chem. Lett.* 18, 6004–6006. doi:10.1016/j.bmcl.2008.09.050
- Budisa, N. (2013). Expanded Genetic Code for the Engineering of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products (RiPPs). *Curr. Opin. Biotechnol.* 24, 591–598. doi:10.1016/j.copbio.2013.02.026
- Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013). A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia coli*. *Biochemistry* 52, 1828–1837. doi:10.1021/bi4000244
- Chatterjee, A., Xiao, H., and Schultz, P. G. (2012). Evolution of Multiple, Mutually Orthogonal Prolyl-tRNA Synthetase/tRNA Pairs for Unnatural Amino Acid Mutagenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14841–14846. doi:10.1073/pnas.1212454109
- Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003). An Expanded Eukaryotic Genetic Code. *Science* 301, 964–967. doi:10.1126/science.1084772
- Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002a). Addition of a Photocrosslinking Amino Acid to the Genetic Code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11020–11024. doi:10.1073/pnas.172226299
- Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002b). Addition of P-Azido-L-Phenylalanine to the Genetic Code of *Escherichia coli*. *J. Am. Chem. Soc.* 124, 9026–9027. doi:10.1021/ja027007w
- Chow, C. M., and Rajbhandary, U. L. (1993). Saccharomyces cerevisiae Cytoplasmic Tyrosyl-tRNA Synthetase Gene. Isolation by Complementation of a Mutant *Escherichia coli* Suppressor tRNA Defective in Aminoacylation and Sequence Analysis. *J. Biol. Chem.* 268, 12855–12863. doi:10.1016/s0021-9258(18)31466-2
- Chung, C. Z., Amikura, K., and Söll, D. (2020). Using Genetic Code Expansion for Protein Biochemical Studies. *Front. Bioeng. Biotechnol.* 8, 598577. doi:10.3389/fbioe.2020.598577
- Cochella, L., Brunelle, J. L., and Green, R. (2007). Mutational Analysis Reveals Two Independent Molecular Requirements during Transfer RNA Selection on the Ribosome. *Nat. Struct. Mol. Biol.* 14, 30–36. doi:10.1038/nsmb1183
- Cowie, D. B., and Cohen, G. N. (1957). Biosynthesis by *Escherichia coli* of Active Altered Proteins Containing Selenium Instead of Sulfur. *Biochimica Biophysica Acta* 26, 252–261. doi:10.1016/0006-3002(57)90003-3
- De La Torre, D., and Chin, J. W. (2021). Reprogramming the Genetic Code. *Nat. Rev. Genet.* 22, 169–184. doi:10.1038/s41576-020-00307-7
- De Simone, A., Acevedo-Rocha, C. G., Hoesl, M. G., and Budisa, N. (2016). Towards Reassignment of the Methionine Codon AUG to Two Different Noncanonical Amino Acids in Bacterial Translation. *Croat. Chem. Acta* 89, 243–253. doi:10.5562/cca2915
- Ding, W., Zhao, H., Chen, Y., Zhang, B., Yang, Y., Zang, J., et al. (2020). Chimeric Design of Pyrrolysyl-tRNA Synthetase/tRNA Pairs and Canonical Synthetase/tRNA Pairs for Genetic Code Expansion. *Nat. Commun.* 11, 3154. doi:10.1038/s41467-020-16898-y
- Feldman, A. W., Dien, V. T., Karadeema, R. J., Fischer, E. C., You, Y., Anderson, B. A., et al. (2019). Optimization of Replication, Transcription, and Translation in a Semi-synthetic Organism. *J. Am. Chem. Soc.* 141, 10644–10653. doi:10.1021/jacs.9b02075
- 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
- 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
- Fujiwara, T., Kimoto, M., Sugiyama, H., Hirao, I., and Yokoyama, S. (2001). Synthesis of 6-(2-thienyl)purine Nucleoside Derivatives that Form Unnatural Base Pairs with Pyridin-2-One Nucleosides. *Bioorg. Med. Chem. Lett.* 11, 2221–2223. doi:10.1016/s0960-894x(01)00415-2
- Futami, K., Kimoto, M., Lim, Y. W. S., and Hirao, I. (2019). Genetic Alphabet Expansion Provides Versatile Specificities and Activities of Unnatural-Base DNA Aptamers Targeting Cancer Cells. *Mol. Ther. Nucleic Acids* 14, 158–170. doi:10.1016/j.omtn.2018.11.011
- 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
- Hashimoto, K., Fischer, E. C., and Romesberg, F. E. (2021). Efforts toward Further Integration of an Unnatural Base Pair into the Biology of a Semisynthetic Organism. *J. Am. Chem. Soc.* 143, 8603–8607. doi:10.1021/jacs.1c03860
- Heck, S. D., Siok, C. J., Krapcho, K. J., Kelbaugh, P. R., Thadeio, P. F., Welch, M. J., et al. (1994). Functional Consequences of Posttranslational Isomerization of Ser 46 in a Calcium Channel Toxin. *Science* 266, 1065–1068. doi:10.1126/science.7973665
- Heckler, T. G., Chang, L. H., Zama, Y., Naka, T., Chorghade, M. S., and Hecht, S. M. (1984). T4 RNA Ligase Mediated Preparation of Novel "chemically Misacylated" tRNAPheS. *Biochemistry* 23, 1468–1473. doi:10.1021/bi00302a020
- Hirao, I., Kimoto, M., Mitsui, T., Fujiwara, T., Kawai, R., Sato, A., et al. (2006). An Unnatural Hydrophobic Base Pair System: Site-specific Incorporation of Nucleotide Analogs into DNA and RNA. *Nat. Methods* 3, 729–735. doi:10.1038/nmeth915
- Hirao, I., and Kimoto, M. (2012). Unnatural Base Pair Systems toward the Expansion of the Genetic Alphabet in the Central Dogma. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 88, 345–367. doi:10.2183/pjab.88.345
- Hirao, I., Mitsui, T., Kimoto, M., and Yokoyama, S. (2007). An Efficient Unnatural Base Pair for PCR Amplification. *J. Am. Chem. Soc.* 129, 15549–15555. doi:10.1021/ja073830m
- Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., et al. (2002). An Unnatural Base Pair for Incorporating Amino Acid Analogs into Proteins. *Nat. Biotechnol.* 20, 177–182. doi:10.1038/nbt0202-177
- Hoernes, T. P., Faserl, K., Juen, M. A., Kremser, J., Gasser, C., Fuchs, E., et al. (2018). Translation of Non-standard Codon Nucleotides Reveals Minimal Requirements for Codon-Anticodon Interactions. *Nat. Commun.* 9, 4865. doi:10.1038/s41467-018-07321-8
- Hohsaka, T., Ashizuka, Y., Murakami, H., and Sisido, M. (1996). Incorporation of Nonnatural Amino Acids into Streptavidin through *In Vitro* Frame-Shift Suppression. *J. Am. Chem. Soc.* 118, 9778–9779. doi:10.1021/ja9614225
- Hohsaka, T., Ashizuka, Y., Sasaki, H., Murakami, H., and Sisido, M. (1999). Incorporation of Two Different Nonnatural Amino Acids Independently into a Single Protein through Extension of the Genetic Code. *J. Am. Chem. Soc.* 121, 12194–12195. doi:10.1021/ja992204p
- Hohsaka, T., and Sisido, M. (2002). Incorporation of Non-natural Amino Acids into Proteins. *Curr. Opin. Chem. Biol.* 6, 809–815. doi:10.1016/s1367-5931(02)00376-9

- Italia, J. S., Addy, P. S., Wrobel, C. J. J., Crawford, L. A., Lajoie, M. J., Zheng, Y., et al. (2017). An Orthogonalized Platform for Genetic Code Expansion in Both Bacteria and Eukaryotes. *Nat. Chem. Biol.* 13, 446–450. doi:10.1038/nchembio.2312
- 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
- Iwane, Y., Kimura, H., Katoh, T., and Suga, H. (2021). Uniform Affinity-Tuning of N-Methyl-Aminoacyl-tRNAs to EF-Tu Enhances Their Multiple Incorporation. *Nucleic Acids Res.* 49, 10807–10817. doi:10.1093/nar/gkab288
- Katoh, T., Sengoku, T., Hirata, K., Ogata, K., and Suga, H. (2020). Ribosomal Synthesis and De Novo Discovery of Bioactive Foldamer Peptides Containing Cyclic  $\beta$ -amino Acids. *Nat. Chem.* 12, 1081–1088. doi:10.1038/s41557-020-0525-1
- Katoh, T., and Suga, H. (2022). *In Vitro* Selection of Foldamer-like Macrocyclic Peptides Containing 2-Aminobenzoic Acid and 3-Aminothiophene-2-Carboxylic Acid. *J. Am. Chem. Soc.* 144, 2069–2072. doi:10.1021/jacs.1c12133
- Kimoto, M., and Hirao, I. (2020). Genetic Alphabet Expansion Technology by Creating Unnatural Base Pairs. *Chem. Soc. Rev.* 49, 7602–7626. doi:10.1039/d0cs00457j
- Kimoto, M., Kawai, R., Mitsui, T., Yokoyama, S., and Hirao, I. (2009). An Unnatural Base Pair System for Efficient PCR Amplification and Functionalization of DNA Molecules. *Nucleic Acids Res.* 37, e14. e14. doi:10.1093/nar/gkn956
- Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S., and Hirao, I. (2013). Generation of High-Affinity DNA Aptamers Using an Expanded Genetic Alphabet. *Nat. Biotechnol.* 31, 453–457. doi:10.1038/nbt.2556
- Klapper, J. A., Downey, S. G., Smith, F. O., Yang, J. C., Hughes, M. S., Kammula, U. S., et al. (2008). High-dose Interleukin-2 for the Treatment of Metastatic Renal Cell Carcinoma. *Cancer* 113, 293–301. doi:10.1002/cncr.23552
- Koh, C. S., and Sarin, L. P. (2018). Transfer RNA Modification and Infection - Implications for Pathogenicity and Host Responses. *Biochimica Biophysica Acta (BBA) - Gene Regul. Mech.* 1861, 419–432. doi:10.1016/j.bbagr.2018.01.015
- Kreil, G. (1994). Conversion of L- to D-Amino Acids: a Posttranslational Reaction. *Science* 266, 996–997. doi:10.1126/science.7973683
- Kreil, G. (1997). D-amino Acids in Animal Peptides. *Annu. Rev. Biochem.* 66, 337–345. doi:10.1146/annurev.biochem.66.1.337
- Krieg, C., Létourneau, S., Pantaleo, G., and Boyman, O. (2010). Improved IL-2 Immunotherapy by Selective Stimulation of IL-2 Receptors on Lymphocytes and Endothelial Cells. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11906–11911. doi:10.1073/pnas.1002569107
- Kubyshev, V., Acevedo-Rocha, C. G., and Budisa, N. (2018). On Universal Coding Events in Protein Biogenesis. *Biosystems* 164, 16–25. doi:10.1016/j.biosystems.2017.10.004
- Lajoie, M. J., Kosuri, S., Mosberg, J. A., Gregg, C. J., Zhang, D., and Church, G. M. (2013a). Probing the Limits of Genetic Recoding in Essential Genes. *Science* 342, 361–363. doi:10.1126/science.1241460
- Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., et al. (2013b). Genomically Recoded Organisms Expand Biological Functions. *Science* 342, 357–360. doi:10.1126/science.1241459
- Lajoie, M. J., Söll, D., and Church, G. M. (2016). Overcoming Challenges in Engineering the Genetic Code. *J. Mol. Biol.* 428, 1004–1021. doi:10.1016/j.jmb.2015.09.003
- Lavergne, T., Degardin, M., Malyshev, D. A., Quach, H. T., Dhami, K., Ordoukhanian, P., et al. (2013). Expanding the Scope of Replicable Unnatural DNA: Stepwise Optimization of a Predominantly Hydrophobic Base Pair. *J. Am. Chem. Soc.* 135, 5408–5419. doi:10.1021/ja312148q
- Ledbetter, M. P., Karadeema, R. J., and Romesberg, F. E. (2018). Reprogramming the Replisome of a Semisynthetic Organism for the Expansion of the Genetic Alphabet. *J. Am. Chem. Soc.* 140, 758–765. doi:10.1021/jacs.7b11488
- Lee, N., Bessho, Y., Wei, K., Szostak, J. W., and Suga, H. (2000). Ribozyme-catalyzed tRNA Aminoacylation. *Nat. Struct. Biol.* 7, 28–33. doi:10.1038/71225
- Magliery, T. J., Anderson, J. C., and Schultz, P. G. (2001). Expanding the Genetic Code: Selection of Efficient Suppressors of Four-Base Codons and Identification of "shifty" Four-Base Codons with a Library Approach in *Escherichia coli* Edited by M. Gottesman. *J. Mol. Biol.* 307, 755–769. doi:10.1006/jmbi.2001.4518
- Malyshev, D. A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J. M., et al. (2014). A Semi-synthetic Organism with an Expanded Genetic Alphabet. *Nature* 509, 385–388. doi:10.1038/nature13314
- Malyshev, D. A., Dhami, K., Quach, H. T., Lavergne, T., Ordoukhanian, P., Torkamani, A., et al. (2012). Efficient and Sequence-independent Replication of DNA Containing a Third Base Pair Establishes a Functional Six-Letter Genetic Alphabet. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12005–12010. doi:10.1073/pnas.1205176109
- Malyshev, D. A., and Romesberg, F. E. (2015). The Expanded Genetic Alphabet. *Angew. Chem. Int. Ed.* 54, 11930–11944. doi:10.1002/anie.201502890
- Malyshev, D. A., Seo, Y. J., Ordoukhanian, P., and Romesberg, F. E. (2009). PCR with an Expanded Genetic Alphabet. *J. Am. Chem. Soc.* 131, 14620–14621. doi:10.1021/ja906186f
- Manandhar, M., Chun, E., and Romesberg, F. E. (2021). Genetic Code Expansion: Inception, Development, Commercialization. *J. Am. Chem. Soc.* 143, 4859–4878. doi:10.1021/jacs.0c11938
- Matsunaga, K., Kimoto, M., and Hirao, I. (2017). High-affinity DNA aptamer generation targeting von Willebrand factor A1-domain by genetic alphabet expansion for systematic evolution of ligands by exponential enrichment using two types of libraries composed of five different bases. *J. Am. Chem. Soc.* 139, 324–334. doi:10.1021/jacs.6b10767
- Matsunaga, K., Kimoto, M., Lim, V. W., Tan, H. P., Wong, Y. Q., Sun, W., et al. (2021). High-affinity Five/six-Letter DNA Aptamers with Superior Specificity Enabling the Detection of Dengue NS1 Protein Variants beyond the Serotype Identification. *Nucleic Acids Res.* 49, 11407–11424. doi:10.1093/nar/gkab515
- Melnikov, S. V., and Söll, D. (2019). Aminoacyl-tRNA Synthetases and tRNAs for an Expanded Genetic Code: What Makes Them Orthogonal? *Int. J. Mol. Sci.* 20, 1929. doi:10.3390/ijms20081929
- Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., et al. (2010). Codon Reassignment in the *Escherichia coli* Genetic Code. *Nucleic Acids Res.* 38, 8188–8195. doi:10.1093/nar/gkq707
- Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., et al. (2015a). Highly Reproductive *Escherichia coli* Cells with No Specific Assignment to the UAG Codon. *Sci. Rep.* 5, 9699. doi:10.1038/srep09699
- Mukai, T., Lajoie, M. J., Englert, M., and Söll, D. (2017). Rewriting the Genetic Code. *Annu. Rev. Microbiol.* 71, 557–577. doi:10.1146/annurev-micro-090816-093247
- Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., et al. (2015b). Reassignment of a Rare Sense Codon to a Non-canonical Amino Acid in *Escherichia coli*. *Nucleic Acids Res.* 43, 8111–8122. doi:10.1093/nar/gkv787
- Murakami, H., Hohsaka, T., Ashizuka, Y., and Sisido, M. (1998). Site-directed Incorporation of P-Nitrophenylalanine into Streptavidin and Site-To-Site Photoinduced Electron Transfer from a Pyrenyl Group to a Nitrophenyl Group on the Protein Framework. *J. Am. Chem. Soc.* 120, 7520–7529. doi:10.1021/ja971890u
- 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., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A., and Chin, J. W. (2009). Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-tRNA Synthetase/tRNACUA Pair and Click Chemistry. *J. Am. Chem. Soc.* 131, 8720–8721. doi:10.1021/ja900553w
- Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001). RNA Tertiary Interactions in the Large Ribosomal Subunit: the A-Minor Motif. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4899–4903. doi:10.1073/pnas.081082398
- Noren, C. J., Anthony-Cahill, 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
- Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001). Recognition of Cognate Transfer RNA by the 30 S Ribosomal Subunit. *Science* 292, 897–902. doi:10.1126/science.1060612
- Ogle, J. M., Murphy, F. V., Tarry, M. J., and Ramakrishnan, V. (2002). Selection of tRNA by the Ribosome Requires a Transition from an Open to a Closed Form. *Cell* 111, 721–732. doi:10.1016/s0092-8674(02)01086-3
- Oh, J., Shin, J., Unarta, I. C., Wang, W., Feldman, A. W., Karadeema, R. J., et al. (2021). Transcriptional Processing of an Unnatural Base Pair by Eukaryotic RNA Polymerase II. *Nat. Chem. Biol.* 17, 906–914. doi:10.1038/s41589-021-00817-3
- Ohtsuki, T., Kawai, G., Watanabe, Y., Kita, K., Nishikawa, K., and Watanabe, K. (1996). Preparation of Biologically Active *Ascaris suum* Mitochondrial tRNAMet with a TV-Replacement Loop by Ligation of Chemically Synthesized RNA Fragments. *Nucleic Acids Res.* 24, 662–667. doi:10.1093/nar/24.4.662

- Ohuchi, M., Murakami, H., and Suga, H. (2007). The Flexizyme System: a Highly Flexible tRNA Aminoacylation Tool for the Translation Apparatus. *Curr. Opin. Chem. Biol.* 11, 537–542. doi:10.1016/j.cbpa.2007.08.011
- Openwetware (2012). *Escherichia coli*/Codon Usage [Online]. Available at: [https://openwetware.org/wiki/Escherichia\\_coli/Codon\\_usage](https://openwetware.org/wiki/Escherichia_coli/Codon_usage) (Accessed August 15, 2012).
- Passioura, T., and Suga, H. (2014). Flexizymes, Their Evolutionary History and Diverse Utilities. *Top. Curr. Chem.* 344, 331–345. doi:10.1007/128\_2013\_421
- Passioura, T., Katoh, T., Goto, Y., and Suga, H. (2014). Selection-based Discovery of Druglike Macrocyclic Peptides. *Annu. Rev. Biochem.* 83, 727–752. doi:10.1146/annurev-biochem-060713-035456
- Piccirilli, J. A., Benner, S. A., Krauch, T., Moroney, S. E., and Benner, S. A. (1990). Enzymatic Incorporation of a New Base Pair into DNA and RNA Extends the Genetic Alphabet. *Nature* 343, 33–37. doi:10.1038/343033a0
- Ptacin, J. L., Caffaro, C. E., Ma, L., San Jose Gall, K. M., Aerni, H. R., Acuff, N. V., et al. (2021). An Engineered IL-2 Reprogrammed for Anti-tumor Therapy Using a Semi-synthetic Organism. *Nat. Commun.* 12, 4785. doi:10.1038/s41467-021-24987-9
- Ramakrishnan, V. (2002). Ribosome Structure and the Mechanism of Translation. *Cell* 108, 557–572. doi:10.1016/s0092-8674(02)00619-0
- Rich, A. (1962). “Problems of Evolution and Biochemical Information Transfer,” in *Horizons Biochem.* Editors M. Kasha and B. Pullman (Academic Press), 103–126.
- Romesberg, F. E. (2022). Creation, Optimization, and Use of Semi-synthetic Organisms that Store and Retrieve Increased Genetic Information. *J. Mol. Biol.* 434, 167331. doi:10.1016/j.jmb.2021.167331
- Saito, H., and Suga, H. (2001). A Ribozyme Exclusively Aminoacylates the 3'-hydroxyl Group of the tRNA Terminal Adenosine. *J. Am. Chem. Soc.* 123, 7178–7179. doi:10.1021/ja015756s
- Seo, Y. J., Malyshev, D. A., Lavergne, T., Ordoukhanian, P., and Romesberg, F. E. (2011). Site-specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs. *J. Am. Chem. Soc.* 133, 19878–19888. doi:10.1021/ja207907d
- Seo, Y. J., Matsuda, S., and Romesberg, F. E. (2009). Transcription of an Expanded Genetic Alphabet. *J. Am. Chem. Soc.* 131, 5046–5047. doi:10.1021/ja9006996
- Shimizu, M., Asahara, H., Tamura, K., Hasegawa, T., and Himeno, H. (1992). The Role of Anticodon Bases and the Discriminator Nucleotide in the Recognition of Some E. coli tRNAs by Their Aminoacyl-tRNA Synthetases. *J. Mol. Evol.* 35, 436–443. doi:10.1007/BF00171822
- 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
- Soma, A., Ikeuchi, Y., Kanemasa, S., Kobayashi, K., Ogasawara, N., Ote, T., et al. (2003). An RNA-Modifying Enzyme that Governs Both the Codon and Amino Acid Specificities of Isoleucine tRNA. *Mol. Cell* 12, 689–698. doi:10.1016/s1097-2765(03)00346-0
- 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, S. B., Schultz, P. G., and Kim, C. H. (2014). Therapeutic Applications of an Expanded Genetic Code. *Chembiochem* 15, 1721–1729. doi:10.1002/cbic.201402154
- Suzuki, T., and Miyauchi, K. (2010). Discovery and Characterization of tRNAIlelysidine Synthetase (TilS). *FEBS Lett.* 584, 272–277. doi:10.1016/j.febslet.2009.11.085
- Switzer, C., Moroney, S. E., and Benner, S. A. (1989). Enzymatic Incorporation of a New Base Pair into DNA and RNA. *J. Am. Chem. Soc.* 111, 8322–8323. doi:10.1021/ja00203a067
- Syed, J., Palani, S., Clarke, S. T., Asad, Z., Bottrill, A. R., Jones, A. M. E., et al. (2019). Expanding the Zebrafish Genetic Code through Site-specific Introduction of Azido-Lysine, Bicyclononyne-Lysine, and Diazirine-Lysine. *Int. J. Mol. Sci.* 20, 2577. doi:10.3390/ijms20102577
- Tanrikulu, I. C., Schmitt, E., Mechulam, Y., Goddard, W. A., 3rd, and Tirrell, D. A. (2009). Discovery of *Escherichia coli* Methionyl-tRNA Synthetase Mutants for Efficient Labeling of Proteins with Azidonorleucine *In Vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15285–15290. doi:10.1073/pnas.0905735106
- Torikai, K., and Suga, H. (2014). Ribosomal Synthesis of an Amphotericin-B Inspired Macrocyclic. *J. Am. Chem. Soc.* 136, 17359–17361. doi:10.1021/ja508648s
- Tsunoda, M., Kusakabe, Y., Tanaka, N., Ohno, S., Nakamura, M., Senda, T., et al. (2007). Structural Basis for Recognition of Cognate tRNA by Tyrosyl-tRNA Synthetase from Three Kingdoms. *Nucleic Acids Res.* 35, 4289–4300. doi:10.1093/nar/gkm417
- Väre, V., Eruysal, E., Narendran, A., Sarachan, K., and Agris, P. (2017). Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. *Biomolecules* 7, 29. doi:10.3390/biom7010029
- Vinogradov, A. A., Yin, Y., and Suga, H. (2019). Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *J. Am. Chem. Soc.* 141, 4167–4181. doi:10.1021/jacs.8b13178
- Wang, K., Fredens, J., Brunner, S. F., Kim, S. H., Chia, T., and Chin, J. W. (2016). Defining Synonymous Codon Compression Schemes by Genome Recoding. *Nature* 539, 59–64. doi:10.1038/nature20124
- 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., and Schultz, P. G. (2001). A General Approach for the Generation of Orthogonal tRNAs. *Chem. Biol.* 8, 883–890. doi:10.1016/s1074-5521(01)00063-1
- Wong, J. T. (1983). Membership Mutation of the Genetic Code: Loss of Fitness by Tryptophan. *Proc. Natl. Acad. Sci. U.S.A.* 80, 6303–6306. doi:10.1073/pnas.80.20.6303
- Yamashige, R., Kimoto, M., Takezawa, Y., Sato, A., Mitsui, T., Yokoyama, S., et al. (2012). Highly Specific Unnatural Base Pair Systems as a Third Base Pair for PCR Amplification. *Nucleic Acids Res.* 40, 2793–2806. doi:10.1093/nar/gkr1068
- Yang, Z., Chen, F., Alvarado, J. B., and Benner, S. A. (2011). Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J. Am. Chem. Soc.* 133, 15105–15112. doi:10.1021/ja204910n
- Yang, Z., Chen, F., Chamberlin, S. G., and Benner, S. A. (2010). Expanded Genetic Alphabets in the Polymerase Chain Reaction. *Angew. Chem. Int. Ed.* 49, 177–180. doi:10.1002/anie.200905173
- Young, D. D., and Schultz, P. G. (2018). Playing with the Molecules of Life. *ACS Chem. Biol.* 13, 854–870. doi:10.1021/acscchembio.7b00974
- Young, T. S., and Schultz, P. G. (2010). Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J. Biol. Chem.* 285, 11039–11044. doi:10.1074/jbc.r109.091306
- Zeng, Y., Wang, W., and Liu, W. R. (2014). Towards Reassigning the Rare AGG Codon in *Escherichia coli*. *Chembiochem* 15, 1750–1754. doi:10.1002/cbic.201400075
- Zhang, Y., Lamb, B. M., Feldman, A. W., Zhou, A. X., Lavergne, T., Li, L., et al. (2017a). A Semisynthetic Organism Engineered for the Stable Expansion of the Genetic Alphabet. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1317–1322. doi:10.1073/pnas.1616443114
- Zhang, Y., Ptacin, J. L., Fischer, E. C., Aerni, H. R., Caffaro, C. E., San Jose, K., et al. (2017b). A Semi-synthetic Organism that Stores and Retrieves Increased Genetic Information. *Nature* 551, 644–647. doi:10.1038/nature24659
- Zhou, A. X.-Z., Dong, X., and Romesberg, F. E. (2020). Transcription and Reverse Transcription of an Expanded Genetic Alphabet *In Vitro* and in a Semisynthetic Organism. *J. Am. Chem. Soc.* 142, 19029–19032. doi:10.1021/jacs.0c09230
- Zhou, A. X.-Z., Sheng, K., Feldman, A. W., and Romesberg, F. E. (2019). Progress toward Eukaryotic Semisynthetic Organisms: Translation of Unnatural Codons. *J. Am. Chem. Soc.* 141, 20166–20170. doi:10.1021/jacs.9b09080

**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.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kimoto and Hirao. 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.