

This is a repository copy of *Genetic Code Expansion: A Brief History and Perspective*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/176009/>

Version: Published Version

---

**Article:**

Shandell, Mia Ann [orcid.org/0000-0002-7864-758X](https://orcid.org/0000-0002-7864-758X), Cornish, Virginia and Tan, Zhongping (2021) Genetic Code Expansion: A Brief History and Perspective. *Biochemistry*. ISSN: 1520-4995

<https://doi.org/10.1021/acs.biochem.1c00286>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NoDerivs (CC BY-ND) licence. This licence allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to the original authors. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# Genetic Code Expansion: A Brief History and Perspective

Mia A. Shandell,\* Zhongping Tan,\* and Virginia W. Cornish\*



Cite This: <https://doi.org/10.1021/acs.biochem.1c00286>



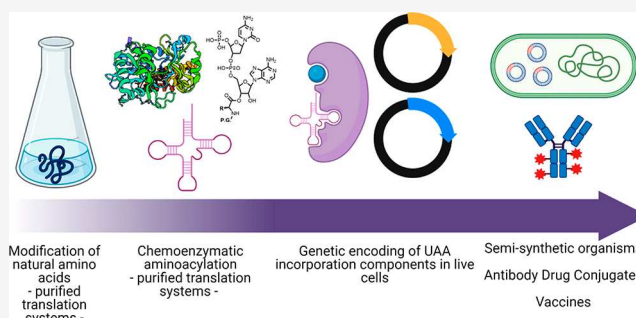
Read Online

ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** Since the establishment of site-specific mutagenesis of single amino acids to interrogate protein function in the 1970s, biochemists have sought to tailor protein structure in the native cell environment. Fine-tuning the chemical properties of proteins is an indispensable way to address fundamental mechanistic questions. Unnatural amino acids (UAAs) offer the possibility to expand beyond the 20 naturally occurring amino acids in most species and install new and useful chemical functions. Here, we review the literature about advances in UAA incorporation technology from chemoenzymatic aminoacylation of modified tRNAs to *in vitro* translation systems to genetic encoding of UAAs in the native cell environment and whole organisms. We discuss innovative applications of the UAA technology to challenges in bioengineering and medicine.



A grand challenge for the field of biochemistry is a general method for tailoring protein structure to address biological mechanism in the native cellular environment and to endow cells with new functions for future engineering applications. Advances in molecular biology in the 1970s made site-specific mutagenesis of single amino acids to probe protein function an everyday reality for researchers. Expanding the mutagenesis repertoire beyond the 20 naturally occurring amino acids in most species, unnatural amino acids (UAAs) enable site-specific installation of new and useful chemical functions, fluorescence, ligand binding, cross-linking, or photocaging, for example. There are still challenges to UAA mutagenesis being an everyday technique for biochemists. In this Perspective, we highlight the 2013 work of Chatterjee and Schultz, key historical papers in the field, and key perspective papers that illustrate future directions being charted by researchers in the field. This Perspective is not meant to stand in for comprehensive reviews published by researchers in the field.<sup>1–4</sup> Furthermore, we acknowledge that there are several exciting technologies that have been developed over the past several decades for chemical modification of proteins; we speak to only the UAA technology here. Chatterjee et al. (Schultz)<sup>5</sup> is a landmark study demonstrating a streamlined plasmid-based system for efficient multisite UAA incorporation in one target protein in live bacterial cells.

Chatterjee et al. (Schultz)<sup>5</sup> integrated conceptual advances in orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA generation, multisite incorporation, and flexibility of codon usage in a minimalist, optimized system for incorporation of UAAs in living cells. UAAs are incorporated into proteins in live cells by bio-orthogonal aaRS enzymes evolved to bind the UAA and its unnatural tRNA but not interact with the naturally occurring amino acids or tRNAs (Figure 1). The unnatural tRNA

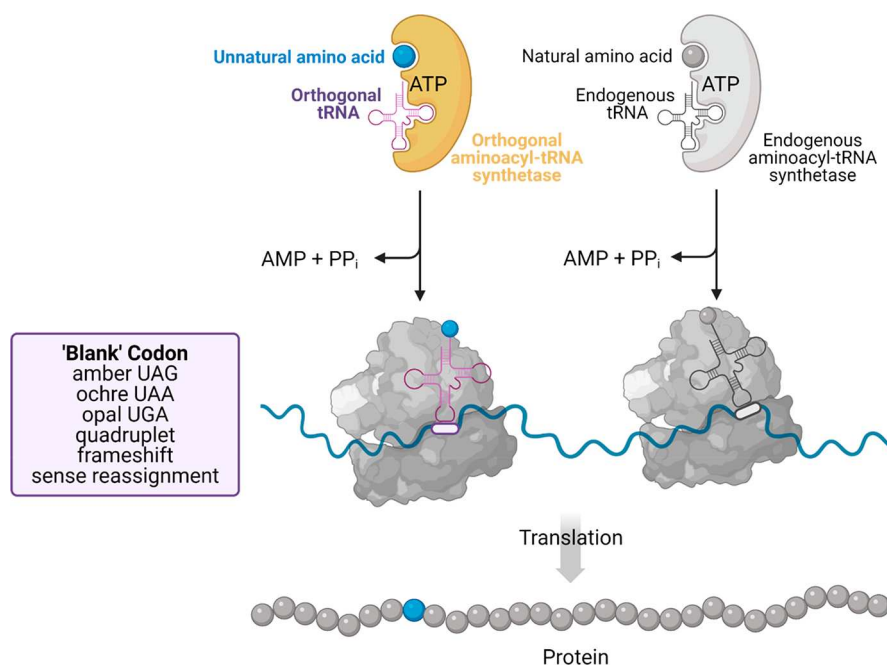
recognizes a stop, quadruplet base pair, or frameshift codon such that this combination manipulates the cell's endogenous translational machinery to incorporate the UAA into the target protein at the specific site of interest.

Chatterjee et al. (Schultz)<sup>5</sup> advanced the field by streamlining multisite UAA incorporation into a simple *Escherichia coli* transformation with two plasmids. One plasmid pUltraII encoded one copy each of two orthogonal aaRS/suppressor tRNA pairs: amber (UAG) suppressing *Methanococcus jannaschii* tyrosyl (*MjTyr*)-derived aaRS/tRNA<sub>CUA</sub> and optimized ochre (UAA) suppressing *Methanosarcina barkeri* pyrrolysyl (*MbPyl*)RS/*Methanosarcina mazei* pyrrolysyl (*MmPyl*)-tRNA<sub>UUA</sub>. The second plasmid encoded target protein green fluorescent protein (GFP) containing amber and ochre nonsense codons (GFP-3TAG-151TAA) to direct the incorporation of two unique UAAs into a single protein. This was previously intractable due to the requirement for multiple copies of aaRS or tRNA expression cassettes to incorporate a single UAA. The predecessor of pUltraII, pEVOL, encoded one copy of a *MjTyr*-derived optimized amber suppressor tRNA and two copies of *MjTyr*RS to incorporate UAAs into GFP151TAG.<sup>6,7</sup> With this construct in the presence of UAA *p*-azido-L-phenylalanine (pAzF), the suppression efficiency for a single amber codon reached approximately 80% of that of wild-type

**Special Issue:** 60th Anniversary Issue

**Received:** April 26, 2021

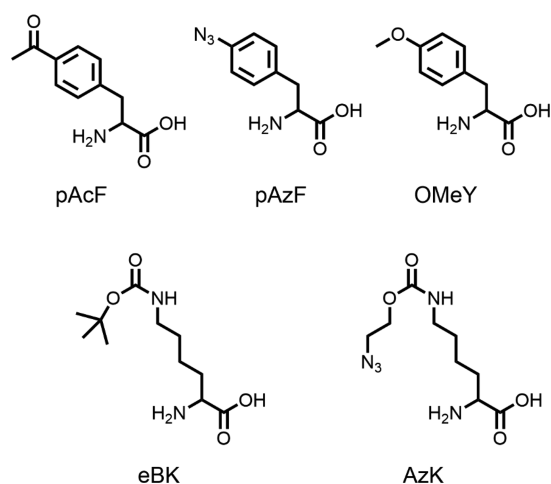
**Revised:** June 16, 2021



**Figure 1.** Unnatural amino acid (UAA) incorporation.

GFP.<sup>7</sup> Incorporation of two UAAs, *p*-acetyl-L-phenylalanine (pAcF) and *N*<sub>ε</sub>-Boc-L-lysine (eBK), into GFP-3TAG-151TAA using pUltraII achieved 20–25% of wild-type GFP expression.<sup>5</sup>

Generality was shown by optimizing incorporation of these UAA pairs simultaneously into GFP using the amber and ochre suppressor aaRS/tRNAs: pAcF and azido-L-lysine (AzK), pAcF and eBK, pAzF, and eBK, and *O*-methyl-L-tyrosine (OMeY) and eBK (Figure 2). UAAs with click handles, for example, pAcF and



**Figure 2.** Chemical structures of UAAs incorporated into GFP-3TAG-151TAA in Chatterjee et al. (Schultz).<sup>5</sup>

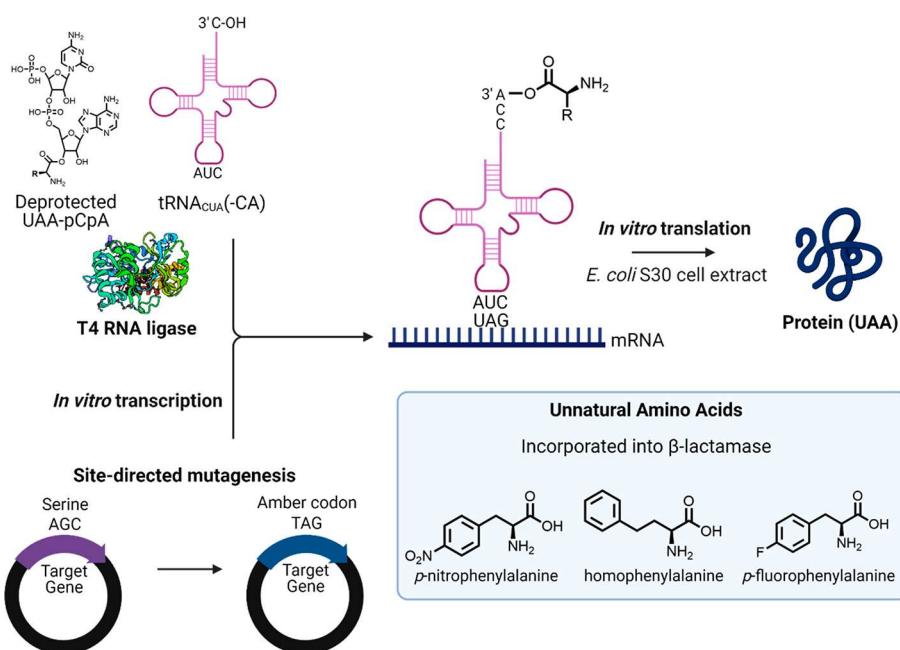
AzK, were incorporated for dual labeling with dyes suitable for in-gel Förster resonance energy transfer (FRET). They applied the dual suppression system to label a nonfluorescent target, ketosteroid isomerase, with a FRET pair by incorporating acetyl and azido click handles and labeling with Alexa Fluor 488-hydroxylamine to label the ketone and Alexa Fluor 594 dibenzocyclooctynol to label the azide, postpurification.<sup>5</sup>

## EARLY RESEARCH WITH UAAS

Researchers began exploring the possibility of using modified tRNAs to incorporate UAAs soon after the discovery of the tRNA adaptor.<sup>8,9</sup> In fact, the tRNA adaptor hypothesis was proven by chemically reducing Cys-tRNA<sup>Cys</sup> to Ala-tRNA<sup>Cys</sup> and showing that Ala would then be incorporated in response to a poly-Cys template.<sup>10</sup> In 1967, it was shown that the translational machinery could utilize D-Tyr-tRNA<sup>Tyr</sup>, prepared enzymatically by tyrosyl-tRNA synthetase (TyrRS), as a substrate.<sup>11</sup> The flexibility of the ribosomal peptidyl transferase center (PTC) to unusual chemistry was further shown by Fahnestock and Rich. They demonstrated that the translational machinery could synthesize oligomers containing multiple ester bonds using chemically converted hydroxyPhe-tRNA<sup>Phe</sup>.<sup>12</sup> However, each of these experiments was only possible because of an idiosyncratic route to the UAA-tRNA: chemical reduction of Cys, enzymatic charging of D-Tyr, and hydroxylation of the aromatic Phe residue. What was missing was a general method for producing the UAA-tRNA.

Right from the start there was significant interest in being able to incorporate biophysical probes into proteins using UAA mutagenesis. There were foundational approaches to the task of incorporating unique side chains. For instance, Johnson and colleagues modified Lys after it had been enzymatically ligated to tRNA. Acylation of the *N*<sub>ε</sub>-amine of Lys-tRNA<sup>Lys</sup> with *N*-hydroxysuccinimide ester-azidobenzoic acid generated *N*<sub>ε</sub>-azidobenzoyl-Lys-tRNA<sup>Lys</sup>.<sup>13</sup> Modified UAA-tRNA interfaced with endogenous translational machinery and was incorporated in place of or in competition with endogenous unmodified Lys in rabbit reticulocyte lysate. Because most target proteins contain multiple Lys residues, the modification could not be restricted to a single site, thus resulting in multisite incorporation of the photoactivatable Lys UAA.<sup>13</sup>

In 1978, Hecht and co-workers established a general procedure for the chemoenzymatic aminoacylation of tRNAs.<sup>14</sup> T4 RNA ligase transfers an aminoacyladenylate moiety from *N*-blocked (with *o*-nitrophenylsulfenyl) amino-



**Figure 3.** General method for site-specific UAA incorporation.

acylated  $P^1, P^2$ -bis( $5'$ -adenosyl)diphosphates to tRNAs lacking the  $3'$ -terminal adenosine. However, a large molar excess ( $>200$ -fold) of aminoacylated nucleotide derivatives were required for good yields, so they optimized the synthesis such that an only 20-fold molar excess was necessary, using  $N$ -acetylaminoacyl pCpA derivatives instead.<sup>15</sup> The modified chemical aminoacylation was used to acylate tRNA<sup>Phe</sup> with both D- and L-Phe, D- and L-Tyr, and  $N$ -acetyl-DL- $\beta$ -Phe. Misacylated tRNAs can participate in peptide bond formation, consistent with the adaptor hypothesis, but efficient dipeptide formation with a poly-Phe message occurred primarily with L-Phe, L-Tyr, and, interestingly,  $\beta$ -Phe, with L-PhetRNA<sup>Phe</sup> as the A-site tRNA.<sup>15</sup>

An alternative approach was demonstrated by Baldini and colleagues in 1988. Prior to this work, due to protection of the amino group during pCpA ligation, the chemically misacylated tRNAs could not bind the ribosomal A site and be incorporated into a growing polypeptide chain; thus, only dipeptides could form. By introducing a transient Boc protection/deprotection into the UAA-tRNA ligation, they demonstrated synthesis of functional *E. coli* tRNA<sup>Phe</sup> charged with a photoactivatable cross-linker UAA, L-4'-[(3-trifluoromethyl)-3H-diazirine-3-yl]-phenylalanine.<sup>16</sup> However, protein yields were often low due to the stoichiometric nature of chemically acylated tRNAs, and modifications were limited because there were no general methods for UAA-tRNA synthesis. Johnson and Brunner's methods both allowed more flexibility in the range of biophysical probes that could be attached to either Lys or Phe; however, they still were not general methods for UAA-tRNA synthesis, and they led to uncontrolled multisite incorporation of the biophysical probes.<sup>13,16</sup>

### ■ A GENERAL METHOD FOR THE SITE-SPECIFIC INCORPORATION OF UAAS *IN VITRO*

A breakthrough was the report of a general method for site-specific *in vitro* incorporation of UAAs by the Schultz lab (Figure 3).<sup>17</sup> Briefly, they developed a general method for synthesizing UAA-tRNAs that recognized UAG stop codons and then demonstrated that this UAA-tRNA could be utilized by a crude

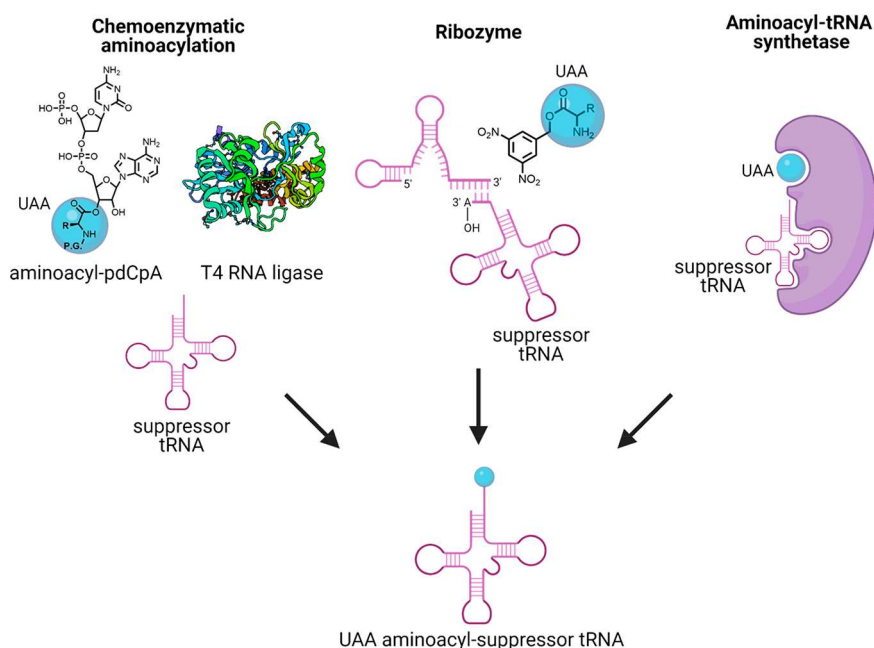
*E. coli* S30 cell extract for site-specific incorporation of the UAA in response to a UAG codon engineered in a protein-coding gene.

The synthesis of the UAA-tRNA was made possible by an efficient chemical synthesis of the UAA-pCpA, the two terminal nucleotides of the tRNA, and by technology previously developed by Sid Hecht and co-workers that showed UAA-pCpA molecules could be efficiently ligated to tRNA missing the terminal dinucleotide pCpA at the  $3'$ -acceptor stem by the natural enzyme T4 RNA ligase.<sup>18,19</sup> Brunner's work enabling misacylated tRNAs to function in the ribosomal A site made it possible to form polypeptides and largely avoid hydrolysis of the amino acyl ester linkage by endogenous aaRSs.<sup>16</sup> Drawbacks of the previous methods were nonselective incorporation of the UAA and size restrictions on the target protein. Noren's UAA mutagenesis approach applied these foundational methods to a generalized system in which in theory a diverse range of UAAs could be used to acylate the suppressor tRNA, the suppressor tRNA could be directed to a specific site by mutagenizing that position to an amber stop codon, and the size of the protein of interest was limited only by what could be encoded on a plasmid.

Noren et al. (Schultz) showed the incorporation of three different UAAs in the active-site residue Phe66 in  $\beta$ -lactamase and kinetic characterization of these variants. They prepared an amber suppressor tRNA using anticodon loop replacement of yeast tRNA<sup>Phe</sup> and demonstrated this tRNA was not recognized by the *E. coli* PheRS in their S30 extract [ $\beta$ -lactamase-(Phe66TAG), non-acylated tRNA<sub>CUA</sub>, and [ $^3$ H]Phe]. No  $\beta$ -lactamase activity was observed, and there was no band corresponding to [ $^3$ H]Phe-incorporated  $\beta$ -lactamase by SDS-PAGE. Significantly, they showed using [ $^3$ H]Phe-tRNA<sub>CUA</sub> and HPLC analysis of trypsin-digested  $\beta$ -lactamase that the [ $^3$ H]Phe was incorporated only at position Phe66, demonstrating not only efficient incorporation of the UAA [ $^3$ H]Phe in response to the UAG codon but also that the UAA [ $^3$ H]Phe was not scrambled with other natural amino acids.<sup>17</sup>

Bain and co-workers used a strategy similar to that of the Schultz group to incorporate L-3-iodo-tyrosine into a 16-residue





**Figure 4.** General methods for misacylating tRNA. P.G., protecting group.

polypeptide.<sup>20</sup> They prepared a semisynthetic, non-hypermodified *E. coli* glycyl tRNA<sub>CUA</sub> nonsense suppressor tRNA acylated with L-3-[<sup>125</sup>I]tyrosine and incubated with the message containing UAG at position 9 in rabbit reticulocyte lysate. The translation product was purified and sequenced to unambiguously determine the site specificity of incorporation. Nonsense suppression was due entirely to the added synthetic suppressor because they could not detect read-through by endogenous aminoacyl-tRNAs (aa-tRNA).<sup>21</sup>

**General Methods for UAA-tRNA Synthesis.** A key issue continued to be the lack of a general method for synthesizing the UAA-pdCpA (Figure 4). A key advance was made by Robertson and Ellman (Schultz) in 1991.<sup>22</sup> Unprotected pdCpA was selectively aminoacylated in high yield with the cyanomethyl ester (CME) of *N*-blocked amino acid and ligated to tRNA. The photolabile nitroveratryl protecting groups for the  $\alpha$ -amine and side chain functional groups enabled the aa-tRNA to be deprotected photochemically. This reaction produces high yields of stable, unblocked aa-tRNA that can be used directly in a purified translation system. The approach greatly simplified the synthesis of UAA-pdCpA to one high-yield (76–87%) step.<sup>22</sup>

An alternative approach to tRNA aminoacylation was catalytic RNA, or ribozymes. Natural ribozymes catalyze transesterification reactions of phosphodiester bonds. Szostak and co-workers isolated catalytic RNAs with acyl transferase activity, like that of the PTC, from pools of random RNA sequences. They selected for enhanced transfer of an *N*-biotinyl-L-methionyl group from the 3'-end of a donor hexanucleotide, 5'-pCAACCA-3', to the 5'-hydroxyl group of the ribozyme.<sup>23,24</sup> Suga, Szostak, and co-workers generated aaRS-like ribozymes with two catalytic domains: one that recognizes the amino acid substrate and self-aminoacylates its 5'-hydroxyl and the other that binds the tRNA and transfers the aminoacyl group to the 3'-end. This ribozyme acts as a synthetase that can charge tRNA<sup>fMet</sup> with Gln or Phe.<sup>25</sup> CME was chosen as a leaving group on the amino acid because it has no hydrogen bond donors or acceptors that could interact with the ribozyme. Active RNAs could be

isolated from the pool by selection with *N*-biotinyl-L-glutamyl-CME and subsequent pull-down with streptavidin.<sup>25</sup>

The Suga lab generalized the ribozyme *de novo* catalyst for tRNA acylation using aaRS-like RNA molecules called Flexizymes (Fx) and mutants thereof.<sup>26</sup> They noticed that Fx recognizes neither the leaving nor the amino group of the substrate, but rather the aromatic functionality of the amino acid side chain and the carbonyl group of the ester. To improve binding, they redesigned substrates incorporating an aromatic ring in the leaving group. They used dinitrobenzyl ester (DBE) and the more activated chlorobenzyl thioester (CBT). Because DBEs are less hydrolytically labile than CBTs, DBEs were used in further experiments. Enhanced interaction between Fx and the substrate significantly enhanced tRNA acylation efficiency and enabled incorporation of citrulline, *N*<sub>ε</sub>-acetyl-L-lysine, *N*<sub>ε</sub>-biotinyl-L-lysine, *p*-iodo-L-phenylalanine, (*S*)-3-isopropylactic acid, and (*S*)-3-phenyllactic acid into short (nine-amino acid) FLAG-tagged peptides expressed in an *E. coli* cell-free translation system, Protein synthesis Using Recombinant Elements (PURE),<sup>27</sup> by amber-programmed frameshift suppression and sense codon reassignment (AGU, AAC, and CAG).<sup>26</sup> This breakthrough represented a powerful tool for enhancing the range of UAAs that could be ligated to tRNAs and incorporated into polypeptides *in vitro*.

**Expanding the Repertoire of UAAs to Address Questions of Protein Structure and Function.** Another focus of the field following Noren et al. (Schultz) in 1989 was pushing the boundaries of what UAA structures and functions could be incorporated and using these UAAs to address important protein structure and function questions. Key questions of enzyme mechanism and protein stability could be addressed by incorporating unnatural isoelectronic or isosteric analogues of natural amino acids at sites of interest and measuring changes in enzyme kinetics and/or protein denaturation.

Building on earlier work with modified nucleosides and tRNAs, Ellman and Mendel working with Schultz explored the tolerance of the translational machinery to changes in the amino

acid structure.<sup>28</sup> Specifically, they tested if different UAAs could be incorporated at position Ala82 in T4 lysozyme (T4L) and how the UAAs affected the stability of T4L. Ala82 is a surface residue located between two helices, distant from the active site. The structure and electronics of the UAA significantly affected its use as a substrate by the translational machinery. The incorporation efficiencies were as follows: none detected (ND) for D-Ala and 30% (suppression efficiency) for lactic acid; N-alkyl amino acids, <5% for azetidine 2-carboxylic acid, 43% for pipercolic acid, 24% for N-methyl-alanine, and <5% for N-ethyl-alanine;  $\alpha,\alpha$ -disubstituted amino acids, 28% for cyclopropylglycine and 23% for  $\alpha$ -aminoisobutyric acid. Interestingly, changing the amino acid structure and electronics changed the apparent yield of protein synthesis; because an *E. coli* S30 crude cell extract was used, the mechanism of this decrease in yield could not be determined at the time. The stabilities of the resulting UAA-substituted T4L proteins were determined by thermal denaturation as measured by circular dichroism.<sup>28</sup> These UAA backbone analogues largely changed the stability of T4L as would be predicted.

Judice and Schultz used UAA incorporation to make more precise changes in amino acid structure than possible with the natural amino acids to probe enzyme mechanism.<sup>29</sup> Staphylococcal nuclease (SNase) accelerates the hydrolysis of phosphodiester bonds in nucleic acids approximately  $10^{16}$ -fold. One hypothesis was that general base catalysis underpins this enormous rate acceleration where Glu43 in SNase acts as a general base to activate a water molecule for attack on the phosphodiester backbone of DNA. However, when Glu43 was replaced with isoelectronic and isosteric analogues, Arg, S-4-nitro-2-aminobutyric acid, S-2-amino-5-hydroxypentanoic acid, aminoethylhomocysteine, and citrulline, differing only by being poorer bases, the kinetics of SNase were virtually unchanged relative to those of the wild-type. A significant accomplishment at the time working with *E. coli* S30 cell extracts, a structure of the enzyme substituted with homoglutamic acid at site 43 was obtained. Combined, the kinetic and structural data suggested that Glu43 may instead play a structural role, fixing the conformation of a nearby loop.<sup>29</sup>

Despite the enormous potential utility of biophysical and other probes, it was becoming clear that the translational machinery places constraints on the size of the amino acid side chain and hence what fluorophores, cross-linking agents, post-translational modification, or other UAAs could be incorporated.<sup>30</sup> Thus, Cornish, Hahn, and Schultz incorporated a small ketone handle that could subsequently be modified to form an oxime or other unnatural linkage to the biophysical probe in what has come to be called bio-orthogonal labeling.<sup>31</sup> They incorporated keto amino acids **1** (5% suppression efficiency) and **2** (30% suppression efficiency) in sites Ser44 and Ala82 in T4L, two sites known generally to give high suppression efficiencies (Figure 5). Subsequently, they showed that electrophilic ketone UAA **2** could be derivatized with fluorescein hydrazide in T4L Ala82  $\rightarrow$  **2**. Fluorescence spectra of purified

T4L Ala82  $\rightarrow$  **2** and wild-type T4L both being subjected to the same labeling conditions with fluorescein hydrazide demonstrated that only the protein containing the ketone handle was labeled with the fluorophore.<sup>31</sup>

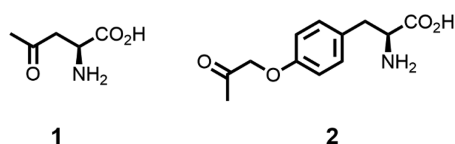
**Alternate Codons for UAA Incorporation.** Another important area was exploring alternate codons for incorporation of UAAs. The main challenge of nonsense suppression is the limited range of nonsense codons: amber (TAG), ochre (TAA), and opal (TGA). Alternate codons would increase the number of UAAs that could be incorporated in a single protein. In the 1990s, frameshift codon suppression was explored as an alternative to nonsense suppression (Figure 6). When the frameshift does not happen, a termination codon UAA appears downstream, resulting in truncated protein. If the reading frame is shifted by suppression of the quadruplet codon with UAA-tRNA, full-length protein is synthesized.

Using this strategy, frameshift suppressor Ala-tRNA<sub>ACCU</sub> and Ala-tRNA<sub>CCUA</sub> incorporated Ala into *E. coli* dihydrofolate reductase (DHFR), shown by a restoration of enzyme activity to 40% and 15% of wild-type activity, respectively.<sup>32</sup> Repurposing of rare codons was another strategy for expanding options for UAA incorporation. AGG was used as an alternative codon for incorporation of the photoactive UAAs *p*-phenylazophenylalanine, 2-anthrylalanine, 1-naphthylalanine, 2-naphthylalanine, and *p*-biphenylalanine into a polypeptide expressed in an *E. coli* S30 extract, where AGG is rare (<3%).<sup>33</sup> The Sisido lab extended this work to incorporate the UAAs nitrophenylalanine, 2-naphthylalanine, *p*-phenylazophenylalanine, and 2-anthrylalanine into streptavidin (Tyr83  $\rightarrow$  AGGU) through frameshift suppression in *E. coli* S30 extracts.<sup>34</sup> CGGG was found to work more efficiently than ACCU in further studies.<sup>35,36</sup>

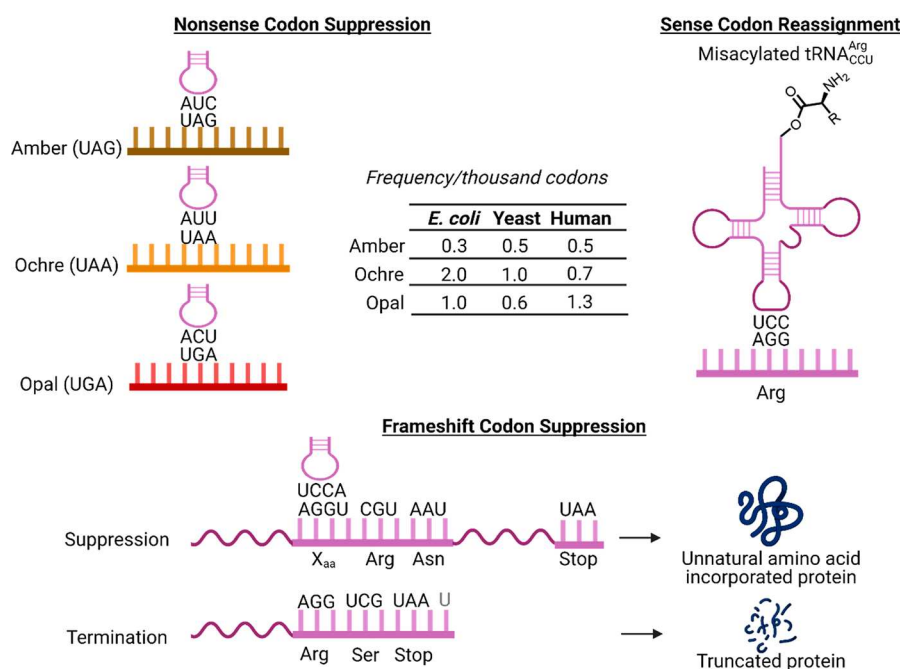
More options for different quadruplet codons were demonstrated in an *E. coli* S30 cell extract with nitrophenylalanine efficiently incorporated into streptavidin using the codons AGGU, CGGU, CCCU, CUCU, CUAU, and GGGU.<sup>37</sup> Further utility of this approach was demonstrated by the incorporation of two UAAs, nitrophenylalanine and 2-naphthylalanine, into streptavidin using the quadruplet codons CGGG and GGGU.<sup>37</sup> Incorporation of two fluorescent UAAs in *E. coli* DHFR was achieved by a combination of quadruplet codon and amber suppression.<sup>38</sup> Specifically, the fluorescent UAA 7-azatryptophan was incorporated with a CGGG quadruplet decoding tRNA and acceptor *N*<sub>β</sub>-dabcy1-1,2-diaminopropionic acid by amber suppression in *E. coli* DHFR.<sup>38</sup>

Sense codon reassignment, or genetic code reprogramming, has been explored as an alternative to nonsense suppression. Nonsense codon suppression is limited to two UAAs because there are only three nonsense codons and one must be used for translation termination. It would be of enormous practical utility and would allow fundamental questions about the genetic code to be addressed if sense codons could be reassigned. Forster, Tan, Cornish, and Blacklow established the concept of genetic code reprogramming of multiple, adjacent sense codons by reassigning three sense codons to UAAs using chemoenzymatically charged tRNAs in a reconstituted translation system lacking aaRSs.<sup>39</sup>

Rather than using an *E. coli* S30 *in vitro* extract with competing aa-tRNAs and aaRSs that could hydrolyze noncognate aa-tRNA pairs and recharge the tRNA with the cognate amino acid, we made a purified *in vitro* translation system ourselves based on published protocols in the ribosome mechanism field. The UAA-tRNAs were prepared using the chemoenzymatic methods



**Figure 5.** Ketone UAAs incorporated in T4 lysozyme in Cornish et al. (Schultz).<sup>31</sup>



**Figure 6.** Alternate codons for UAA incorporation. Frequency data from <https://www.genscript.com/tools/codon-frequency-table>.

being used by the Schultz group at the time. Because it was the first attempt to modify multiple sense codons, we started with three conservative side chain modifications: *O*-methyl-L-serine, 2-amino-4-pentenoic acid (allylglycine), and 2-amino-4-pentynoic acid (propargyl glycine). Tracking the peptide synthesis using [<sup>35</sup>S]Met-tRNA, [<sup>3</sup>H]Glu-tRNA, and authentic peptide markers prepared by solid-phase peptide synthesis, translation of a peptide with five of the same UAAs in a row was demonstrated. Finally, translation of a peptide with three different UAAs in a row each in response to a different sense codon was achieved. Together, this work showed for the first time that multiple sense codons could be reassigned allowing for translation of unnatural oligomers, a direction we argue below will become increasingly important to the field.

Around the same time, Josephson, Hartman, and Szostak established ribosomal synthesis of nonribosomal peptide-like molecules containing 10 UAA side chain analogues by sense codon reassignment, significantly increasing the number of UAAs that can be incorporated into a single polypeptide.<sup>40</sup> The Suga lab established the mRNA-encoded incorporation of multiple, consecutive amino acid analogues for the *in vitro* synthesis of unnatural polypeptides using a combination of frameshift suppression and sense codon reassignment (AGU, AAC, and CAG) in the Fx/PURE reconstituted translation system with depleted aaRSs and cognate amino acids.<sup>26</sup> The ability to control installation of multiple UAAs significantly increased the diversity of peptides that can be synthesized and screened for therapeutic properties. Suga has gone on to do a lot of work in this area, for example, incorporating multiple, consecutive amino acid backbone analogues into the peptide backbone, including  $\alpha$ -hydroxy amino acids,<sup>41</sup> *N* $\alpha$ -methylated amino acids,<sup>42</sup> and, more recently, D-amino acids<sup>43</sup> and  $\beta$ -amino acids.<sup>44</sup> However, to date, sense codon reassignment is limited to *in vitro* translation systems.

## ■ A GENERAL METHOD FOR THE SITE-SPECIFIC INCORPORATION OF UAAS *IN VIVO*

To date, UAAs were incorporated into *E. coli* S30, wheat germ, or rabbit reticulocyte extracts. Methods developed in the laboratories of Dougherty and Lester pioneered the use of UAAs in *Xenopus* oocytes, where they injected UAA-pdCpA-ligated tRNA and mRNA encoding the protein of interest with an amber codon, with a particular focus on eukaryotic ion channels.<sup>45</sup> Specifically, Lummis and co-workers explored the role of a highly conserved Pro at site 8 (Pro8) in cation-selective 5-hydroxytryptamine type 3 receptors. Pro8 acts as a hinge in the loop between the second and third transmembrane helices, a region that interacts with the extracellular ligand binding domain and was hypothesized to play an important role linking neurotransmitter binding to channel gating through *cis*–*trans* isomerization of the protein backbone. Incorporation of Pro analogues favoring the *cis* conformer produced functional channels, while those favoring the *trans* conformer produced nonfunctional channels. Importantly, the *cis*–*trans* energy gap of the Pro analogue was strongly correlated with channel activation, suggesting *cis*–*trans* isomerization of this single Pro acts as a gating switch between open and closed channel states.<sup>45</sup> These experiments built upon their earlier work optimizing the pdCpA ligation chemistry for backbone analogues, such as  $\alpha$ -hydroxy amino acids.<sup>46</sup>

With respect to expanding the repertoire of codons for incorporation of multiple UAAs in *Xenopus*, Rodriguez and co-workers demonstrated multisite incorporation of UAAs into nicotinic acetylcholine receptors by combining nonsense and frameshift codon suppression.<sup>47</sup> A limitation of chemoenzymatically charged tRNAs is that they cannot be reacylated once inside the cell, capping the amount of protein that can be generated. An ideal system would include all of the necessary components genetically encoded in the cell. A main challenge of eukaryotic genetic code expansion is the fact that translational machinery is not well conserved between prokaryotes and eukaryotes.



The next big breakthrough in the field was full genetic encoding of UAA incorporation components in live cells. In 2001, Wang et al. (Schultz) reported a general method for incorporating UAAs into proteins in *E. coli* through directed evolution of an orthogonal aaRS/tRNA pair.<sup>48</sup> This advance addressed multiple technical challenges in the field. One, it addressed the technical difficulty of preparing the aminoacyl-tRNA chemoenzymatically. Two, it removed the limit on protein yield imposed by use of an *in vitro* translation extract. Significantly, the UAA technology could now be adopted by a non-expert in the technology.

The key conceptual advance was the positive and negative selection strategy for evolving the orthogonal tRNA and aaRS.<sup>49</sup> The naturally orthogonal *Mjt*tRNA<sup>Tyr</sup><sub>CUA</sub> and *MjTyrRS* were used as the starting point. First, they used a negative selection based on suppression of a UAG codon in the toxic RNase Barnase in the absence of *MjTyrRS* to select for *Mjt*tRNA<sup>Tyr</sup><sub>CUA</sub> variants that could not be aminoacylated by the endogenous natural aaRS enzymes.<sup>49</sup> Next, the winners were subject to a positive selection for *Mjt*tRNA<sup>Tyr</sup><sub>CUA</sub> variants that could suppress amber mutations in  $\beta$ -lactamase in the presence of *MjTyrRS*.<sup>49</sup> Finally, to generate the orthogonal *MjTyrRS*, *E. coli* cells were transformed with a library of *MjTyrRS* genes and subjected to a positive selection for suppression of a UAG codon in chloramphenicol acetyltransferase (CAT). The library was subject to both positive (+UAA, +chloramphenicol) and negative selection (−UAA, +chloramphenicol) to yield pairs that incorporate the UAA but not natural amino acids in response to the UAG codon. They used the resulting orthogonal aaRS/tRNA pair to incorporate *O*-methyl-L-tyrosine into DHFR expressed in *E. coli*.<sup>48</sup>

While still limited by the need to purify the UAA-incorporated protein out of the cell, this work eliminated the problem of low protein yields with *in vitro* cell extracts and broadly enabled engineering UAA biosynthetic machinery in live cells.<sup>48</sup> There were some hurdles in getting there, with different selection approaches being less successful.<sup>50</sup> Once a general method for generating mutant aaRS-tRNA pairs was established, issues that needed to be addressed were how to generate new orthogonal aaRS-tRNA pairs for UAA incorporation in model systems where large mutant libraries cannot be made, such as mammalian cells, and the lack of suitable starter aaRS-tRNA pairs orthogonal in these systems. The transfer of *E. coli* TyrRS-tRNA<sub>CUA</sub> and LeuRS-tRNA<sub>CUA</sub> to yeast and mammalian cells was feasible; however, tRNA expression was a major hurdle at the time. The Wang lab developed a general method for expressing orthogonal tRNAs in mammalian cells using type 3 Pol III promoters, and this is the general method currently being used in the field for UAA incorporation in yeast, mammalian cells, and various animals.<sup>51</sup> Another challenge was that the size of the aaRS binding pocket limited the stereochemical diversity of UAAs that could be incorporated. A breakthrough in the field was made by the Wang lab with the finding that mutation of the *Methanosarcina* PylRS binding pocket can generate more flexible substrate specificity for Phe and Tyr analogues with bulky conjugated rings or long azobenzene side chains.<sup>52</sup>

The Schultz lab demonstrated the first fully genetically encoded UAA incorporation system in eukaryotes when they incorporated UAAs in *Saccharomyces cerevisiae*.<sup>53</sup> They exploited the fact that *E. coli* tyrosyl tRNA<sub>CUA</sub> can be expressed in yeast and is a poor substrate for the *S. cerevisiae* aaRSs. They evolved TyrRS in yeast for incorporation of acetyl, benzoyl, azido, and iodo-Phe analogues, as well as *O*-methyl-L-tyrosine, into human

superoxide dismutase.<sup>53</sup> This system became the basis for directed evolution of aaRSs for UAAs that could be readily transferred to mammalian cells. Genetic encoding of orthogonal synthetase-tRNA pairs in mammalian cells and animals, including *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Mus musculus*, followed suit.<sup>54–60</sup>

**Incorporation of UAAs into Mammalian Cells and Animals.** As early as 2006, Tirrell and Schuman established metabolic labeling with click handle UAAs as a nonspecific, heterogeneous multisite UAA incorporation method for tagging newly synthesized proteins in mammalian cells.<sup>61</sup> Azidohomoalanine (AHA) is an azide-bearing methionine analogue that can be incorporated at methionine codons and tagged with an alkyne-affinity tag using copper-catalyzed azide-alkyne [3+2] cycloaddition for the identification of AHA-labeled proteins via mass spectrometry.<sup>61</sup> Another methionine surrogate, alkyne-bearing homopropargylglycine (HPG), was used in tandem with AHA to fluorescently label newly synthesized proteins in rat hippocampal neurons by strain-promoted cycloaddition.<sup>62</sup> Pulse-chase experiments enabled fluorescent labeling of two distinct proteomes synthesized sequentially in time such that the dynamics of protein synthesis and fate could be monitored in neurons.<sup>62</sup> Thus, they could address important questions about the role of newly synthesized proteins in neuron function without the need for selective incorporation of the UAA.

Efficient incorporation of UAAs into proteins in animals has also been a critical challenge. To improve the efficiency, many different methods have been researched, including using the type 3 polymerase III promoter to more efficiently express orthogonal prokaryotic tRNAs,<sup>51</sup> UAA esterification to increase UAA bioavailability,<sup>63</sup> and optimizing tRNA/synthetase affinity to increase the level of UAA incorporation.<sup>64</sup> Encouraging progress was achieved in this research area through the combined use of these optimized methods.<sup>4</sup> In their letter to the editor of *Cell Research*, Ye, Wang, Li and co-workers reported the introduction, maintenance, and transmission of the genetic material for code expansion in mice. In this work, they integrated the orthogonal pAzFRS/tRNA<sub>CUA</sub> pair into the mouse genome.<sup>59</sup> They demonstrated that, in the presence of pAzF, the suppressor tRNA can decode the UAG amber codon to express a dual fluorescent reporter eGFP-TAG-mCherry in neurons and bone marrow cells of mice.<sup>59</sup>

**Expanding the Repertoire of UAAs That Can Be Incorporated *In Vivo*.** One of the most exciting classes of UAAs continues to be backbone analogues because of the potential to extend the power of genetic encoding to oligomers other than  $\alpha$ -L-polypeptides. The logic of making backbone analogue UAA incorporation work *in vivo* began with the discovery or generation of ribosomes that can accommodate these UAAs followed by testing of known aaRS/tRNA pairs for charging them. In 2016, Schepartz and Söll incorporated  $\beta^3$ -amino acids into full-length DHFR in *E. coli*.<sup>65</sup> Previously, Dedkova and Hecht had found that ribosomes from some erythromycin-resistant *E. coli* mutants could tolerate the incorporation of  $\beta^3$ -amino acids into full-length DHFR *in vitro*.<sup>66</sup> Building on this work, Schepartz and Söll highlighted the unexpected flexibility of the endogenous *E. coli* translational machinery to  $\beta^3$ -amino acid backbone analogues when they demonstrated incorporation of  $\beta^3$ -amino acids into DHFR expressed in *E. coli* harboring a plasmid encoding mutant ribosomes from erythromycin-resistant strains.<sup>65</sup> Significantly, they demonstrated that  $\beta^3$ -(*p*-Br)Phe and  $\beta^3$ -Gly could be charged by endogenous aaRS enzymes, with PheRS being the



most tolerant of these substrates. Furthermore, wild-type EF-Tu interacted efficiently with  $\beta^3$ -Phe-tRNA<sup>Phe</sup> for delivery to the ribosome. To improve the efficiency and selectivity for  $\beta^3$ -amino acid incorporation, a library of peptidyl transferase center 23S rRNA mutant ribosomes were screened for erythromycin resistance and  $\beta$ -puromycin sensitivity, resulting in a new mutant P7A7 that imparted 3-fold higher levels of  $\beta^3$ -amino acid incorporation over those of the previously discovered mutants.<sup>65</sup> Tryptic digest of DHFR peptide fragments containing either  $\alpha$ -Phe or  $\beta^3$ -(*p*-Br)Phe at F128 showed a 10-fold lower level of incorporation of the  $\beta^3$ -amino acid versus  $\alpha$ -Phe.<sup>65</sup>

Underscoring the importance of recent work incorporating backbone analogue UAAs *in vivo*, it has been difficult for scientists to produce peptides containing D- and  $\beta$ -amino acids by UAA incorporation *in vitro*. By tuning tRNA sequence and concentrations of native initiation (IF<sub>2</sub>) and elongation factors (EF-Tu/Ts and EF-G), in 2017, the Suga lab increased the yield of a D-Ala-D-Ala-containing peptide by >5-fold and incorporated 10 consecutive D-Ser residues into a peptide chain.<sup>43</sup> The existence of mutant *E. coli* ribosomes that enhance D-amino acid incorporation *in vitro* indicates potential for *in vivo* incorporation of D-amino acids.<sup>67</sup> Similarly, in 2018, the Murakami lab improved incorporation of multiple  $\beta$ -amino acids, producing peptides with natural amino acid spacers between two or three  $\beta$ -amino acids in their optimized translation system.<sup>68</sup> Translating the *in vitro* advances in incorporation of backbone analogue UAAs to cells should catch on as more suitable aaRS/tRNA pairs are established. In 2019, Dedkova and Hecht found that wild-type PylRS could incorporate a fluorescent oxazole UAA lacking an asymmetric center or  $\alpha$ -amino group. MreB (Leu13TAG) was co-expressed with PylRS in an *E. coli* strain with modified ribosomes that could incorporate dipeptides and dipeptidomimetics.<sup>66,69</sup> It is becoming clearer how malleable the translational machinery is for incorporation of more exotic UAAs. Importantly, this research signals the possibility of finding more mutant ribosomes and aaRS/tRNA pairs for backbone analogue UAA incorporation *in vivo*. One direction the field is going is to combine an expanded pool of UAA backbone analogues with an expanded pool of orthogonal codons and engineered orthogonal ribosomes for genetic encoding of unnatural oligomers *in vivo*.

Another class of UAA analogues that could be highly impactful is epigenetic protein modifications. Research on the functional effects of specific epigenetic protein modifications is hindered by the difficulty of synthesizing post-translationally modified target proteins in cells. O-phosphoserine (Sep) is the most abundant phosphorylated amino acid in eukaryotes. It is synthesized post-translationally by acylation of tRNA<sup>Cys</sup> with Sep by SepRS, an aaRS unique to methanogenic archaeobacteria. Park et al. (Söll, Noren, and Rinehart) made an amber suppressing tRNA<sup>Sep</sup> and, critically, evolved an EF-Tu mutant, EF-Sep, that could bind Sep-tRNA<sup>Sep</sup> for site-specific incorporation of Sep into proteins in *E. coli*.<sup>70</sup> SepRS, tRNA<sup>Sep</sup>, and EF-Sep together allow *E. coli* to read UAG as a Sep codon; they synthesized mitogen-activated ERK activating kinase 1 (MEK1) with Sep incorporated at a key modified residue Ser218 by amber suppression *in vivo*.<sup>70</sup> In addition, it would need to be established that Sep was incorporated at only the intended UAG residue. Similarly, tyrosine phosphorylation is a critical PTM in cellular signal transduction. In 2017, Luo et al. (Schultz and Wang) incorporated O-phosphotyrosine (pTyr) and its non-hydrolyzable analogue, 4-phosphomethylphenylalanine, into recombinant proteins by amber suppression in *E. coli*.<sup>71</sup> Around

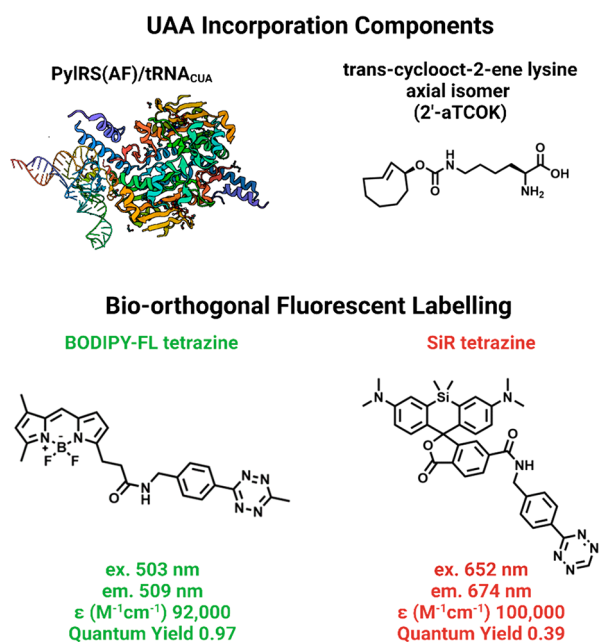
the same time, Hoppmann and co-workers (Hunter, Shokat, and Wang) incorporated a neutral pTyr analogue into recombinant proteins in *E. coli* by amber suppression; deprotection results in a native, negatively charged pTyr at the desired site.<sup>72</sup> Multisite incorporation of these UAAs would enable modeling of multiple phosphorylated residues of a protein.<sup>71,72</sup>

Low yield is a significant hurdle to studying proteins modified with UAAs by amber suppression *in vivo*. Wild-type expression levels of UAA-modified protein cannot be realized because suppressor tRNAs compete with release factors (RFs) for the stop codon. Church, Isaacs, and co-workers have pioneered the breakthrough in this area by developing genomically recoded organisms (GROs). Upon recoding of the entire genome of *E. coli* such that all UAG stop codons are mutated to the UAA stop codon and deletion of RF1, the “blank” UAG codon could then be reintroduced as a sense codon for highly efficient incorporation of a UAA.<sup>73</sup> The new genome enabled this new strain of *E. coli* C321.ΔA to exhibit increased resistance to viral infection by blocking the translation of viral proteins.<sup>73</sup> The Church lab has led this area of research toward creating GROs with expanded capabilities. Ostrov and co-workers (Church) created *E. coli* with a 57-codon genome in which all 62214 instances of seven codons were replaced with synonymous codons in all protein-encoding genes.<sup>74</sup> When the recoded codons' respective tRNAs and release factor are removed, up to four orthogonal UAAs could be introduced into the organism.<sup>74</sup> The increased yield and specificity of UAA incorporation in GROs should empower efforts for industrial UAA-modified protein production and more representative *in vivo* experiments with UAA-modified proteins.

The genomically recoded *E. coli* strain C321.ΔA has been used subsequently to advance techniques for studying post-translational modifications with UAAs. Isaacs and Rinehart conducted a proteome-wide investigation of the role of phosphorylation of human proteins *in vivo*.<sup>75</sup> They genetically encoded Sep<sup>70</sup> in a synthetic human phosphopeptide library expressed in C321.ΔA and identified proteome-wide phosphorylation-dependent interactions using bimolecular fluorescence complementation in cells.<sup>75</sup> In contrast, for *in vitro* studies of the regulation by phosphorylation of 26S proteasome subunit RpnI, Sep was incorporated into RpnI(361TAG) expressed in *E. coli* strain C321.ΔA, allowing for purification of homogeneously phosphorylated RpnI.<sup>76</sup> To transfer this technology to mammalian cells, Chin and co-workers demonstrated orthogonality in mammalian cells of an evolved SepRS/tRNA<sup>CUA</sup> pair<sup>77</sup> based on the original by Park and co-workers.<sup>70,78</sup> The Sep incorporation system in mammalian cells was completed by co-expression with eRF1(E55D) (a eukaryotic RF more permissive of UAG read-through), creation of a eukaryotic elongation factor variant EF-1 $\alpha$ -Sep containing mutations analogous to those of the prokaryotic EF-Sep, and deletion of phosphoserine phosphatase to increase the intracellular Sep concentration.<sup>78</sup> If the UAA technology can be used to selectively introduce Sep (and ultimately other epigenetic modifications simultaneously) at multiple, defined positions in a single protein in living cells, this will be a very powerful tool for biologists.

Fluorescent UAAs remain one of the most sought after yet challenging classes of analogues. Tagging with fluorescent proteins (FP) is an indispensable technique for localization and mechanistic studies of protein targets inside cells. However, the large size of FPs (27 kDa) limits tagging to the protein termini and the targets that can be studied. The most efficient chemical tags also require protein tags, and the peptide chemical tags rely

on two-step labeling with bio-orthogonal chemistry. The holy grail of fluorescent labeling would be direct and selective incorporation of a fluorescent UAA with a high photon output in mammalian cells. The challenges are having the ribosome accept a large fluorophore as a substrate, minimal nonspecific labeling of the fluorescent amino acid in the cell or incorporation into other UAG codons, or rapid and selective bio-orthogonal chemistry; one indication that the labeling technology meets these criteria is that it can be used to image a typical cellular protein at  $\sim 1 \mu\text{M}$  concentrations freely diffusing intracellularly with single-molecule resolution. Comprehensive reviews of this literature have been published,<sup>79,80</sup> and we have highlighted advances in this field and the related chemical tagging field previously.<sup>81</sup> Here we focus on the significant challenge of efficient and robust fluorescent labeling of intracellular proteins in mammalian cells, highlighting the systematic comparison of fluorescent UAA labeling by bio-orthogonal click chemistry technologies from Peng and Hang.<sup>82</sup> While there has been success in labeling intracellular proteins in live mammalian cells using bio-orthogonal strain-promoted click chemistry,<sup>83,84</sup> the majority of published work still focuses on cell surface proteins. Peng and Hang wanted to observe the localization and trafficking of the small membrane-associated protein interferon-inducible transmembrane protein 3 (IFITM3) in cells. Given that GFP is twice the size of IFITM3, they took advantage of the small size and modularity of organic fluorophores afforded by the UAA technology. Site-specific incorporation of the commercially available *trans*-cyclooct-2-ene lysine (2'-aTCOK) into IFITM3 by amber suppression with the *Mm*PylRS(Y306A, Y384F)AF/tRNA pair<sup>85–87</sup> led to the most efficient and specific labeling with monosubstituted tetrazine (Tz) fluorophores, e.g., tetrazine silicon rhodamine (H-Tz-SiR), in live mammalian cells (Figure 7).<sup>82</sup> H-Tz-SiR was the best all around for intracellular labeling; H-Tz-BODIPY-FL was good but is more appropriate for membrane protein labeling due to its relative hydrophobicity.<sup>82</sup>



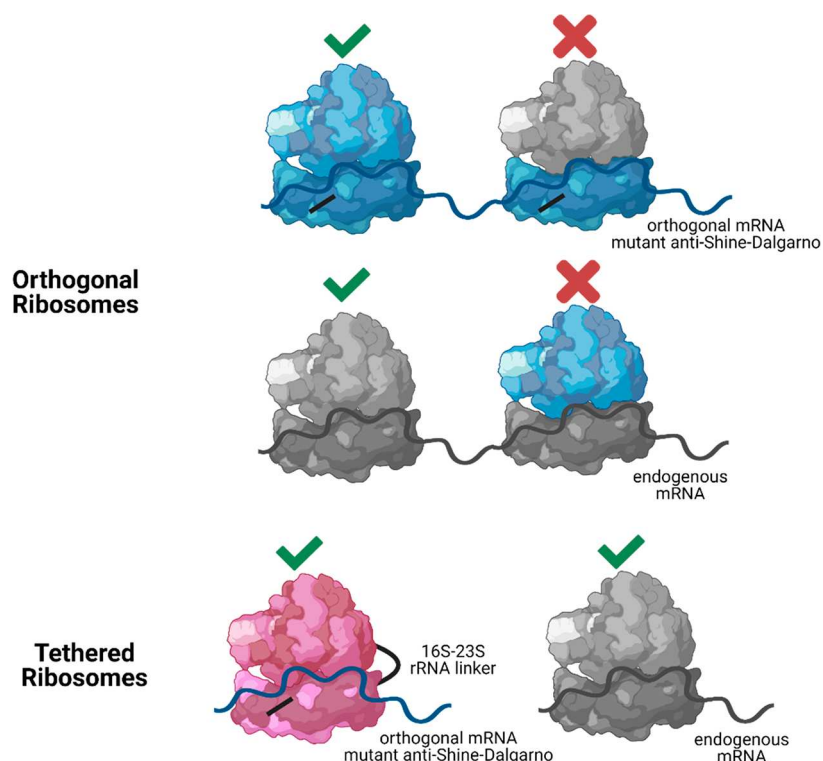
**Figure 7.** Ideal UAA incorporation components and fluorescent click reagents for bio-orthogonal intracellular protein labeling in mammalian cells.

The alternative to bio-orthogonal click chemistry labeling is direct incorporation of a fluorescent UAA. The benefit of this approach is ease. All that is required is transfection of the aaRS/tRNA and target constructs, incubation with the UAA, and washing out excess UAA. The drawbacks are the limitations on the size of the fluorescent side chain due to the aaRS binding pocket and, thus, the tendency for these fluorophores to be relatively blue-shifted and dim. Nonetheless, site-specific fluorescent labeling of proteins with fluorescent UAAs is built on the foundation of direct incorporation of a single fluorescent amino acid. Jan and Cohen established the state of the art with their incorporation of environmentally sensitive fluorescent UAA Aladan into the B1 domain of streptococcal protein G (GB1) by solid-phase synthesis.<sup>88</sup> The small size, the flexibility of the incorporation site, and the keen environmental sensitivity of Aladan provide unparalleled spatial resolution for detecting the electrostatic properties of various regions of this protein.<sup>88</sup>

*In vivo* incorporation of Aladan analogues, such as Anap, was enabled by the establishment of aaRS/tRNA pairs for this UAA, derived from *E. coli* LeuRS/tRNA<sub>CUA</sub>.<sup>89</sup> Anap undergoes a blue-shift in fluorescence emission in increasingly hydrophobic environments, making it promising as a sensor of protein conformational change. Shandell, Cornish, and Kass demonstrated the feasibility of sensing the conformational change of a population of UAA-modified cardiac voltage-gated sodium channels expressed in live mammalian cells through incorporation of Anap into the inactivation gate, a dynamic  $\sim 50$ -amino acid intracellular linker.<sup>90</sup> Since the foundational work of Dougherty and Lester, ion channel physiologists have embraced the UAA technology in oocytes.<sup>91,92</sup> Significantly, Puljung and co-workers incorporated Anap into  $K_{\text{ATP}}$  channels in live mammalian cells, enabling voltage-clamp fluorimetry experiments in this new, possibly more physiologically relevant context.<sup>93</sup> Ligand binding or conformational change measured by Anap environmental sensitivity can now be coupled to functional changes in channel gating measured by electrophysiology in mammalian cells.<sup>93</sup> More impactful applications of the technology are being published as technical challenges are overcome and efficiency and ease of use improved. When multiple fluorophores with single-molecule resolution can be incorporated selectively into mammalian cells, the UAA technology will be a powerful tool for studying biological mechanism in living cells.<sup>94</sup>

## NEW CONCEPTUAL APPLICATIONS OF UAA INCORPORATION

Wonderfully creative, the UAA technology will inspire myriad new directions. One particularly expansive direction is to engineer not only the amino acids but also the nucleic acids, organelles, and other parts of the cell to give rise to unnatural, chimeric, and semisynthetic organisms (SSOs). In some ways, this is the counterpart to building an artificial cell ground up from artificial RNA parts.<sup>95</sup> To model biological conditions that could explain the transition from an RNA world, Schultz and co-workers engineered chimeric *E. coli* in which 40% ribonucleotide versus deoxyribonucleotide could be incorporated into the genome when the size of the pool of deoxyribonucleotide triphosphates in the cell was significantly decreased in concert with defects in DNA repair.<sup>96</sup> In a similar fashion, Schultz and co-workers modeled the central hypothesis of endosymbiotic theory that mitochondria could have evolved from prokaryotes entering host cells and being maintained as endosymbionts. They engineered chimeras of *E. coli* and *S. cerevisiae* in which



**Figure 8.** Bio-orthogonal translation with orthogonal and tethered ribosomes.

mutant *E. coli* live in the cytosol of and provide ATP to a respiration-deficient yeast mutant or yeast provide thiamin to a resident *E. coli* thiamin auxotroph.<sup>97</sup> SSOs can be generated by codons containing unnatural base pairs or through the engineering of sense codon usage. Zhang et al. described an optimized SSO that stores genetic information using DNA containing two additional letters, which form a third, unnatural base pair, dNaM (mRNA codon) and dTPT3 (tRNA anticodon).<sup>98</sup> This expanded genetic code enables decoding of new codons to direct site-specific incorporation of UAAs into proteins in *E. coli*.<sup>98</sup> Recently, the Chin lab announced the creation of a SSO with a 61-codon genome. Creation of such a synthetic, minimally recoded *E. coli* genome by compressing synonymous codons addresses origins of life questions and biological mechanism and is enormously useful for therapeutic applications.<sup>99</sup>

**Mechanistic Studies of Translation with UAAs.** Somewhat surprisingly, there has been little integration of the incredible advances in our understanding of the structure and mechanism of the translational machinery and the UAA technology since the breakthrough publication by Noren et al. (Schultz) in 1989. It has often been thought that natural limitations of the translational machinery underlie the difficulty in incorporation of D- and  $\beta$ -amino acids. It turned out that the translational machinery may be more tolerant to unnatural substrates than previously thought. Mechanistic insight into how these analogues interfere with translational machinery was needed. Leyh, Gonzalez, Cornish, and co-workers further clarified the mechanism by which D-amino acids disrupt translation in a purified translation system, finding that while D-aa-tRNA can be accepted at the A site, act as a peptidyl-transfer acceptor, and translocate the peptidyl-D-aa-tRNA into the P site, this process occurs slowly.<sup>100</sup> Furthermore, the peptidyl-D-aa-tRNA at the P site partitions ribosomes into

arrested and non-arrested subpopulations. Chemical protection and molecular dynamics simulations demonstrated that P-site-bound peptidyl-D-aa-tRNA traps the PTC in a conformation that is not conducive to peptidyl transfer, providing insight into how the ribosome discriminates between L- and D-amino acids.<sup>100</sup> This mechanism of discrimination against D-amino acids appears to be similar to other peptide stalling mechanisms of the ribosome and may suggest the mechanism by which it discriminates against UAAs generally. Further mechanistic work with UAAs no doubt would lead to additional mechanistic surprises and could significantly inform efforts to improve the efficiency and breadth of the technology.

#### Engineering Multisite UAA Incorporation *In Vivo*.

Whether the objective is to incorporate different epigenetic modifications or to synthesize an unnatural backbone oligomer, it will be critical to be able to incorporate multiple different UAAs *in vivo* using alternate codons. Yields of UAA-incorporated proteins expressed in mammalian cells are significantly lower than those expressed in *E. coli*. Thus, it is an uphill battle to yield multiply UAA incorporated proteins in mammalian cells. In 2013, Chatterjee et al. (Schultz) developed a baculovirus-based delivery system for efficient incorporation of UAAs into proteins in mammalian cells.<sup>101</sup> Later that year, the same mammalian suppression system was applied to incorporate two distinct UAAs (eBK and OMeY) into EGFP in HEK293T cells using TAG and TAA suppression with *EcTyrRS*/tRNA<sub>CUA</sub> and *MbPylRS*/*MmtRNA*<sub>UUA</sub> pairs, respectively. They also demonstrated the application of dual suppression to fluorescent labeling of antibody–drug conjugates (anti-Her2-IgG-nAF) purified from HEK293F cells.<sup>102</sup> In mammalian cells, nonsense suppression suffers from a low level of expression of suppressor tRNAs and competition with endogenous release factors seeking to truncate target protein. In 2011, Johnson and co-workers established that knockout of release factor 1 in *E. coli* enables



incorporation of UAAs at multiple TAG sites in the same gene.<sup>103</sup> In mammalian cells, by optimizing the PylRS/tRNA<sub>CUA</sub> expression system and engineering eukaryotic release factor 1, Schmied and co-workers were able to increase the yield of protein containing UAAs at three sites by 2–4-fold.<sup>104</sup> Multisite incorporation at unique codons selectively *in vivo* remains a challenge for the field.

**Engineering the Ribosome for Improved Incorporation of UAAs.** One strategy toward multisite UAA incorporation at unique codons is to take advantage of our growing structural and mechanistic understanding of the translational machinery to engineer the ribosome itself. When the toolbox is expanded to include both quadruplet and amber codons, the ability of the natural ribosome to decode such codons limits the efficiency of UAA incorporation and the resulting yield of UAA-incorporated protein. It has been an outstanding challenge incorporating multiple UAAs in a single protein even in *E. coli* expression systems due to the lack of several blank codons and mutually orthogonal aaRS/tRNA pairs. In a breakout publication, Rackham and Chin addressed this challenge by engineering an orthogonal ribosome in *E. coli* via an engineered duplicate, orthogonal Shine-Dalgarno mRNA sequence/16S small ribosomal subunit RNA pair.<sup>105</sup> Building on this work, Chin and co-workers have evolved an orthogonal ribosome for quadruplet and amber decoding.<sup>106</sup> Orthogonal ribosomes can be devoted to efficient decoding of alternative codons because they are directly targeted to a corresponding orthogonal mRNA and thus do not synthesize the proteome. Through the combined use of mutually orthogonal aaRS/tRNA pairs, the evolved ribosome, and corresponding orthogonal mRNA, two UAAs were incorporated into single recombinant fusion proteins in *E. coli*. The cross-linking UAA *p*-benzoylphenylalanine (Bpa) and click handle pAzF were incorporated into glutathione *S*-transferase (GST)-maltose binding protein in response to a quadruplet and amber codon. The alkyne N6-[(2-propynyloxy)carbonyl]lysine (CAK) and pAzF were incorporated into GST-calmodulin in response to a quadruplet and amber codon.<sup>106</sup> Incorporation of pAzF and CAK into calmodulin enabled formation of a triazole cross-link by copper-catalyzed click chemistry, demonstrating precise genetic control of protein conformation with the UAA technology.

Early strategies for achieving orthogonality involved the development of orthogonal aaRS/tRNA pairs, building of orthogonal genetic codes, and creation of orthogonal ribosome–mRNA pairs by engineering the 16S rRNA and complementary mRNA Shine-Dalgarno sequences.<sup>107</sup> However, the continuous exchange of the subunits of the ribosome still limits the establishment of complete orthogonality with native translation (Figure 8). To address this gap, Jewett and Mankin used a circular permutation approach to systematically generate linked 16S–23S rRNA variants that could assemble functional tethered ribosomes in cells.<sup>108</sup> They demonstrated that the engineered ribosome with inseparable tethered subunits (Ribo-T) is capable of supporting the growth of *E. coli* cells, wholly orthogonal, and does not interfere with wild-type ribosomes.<sup>108,109</sup> They demonstrated the unique utility of Ribo-T in studying dominant lethal mutations of rRNA, a nearly impossible task in other cell models.<sup>108</sup> The Chin lab used a similar approach to covalently link the small and large ribosomal subunits by RNA linkers.<sup>110,111</sup> The compatibility of the tethered ribosomal complexes with the multisite incorporation of UAAs was evaluated by Jewett and Mankin through the fluorescence analysis of a super folder GFP (sfGFP) variant

containing five TAG codons, finding that the tethered translation system is effective in incorporating five pAzF click handles into sfGFP.<sup>112</sup> What remains is to show incorporation of multiple different UAAs in a row, each in response to a different codon.

**Application of UAA Technology to Biomedicine.** The UAA technology is ideal for antibody–drug conjugate (ADC) generation and other applications in biomedicine. The problem is production of homogeneous titers of ADCs for targeted cancer chemotherapy. ADCs are anticancer therapies designed to target tumors with cytotoxic small molecule drugs. Bypassing normal tissues reduces the toxic side effects of chemotherapy. Site-specific conjugation of the small molecule drug to the antibody homogenizes the mixture of ADCs, thus providing more reliable pharmacokinetic properties, efficacy, and safety profiles. In 2014, Tian and co-workers produced gram per liter scale titers of UAA-incorporated ADCs from stable CHO cells using antibodies targeting common antigens on colorectal/gastric and breast cancers.<sup>113</sup> Ambrx, Inc., developed an anti-HER2 ADC product ARX788 using the UAA incorporation strategy. ARX788 is generated by the formation of a highly stable oxime bond between a noncleavable Amberstatin (AS269) drug linker and a ketone-bearing UAA, pAcF, which was incorporated into the primary sequence of the antibody through amber codon suppression. The Food and Drug Administration has granted fast track designation to ARX788 for the treatment of patients with advanced or metastatic HER2-positive breast cancer in 2021.<sup>114</sup>

UAAs can also be used for vaccine development. In 2014, Wang and co-workers developed a theoretically safe and effective HIV-1 vaccine by making viral replication dependent on the presence of UAA and the aaRS/tRNA pair.<sup>115</sup> In 2016, the Zhou lab developed a live attenuated influenza A vaccine strain containing multiple amber codons in its genome.<sup>116</sup> The strain can be replicated only in a transgenic 293T cell line that harbors an orthogonal amber suppressor aaRS/tRNA<sub>CUA</sub> pair and the cognate UAA, making it replication incompetent in normal human cells and thus useful for immunization.<sup>116</sup> Undoubtedly, the UAA technology can be exploited in other modalities for therapeutics.

**Studying Fundamental Biological Processes with UAA Technology.** Once multiple different UAAs can be incorporated at unique codons *in vivo* with no cross reactivity, the UAA technology will be a powerful tool for systems biology that enables biological mechanism to be studied in living cells. As an example, biologists would like to control various protein signaling states by turning on and off receptor–ligand interactions and intraprotein interactions. Such control would enable clear-cut conclusions about the functional effects of specific protein structures in the native cell environment. This was recently illustrated in a study by the Chen lab exploiting transition metal-based bio-orthogonal cleavage reactions for on-demand release of toxic drugs from ADCs and precise control of ligand–receptor interactions at the cell surface.<sup>117</sup> In this study, they used the genetic code expansion strategy to incorporate chemically caged Tyr and Lys analogues into eight different sites of Z<sub>Her2</sub>, a small protein with a high affinity for the membrane protein HER2. The UAA-modified Z<sub>Her2</sub> mutants were expressed in *E. coli*, purified, and fluorescently labeled. Fluorescent Z<sub>Her2</sub> and its UAA-modified mutants were incubated with SK-BR-3 cells, a human breast cancer cell line that overexpresses HER2, and analyzed by flow cytometry. Strong fluorescence was observed when wild-type Z<sub>Her2</sub> and HER2 interacted. A decrease



in fluorescence was observed with caged UAA-modified Z<sub>Her2</sub> mutants that could no longer interact. Fluorescence was rescued upon decaging to the native amino acid. This allowed them to directly probe the functional role of each amino acid residue in the interaction between Z<sub>Her2</sub> and Her2.<sup>117</sup> It should be emphasized that the UAA technology is at a sufficiently mature stage where it can be adopted by biologists not expert in the methodology with similar ease and with even more diverse applications, perhaps, than GFP.

## FUTURE DIRECTIONS

Considerable progress has been made in the field of genetic code expansion since the publication of the study by Chatterjee et al. in 2013. However, many technical challenges remain, and the application of the technology to drug development and basic science is only just beginning.

Significantly changing the structure of the amino acid and still having it be accepted as a substrate by the translational machinery remains difficult. For example, there are limitations to what backbone analogues can be incorporated for unnatural oligomer synthesis or powerful fluorophores for biophysical studies of proteins. There have been incredible advances in our understanding of ribosome structure and function since Schultz, Noren, and co-workers published their seminal paper in 1989. In collaboration, we demonstrated that this mechanistic knowledge could be exploited to gain insight into how the translational machinery discriminated the structure and electronics of the amino acid. More mechanistic work with UAAs is needed. This mechanistic work can provide insight into the mechanism of translation and guide future engineering efforts. It also is now possible to engineer translation and the ribosome at a scale not previously possible with GROs, stapled ribosomes *in vivo*, and a growing arsenal of UAAs. Together, these advances allow direct selection for orthogonal ribosomes that work with an expanded set of UAAs and codons.

Optimization has made expression and purification of UAA-incorporated proteins from *E. coli* a robust method accessible to scientists not expert in the field. Biochemists and biologists are encouraged to adopt the technology and apply it to a variety of mechanistic questions. With the ability to produce a broad range of UAA-containing proteins in *E. coli* in high yield and advances in bio-orthogonal labeling methods, the field is at an exciting moment to apply the technology to challenging problems in drug discovery and biotechnology. Antibody–drug conjugates and vaccine development likely are just the start for therapeutic applications of the technology. Undoubtedly, an important next step for the field will be synthesizing unnatural oligomers directly in *E. coli*.

The next challenge is to develop methods for incorporating multiple, different UAAs in response to different codons in live mammalian cells. This would allow mechanistic questions, like the role of different epigenetic modifications, to be addressed directly in the cell. Technically, this likely will require (1) using GROs so the UAAs are not incorporated at endogenous codons, (2) engineering tethered ribosomes in mammalian cells, and (3) evolving tethered ribosomes to work with different UAAs and different codons. If successful, like GFP before it, UAAs could be a powerful tool for studying biological mechanism in live cells, but with a much broader repertoire of chemical functionality.

## CONCLUSION

The incorporation of unnatural amino acids by the translational machinery using artificial UAA-tRNAs is a wonderfully creative idea bringing together organic chemistry, molecular biology, and synthetic biology of the translational machinery. It is now a robust field with many who trained in the technology leading their own exciting advances for the tools in the UAA repertoire, in the aaRS/tRNA orthogonal pairs, in moving to higher organisms, and even in tethered orthogonal ribosomes and *de novo* GROs. Looking forward, as the tools continue to progress, UAA mutagenesis no doubt will become an essential tool for asking fundamental questions in systems biology and will be further adapted as a new strategy for drug development.

## AUTHOR INFORMATION

### Corresponding Authors

Mia A. Shandell — York Structural Biology Laboratory, University of York, York YO10 5DD, U.K.; [orcid.org/0000-0002-7864-758X](https://orcid.org/0000-0002-7864-758X); Email: [mia.shandell@york.ac.uk](mailto:mia.shandell@york.ac.uk)

Zhongping Tan — State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China; [orcid.org/0000-0002-9302-150X](https://orcid.org/0000-0002-9302-150X); Email: [zhongping.tan@imm.pumc.edu.cn](mailto:zhongping.tan@imm.pumc.edu.cn)

Virginia W. Cornish — Department of Chemistry and Department of Systems Biology, Columbia University, New York 10027, United States; Email: [vc114@columbia.edu](mailto:vc114@columbia.edu)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.biochem.1c00286>

### Funding

Z.T. is funded by The National Key R&D Program of China (Grant 2018YFE0111400), the Training Program of the Major Research Plan of National Natural Science Foundation of China (Grant 91853120), the National Major Scientific and Technological Special Project of China (Grants 2018ZX09711001-005 and 2018ZX09711001-013), and the National Institutes of Health (NIH) Research Project Grant Program (R01 EB025892). V.W.C. is funded by the NIH Research Project Grant Program (R01 GM090126) and Columbia University.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank Lei Wang for insightful comments on the manuscript. V.W.C. thanks P. G. Schultz for the opportunity to earn her Ph.D. working for him on the UAA technology. Figures were created with [BioRender.com](https://BioRender.com).

## REFERENCES

- (1) Cornish, V. W., Mendel, D., and Schultz, P. G. (1995) Probing Protein Structure and Function with an Expanded Genetic Code. *Angew. Chem., Int. Ed. Engl.* 34, 621–633.
- (2) Liu, C. C., and Schultz, P. G. (2010) Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* 79, 413–444.
- (3) de la Torre, D., and Chin, J. W. (2021) Reprogramming the genetic code. *Nat. Rev. Genet.* 22, 169–184.
- (4) Wang, L. (2017) Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. *Acc. Chem. Res.* 50, 2767–2775.

- (5) 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.
- (6) Guo, J., Melançon, C. E., 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.* 48, 9148–9151.
- (7) Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010) An Enhanced System for Unnatural Amino Acid Mutagenesis in *E. coli*. *J. Mol. Biol.* 395, 361–374.
- (8) Crick, F. H. (1958) On protein synthesis. *Symp. Soc. Exp. Biol.* 12, 138–163.
- (9) Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., and Zamecnik, P. C. (1958) A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* 231, 241–257.
- (10) Chapeville, F., Lipmann, F., von Ehrenstein, G., Weisblum, B., Ray, W. J., and Benzer, S. (1962) On the Role of Soluble Ribonucleic Acid in Coding for Amino Acids. *Proc. Natl. Acad. Sci. U. S. A.* 48, 1086–1092.
- (11) Calendar, R., and Berg, P. (1967) D-Tyrosyl RNA: formation, hydrolysis and utilization for protein synthesis. *J. Mol. Biol.* 26, 39–54.
- (12) Fahnstock, S., and Rich, A. (1971) Ribosome-Catalyzed Polyester Formation. *Science* 173, 340–343.
- (13) Johnson, A. E., Woodward, W. R., Herbert, E., and Menninger, J. R. (1976) Nepsilon-acetyllysine transfer ribonucleic acid: a biologically active analogue of aminoacyl transfer ribonucleic acids. *Biochemistry* 15, 569–575.
- (14) Hecht, S. M., Alford, B. L., Kuroda, Y., and Kitano, S. (1978) Chemical aminoacylation of tRNA's. *J. Biol. Chem.* 253, 4517–4520.
- (15) Heckler, T. G., Zama, Y., Naka, T., and Hecht, S. M. (1983) Dipeptide formation with misacylated tRNAPheS. *J. Biol. Chem.* 258, 4492–4495.
- (16) Baldini, G., Martoglio, B., Schachenmann, A., Zugliani, C., and Brunner, J. (1988) Mischarging *Escherichia coli* tRNAPhe with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]phenylalanine, a photoactivatable analogue of phenylalanine. *Biochemistry* 27, 7951–7959.
- (17) 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.
- (18) 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.
- (19) Heckler, T. G., Chang, L. H., Zama, Y., Naka, T., and Hecht, S. M. (1984) Preparation of 2,3-O-Acyl-pCpA derivatives as substrates for T4 RNA ligase-mediated "chemical aminoacylation. *Tetrahedron* 40, 87–94.
- (20) 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.
- (21) Bain, J. D., Diala, E. S., Glabe, C. G., Wacker, D. A., Lyttle, M. H., Dix, T. A., and Chamberlin, A. R. (1991) Site-specific incorporation of nonnatural residues during in vitro protein biosynthesis with semi-synthetic aminoacyl-tRNAs. *Biochemistry* 30, 5411–5421.
- (22) Robertson, S. A., Ellman, J. A., and Schultz, P. G. (1991) A general and efficient route for chemical aminoacylation of transfer RNAs. *J. Am. Chem. Soc.* 113, 2722–2729.
- (23) Bartel, D. P., and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* 261, 1411–1418.
- (24) Lohse, P. A., and Szostak, J. W. (1996) Ribozyme-catalysed amino acid transfer reactions. *Nature* 381, 442–444.
- (25) Lee, N., Bessho, Y., Wei, K., Szostak, J. W., and Suga, H. (2000) Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.* 7, 28–33.
- (26) 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.
- (27) Shimizu, Y., Kanamori, T., and Ueda, T. (2005) Protein synthesis by pure translation systems. *Methods* 36, 299–304.
- (28) Ellman, J. A., Mendel, D., and Schultz, P. G. (1992) Site-specific incorporation of novel backbone structures into proteins. *Science* 255, 197–200.
- (29) Judice, J. K., Gamble, T. R., Murphy, E. C., de Vos, A., and Schultz, P. G. (1993) Probing the mechanism of staphylococcal nuclease with unnatural amino acids: kinetic and structural studies. *Science* 261, 1578–1581.
- (30) Cornish, V. W., Benson, D. R., Altenbach, C. A., Hideg, K., Hubbell, W. L., and Schultz, P. G. (1994) Site-specific incorporation of biophysical probes into proteins. *Proc. Natl. Acad. Sci. U. S. A.* 91, 2910–2914.
- (31) Cornish, V. W., Hahn, K. M., and Schultz, P. G. (1996) Site-Specific Protein Modification Using a Ketone Handle. *J. Am. Chem. Soc.* 118, 8150–8151.
- (32) Ma, C., Kudlicki, W., Odom, O. W., Kramer, G., and Hardesty, B. (1993) In vitro protein engineering using synthetic tRNA(Ala) with different anticodons. *Biochemistry* 32, 7939–7945.
- (33) Hoshaka, T., Sato, K., Sisido, M., Takai, K., and Yokoyama, S. (1994) Site-specific incorporation of photofunctional nonnatural amino acids into a polypeptide through in vitro protein biosynthesis. *FEBS Lett.* 344, 171–174.
- (34) Hoshaka, 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.
- (35) Murakami, H., Hoshaka, 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.
- (36) Hoshaka, 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.
- (37) Hoshaka, T., Ashizuka, Y., Taira, H., Murakami, H., and Sisido, M. (2001) Incorporation of nonnatural amino acids into proteins by using various four-base codons in an *Escherichia coli* in vitro translation system. *Biochemistry* 40, 11060–11064.
- (38) Anderson, R. D., Zhou, J., and Hecht, S. M. (2002) Fluorescence Resonance Energy Transfer between Unnatural Amino Acids in a Structurally Modified Dihydrofolate Reductase. *J. Am. Chem. Soc.* 124, 9674–9675.
- (39) Forster, A. C., Tan, Z., Nalam, M. N. L., Lin, H., Qu, H., Cornish, V. W., and Blacklow, S. C. (2003) Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6353–6357.
- (40) Josephson, K., Hartman, M. C. T., and Szostak, J. W. (2005) Ribosomal Synthesis of Unnatural Peptides. *J. Am. Chem. Soc.* 127, 11727–11735.
- (41) Ohta, A., Murakami, H., Higashimura, E., and Suga, H. (2007) Synthesis of Polyester by Means of Genetic Code Reprogramming. *Chem. Biol.* 14, 1315–1322.
- (42) 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.
- (43) Katoh, T., Tajima, K., and Suga, H. (2017) Consecutive Elongation of D-Amino Acids in Translation. *Cell Chem. Biol.* 24, 46–54.
- (44) Katoh, T., and Suga, H. (2018) Ribosomal Incorporation of Consecutive  $\beta$ -Amino Acids. *J. Am. Chem. Soc.* 140, 12159–12167.
- (45) Lummis, S. C. R., Beene, D. L., Lee, L. W., Lester, H. A., Broadhurst, R. W., and Dougherty, D. A. (2005) Cis – trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* 438, 248–252.
- (46) England, P. M., Lester, H. A., and Dougherty, D. A. (1999) Mapping Disulfide Connectivity Using Backbone Ester Hydrolysis. *Biochemistry* 38, 14409–14415.
- (47) Rodriguez, E. A., Lester, H. A., and Dougherty, D. A. (2006) In vivo incorporation of multiple unnatural amino acids through nonsense

and frameshift suppression. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8650–8655.

(48) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the Genetic Code of *Escherichia coli*. *Science* 292, 498–500.

(49) Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* 8, 883–890.

(50) Liu, D. R., and Schultz, P. G. (1999) Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4780–4785.

(51) Wang, W., Takimoto, J. K., Louie, G. V., Baiga, T. J., Noel, J. P., Lee, K.-F., Slesinger, P. A., and Wang, L. (2007) Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci.* 10, 1063–1072.

(52) Wang, L. (2017) Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. *Acc. Chem. Res.* 50, 2767–2775.

(53) 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.

(54) Liu, W., Brock, A., Chen, S., Chen, S., and Schultz, P. G. (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* 4, 239–244.

(55) Greiss, S., and Chin, J. W. (2011) Expanding the Genetic Code of an Animal. *J. Am. Chem. Soc.* 133, 14196–14199.

(56) Bianco, A., Townsley, F. M., Greiss, S., Lang, K., and Chin, J. W. (2012) Expanding the genetic code of *Drosophila melanogaster*. *Nat. Chem. Biol.* 8, 748–750.

(57) Ernst, R. J., Krogager, T. P., Maywood, E. S., Zanchi, R., Beránek, V., Elliott, T. S., Barry, N. P., Hastings, M. H., and Chin, J. W. (2016) Genetic code expansion in the mouse brain. *Nat. Chem. Biol.* 12, 776–778.

(58) Kang, J.-Y., Kawaguchi, D., Coin, I., Xiang, Z., O'Leary, D. D. M., Slesinger, P. A., and Wang, L. (2013) In Vivo Expression of a Light-Activatable Potassium Channel Using Unnatural Amino Acids. *Neuron* 80, 358–370.

(59) Chen, Y., Ma, J., Lu, W., Tian, M., Thauvin, M., Yuan, C., Volovitch, M., Wang, Q., Holst, J., Liu, M., Vriz, S., Ye, S., Wang, L., and Li, D. (2017) Heritable expansion of the genetic code in mouse and zebrafish. *Cell Res.* 27, 294–297.

(60) Parrish, A. R., She, X., Xiang, Z., Coin, I., Shen, Z., Briggs, S. P., Dillin, A., and Wang, L. (2012) Expanding the genetic code of *Caenorhabditis elegans* using bacterial aminoacyl-tRNA synthetase/tRNA pairs. *ACS Chem. Biol.* 7, 1292–1302.

(61) Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* 103, 9482–9487.

(62) Dieterich, D. C., Hodas, J. J. L., Gouzer, G., Shadrin, I. Y., Ngo, J. T., Triller, A., Tirrell, D. A., and Schuman, E. M. (2010) In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.* 13, 897–905.

(63) Takimoto, J. K., Xiang, Z., Kang, J.-Y., and Wang, L. (2010) Esterification of an Unnatural Amino Acid Structurally Deviating from Canonical Amino Acids Promotes Its Uptake and Incorporation into Proteins in Mammalian Cells. *ChemBioChem* 11, 2268–2272.

(64) Takimoto, J. K., Adams, K. L., Xiang, Z., and Wang, L. (2009) Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol. Biosyst.* 5, 931.

(65) Melo Czekster, C., Robertson, W. E., Walker, A. S., Söll, D., and Schepartz, A. (2016) In Vivo Biosynthesis of a  $\beta$ -Amino Acid-Containing Protein. *J. Am. Chem. Soc.* 138, 5194–5197.

(66) Dedkova, L. M., Fahmi, N. E., Paul, R., del Rosario, M., Zhang, L., Chen, S., Feder, G., and Hecht, S. M. (2012)  $\beta$ -Puromycin selection of modified ribosomes for in vitro incorporation of  $\beta$ -amino acids. *Biochemistry* 51, 401–415.

(67) 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.

(68) Fujino, T., and Murakami, H. (2016) In Vitro Selection Combined with Ribosomal Translation Containing Non-proteinogenic Amino Acids. *Chem. Rec.* 16, 365–377.

(69) 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.

(70) Park, H.-S., Hohn, M. J., Umehara, T., Guo, L.-T., Osborne, E. M., Benner, J., Noren, C. J., Rinehart, J., and Soll, D. (2011) Expanding the Genetic Code of *Escherichia coli* with Phosphoserine. *Science* 333, 1151–1154.

(71) Luo, X., Fu, G., Wang, R. E., Zhu, X., Zambaldo, C., Liu, R., Liu, T., Lyu, X., Du, J., Xuan, W., Yao, A., Reed, S. A., Kang, M., Zhang, Y., Guo, H., Huang, C., Yang, P.-Y., Wilson, I. A., Schultz, P. G., and Wang, F. (2017) Genetically encoding phosphotyrosine and its non-hydrolyzable analog in bacteria. *Nat. Chem. Biol.* 13, 845–849.

(72) Hoppmann, C., Wong, A., Yang, B., Li, S., Hunter, T., Shokat, K. M., and Wang, L. (2017) Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nat. Chem. Biol.* 13, 842–844.

(73) Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M., and Isaacs, F. J. (2013) Genomically Recoded Organisms Expand Biological Functions. *Science* 342, 357–360.

(74) Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M. G., Moosburner, M., Shrock, E., Pruitt, B. W., Conway, N., Goodman, D. B., Gardner, C. L., Tyree, G., Gonzales, A., Wanner, B. L., Norville, J. E., Lajoie, M. J., and Church, G. M. (2016) Design, synthesis, and testing toward a 57-codon genome. *Science* 353, 819–822.

(75) Barber, K. W., Muir, P., Szeligowski, R. V., Rogulina, S., Gerstein, M., Sampson, J. R., Isaacs, F. J., and Rinehart, J. (2018) Encoding human serine phosphopeptides in bacteria for proteome-wide identification of phosphorylation-dependent interactions. *Nat. Biotechnol.* 36, 638–644.

(76) Liu, X., Xiao, W., Zhang, Y., Wiley, S. E., Zuo, T., Zheng, Y., Chen, N., Chen, L., Wang, X., Zheng, Y., Huang, L., Lin, S., Murphy, A. N., Dixon, J. E., Xu, P., and Guo, X. (2020) Reversible phosphorylation of Rpn1 regulates 26S proteasome assembly and function. *Proc. Natl. Acad. Sci. U. S. A.* 117, 328–336.

(77) Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaitė, A., Hancock, S. M., Huguenin-Dezot, N., Muqit, M. M. K., Fry, A. M., Bayliss, R., and Chin, J. W. (2015) Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* 11, 496–503.

(78) Beránek, V., Reinkemeier, C. D., Zhang, M. S., Liang, A. D., Kym, G., and Chin, J. W. (2018) Genetically Encoded Protein Phosphorylation in Mammalian Cells. *Cell Chem. Biol.* 25, 1067–1074 e5.

(79) Xiao, H., and Schultz, P. G. (2016) At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harbor Perspect. Biol.* 8, a023945.

(80) Young, D. D., and Schultz, P. G. (2018) Playing with the Molecules of Life. *ACS Chem. Biol.* 13, 854–870.

(81) Romanini, D. W., and Cornish, V. W. (2012) Playing tag with proteins. *Nat. Chem.* 4, 248–250.

(82) Peng, T., and Hang, H. C. (2016) Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J. Am. Chem. Soc.* 138, 14423–14433.

(83) Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A., and Chin, J. W. (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat. Chem.* 4, 298–304.

(84) Uttamapinant, C., Howe, J. D., Lang, K., Beránek, V., Davis, L., Mahesh, M., Barry, N. P., and Chin, J. W. (2015) Genetic Code Expansion Enables Live-Cell and Super-Resolution Imaging of Site-Specifically Labeled Cellular Proteins. *J. Am. Chem. Soc.* 137, 4602–4605.



- (85) Nikić, I., Plass, T., Schraidt, O., Szymański, J., Briggs, J. A. G., Schultz, C., and Lemke, E. A. (2014) Minimal Tags for Rapid Dual-Color Live-Cell Labeling and Super-Resolution Microscopy. *Angew. Chem., Int. Ed.* 53, 2245–2249.
- (86) Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzoyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem. Biol.* 15, 1187–1197.
- (87) Yanagisawa, T., Hino, N., Iraha, F., Mukai, T., Sakamoto, K., and Yokoyama, S. (2012) Wide-range protein photo-crosslinking achieved by a genetically encoded Nε-(benzoyloxycarbonyl)lysine derivative with a diazirinyl moiety. *Mol. Biosyst.* 8, 1131.
- (88) Cohen, B. E., McAnaney, T. B., Park, E. S., Jan, Y. N., Boxer, S. G., and Jan, L. Y. (2002) Probing Protein Electrostatics with a Synthetic Fluorescent Amino Acid. *Science* 296, 1700–1703.
- (89) Chatterjee, A., Guo, J., Lee, H. S., and Schultz, P. G. (2013) A Genetically Encoded Fluorescent Probe in Mammalian Cells. *J. Am. Chem. Soc.* 135, 12540–12543.
- (90) Shandell, M. A., Quejada, J. R., Yazawa, M., Cornish, V. W., and Kass, R. S. (2019) Detection of Nav1.5 Conformational Change in Mammalian Cells Using the Noncanonical Amino Acid ANAP. *Biophys. J.* 117, 1352–1363.
- (91) Gordon, S. E., Senning, E. N., Aman, T. K., and Zagotta, W. N. (2016) Transition metal ion FRET to measure short-range distances at the intracellular surface of the plasma membrane. *J. Gen. Physiol.* 147, 189–200.
- (92) Zagotta, W. N., Gordon, M. T., Senning, E. N., Munari, M. A., and Gordon, S. E. (2016) Measuring distances between TRPV1 and the plasma membrane using a noncanonical amino acid and transition metal ion FRET. *J. Gen. Physiol.* 147, 201–216.
- (93) Puljung, M., Vedovato, N., Usher, S., and Ashcroft, F. (2019) Activation mechanism of ATP-sensitive K<sup>+</sup> channels explored with real-time nucleotide binding. *eLife* 8, e41103.
- (94) Hu, X., Jing, C., Xu, X., Nakazawa, N., Cornish, V. W., Margadant, F. M., and Sheetz, M. P. (2016) Cooperative Vinculin Binding to Talin Mapped by Time-Resolved Super Resolution Microscopy. *Nano Lett.* 16, 4062–4068.
- (95) Blain, J. C., and Szostak, J. W. (2014) Progress Toward Synthetic Cells. *Annu. Rev. Biochem.* 83, 615–640.
- (96) Mehta, A. P., Wang, Y., Reed, S. A., Supekova, L., Javahishvili, T., Chaput, J. C., and Schultz, P. G. (2018) Bacterial Genome Containing Chimeric DNA–RNA Sequences. *J. Am. Chem. Soc.* 140, 11464–11473.
- (97) Mehta, A. P., Supekova, L., Chen, J.-H., Pestonjamas, K., Webster, P., Ko, Y., Henderson, S. C., McDermott, G., Supek, F., and Schultz, P. G. (2018) Engineering yeast endosymbionts as a step toward the evolution of mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 115, 11796–11801.
- (98) Zhang, Y., Lamb, B. M., Feldman, A. W., Zhou, A. X., Lavergne, T., Li, L., and Romesberg, F. E. (2017) A semisynthetic organism engineered for the stable expansion of the genetic alphabet. *Proc. Natl. Acad. Sci. U. S. A.* 114, 1317–1322.
- (99) Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., Chia, T., Schmied, W. H., Dunkelmann, D. L., Beránek, V., Uttamapinant, C., Llamazares, A. G., Elliott, T. S., and Chin, J. W. (2019) Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 569, 514–518.
- (100) Englander, M. T., Avins, J. L., Fleisher, R. C., Liu, B., Effraim, P. R., Wang, J., Schulten, K., Leyh, T. S., Gonzalez, R. L., and Cornish, V. W. (2015) The ribosome can discriminate the chirality of amino acids within its peptidyl-transferase center. *Proc. Natl. Acad. Sci. U. S. A.* 112, 6038–6043.
- (101) Chatterjee, A., Xiao, H., Bollong, M., Ai, H.-W., and Schultz, P. G. (2013) Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11803–11808.
- (102) Xiao, H., Chatterjee, A., Choi, S., Bajjuri, K. M., Sinha, S. C., and Schultz, P. G. (2013) Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells. *Angew. Chem., Int. Ed.* 52, 14080–14083.
- (103) Johnson, D. B. F., Xu, J., Shen, Z., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Xiang, Z., Ecker, J. R., Briggs, S. P., and Wang, L. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* 7, 779–786.
- (104) Schmied, W. H., Elsässer, S. J., Uttamapinant, C., and Chin, J. W. (2014) Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1. *J. Am. Chem. Soc.* 136, 15577–15583.
- (105) Rackham, O., and Chin, J. W. (2005) A network of orthogonal ribosome-mRNA pairs. *Nat. Chem. Biol.* 1, 159–166.
- (106) 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.
- (107) Hammerling, M. J., Fritz, B. R., Yoesep, D. J., Kim, D. S., Carlson, E. D., and Jewett, M. C. (2020) In vitro ribosome synthesis and evolution through ribosome display. *Nat. Commun.* 11, 1108.
- (108) 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.
- (109) Aleksashin, N. A., Szal, T., d'Aquino, A. E., Jewett, M. C., Vázquez-Laslop, N., and Mankin, A. S. (2020) A fully orthogonal system for protein synthesis in bacterial cells. *Nat. Commun.* 11, 1858.
- (110) Fried, S. D., Schmied, W. H., Uttamapinant, C., and Chin, J. W. (2015) Ribosome Subunit Stapling for Orthogonal Translation in *E. coli*. *Angew. Chem., Int. Ed.* 54, 12791–12794.
- (111) 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.
- (112) Carlson, E. D., d'Aquino, A. E., Kim, D. S., Fulk, E. M., Hoang, K., Szal, T., Mankin, A. S., and Jewett, M. C. (2019) Engineered ribosomes with tethered subunits for expanding biological function. *Nat. Commun.* 10, 3920.
- (113) Tian, F., Lu, Y., Manibusan, A., Sellers, A., Tran, H., Sun, Y., Phuong, T., Barnett, R., Hehli, B., Song, F., DeGuzman, M. J., Ensari, S., Pinkstaff, J. K., Sullivan, L. M., Biroc, S. L., Cho, H., Schultz, P. G., DiJoseph, J., Dougher, M., Ma, D., Dushin, R., Leal, M., Tchistiakova, L., Feyfant, E., Gerber, H.-P., and Sapra, P. (2014) A general approach to site-specific antibody drug conjugates. *Proc. Natl. Acad. Sci. U. S. A.* 111, 1766–1771.
- (114) Kang, M., Lu, Y., Chen, S., and Tian, F. (2018) Harnessing the power of an expanded genetic code toward next-generation biopharmaceuticals. *Curr. Opin. Chem. Biol.* 46, 123–129.
- (115) Wang, N., Li, Y., Niu, W., Sun, M., Cerny, R., Li, Q., and Guo, J. (2014) Construction of a Live-Attenuated HIV-1 Vaccine through Genetic Code Expansion. *Angew. Chem.* 126, 4967–4971.
- (116) Si, L., Xu, H., Zhou, X., Zhang, Z., Tian, Z., Wang, Y., Wu, Y., Zhang, B., Niu, Z., Zhang, C., Fu, G., Xiao, S., Xia, Q., Zhang, L., and Zhou, D. (2016) Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science* 354, 1170–1173.
- (117) Wang, X., Liu, Y., Fan, X., Wang, J., Ngai, W. S. C., Zhang, H., Li, J., Zhang, G., Lin, J., and Chen, P. R. (2019) Copper-Triggered Bioorthogonal Cleavage Reactions for Reversible Protein and Cell Surface Modifications. *J. Am. Chem. Soc.* 141, 17133–17141.