Journal of Zhejiang University-SCIENCE B in press
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Review

https://doi.org/10.1631/jzus.B2600331

Sense codon reassignment: from global proteome profiling to precision protein engineering

Yiru ZHANG[1*], Wei YU[1*], Wenlong DING[2✉], Shixian LIN[1,2,3✉]

[1]*Zhejiang Key Laboratory of Molecular Cancer Biology, Life Sciences Institute, Zhejiang University, Hangzhou 310058, China*
[2]*Department of Respiratory and Critical Care Medicine, Center for Oncology Medicine, the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu 322000, China*
[3]*Institute of Fundamental and Transdisciplinary Research, Zhejiang University, Hangzhou 310058, China*

Abstract: The incorporation of non-canonical amino acids (ncAAs) has expanded the chemical diversity and functional repertoire of proteins beyond the 20 canonical amino acids. While stop codon suppression introduces a single orthogonal decoding event, sense codon reassignment actively manipulates endogenous translational competition within the ribosomal decoding center. This strategy enables both proteome-wide stochastic labeling and precise site-specific modification of target proteins. In this review, we highlight recent advances in sense codon recoding strategies and their applications in nascent proteome profiling and precision protein engineering. We first discuss global metabolic labeling using endogenous aminoacyl-tRNA synthetases and the programmable Stochastic Orthogonal Recoding of Translation (SORT) platform. We then examine site-specific approaches, including competitive rare codon recoding in mammalian cells, genome-scale codon compression in *Escherichia coli*, and the creation of blank coding channels through total genome synthesis. These complementary paradigms have enabled high-resolution spatiotemporal mapping of nascent proteomes in live animals, multiplexed incorporation of up to five distinct ncAAs in a single mammalian protein, and the *in vivo* biosynthesis of complex sequence-defined polymers. By converting translational competition into an engineerable parameter, sense codon reassignment offers new opportunities to transform our ability to decode complex biological networks and develop next-generation molecular medicines.

Key words: Sense codon reassignment; Genetic code expansion; Stochastic orthogonal recoding of translation; Rare codon; Orthogonal translation system

1 Introduction

The universal genetic code is among the most highly conserved features of biology. It uses 61 sense codons to encode the 20 canonical amino acids (cAAs) and 3 stop codons to terminate translation. Although this decoding system underpins biological inheritance and protein evolution, the chemical functionalities offered by the 20 cAAs are inherently limited (Huang *et al.*, 2025). As modern chemical biology seeks to interrogate complex cellular processes and engineer proteins with new chemical or biophysical properties, expanding the coding capacity of living systems has emerged as a central challenge.

✉ Shixian LIN, sxlin@zju.edu.cn
Wenlong DING, longwending@163.com
* The two authors contributed equally to this work.
Shixian LIN, https://orcid.org/0000-0001-9718-9986
Wenlong DING, https://orcid.org/0000-0001-8385-7488
Yiru ZHANG, https://orcid.org/0009-0008-0664-6845
WeiYu, https://orcid.org/0009-0002-2086-9830

Received May 11, 2026; Revision accepted Aug. 11, 2026;
Crosschecked xxx. xx, 20xx

Nature itself provides precedents for genetic code expansion through the reassignment of translational termination signals. In certain organisms, the stop codons UGA and UAG are naturally repurposed to encode selenocysteine and pyrrolysine, respectively (Chambers *et al.*, 1986; Srinivasan *et al.*, 2002). Inspired by these natural recoding events, extensive efforts have focused on reprogramming stop codons, particularly the amber codon, to enable the site-specific incorporation of noncanonical amino acids (ncAAs) (Liu and Schultz, 1999; Noren *et al.*, 1989). These stop codon suppression strategies have laid the technological foundation of modern genetic code expansion (GCE) and have driven transformative advances in protein engineering (Drienovská and Roelfes, 2020; Zhao *et al.*, 2023), chemical biology (Ding *et al.*, 2024b; Yi *et al.*, 2024), and therapeutic development (Axup *et al.*, 2012; Si *et al.*, 2016). Because stop codon suppression has matured into a well-established field and has been comprehensively reviewed elsewhere (Huang *et al.*, 2025; Niu and Guo, 2024; Osgood *et al.*, 2025), it is not the focus of this review.

Instead, we focus on the emerging field of sense codon reassignment. Beyond the three stop codons, the genetic code contains a much larger and largely untapped coding landscape represented by the 61 sense codons (Jones and Hartman, 2024). Reprogramming sense codons, however, presents a fundamentally distinct challenge. Unlike stop codons, which are normally excluded from translation elongation, sense codons are actively decoded by endogenous aminoacylated tRNAs. Consequently, sense codon reassignment requires the active manipulation of translational competition within the ribosomal decoding center. This competition-centric framework has given rise to two complementary paradigms: (1) harnessing endogenous competition to achieve stochastic proteome-wide incorporation of ncAAs for metabolic labeling and systems-level interrogation, and (2) overcoming or eliminating endogenous competition to establish dedicated decoding pathways for precise, site-specific protein modification.

In this review, we first examine strategies that exploit endogenous translational competition for global proteome profiling and dynamic interrogation of cellular processes. We then discuss advances in site-specific sense codon reassignment, focusing on competitive rare codon recoding, genome-scale codon compression, and the development of orthogonal recoding channels for multiplexed ncAA incorporation. Ultimately, we highlight how transforming translational competition from a biological constraint into an engineerable parameter is expanding the functional scope of the genetic code and opening versatile new opportunities for basic biological discovery and next-generation molecular medicine.

2 Sense Codon Reassignment Strategies for Global Proteome Labeling

The cellular proteome dictates cell state and function. Dynamic changes in protein expression drive critical processes such as cell division, differentiation, stress adaptation, and disease progression (Chen *et al.*, 2016; Phillips *et al.*, 2023). Traditional proteomics captures mainly steady-state protein abundance, leaving the low-abundance nascent proteome masked by highly abundant pre-existing proteins (Tang and Chen, 2023; Zhang *et al.*, 2014). Global proteome labeling via sense codon reassignment elegantly overcomes this limitation by chemically tagging actively translating proteins, thereby distinguishing the nascent proteome from the pre-existing pool. Current technologies diverge into two conceptual frameworks depending on how the host translational machinery is manipulated: (1) global metabolic labeling that leverages endogenous systems, and (2) programmable orthogonal translation systems for stochastic recoding (Fig. 1a and 1b).

2.1 Global metabolic labeling via endogenous translation machinery

The earliest approaches relied on selective pressure incorporation (SPI), an auxotroph-based strategy advanced by Budisa, Tirrell, and others (Budisa et al., 1995; Kiick et al., 2002). SPI exploits the substrate promiscuity of endogenous aminoacyl-tRNA synthetases (aaRSs), allowing ncAAs to compete with cAAs for incorporation at specific sense codons (Fig. 1a). This framework established endogenous aaRS promiscuity as a practical entry point for residue-specific and proteome-scale ncAA incorporation (Budisa, 2004; Lepthien et al., 2010). Although effective, reliance on natural aaRS promiscuity imposed severe metabolic stress, low

incorporation efficiency, and a narrow repertoire of compatible ncAAs. These limitations prompted the engineering of synthetases with enhanced ncAA selectivity, greatly expanding the accessible chemical space and enabling robust labeling under near-physiological conditions.

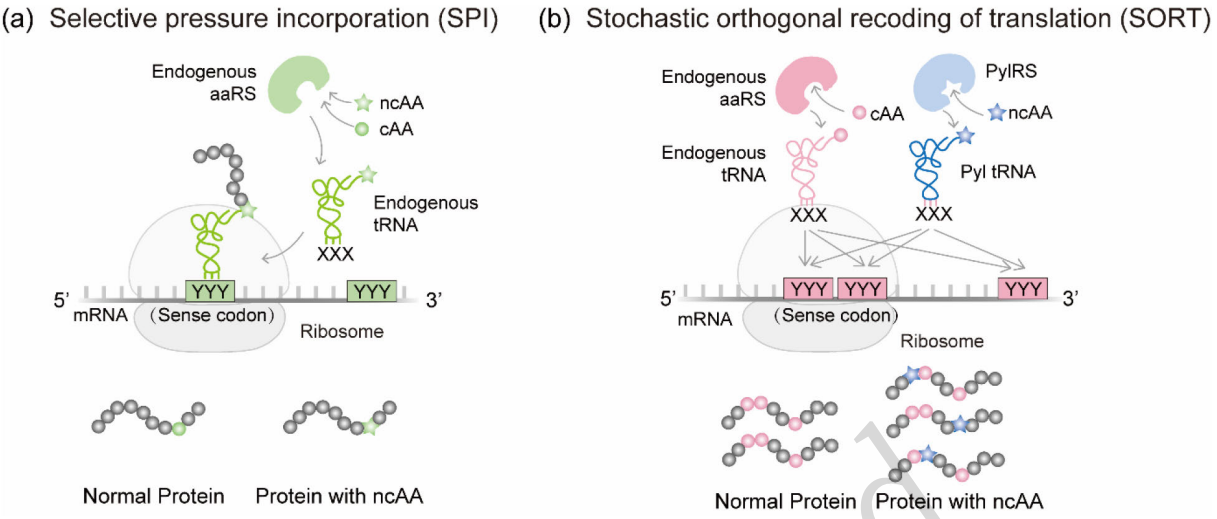


Fig. 1 Strategies for global proteome labeling via sense codon recoding. (a) Selective pressure incorporation (SPI). SPI exploits the promiscuity of endogenous synthetases (or mutated endogenous synthetases) to charge endogenous cognate tRNAs (with anticodon XXX) with a structurally similar non-canonical amino acid (ncAA) instead of the canonical amino acid (cAA). The ncAA competes directly with the cAA during translation, leading to global substitution at target sense codons (YYY) and yielding a mixture of normal and ncAA-modified proteins. (b) Stochastic orthogonal recoding of translation (SORT). An orthogonal PylRS/tRNA pair is introduced and engineered with a mutated anticodon (XXX) to target specific sense codons (YYY). The ncAA-charged orthogonal tRNA competes stochastically with the endogenous cAA-charged tRNA for the same target sense codon, resulting in a proteome-wide statistical mixture of normal and ncAA-modified proteins. aaRS: aminoacyl-RNA Synthetase; tRNA: transfer RNA; mRNA: messenger RNA.

2.1.1 Global metabolic labeling based on endogenous synthetases

Bioorthogonal non-canonical amino acid tagging (BONCAT) pioneered the efficient profiling of the global nascent proteome by exploiting the substrate promiscuity of wild-type methionyl-tRNA synthetase (MetRS) (Dieterich *et al.*, 2006). Methionine analogues such as azidohomoalanine (AHA) or homopropargylglycine (HPG) are incorporated at AUG codons, enabling large-scale co-translational labeling without complex genetic manipulation of the host (Fig. 2a) (Ngo and Tirrell, 2011).

Coupling BONCAT with downstream detection methods has generated powerful multidimensional platforms. Fluorescent non-canonical amino acid tagging (FUNCAT) uses click chemistry with fluorescent probes for *in situ* spatial visualization of nascent proteins within cellular or tissue microenvironments (Dieterich *et al.*, 2010). Quantitative non-canonical amino acid tagging (QuaNCAT) integrates BONCAT with stable isotope labeling with amino acids in cell culture (SILAC) to selectively enrich and quantify nascent polypeptides while eliminating interference from pre-existing proteins (Howden *et al.*, 2013). These approaches have successfully mapped dynamic proteomic shifts in diverse biological contexts, including pathological cardiac remodeling and embryonic development (Liu *et al.*, 2016).

A major advantage of wild-type based labeling is its direct applicability *in vivo* without genetic engineering. Systemic administration of AHA or HPG in wild-type mice achieves whole-animal labeling, including efficient crossing of the placental barrier for high-contrast imaging of nascent proteomes in developing embryonic tissues (Calve *et al.*, 2016).

Nevertheless, wild-type MetRS strongly prefers natural methionine (about a 400-fold substrate preference) (Hinz *et al.*, 2012), necessitating methionine depletion in culture (Anim *et al.*, 2025) or excessively high analogue concentrations *in vivo* (Calve *et al.*, 2016). These non-physiological conditions perturb cellular metabolic homeostasis, trigger stress responses, alter translation rates, and can distort the very physiological states under investigation (Michels *et al.*, 2021). Prolonged exposure to high concentrations of methionine analogues also induces developmental toxicity in animal and plant models, limiting long-term *in vivo* utility (Tivendale *et al.*, 2021).

2.1.2 Engineered synthetases for enhanced profiling

To overcome these limitations, researchers engineered the substrate-binding pocket of synthetases to improve ncAA selectivity. A landmark example is the *E. coli* MetRS mutant (NLL-MetRS, harboring L13N/Y260L/H301L), which exclusively recognizes azidonorleucine (ANL) and remains inert to the endogenous host MetRS (Tanrikulu *et al.*, 2009) (Fig. 2b). This cell-selective labeling enabled the specific capture of bacterial virulence effectors during infection of macrophages, revealing distinct secretion profiles of the type III secretion system (T3SS) (Mahdavi *et al.*, 2014).

Moreover, engineered synthetase strategies have been adapted to mammalian systems. A mutant *Mus musculus* MetRS (*Mm*MetRS, L274G) was developed for efficient ANL incorporation in mammalian cell lines under standard culture conditions, completely obviating the need for methionine depletion (Mahdavi *et al.*, 2016). Building on this foundation, a recently developed transgenic mouse line expressing multiple copies of this mutant enzyme (3×MetRS, L274G) achieves highly efficient global proteome labeling *in vivo*. By fundamentally mitigating metabolic stress, this engineered synthetase approach broadens the scope of application for studying physiological and pathological homeostasis in complex living organisms (Alvarez-Pardo *et al.*, 2025).

2.1.3 Multiplexed profiling via expanded codon channels

Although AUG-targeted BONCAT has been widely used, relying exclusively on this single sense codon intrinsically limits the comprehensiveness of the captured proteome. Methionine residues are frequently subject to post-translational modifications, such as N-terminal excision, which can irreversibly remove the incorporated chemical tags (Giglione *et al.*, 2004; Wang *et al.*, 2008). Furthermore, forcing analogue incorporation exclusively at methionine sites can occasionally perturb local protein folding and stability, causing certain protein subsets to be systematically missed (Yang *et al.*, 2018). Consequently, any single-codon labeling system inevitably captures a biased subproteome, leaving significant blind spots. To achieve unbiased, global proteomic coverage, researchers have had to develop alternative metabolic labeling strategies targeting other sense codons. This pursuit has driven the successful expansion of the labeling repertoire to threonine, tyrosine, phenylalanine, and tryptophan codons (Ignacio *et al.*, 2023; Loynd *et al.*, 2026; Yang *et al.*, 2018), paving the way for both deeper proteome capture and advanced multiplexing.

Parallel to these efforts, engineered synthetases have further expanded the codon pool to capture unique proteomic subsets. Researchers developed a *Saccharomyces cerevisiae* tyrosyl-tRNA synthetase (*Sc*TyrRS, Y43G) and a *Mus musculus* phenylalanyl-tRNA synthetase (*Mm*PheRS, T413G) to redirect metabolic labeling toward tyrosine and phenylalanine sense codons (Yang *et al.*, 2018) (Fig. 2c). These variants preferentially activate the corresponding azide-bearing ncAAs over cAAs under the reported labeling conditions, enabling residue-specific bioorthogonal tagging in live tumor models. By labeling the melanoma tumor proteome and plasma secretome in mice, *Sc*Tyr-Y43G and *Mm*Phe-T413G achieved the first global proteome labeling in a tumor model and identified plasma factors secreted from specific cell types. Furthermore, integrating these distinct synthetases captured unique proteomic subsets that single channel labeling inherently missed.

Most recently, these expansion efforts have culminated in highly sophisticated spatial and temporal control. Multiplexed BONCAT (mBONCAT) introduced engineered *E. coli* tryptophanyl (*Ec*TrpRS-h14) and tyrosyl tRNA synthetases (*Ec*TyrRS-VSMA) (Fig. 2c). A defining advantage of these engineered enzymes is their

broad substrate compatibility, accommodating both click-reactive ncAAs and inert analogues (*e.g.*, 5OMeW or OMeY). This distinct property facilitates a precise pulse-chase modality: high-affinity inert ncAAs can competitively terminate the reactive pulse, granting exceptional temporal resolution. Coupled with distinct fluorescent probes, the strict orthogonality of these integrated systems enables the simultaneous tracking of preexisting and nascent protein dynamics within the same cellular environment (Loynd *et al.*, 2026).

Despite these optimization strategies effectively mitigating metabolic stress and expanding targeted codons from AUG to the sense codons of tryptophan, tyrosine, phenylalanine, and threonine, such systems still face fundamental mechanistic constraints. Although engineered synthetases have successfully accommodated a broad array of chemically diverse ncAAs, these approaches remain inextricably tethered to the specific codons dictated by their cognate host tRNAs. They completely lack the modular flexibility required for the customized recoding of arbitrary codons, as altering the anticodon loop of endogenous tRNAs to redirect them to new codons would catastrophically disrupt the host's natural genetic code and global proteostasis.

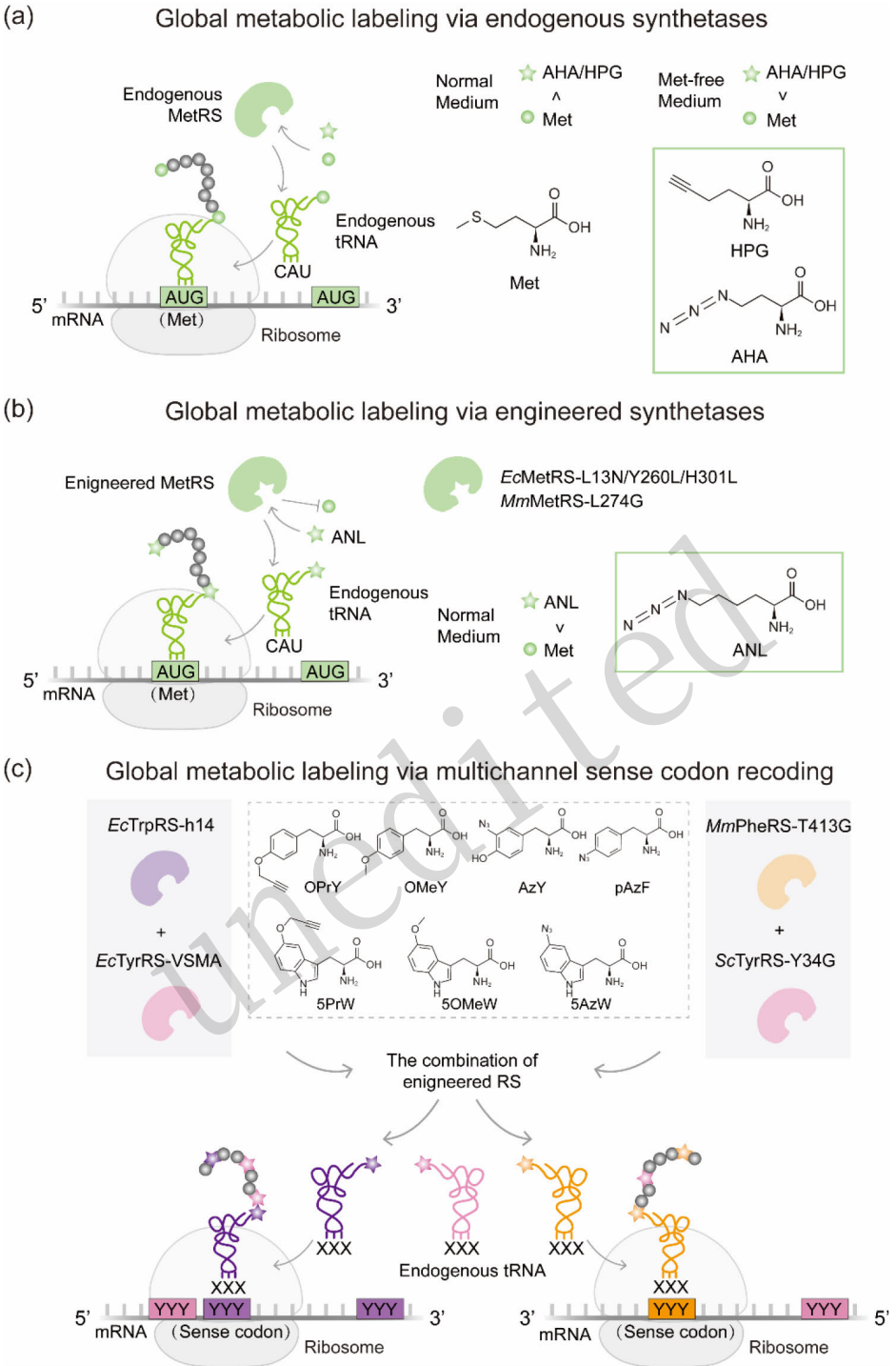


Fig. 2 Global metabolic labeling and multichannel sense codon recoding strategies. (a) Global metabolic labeling via endogenous synthetases. Endogenous MetRS charges its cognate tRNA with ncAA analogues like azidohomoalanine (AHA) or homopropargylglycine (HPG). Due to competitive inhibition by natural Met, the incorporation of AHA/HPG requires metabolic starvation in a Met-free medium. (b) Global metabolic labeling via engineered synthetases. Mutant synthetases (*Ec*MetRS-L13N/Y260L/H301L and *Mm*MetRS-L274G) specifically recognize and activate analogues like azidonorleucine (ANL), enabling direct ncAA incorporation in normal medium. (c) Global metabolic labeling via multichannel sense codon recoding. Co-expressing combinations of engineered synthetases (*Ec*TrpRS-h14, *Ec*TyrRS-VSMA, *Mm*PheRST413G, and *Sc*TyrRS-Y34G) enables the simultaneous targeting of multiple sense codons. The chemical structures of representative ncAAs used for multiplexed labeling, including analogues of tyrosine (OPrY, OMeY, AzY), phenylalanine (pAzF),

tryptophan (5PrW, 5OMeW, 5AzW), are shown. Met: Methionine; OPrY: O-Propargyl-L-tyrosine; OMeY: O-Methyl-L-tyrosine; AzY: Azido-L-tyrosine; pAzF: para-Azido-L-phenylalanine; 5PrW: 5-Propargyl-L-tryptophan; 5OMeW: 5-Methoxy-L-tryptophan; 5AzW: 5-Azido-L-tryptophan.

2.2 SORT as a programmable platform for global proteome labeling

To establish an orthogonal and versatile strategy for ncAA-mediated proteome labeling, Chin and colleagues repurposed the highly orthogonal pyrrolysyl-tRNA synthetase (PylRS)/tRNA pair, originally developed for site-specific ncAA incorporation, to create Stochastic Orthogonal Recoding of Translation (SORT) for global proteomic applications (Fig. 1b) (Elliott *et al.*, 2014a). The defining feature of SORT is the largely anticodon-independent aminoacylation activity of PylRS, which enables PylT anticodons to be reprogrammed toward a broad range of compatible sense codons while retaining aminoacylation activity. In practice, however, successful recoding efficiency is dictated by multiple factors, including endogenous tRNA abundance (Schwark *et al.*, 2020), mRNA sequence context (Ding *et al.*, 2024a), aminoacylation efficiency (Bruce and Uhlenbeck, 1982), and necessary tRNA engineering. Upon expression in cells, these optimized orthogonal tRNAs compete with endogenous tRNAs to stochastically decode selected codons with ncAAs during translation.

This unique architecture transformed stochastic proteome labeling from a largely amino acid–restricted strategy into a programmable codon-level recoding platform. By altering the orthogonal tRNA anticodon, ncAA incorporation can be redirected toward distinct codon populations while preserving efficient aminoacylation. This programmability enables flexible tuning of labeling density, proteome coverage, and subproteome selectivity according to the abundance and positional distribution of different amino acids across the proteome.

2.2.1 SORT enabling programmable multi-codon proteome labeling and enrichment

Building on this programmable foundation, the SORT toolkit has rapidly evolved. The following sections detail its expansion from initial global proteome visualization (SORT-M) and deep nascent protein identification (SORT-E) to spatiotemporal mapping in live mammalian disease models, establishing SORT as a robust engine for *in vivo* chemical proteomics.

In 2014, Chin and colleagues introduced SORT with Chemoselective Modification (SORT-M) (Elliott *et al.*, 2014b). In this approach, engineered PylRS/tRNA pairs were used to stochastically incorporate ncAAs bearing bioorthogonal chemical handles into proteins throughout the proteome at designated sense codons. The incorporated handles, including cyclopropene-containing ncAAs, subsequently reacted with tetrazine-conjugated probes through rapid inverse electron-demand Diels–Alder (IEDDA) cycloaddition (Blackman *et al.*, 2008; Patterson *et al.*, 2012). Owing to its exceptionally fast kinetics and high selectivity, this reaction enabled efficient visualization and enrichment of nascent proteins while minimally perturbing cellular physiology.

The versatility of SORT-M was systematically demonstrated across multiple biological systems. Robust proteome-wide labeling with favorable signal-to-noise ratios was achieved in both *E. coli* and human HEK293T cells. To establish the generality of the platform, researchers engineered orthogonal tRNAs targeting a broad range of sense codons, including those encoding Met (AUG), Lys (AAA and AAG), Ser (AGU and AGC), Ala (GCA), Leu (CUG), Val (GUG), Glu (GAG), Tyr (UAU and UAC), and Trp (UGG), all of which successfully supported ncAA incorporation and proteome labeling. Importantly, comprehensive analyses in both bacterial and mammalian systems demonstrated that systemic sense codon recoding under optimized conditions caused minimal cytotoxicity and preserved overall cellular fitness (Elliott *et al.*, 2014b).

The application of SORT was subsequently extended to multicellular organisms through a modular transgenic platform in *Drosophila melanogaster* using the UAS/GAL4 system (Brand and Perrimon, 1993). Tissue-specific expression of PylRS enabled spatiotemporally resolved imaging of nascent proteomes in defined cell populations, including ovarian germline cells and wing imaginal discs. For example, expression driven by

nos-vp16-GAL4 produced strong labeling in stage 5 egg chambers with minimal signal in surrounding follicle cells, while causing no detectable developmental abnormalities (Elliott *et al.*, 2014b).

To further enhance nascent proteome identification, researchers subsequently developed SORT with Enrichment (SORT-E), which incorporated a cleavable tetrazine diazobenzene biotin (TDB) probe into the workflow (Elliott *et al.*, 2016; Szychowski *et al.*, 2010; Verhelst *et al.*, 2007). Leveraging the ultrafast kinetics of the IEDDA reaction (Dommerholt *et al.*, 2014), TDB enabled direct capture of labeled proteins from crude lysates at low reagent concentrations without precipitation-based processing. This streamlined workflow minimized sample loss and substantially improved recovery of low-abundance proteins that are frequently inaccessible in conventional whole-cell proteomics.

In *E. coli*, SORT-E achieved a highly specific enrichment of tagged proteins, with a specificity exceeding 92%. Quantitative comparisons against comprehensive protein abundance databases (such as PaxDb) (Wang *et al.*, 2015; Wang *et al.*, 2012) confirmed its preferential enrichment of low-abundance proteins frequently lost in standard whole-cell analyses. Importantly, distinct proteomic subsets exhibit codon-specific enrichment preferences. While earlier methodologies suggested that tagging exclusively at methionine residues might be sufficient for proteome-wide analysis (Yuet and Tirrell, 2014), these codon-specific biases highlight the critical need for multi-codon strategies. By co-expressing multiple orthogonal tRNAs targeting codons for Ala, Ser, and Met, labeling coverage and detection sensitivity were substantially improved, thereby enabling high-resolution proteomic mapping in complex tissues. When applied *in vivo* to *Drosophila* ovarian germline cells via the *nos-vp16-GAL4* system, this combinatorial approach dramatically increased nascent protein identification compared to single-codon controls (Elliott *et al.*, 2016). Overall, SORT-E offers a powerful and necessary tool for decoding complex, cell-specific proteomic dynamics during development.

2.2.2 Spatiotemporal mapping in complex mammalian organs and disease models

The extension of SORT to live animal models represents a major milestone in *in vivo* chemical proteomics. In 2018, Krogager *et al*. delivered SORT components via adeno-associated viruses (AAVs) to the mouse brain using tissue specific promoters (*hSyn1*, *GFAP*, *Vgat*, and *Vglut1*) to drive expression of the engineered translation system in specific neural populations including excitatory neurons, inhibitory GABAergic neurons, and glia (Krogager *et al.*, 2018) (Fig. 3a). Non-invasive administration of the alkynyl ncAA, Ne-(propargyloxycarbonyl)-l-lysine (AlkK), through drinking water enabled *in vivo* tagging of nascent proteins without perturbing physiological homeostasis (Ernst *et al.*, 2016; Nguyen *et al.*, 2009). This approach identified nearly two thousand proteins from striatal neurons with a high enrichment specificity of 96%. This study established the feasibility of using SORT in live mammals and highlighted its utility for analyzing neural circuits and cell-specific proteomic dynamics.

Building upon the success of AAV-mediated delivery in the nervous system, recent advancements have expanded the platform to address intricate pathological environments. In 2025, Lin and colleagues developed SORT-AC (SORT induced by AAV-delivered Cre) (Gu *et al.*, 2025) (Fig. 3b). This technology integrated a CRISPR-Cas9-mediated knock-in mouse model with AAV-delivered Cre recombinase driven by the liver-specific transthyretin promoter (TTR-Cre), enabling precise and inducible expression of the AlkKRS/PylT_{KASM} system in the live liver. To optimize the labeling of the nascent membrane proteome, researchers used a rational codon selection strategy (KASM). The combination of Lys (K), Ala (A), Ser (S), and Met (M) was selected based on their high abundance, enrichment in transmembrane regions, and preferential distribution on protein surfaces, which minimizes steric hindrance during subsequent chemical labeling.

By simultaneously targeting these four codons and optimizing hepatic click chemistry, the system successfully generated a high-resolution nascent membrane proteome map, termed SORT-AC Print, where membrane proteins accounted for about 50% of the newly identified liver proteome. In a model of alcohol-associated liver disease, SORT-AC further revealed the regulatory role of *Phb1/2* in alcohol metabolism and identified *Acsl1/5* as potential therapeutic targets for lipid accumulation (Gu *et al.*, 2025). This evolution toward multi-codon rational design provides a robust toolkit for identifying low-abundance regulatory factors and

screening therapeutic interventions in various disease states.

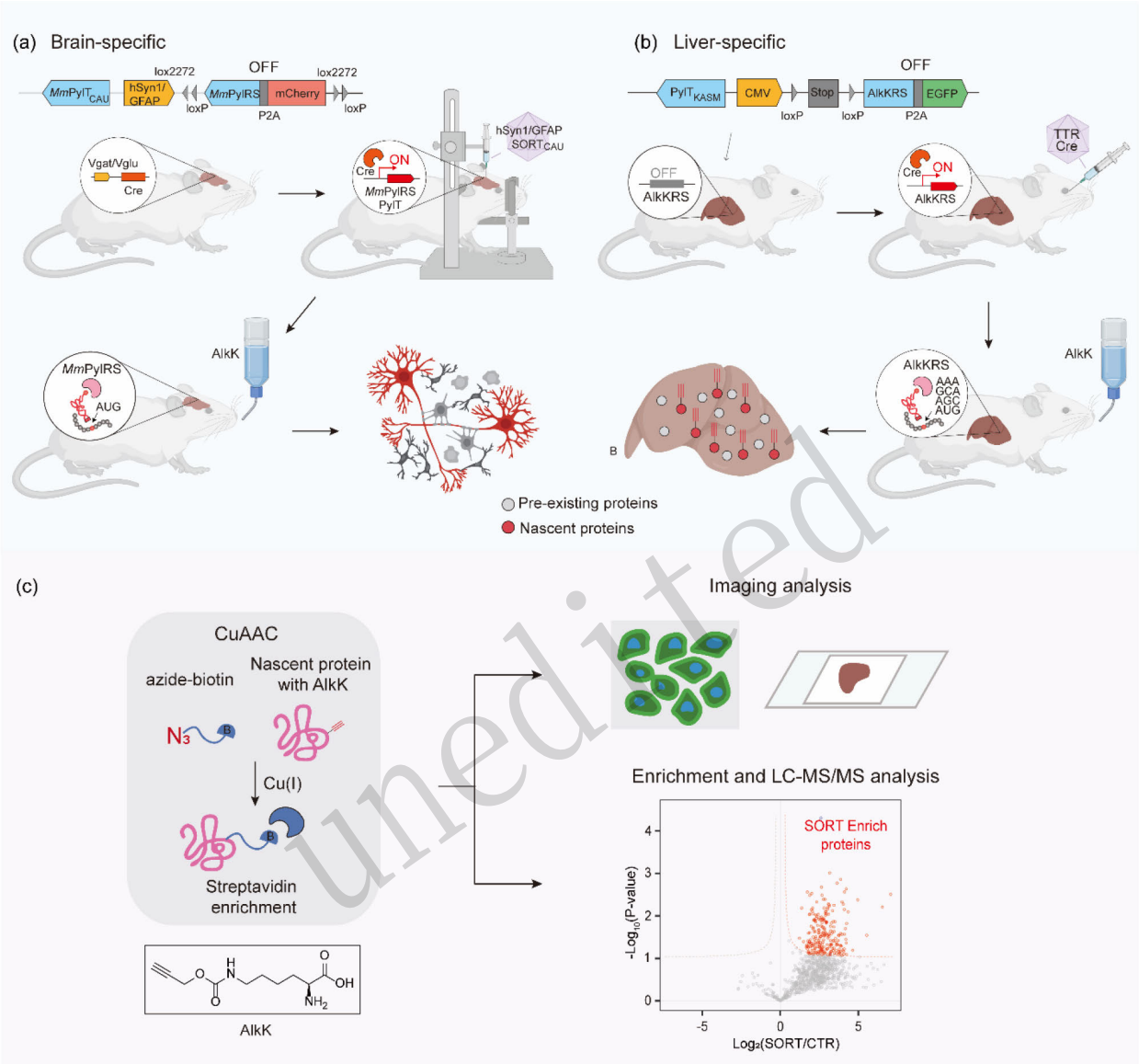


Fig. 3 *In vivo* tissue- and cell-type-specific nascent proteome labeling utilizing Stochastic Orthogonal Recoding of Translation (SORT) in live mammalian systems. (a) Brain-specific proteome tagging. Cre-expressing transgenic mice (*Vgat*/*Vglu*-Cre) receive stereotaxic injections of a Cre-dependent AAV vector. Cre-mediated recombination of the FLEX cassette irreversibly activates *Mm*PylRS and mCherry expression. Subsequent oral Ne- (propargyloxycarbonyl)-l-lysine (AlkK) administration enables stochastic incorporation of AlkK at Met codons (AUG) within targeted neural populations. (b) Liver-specific proteome tagging. Transgenic mice harboring a genomically integrated, LoxP-STOP-LoxP controlled SORT cassette receive an orbital injection of AAV-TTR-Cre. Cre-mediated excision of the STOP cassette selectively activates AlkKRS and EGFP expression in hepatocytes. Upon AlkK administration, the optimized AlkKRS/tRNA$_{KASM}$ system incorporates AlkK across four sense codons (AAA, GCA, AGC, and AUG). (c) Bioorthogonal labeling and downstream analysis. Harvested tissue lysates undergo copper-catalyzed azide-alkyne cycloaddition (CuAAC) click chemistry to conjugate azide-biotin probes to AlkK-tagged proteins. The biotin-tagged proteome supports high-resolution *in situ* imaging alongside subsequent streptavidin enrichment. Enriched nascent proteomes are subsequently quantified

via LC-MS/MS to identify upregulated newly synthesized proteins (red dots in the volcano plot).

3 Sense Codon Reassignment Strategies for Site Specific Incorporation

While the SPI and SORT platform provides a powerful mechanism for proteome-wide reassignment, it is inherently unsuitable for the targeted modification of a specific protein of interest (POI), as its global deployment would lead to massive, unpredictable off-target incorporation. However, achieving precise, site-specific incorporation of ncAAs is crucial (Table 1). It has revolutionized protein engineering by endowing proteins with novel chemical functionalities, enabling the real-time visualization of dynamic biological processes, the targeted regulation of protein functions, and the accelerated development of next-generation biotherapeutics.

Historically, achieving such site-directed incorporation has relied mainly on stop codon suppression (particularly the amber codon, UAG). However, this approach is intrinsically limited by intense competition with endogenous release factors, the potential disruption of natural translational termination, and a restricted coding space that severely hinders the simultaneous encoding of multiple distinct ncAAs. To circumvent these bottlenecks, sense codon reassignment offers a compelling alternative. This strategy is fundamentally underpinned by the degeneracy of the genetic code. Because 18 of the 20 canonical amino acids are encoded by multiple synonymous codons (*e.g.*, six synonymous codons for serine, and six for arginine), reassigning one specific sense codon provides a vast potential codon reservoir for site-specific ncAA encoding without abolishing the natural amino acid's incorporation across the proteome.

Despite this potential, sense codon reassignment faces three main challenges. First, many sense codons occur frequently in the genome, so their reassignment can lead to off-target ncAA incorporation at endogenous protein sites. Second, at the engineered codon in the target mRNA, the orthogonal tRNA must compete with endogenous isoacceptor tRNAs that normally decode the same codon. Third, maintaining aminoacylation orthogonality becomes problematic in multiplexed ncAA encoding. Because the repertoire of tRNA identity elements that can be used to build mutually orthogonal pairs is limited (Saint-Leger *et al.*, 2016), different orthogonal aaRS/tRNA systems may cross-react with one another or with endogenous translation components. Furthermore, engineering aaRS substrate-recognition pockets to accommodate ncAAs can inadvertently broaden substrate specificity or compromise catalytic efficiency, resulting in mischarging with canonical amino acids or unintended ncAAs.

Targeting rare sense codons offers a practical strategy to overcome these decoding related barriers: their low genomic usage and reliance on low abundance endogenous tRNAs naturally reduce proteome wide background and native decoding competition. However, ensuring aminoacylation orthogonality remains a distinct prerequisite for efficient and faithful multiplexed ncAA incorporation, necessitating further investigation into its mechanistic basis and engineering strategies. The following sections discuss how these principles have been implemented in *E. coli,* expanded via genome recoding and total synthesis, and ultimately adapted to complex mammalian systems.

Table 1 Mechanistic comparison of sense codon reassignment strategies

Strategies	BONCAT/SPI	SORT	Rare codon recoding	Genome-recoded organisms
Underlying Mechanism	Residue-level analogue incorporation via endogenous/engineered synthetases	Stochastic codon competition through engineered orthogonal aaRS/tRNAs	Site-specific competitive reassignment at inherently rare codons	True blank-codon reassignment after genome recoding and cognate tRNA deletion
Target codon	Typically AUG (Met), or specific codons dictated by endogenous host tRNAs	Programmable across various sense codons by altering orthogonal tRNA anticodons	Rare sense codons with lowest genomic usage and low endogenous tRNA abundance (*e.g.*, TCG in mammalian cells)	Compressed/liberated sense codons completely removed from the genome (*e.g.*, TCG, TCA in Syn61)
Required Components	Analogue (*e.g.*, AHA/HPG/ANL) + corresponding endogenous or engineered aaRS	Orthogonal aaRS/tRNA pair (*e.g.*, PylRS/tRNA) + ncAAs	Orthogonal aaRS/tRNA pairs + ncAAs	Synthetic genome + deleted endogenous tRNAs + Orthogonal aaRS/tRNA+ ncAAs
Host system	Bacteria, mammalian cells, zebrafish, mice, transgenic/viral animal model	*E. coli*, HEK293T, *Drosophila*, mouse models	Mammalian cells (where whole-genome synthesis is currently impractical)	Prokaryotes (*e.g.*, highly engineered *E. coli* variants like Syn61)
Incorporation mode	Global metabolic labeling	Proteome-wide stochastic labeling	Highly efficient target incorporation overriding weak endogenous competition	Absolute incorporation without endogenous competition
Site specificity	Global/Stochastic (Residue-specific but not protein-specific)	Proteome-wide (Codon-specific but stochastic across the proteome)	High (can be promoted by context-sequence optimization)	Absolute (The reassigned codon is absent from the host genome and is introduced solely into the target mRNA)
Off-target risk	Irrelevance	Irrelevance	Moderate (residual background due to endogenous tRNA competition)	None (completely interference-free)
Major limitations	Metabolic stress or amino acid starvation; residue-level bias; limited codon programmability	Context-dependency of recoding efficiency	Context-dependency of recoding efficiency; residual background in sensitive systems	Unattainable in complex mammalian systems due to vast scale and regulatory complexity
Applications	Nascent proteome labeling; cell-type-specific proteomics; *in vivo* metabolic profiling	Programmable proteome-wide labeling; low-abundance protein enrichment; *in vivo* proteome mapping	Precision protein engineering; multiplexed ncAA incorporation	Programmable biosynthesis of complex sequence-defined polymers and macrocycles

BONCAT, bioorthogonal noncanonical amino acid tagging; SPI, selective pressure incorporation; SORT, stochastic orthogonal recoding of translation; ncAA, noncanonical amino acid; AHA, azidohomoalanine; HPG, homopropargylglycine; ANL, azidonorleucine; Syn61: an *E. coli* strain with a 61-codon genome, designed to enable GCE by providing blank codons for ncAAs.

3.1 Establishment and evolution of sense codon reassignment in *E. coli*

Before delving into the mechanistic strategies of sense codon reassignment, it is crucial to establish a robust evaluation methodology. Unlike traditional nonsense suppression, where successful incorporation is easily indicated by full-length protein expression versus a truncated byproduct, sense codon reassignment presents a unique challenge: a sense codon inevitably yields a full-length protein, regardless of whether it is decoded by the introduced ncAA or an endogenous cAA. Therefore, accurately distinguishing targeted ncAA encoding from the cAA background is foundational. To address this challenge, a general screen was developed using permissive active site mutations in green fluorescent protein (GFP). When a critical fluorophore-maturation residue is mutated to a target sense codon, the resulting fluorescence becomes strictly dependent on the successful incorporation of the specific ncAA, thereby providing a direct optical readout for reassignment efficiency (Biddle *et al.*, 2015). Furthermore, the development of a fluorescence-activated cell sorting (FACS) dual-fluorescence pipeline has significantly advanced this process. This platform uses two distinct reporters to simultaneously monitor global protein expression and specific ncAA incorporation, facilitating the high-throughput screening and directed evolution of highly active orthogonal aminoacyl-tRNA synthetases (Biddle *et al.*, 2022). Together, these platforms provide the essential quantitative tools required to rigorously evaluate the desired target incorporation.

Equipped with these quantitative screening tools, researchers can effectively dissect and manipulate the core mechanism underlying sense codon reassignment: translational competition. Fundamentally, the introduced orthogonal translation system (OTS) must actively compete against the endogenous host translational machinery to decode a specific mRNA codon. Early proof-of-concept studies successfully validated this competitive strategy by targeting rare codons. For instance, by engineering the anticodon of the PylRS/tRNA and coupling it with highly optimized cultivation conditions, such as using specific minimal media (GMML) and precise induction timing, the translation of the rare arginine codon AGG was forcefully hijacked (Zeng *et al.*, 2014). Following a similar paradigm, another rare arginine codon, AGA, was also successfully reassigned using an engineered Pyl system under optimized cultivation (Wang and Tsao, 2016).

As these competitive strategies yielded initial successes, the underlying biological mechanisms governing this competition were further elucidated. For AGG reassignment, Lee *et al.* demonstrated that deletion of *argW* removed the major endogenous $tRNA_{CCU}$ that normally decodes AGG. Endogenous AGG decoding then depends mainly on the weaker wobble pairing by $tRNA_{UCU}$, giving the engineered orthogonal tRNA a stronger competitive advantage and thereby improving ncAA incorporation efficiency (Lee *et al.*, 2015). This observation highlighted a critical strategic principle that reducing the abundance or activity of competing endogenous tRNAs in the host system significantly enhances the competitive advantage of the OTS. This paradigm was further refined by the demonstration that endogenous tRNA abundance, rather than mere codon rarity, is the definitive quantitative determinant of competitive success (Schwark *et al.*, 2020). While codon rarity correlates with lower global toxicity, the local decoding competition is directly dictated by the cellular concentration of the competing endogenous tRNAs; theoretically, a lower abundance allows for higher ncAA incorporation efficiency. Beyond passively selecting low-competition codons, actively attenuating the decoding capacity of host's endogenous translation machinery offers a powerful alternative. For example, knocking out the host enzyme responsible for queuosine (Q) modification on endogenous tRNAs effectively reduces their decoding efficiency, thereby attenuating the endogenous machinery and shifting the translational balance in favor of the OTS (Tittle *et al.*, 2022).

Empowered by this deeper mechanistic understanding of translational competition, researchers have rapidly expanded the boundaries of sense codon reassignment far beyond rare codons. Demonstrating the robust potential of this competitive strategy, a series of highly frequent codons in *E. coli*, including AGU, AGC, UCG (encoding serine), as well as CUG (encoding leucine), was successfully reassigned by massively overexpressing the OTS, proving that extreme competitive pressure can overcome even highly abundant endogenous tRNAs (Ho *et al.*, 2016). Expanding the scope further, the target was shifted from the elongation phase to the initiation

phase, successfully reprogramming the initiator codon (AUG) to achieve N-terminal ncAA incorporation (Tharp *et al.*, 2021). Remarkably, by combining sense codon reassignment with traditional stop codon suppression, the co-translational incorporation of up to three distinct ncAAs within a single complex protein has been realized (Cui *et al.*, 2017; Tharp *et al.*, 2021).

In summary, driven by increasingly sophisticated competitive strategies, a diverse array of codons, including AGG, AGA, AGU, AGC, UCG, CUG, and the initiator AUG, have been successfully reassigned to ncAAs in *E. coli*. However, while optimizing competitive dynamics can maximize targeted incorporation efficiency, it cannot fundamentally eliminate the proteome-wide background misincorporation and associated systemic toxicity that arise from the pervasive distribution of sense codons throughout the host genome.

3.2 Overcoming endogenous competition through genomic recoding and total synthesis

Due to the widespread distribution of sense codons across the genome, achieving targeted ncAA encoding via sense codon reassignment inevitably results in the erroneous misincorporation of ncAAs at natural synonymous sites throughout the host proteome. To address this fundamental limitation, genomic recoding and total synthesis have been developed. These approaches systematically remove all instances of a target codon from the host genome, creating a vacant coding channel that can be reassigned without interference. The technical feasibility of large-scale genomic recoding was first demonstrated through the total replacement of all 321 UAG stop codons with synonymous UAA codons in the *E. coli* genome (Lajoie *et al.*, 2013). By successfully removing every annotated occurrence of UAG, release factor 1 was deleted, converting UAG into a blank coding slot with zero competition from the host machinery. Building upon this, the exploration expanded toward sense codons, which present greater complexity due to their high frequency. A systematic evaluation of synonymous codon compression schemes established critical design rules for sense codon replacement (Wang *et al.*, 2016). It was determined that while many synonymous substitutions are well-tolerated, specific positions are highly sensitive to changes that disrupt mRNA secondary structure, induce ribosome stalling, or perturb other regulatory elements, ultimately providing the essential blueprint and multi-omics troubleshooting framework for whole-genome synthesis (Nyerges *et al.*, 2024).

Guided by these compression rules, genomic recoding progressed from localized editing to bottom-up total synthesis. This monumental shift was realized with the creation of Syn61, an *E. coli* variant harboring a fully synthetic 4-megabase genome (Fredens *et al.*, 2019). In this synthetic genome, over 18,000 target codons were deterministically replaced, successfully eliminating two sense codons (UCG and UCA, encoding serine) and one stop codon (UAG). Crucially, the mere removal of these codons from the genome is insufficient for orthogonal reassignment; the corresponding decoding machinery must also be eliminated. Therefore, the endogenous tRNAs responsible for decoding UCG and UCA were subsequently deleted. This dual knockout strategy, erasing both the genomic codons and their cognate tRNAs, yielded pure, blank sense codons. By introducing orthogonal translation systems, these emancipated sense codons were harnessed to achieve viral resistance and the efficient, zero-background synthesis of complex sequence-defined polymers and macrocycles comprising multiple distinct non-canonical monomers *in vivo* (Robertson *et al.*, 2021).

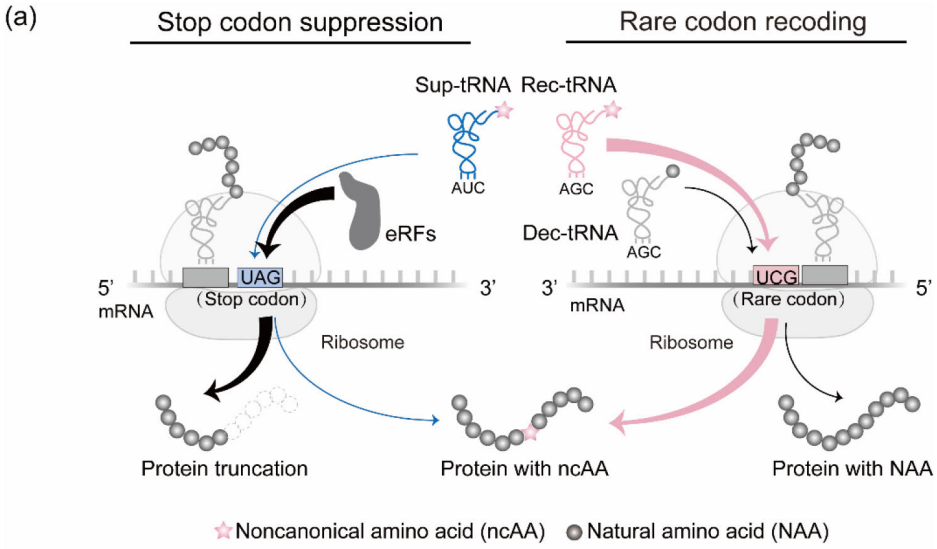
While Syn61 provided three blank codons, the demand for increasingly complex chemically modified biologics necessitates a broader repertoire of orthogonal channels. Consequently, efforts have focused on pushing the limits of genetic code compression. Recent advancements have demonstrated the feasibility of compressing the three natural stop codons into a single UAA stop codon, completely liberating the remaining two for genetic isolation and reassignment (Grome *et al.*, 2025). Pushing this compression to its absolute current limit, a highly engineered *E. coli* strain operating on a 57-codon genetic code was successfully synthesized by removing seven distinct codons from the genome (Ostrov *et al.*, 2016; Robertson *et al.*, 2025). The emergence of such highly compressed genomes marks a profound paradigm shift in sense codon reassignment. Rather than struggling against endogenous competition for a single target site, researchers will soon have up to seven entirely interference-free blank codons. This expansive orthogonal coding space will empower the programmable biosynthesis of highly complex proteins and novel materials bearing extensive, multiplexed

chemical modifications.

3.3 Implementation of sense codon reassignment in complex mammalian systems

While total genome synthesis and compression have successfully liberated sense codons in prokaryotes, these disruptive strategies have largely failed to translate into mammalian systems due to the vast scale and regulatory complexity of the eukaryotic genome. Unlike the highly programmable nature of synthetic bacterial genomes, the mammalian translational landscape is governed by a dense network of endogenous tRNAs and intricate mRNA structures, making *de novo* genome synthesis currently impractical for generating blank coding channels. This limitation has prompted a methodological transition from comprehensively rewriting the host genome to exploiting the inherent non-uniformity of natural codon usage through a strategy known as rare codon recoding (RCR) (Fig. 4a).

Translating this concept into practice, RCR was first successfully implemented in mammalian cells by leveraging the natural scarcity of the TCG codon. Because TCG exhibits both the lowest genomic usage frequency and the lowest abundance of its cognate endogenous tRNA, it serves as an ideal candidate to achieve high-efficiency ncAA incorporation into a target POI alongside minimal background misincorporation across the proteome. The development of a systematic engineering framework and a big-data model, specifically the sequence context for TCG recoding (SCTR) model, allowed for the precise prediction and optimization of ncAA incorporation sites (Ding *et al.*, 2024a). By screening mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs and optimizing their expression levels, this approach achieved recoding efficiencies comparable to wild type protein expression. Ultimately, this breakthrough demonstrated that even without massive genomic alterations, specific rare sense codons could be repurposed into highly selective and efficient vehicles for ncAA incorporation.



(b)

Genetic Code Table

Recoded with ncAAs Capable for ncAAs

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC		UCC		UAC		UGC	
UUA	Leu	UCA		UAA	Stop	UGA	Stop
UUG		UCG		UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC		CCC		CAC		CGC	
CUA		CCA		CAA	Gln	CGA	
CUG		CCG		CAG		CGG	
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC		ACC		AAC		AGC	
AUA		ACA		AAA	Lys	AGA	Arg
AUG	Met	ACG		AAG		AGG	
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC		GCC		GAC		GGC	
GUA		GCA		GAA	Glu	GGA	
GUG		GCG		GAG		GGG	

Fig. 4 Strategies for expanding the genetic code via rare codon recoding. (a) Schematic of the rare codon recoding strategy for integrating ncAAs in mammalian cells. Within this framework, Rec-tRNA refers to the engineered tRNA for rare codon recoding. Sup-tRNA represents the tRNA used in conventional stop codon suppression. Dec-tRNA designates the native endogenous tRNA that decodes natural amino acids. In stop codon suppression, the decoding of aminoacylated Sup-tRNA is heavily impeded by the translation termination machinery (such as eukaryotic Release Factors (eRFs)), limiting ncAA incorporation. Conversely, the rare codon recoding approach enables the aminoacylated Rec-tRNA to successfully outcompete the inherently weak endogenous Dec-tRNA, thereby facilitating highly efficient ncAA incorporation. The width of the lines visually indicates the intensity of the process flow. (b) Codon table schematic illustrating the expansion of available coding channels for multi-type ncAA incorporation. Experimentally recoded codons are highlighted in blue,

and candidate codons with potential for future ncAA reassignment are highlighted in light green.

Building upon this foundational TCG recoding framework, recent advancements have pushed the technology toward the simultaneous incorporation of multiple distinct ncAAs. By systematically evaluating additional rare codons and engineering highly orthogonal translation machinery, it is now possible to incorporate up to five distinct ncAAs into a single protein within mammalian cells (Fig. 4b) (Fang *et al.*, 2026). This multiplexed recoding strategy achieves high fidelity and site-specificity, enabling the creation of complex protein architectures with diverse chemical functionalities. This progress reveals the redefinable nature of the eukaryotic genetic code and provides a robust platform for engineering sophisticated biological tools.

Ultimately, the successful implementation of such multiplexed sense codon reassignment in mammalian cells expands the scope of site-specific protein engineering. This technology facilitates the precise introduction of bioorthogonal handles, structural probes, or post-translational modification mimics into selected POI. In future studies, these engineered systems may help probe defined protein states and interactions in carefully validated settings. As the repertoire of available rare codons expands, this platform may support the engineering of therapeutic proteins with defined chemical modifications and enable more targeted studies of selected disease-relevant protein functions.

4 Conclusions and outlook

In conclusion, sense codon reassignment leverages selected reassignment-permissive codons to expand control over protein composition and function. This is achieved by exploiting the redundancy of the genetic code: 61 sense codons encode 20 canonical amino acids, and theoretical analyses suggest that roughly 30-40 codons may suffice to encode an organism's proteome, leaving over 20 codons as potential ncAA-reassignment targets (Budisa, 2025; Krishnakumar *et al.*, 2013). By actively manipulating translational competition within the ribosomal decoding center, this paradigm has successfully evolved into two complementary technological routes. The first route harnesses endogenous competition for global proteome profiling, facilitating dynamic imaging and deep coverage of the nascent proteome under physiological conditions. The second route focuses on establishing dedicated decoding pathways for the precise, site-specific, and multiplexed incorporation of ncAAs into target proteins. Importantly, the efficiency and fidelity of both routes are not determined solely by competition at the ribosomal decoding center, but also by aaRS/tRNA orthogonality, endogenous tRNA abundance and modification status, aminoacylation kinetics, elongation factor compatibility, mRNA and codon-pair context, ncAA uptake, and cellular toxicity. In prokaryotes, coupling this strategy with genomic compression and total synthesis has generated dedicated blank codons, empowering the *in vivo* programmable synthesis of complex sequence-defined polymers and macrocycles. Meanwhile, in mammalian systems, rare sense codon recoding has enabled the incorporation of up to five distinct ncAAs into a single protein. In addition, future studies may leverage rare sense codon recoding to install defined chemical probes or post-translational modification (PTM) mimics into selected proteins, providing a useful platform for hypothesis-driven dissection and manipulation of protein function.

Despite these advancements, the broader implementation of this technology encounters several critical limitations. First, simultaneous incorporation of multiple ncAAs can lead to inter-system crosstalk and cross-reactions with the endogenous translation machinery, resulting in deacylation or misreading. Second, in mammalian systems where whole-genome synthesis remains currently unattainable, strategies that rely on competition with endogenous tRNAs inevitably produce background incorporation at the proteome-wide level. This noise can interfere with biological observations in sensitive cell lines or delicate *in vivo* microenvironments. Third, orthogonal components derived from prokaryotes or archaea may misfold within eukaryotic environments or display suboptimal substrate affinity. Furthermore, their prolonged expression risks inciting immunogenicity or cellular stress.

To resolve these constraints, future research must pivot toward several strategic breakthroughs. A primary

imperative is the development of more multiplexed orthogonal translation systems in mammalian cells. Through directed evolution and rational design, mutually orthogonal pairs could be constructed that do not crosstalk with each other or with the endogenous system, thereby supporting the coordinated encoding of multiple distinct ncAAs (Beattie *et al.*, 2023; Fang *et al.*, 2026; Willis and Chin, 2018). Concurrently, using CRISPR-mediated genome editing to ablate or replace endogenous transfer RNA genes corresponding to target rare codons will establish a pristine genetic background, effectively eradicating competitive recognition (Ding *et al.*, 2024a). Furthermore, integrating machine learning algorithms with sequence-context models is essential to systematically decipher the determinants of recoding efficiency. While rare codon recoding efficiencies are generally robust, ranging from 40% to over 90%, they are profoundly dictated by the flanking nucleotide sequences (Ding *et al.*, 2024a). Integrating structural biology, high-throughput assays, and deep learning will elucidate the molecular mechanisms governing this context dependency. This integration will drive the development of universal predictive tools capable of pre-evaluating and optimizing flanking sequences for any target site, elevating recoding efficiency to its theoretical maximum while maintaining a low off-target background. Finally, the field must transition from global overexpression paradigms toward precise spatiotemporal and localized regulation (Koehler *et al.*, 2020). Deploying chemically induced dimerization systems (Jiang *et al.*, 2023) or optogenetic switches (Charette *et al.*, 2025) can strictly limit the incorporation time window, while anchoring orthogonal translation components to specific subcellular compartments or transcripts can effectively confine ncAA incorporation to the desired spatial domains (Ding *et al.*, 2024a; Reinkemeier *et al.*, 2019).

Supported by the precise regulatory capabilities afforded by these technical advancements, sense codon reassignment presents a broad landscape of transformative applications. In basic life sciences, researchers could precisely introduce photo-crosslinking groups and bioorthogonal handles into distinct domains of a single protein, facilitating a sequential capture and labeling methodology to identify transient interacting partners. In translational medicine and next-generation biotherapeutic development, this technology provides a foundation for engineering advanced safety circuits in chimeric antigen receptor T (CAR-T) cells. By rendering anti-apoptotic proteins and metabolic enzymes strictly dependent on distinct ncAAs encoded by separate rare codons, cellular activation can be restricted to a precise logical AND-gate condition. Moreover, embedding rare codons into essential sites of viral structural proteins can generate highly safe vaccines that absolutely require dual ncAAs for packaging, mechanistically eliminating the risk of reversion mutations. Ultimately, sense codon reassignment has the potential to become a versatile tool for decoding the complex molecular logic of life to significantly advance the future landscape of synthetic biology and precision molecular medicine.

Data availability statement

It is not applicable to this article as no new data were created or analyzed in this study.

Acknowledgments

We thank the National Key R&D Program of China (Grant 2025YFC2608800 and 2024YFC3407200), the Scientific Research Innovation Capability Support Project for Young Faculty SRICSPYF-ZY2025099, and the National Natural Science Foundation of China (Grant 22437004, 22361142828, 22625704 and 22577114) for financial support.

Author contributions

Yiru ZHANG, Wei YU and Wenlong DING were involved in writing – original draft and editing. Wenlong DING and Shixian LIN were involved in writing – review and editing. All authors have read and approved the final manuscript.

Compliance with ethics guidelines

Yiru ZHANG, Wei YU, Wenlong DING, Shixian LIN declare that they have no conflicts of interest.

This review does not contain any studies with human or animal subjects performed by any of the authors.

Declaration on the use of generative AI tools

During the preparation of this work, the authors used ChatGPT for language polishing, readability enhancement, and grammar

checking. The generated output was thoroughly reviewed and edited by the authors to ensure accuracy, originality, and compliance with journal standards. The authors take full responsibility for the final content of this manuscript.

References

Alvarez-Pardo R, Tom Dieck S, Desch K*, et al.*, 2025. Cell type-specific in vivo proteomes with a multicopy mutant methionyl trna synthetase mouse line. *Lab Anim (NY)*, **54**(9):228-237. https://doi.org/10.1038/s41684-025-01589-2

Anim EP, Mezzanotte J, Chu SW*, et al.*, 2025. Bioorthogonal non-canonical amino acid tagging (boncat) to detect newly synthesized proteins in cells and their secretome. *Plos One*, **20**(8):e0329857. https://doi.org/10.1371/journal.pone.0329857

Axup JY, Bajjuri KM, Ritland M*, et al.*, 2012. Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc Natl Acad Sci USA*, **109**(40):16101-16106. https://doi.org/10.1073/pnas.1211023109

Beattie AT, Dunkelmann DL, Chin JW, 2023. Quintuply orthogonal pyrrolysyl-trna synthetase/trna(pyl) pairs. *Nat Chem*, **15**(7):948-959. https://doi.org/10.1038/s41557-023-01232-y

Biddle W, Schmitt MA, Fisk JD, 2015. Evaluating sense codon reassignment with a simple fluorescence screen. *Biochemistry*, **54**(50):7355-64. https://doi.org/10.1021/acs.biochem.5b00870

Biddle W, Schwark DG, Schmitt MA*, et al.*, 2022. Directed evolution pipeline for the improvement of orthogonal translation machinery for genetic code expansion at sense codons. *Front Chem*, **10**:815788. https://doi.org/10.3389/fchem.2022.815788

Blackman ML, Royzen M, Fox JM, 2008. Tetrazine ligation: Fast bioconjugation based on inverse-electron-demand diels-alder reactivity. *J Am Chem Soc*, **130**(41):13518-9. https://doi.org/10.1021/ja8053805

Brand AH, Perrimon N, 1993. Targeted gene-expression as a means of altering cell fates and generating dominant phenotypes. *Development*, **118**(2):401-415. https://doi.org/10.1242/dev.118.2.401

Bruce AG, Uhlenbeck OC, 1982. Specific interaction of anticodon loop residues with yeast phenylalanyl-trna synthetase. *Biochemistry*, **21**(17):3921-6. https://doi.org/10.1021/bi00260a003

Budisa N, 2004. Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angew Chem Int Ed Engl*, **43**(47):6426-63. https://doi.org/10.1002/anie.200300646

Budisa N, 2025. Introduction: "Noncanonical amino acids". *Chem Rev*, **125**(4):1659-1662. https://doi.org/10.1021/acs.chemrev.5c00065

Budisa N, Steipe B, Demange P*, et al.*, 1995. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in escherichia coli. *Eur J Biochem*, **230**(2):788-96. https://doi.org/10.1111/j.1432-1033.1995.tb20622.x

Calve S, Witten AJ, Ocken AR*, et al.*, 2016. Incorporation of non-canonical amino acids into the developing murine proteome. *Scientific Reports*, **6**:32377. https://doi.org/10.1038/srep32377

Chambers I, Frampton J, Goldfarb P*, et al.*, 1986. The structure of the mouse glutathione-peroxidase gene - the selenocysteine in the active-site is encoded by the termination codon, tga. *EMBO J*, **5**(6):1221-1227. https://doi.org/10.1002/j.1460-2075.1986.tb04350.x

Charette M, Rosenblum C, Shade O*, et al.*, 2025. Optogenetics with atomic precision-a comprehensive review of optical control of protein function through genetic code expansion. *Chemical Reviews*, **125**(4):1663-1717. https://doi.org/10.1021/acs.chemrev.4c00224

Chen W, Smeekens JM, Wu R, 2016. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem Sci*, **7**(2):1393-1400. https://doi.org/10.1039/c5sc03826j

Cui Z, Mureev S, Polinkovsky ME*, et al.*, 2017. Combining sense and nonsense codon reassignment for site-selective protein modification with unnatural amino acids. *ACS Synth Biol*, **6**(3):535-544. https://doi.org/10.1021/acssynbio.6b00245

Dieterich DC, Hodas JJ, Gouzer G*, et al.*, 2010. In situ visualization and dynamics of newly synthesized proteins in rat

hippocampal neurons. *Nat Neurosci*, **13**(7):897-905.
https://doi.org/10.1038/nn.2580

Dieterich DC, Link AJ, Graumann J*, et al.*, 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**(25):9482-7.
https://doi.org/10.1073/pnas.0601637103

Ding W, Yu W, Chen Y*, et al.*, 2024a. Rare codon recoding for efficient noncanonical amino acid incorporation in mammalian cells. *Science*, **384**(6700):1134-1142.
https://doi.org/10.1126/science.adm8143

Ding WL, Liu C, Chen YL*, et al.*, 2024b. Computational design and genetic incorporation of lipidation mimics in living cells. *Nat Chem Biol*, **20**(1):42-+.
https://doi.org/10.1038/s41589-023-01400-8

Dommerholt J, van Rooijen O, Borrmann A*, et al.*, 2014. Highly accelerated inverse electron-demand cycloaddition of electron-deficient azides with aliphatic cyclooctynes. *Nat Commun*, **5**:5378.
https://doi.org/10.1038/ncomms6378

Drienovská I, Roelfes G, 2020. Expanding the enzyme universe with genetically encoded unnatural amino acids. *Nature Catalysis*, **3**(3):193-202.
https://doi.org/10.1038/s41929-019-0410-8

Elliott TS, Bianco A, Chin JW, 2014a. Genetic code expansion and bioorthogonal labelling enables cell specific proteomics in an animal. *Curr Opin Chem Biol*, **21**:154-60.
https://doi.org/10.1016/j.cbpa.2014.07.001

Elliott TS, Bianco A, Townsley FM*, et al.*, 2016. Tagging and enriching proteins enables cell-specific proteomics. *Cell Chem Biol*, **23**(7):805-815.
https://doi.org/10.1016/j.chembiol.2016.05.018

Elliott TS, Townsley FM, Bianco A*, et al.*, 2014b. Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat Biotechnol*, **32**(5):465-72.
https://doi.org/10.1038/nbt.2860

Ernst RJ, Krogager TP, Maywood ES*, et al.*, 2016. Genetic code expansion in the mouse brain. *Nat Chem Biol*, **12**(10):776-778.
https://doi.org/10.1038/nchembio.2160

Fang Y, Yu W, Li J*, et al.*, 2026. Recoding multiple rare codons enables the simultaneous incorporation of up to five distinct noncanonical amino acids. *Nat Chem*, **18**:695–706.
https://doi.org/10.1038/s41557-026-02084-y

Fredens J, Wang K, de la Torre D*, et al.*, 2019. Total synthesis of escherichia coli with a recoded genome. *Nature*, **569**(7757):514-518.
https://doi.org/10.1038/s41586-019-1192-5

Giglione C, Boularot A, Meinnel T, 2004. Protein n-terminal methionine excision. *Cell Mol Life Sci*, **61**(12):1455-74.
https://doi.org/10.1007/s00018-004-3466-8

Grome MW, Nguyen MTA, Moonan DW*, et al.*, 2025. Engineering a genomically recoded organism with one stop codon. *Nature*, **639**(8054):512-521.
https://doi.org/10.1038/s41586-024-08501-x

Gu J, Lao L, Hu L*, et al.*, 2025. Nascent liver proteome reveals enzymes and transcription regulators under physiological and alcohol exposure conditions. *Nat Commun*, **16**(1):7945.
https://doi.org/10.1038/s41467-025-63212-9

Hinz FI, Dieterich DC, Tirrell DA*, et al.*, 2012. Non-canonical amino acid labeling in vivo to visualize and affinity purify newly synthesized proteins in larval zebrafish. *ACS Chem Neurosci*, **3**(1):40-49.
https://doi.org/10.1021/cn2000876

Ho JM, Reynolds NM, Rivera K*, et al.*, 2016. Efficient reassignment of a frequent serine codon in wild-type escherichia coli. *ACS Synth Biol*, **5**(2):163-71.
https://doi.org/10.1021/acssynbio.5b00197

Howden AJ, Geoghegan V, Katsch K*, et al.*, 2013. Quancat: Quantitating proteome dynamics in primary cells. *Nat Methods*, **10**(4):343-6.
https://doi.org/10.1038/nmeth.2401

Huang Y, Zhang P, Wang H*, et al.*, 2025. Genetic code expansion: Recent developments and emerging applications. *Chem Rev*, **125**(2):523-598.
https://doi.org/10.1021/acs.chemrev.4c00216

Ignacio BJ, Dijkstra J, Mora N*, et al.*, 2023. Throncat: Metabolic labeling of newly synthesized proteins using a bioorthogonal threonine analog. *Nat Commun*, **14**(1):3367.

https://doi.org/10.1038/s41467-023-39063-7

Jiang HK, Ambrose NL, Chung CZ*, et al.*, 2023. Split aminoacyl-trna synthetases for proximity-induced stop codon suppression. *Proc Natl Acad Sci U S A*, **120**(8):e2219758120.
https://doi.org/10.1073/pnas.2219758120

Jones CA, Hartman MCT, 2024. Breaking the degeneracy of sense codons - how far can we go? *Israel Journal of Chemistry*, **64**(8-9):e202400026.
https://doi.org/10.1002/ijch.202400026

Kiick KL, Saxon E, Tirrell DA*, et al.*, 2002. Incorporation of azides into recombinant proteins for chemoselective modification by the staudinger ligation. *Proc Natl Acad Sci USA*, **99**(1):19-24.
https://doi.org/DOI 10.1073/pnas.012583299

Koehler C, Estrada Girona G, Reinkemeier CD*, et al.*, 2020. Inducible genetic code expansion in eukaryotes. *ChemBioChem*, **21**(22):3216-3219.
https://doi.org/10.1002/cbic.202000338

Krishnakumar R, Prat L, Aerni HR*, et al.*, 2013. Transfer rna misidentification scrambles sense codon recoding. *ChemBioChem*, **14**(15):1967-72.
https://doi.org/10.1002/cbic.201300444

Krogager TP, Ernst RJ, Elliott TS*, et al.*, 2018. Labeling and identifying cell-specific proteomes in the mouse brain. *Nat Biotechnol*, **36**(2):156-159.
https://doi.org/10.1038/nbt.4056

Lajoie MJ, Rovner AJ, Goodman DB*, et al.*, 2013. Genomically recoded organisms expand biological functions. *Science*, **342**(6156):357-60.
https://doi.org/10.1126/science.1241459

Lee BS, Shin S, Jeon JY*, et al.*, 2015. Incorporation of unnatural amino acids in response to the agg codon. *ACS Chem Biol*, **10**(7):1648-53.
https://doi.org/10.1021/acschembio.5b00230

Lepthien S, Merkel L, Budisa N, 2010. In vivo double and triple labeling of proteins using synthetic amino acids. *Angew Chem Int Ed Engl*, **49**(32):5446-50.
https://doi.org/10.1002/anie.201000439

Liu DR, Schultz PG, 1999. Progress toward the evolution of an organism with an expanded genetic code. *Proc Natl Acad Sci U S A*, **96**(9):4780-5.
https://doi.org/10.1073/pnas.96.9.4780

Liu R, Kenney JW, Manousopoulou A*, et al.*, 2016. Quantitative non-canonical amino acid tagging (quancat) proteomics identifies distinct patterns of protein synthesis rapidly induced by hypertrophic agents in cardiomyocytes, revealing new aspects of metabolic remodeling. *Mol Cell Proteomics*, **15**(10):3170-3189.
https://doi.org/10.1074/mcp.M115.054312

Loynd C, Singha Roy SJ, Canarelli SE*, et al.*, 2026. Cell-selective multiplexed bioorthogonal noncanonical amino acid tagging for nascent proteomics. *Nat Chem Biol*, **22**(2):318-327.
https://doi.org/10.1038/s41589-025-02039-3

Mahdavi A, Hamblin GD, Jindal GA*, et al.*, 2016. Engineered aminoacyl-trna synthetase for cell-selective analysis of mammalian protein synthesis. *J Am Chem Soc*, **138**(13):4278-81.
https://doi.org/10.1021/jacs.5b08980

Mahdavi A, Szychowski J, Ngo JT*, et al.*, 2014. Identification of secreted bacterial proteins by noncanonical amino acid tagging. *Proc Natl Acad Sci USA*, **111**(1):433-438.
https://doi.org/10.1073/pnas.1301740111

Michels DE, Lomenick B, Chou TF*, et al.*, 2021. Amino acid analog induces stress response in marine. *Appl Environ Microbiol*, **87**(15):e0020021.
https://doi.org/10.1128/AEM.00200-21

Ngo JT, Tirrell DA, 2011. Noncanonical amino acids in the interrogation of cellular protein synthesis. *Acc Chem Res*, **44**(9):677-85.
https://doi.org/10.1021/ar200144y

Nguyen DP, Lusic H, Neumann H*, et al.*, 2009. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins a pyrrolysyl-trna synthetase/trna pair and click chemistry. *J Am Chem Soc*, **131**(25):8720-+.
https://doi.org/10.1021/ja900553w

Niu W, Guo J, 2024. Cellular site-specific incorporation of noncanonical amino acids in synthetic biology. *Chem Rev*, **124**(18):10577-10617.
https://doi.org/10.1021/acs.chemrev.3c00938

Noren CJ, Anthony-Cahill SJ, Griffith MC*, et al.*, 1989. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science*, **244**(4901):182-8.
https://doi.org/10.1126/science.2649980

Nyerges A, Chiappino-Pepe A, Budnik B*, et al.*, 2024. Synthetic genomes unveil the effects of synonymous recoding. *bioRxiv*: .
https://doi.org/10.1101/2024.06.16.599206

Osgood AO, Huang ZY, Szalay KH*, et al.*, 2025. Strategies to expand the genetic code of mammalian cells. *Chemical Reviews*, **125**(4):2474-2501.
https://doi.org/10.1021/acs.chemrev.4c00730

Ostrov N, Landon M, Guell M*, et al.*, 2016. Design, synthesis, and testing toward a 57-codon genome. *Science*, **353**(6301):819-22.
https://doi.org/10.1126/science.aaf3639

Patterson DM, Nazarova LA, Xie B*, et al.*, 2012. Functionalized cyclopropenes as bioorthogonal chemical reporters. *J Am Chem Soc*, **134**(45):18638-18643.
https://doi.org/10.1021/ja3060436

Phillips NJ, Vinaithirthan BM, Oses-Prieto JA*, et al.*, 2023. Capture, release, and identification of newly synthesized proteins for improved profiling of functional translatomes. *Mol Cell Proteomics*, **22**(3):100497.
https://doi.org/10.1016/j.mcpro.2023.100497

Reinkemeier CD, Girona GE, Lemke EA, 2019. Designer membraneless organelles enable codon reassignment of selected mrnas in eukaryotes. *Science*, **363**(6434) https://doi.org/10.1126/science.aaw2644

Robertson WE, Funke LFH, de la Torre D*, et al.*, 2021. Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science*, **372**(6546):1057-1062.
https://doi.org/10.1126/science.abg3029

Robertson WE, Rehm FBH, Spinck M*, et al.*, 2025. Escherichia coli with a 57-codon genetic code. *Science*, **390**(6771):eady4368.
https://doi.org/10.1126/science.ady4368

Saint-Leger A, Bello C, Dans PD*, et al.*, 2016. Saturation of recognition elements blocks evolution of new trna identities. *Sci Adv*, **2**(4):e1501860.
https://doi.org/10.1126/sciadv.1501860

Schwark DG, Schmitt MA, Biddle W*, et al.*, 2020. The influence of competing trna abundance on translation: Quantifying the efficiency of sense codon reassignment at rarely used codons. *ChemBioChem*, **21**(16):2274-2286.
https://doi.org/10.1002/cbic.202000052

Si LL, Xu H, Zhou XY*, et al.*, 2016. Generation of influenza a viruses as live but replication-incompetent virus vaccines. *Science*, **354**(6316):1170-1173.
https://doi.org/10.1126/science.aah5869

Srinivasan G, James CM, Krzycki JA, 2002. Pyrrolysine encoded by uag in archaea: Charging of a uag-decoding specialized trna. *Science*, **296**(5572):1459-62.
https://doi.org/10.1126/science.1069588

Szychowski J, Mahdavi A, Hodas JJ*, et al.*, 2010. Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J Am Chem Soc*, **132**(51):18351-60.
https://doi.org/10.1021/ja1083909

Tang Q, Chen X, 2023. Nascent proteomics: Chemical tools for monitoring newly synthesized proteins. *Angew Chem Int Ed Engl*, **62**(40):e202305866.
https://doi.org/10.1002/anie.202305866

Tanrikulu IC, Schmitt E, Mechulam Y*, et al.*, 2009. Discovery of escherichia coli methionyl-trna synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci U S A*, **106**(36):15285-90.
https://doi.org/10.1073/pnas.0905735106

Tharp JM, Vargas-Rodriguez O, Schepartz A*, et al.*, 2021. Genetic encoding of three distinct noncanonical amino acids using reprogrammed initiator and nonsense codons. *ACS Chem Biol*, **16**(4):766-774.
https://doi.org/10.1021/acschembio.1c00120

Tittle JM, Schwark DG, Biddle W*, et al.*, 2022. Impact of queuosine modification of endogenous e. Coli trnas on sense codon reassignment. *Front Mol Biosci*, **9**:938114.
https://doi.org/10.3389/fmolb.2022.938114

Tivendale ND, Fenske R, Duncan O*, et al.*, 2021. Homopropargylglycine incorporation enables sampling, isolation and characterization of nascent proteins from. *Plant J*, **107**(4):1260-1276.
https://doi.org/10.1111/tpj.15376

Verhelst SH, Fonovic M, Bogyo M, 2007. A mild chemically cleavable linker system for functional proteomic applications. *Angew Chem Int Ed Engl*, **46**(8):1284-6.

https://doi.org/10.1002/anie.200603811

Wang A, Winblade Nairn N, Johnson RS*, et al.*, 2008. Processing of n-terminal unnatural amino acids in recombinant human interferon-beta in escherichia coli. *ChemBioChem*, **9**(2):324-30.
https://doi.org/10.1002/cbic.200700379

Wang K, Fredens J, Brunner SF*, et al.*, 2016. Defining synonymous codon compression schemes by genome recoding. *Nature*, **539**(7627):59-64.
https://doi.org/10.1038/nature20124

Wang M, Herrmann CJ, Simonovic M*, et al.*, 2015. Version 4.0 of paxdb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics*, **15**(18):3163-8.
https://doi.org/10.1002/pmic.201400441

Wang M, Weiss M, Simonovic M*, et al.*, 2012. Paxdb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics*, **11**(8):492-500.
https://doi.org/10.1074/mcp.O111.014704

Wang Y, Tsao ML, 2016. Reassigning sense codon aga to encode noncanonical amino acids in escherichia coli. *ChemBioChem*, **17**(23):2234-2239.
https://doi.org/10.1002/cbic.201600448

Willis JCW, Chin JW, 2018. Mutually orthogonal pyrrolysyl-trna synthetase/trna pairs. *Nat Chem*, **10**(8):831-837.
https://doi.org/10.1038/s41557-018-0052-5

Yang AC, du Bois H, Olsson N*, et al.*, 2018. Multiple click-selective trna synthetases expand mammalian cell-specific proteomics. *J Am Chem Soc*, **140**(23):7046-7051.
https://doi.org/10.1021/jacs.8b03074

Yi HB, Lee S, Seo K*, et al.*, 2024. Cellular and biophysical applications of genetic code expansion. *Chemical Reviews*, **124**(11):7465-7530.
https://doi.org/10.1021/acs.chemrev.4c00112

Yuet KP, Tirrell DA, 2014. Chemical tools for temporally and spatially resolved mass spectrometry-based proteomics. *Ann Biomed Eng*, **42**(2):299-311.
https://doi.org/10.1007/s10439-013-0878-3

Zeng Y, Wang W, Liu WR, 2014. Towards reassigning the rare agg codon in escherichia coli. *ChemBioChem*, **15**(12):1750-4.
https://doi.org/10.1002/cbic.201400075

Zhang G, Bowling H, Hom N*, et al.*, 2014. In-depth quantitative proteomic analysis of de novo protein synthesis induced by brain-derived neurotrophic factor. *J Proteome Res*, **13**(12):5707-14.
https://doi.org/10.1021/pr5006982

Zhao HX, Tang L, Fang Y*, et al.*, 2023. Manipulating cation-π interactions of reader proteins in living cells with genetic code expansion. *J Am Chem Soc*, **145**(30):16406-16416.
https://doi.org/10.1021/jacs.3c02293