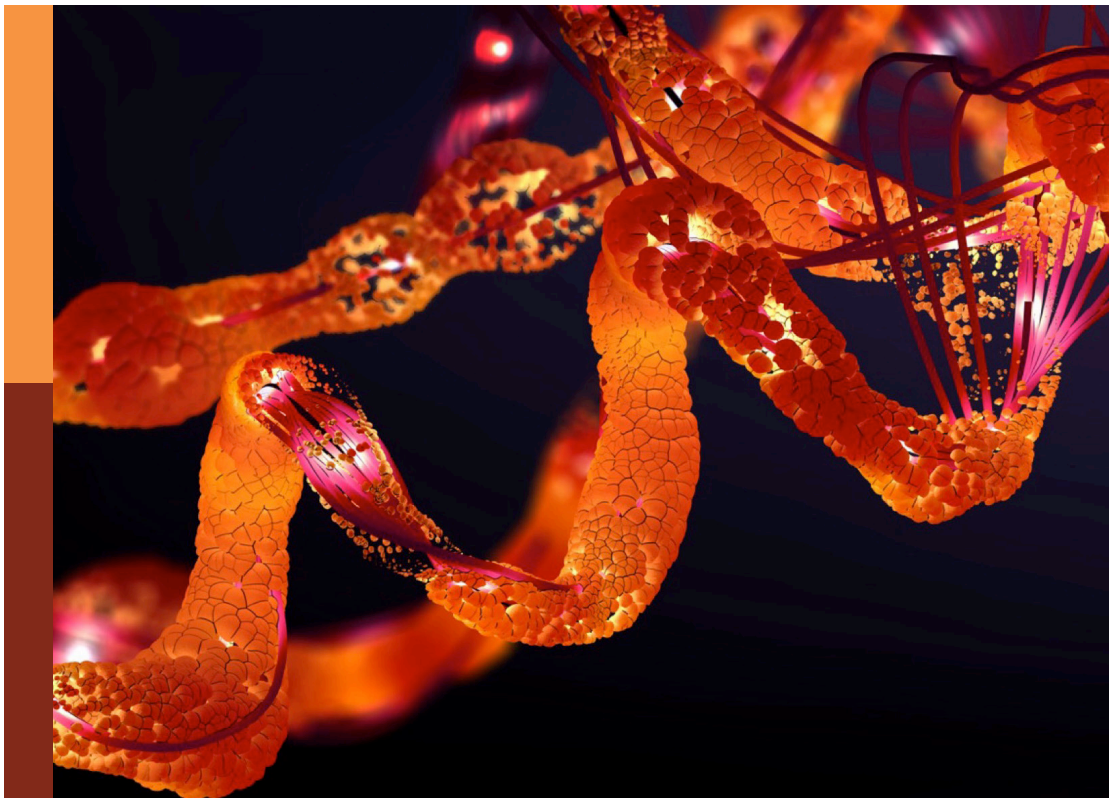
frontiers | Research Topics

Exploring and expanding the protein universe with non-canonical amino acids

Edited by
Gustavo Fuertes, Kensaku Sakamoto and Ned Budisa

Published in
Frontiers in Molecular Biosciences
Frontiers in Bioengineering and Biotechnology



ISSN 1664-8714
ISBN 978-2-8325-3802-9
DOI 10.3389/978-2-8325-3802-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Exploring and expanding the protein universe with non-canonical amino acids

Topic editors

Gustavo Fuertes — Institute of Biotechnology (ASCR), Czechia

Kensaku Sakamoto — RIKEN, Japan

Ned Budisa — University of Manitoba, Canada

Citation

Fuertes, G., Sakamoto, K., Budisa, N., eds. (2023). *Exploring and expanding the protein universe with non-canonical amino acids*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3802-9

Table of contents

04 Editorial: Exploring and expanding the protein universe with non-canonical amino acids
Gustavo Fuertes, Kensaku Sakamoto and Nediljko Budisa

07 Directed Evolution of *Methanomethylophilus alvus* Pyrrolysyl-tRNA Synthetase Generates a Hyperactive and Highly Selective Variant
Jonathan T. Fischer, Dieter Söll and Jeffery M. Tharp

18 Cell Engineering and Cultivation of Chinese Hamster Ovary Cells for the Development of Orthogonal Eukaryotic Cell-free Translation Systems
Jeffrey L. Schloßhauer, Niño Cavak, Anne Zemella, Lena Thoring and Stefan Kubick

31 Genetic Code Engineering by Natural and Unnatural Base Pair Systems for the Site-Specific Incorporation of Non-Standard Amino Acids Into Proteins
Michiko Kimoto and Ichiro Hirao

46 Tuning the Properties of Protein-Based Polymers Using High-Performance Orthogonal Translation Systems for the Incorporation of Aromatic Non-Canonical Amino Acids
Osher Gueta, Ortal Sheinenzon, Rotem Azulay, Hadas Shalit, Daniela S. Strugach, Dagan Hadar, Sigal Gelkop, Anat Milo and Miriam Amiram

58 Impact of queuosine modification of endogenous *E. coli* tRNAs on sense codon reassignment
Jillyn M. Tittle, David G. Schwark, Wil Biddle, Margaret A. Schmitt and John D. Fisk

70 Delivery of the selenoprotein thioredoxin reductase 1 to mammalian cells
David E. Wright, Tarana Siddika, Ilka U. Heinemann and Patrick O'Donoghue

82 Engineered bacterial host for genetic encoding of physiologically stable protein nitration
Nikolaj G. Koch, Tobias Baumann, Jessica H. Nickling, Anna Dziegielewski and Nediljko Budisa

98 Dual incorporation of non-canonical amino acids enables production of post-translationally modified selenoproteins
Pearl Morosky, Cody Comyns, Lance G. A. Nunes, Christina Z. Chung, Peter R. Hoffmann, Dieter Söll, Oscar Vargas-Rodriguez and Natalie Krahn

108 Regulation of IL-24/IL-20R2 complex formation using photocaged tyrosines and UV light
Phuong Ngoc Pham, Jiří Zahradník, Lucie Kolářová, Bohdan Schneider and Gustavo Fuertes

frontiers | Frontiers in Molecular Biosciences

TYPE Editorial
PUBLISHED 12 October 2023
DOI 10.3389/fmolb.2023.1303286

OPEN ACCESS

EDITED AND REVIEWED BY
Andrea Mozzarelli,
University of Parma, Italy

*CORRESPONDENCE
Gustavo Fuertes,
gustavo.fuertes@ibt.cas.cz
Kensaku Sakamoto,
kensaku.sakamoto@riken.jp
Nediljko Budisa,
nediljko.budisa@umanitoba.ca

RECEIVED 27 September 2023
ACCEPTED 02 October 2023
PUBLISHED 12 October 2023

CITATION
Fuertes G, Sakamoto K and Budisa N (2023), Editorial: Exploring and expanding the protein universe with non-canonical amino acids.
Front. Mol. Biosci. 10:1303286.
doi: 10.3389/fmolb.2023.1303286

Editorial: Exploring and expanding the protein universe with non-canonical amino acids

Gustavo Fuertes[1]*, Kensaku Sakamoto[2]* and Nediljko Budisa[3]*

[1]Laboratory of Biomolecular Recognition, Institute of Biotechnology of the Czech Academy of Sciences, Vestec, Czechia, [2]Laboratory for Nonnatural Amino Acid Technology, RIKEN Center for Biosystems Dynamics Research, Yokohama, Japan, [3]Chemical Synthetic Biology Group, Department of Chemistry, University of Manitoba, Winnipeg, Canada

KEYWORDS

non-canonical amino acid (ncAA), genetic code expansion (GCE), stop codon suppression (SCS), sense codon reassignment (SCR), unnatural base pair (UBP), cell-free translation system (CFTS), aminoacyl tRNA synthetase (aaRS), orthogonal life

Editorial on the Research Topic
Exploring and expanding the protein universe with non-canonical amino acids

In just 30 years, genetic code engineering has allowed us to expand the repertoire of amino acids in proteins from nature's 20 canonical amino acids to more than 250 non-canonical amino acids (ncAAs), including non-α-amino acids (Fricke et al., 2023), and the trend is increasing. It is now possible to incorporate many types of amino acid substrates (natural and unnatural, long/heavy/bulky, aliphatic, aromatic, halogenated, etc.) into recombinant proteins. Such a field development, especially genetic code expansion (GCE) by orthogonal pairs and expanded genetic alphabet, is unprecedented in natural sciences. Applications of ncAAs are diverse, ranging from biochemical studies of protein-activity relationships with atomic precision to the generation of protein-based polymers with novel functions and ultimately novel life forms. Despite the advantages of ncAAs, the understanding, manipulation, and design of protein structure, dynamics, and function still largely rely on canonical amino acids. Our long-term mission is to transform GCE into a routine toolbox for many laboratories and industries.

The present Research Topic "*Exploring and Expanding the Protein Universe with Non-canonical Amino Acids*" aims to provide the reader with the fundamentals of GCE along with the latest advances. The Research Topic contains 1 review and 8 original research articles, with contributions from both experts and newcomers in the field.

An excellent starting point for anyone interested in genetic code engineering is the review paper by Kimoto and Hirao, who discuss both natural base pairs (NBP) and unnatural base pairs (UBP) from the perspective of codon-anticodon interactions. The NBP system includes stop codon suppression (SCS), four-base codon-anticodon interactions, and sense codon reprogramming (SCR). SCS, and in particular, in-frame amber (UAG) codon suppression, is by far the most popular implementation of GCE. In fact, only one original research paper ventures to do SCR, the work by Tittle et al. The authors of this study conclude that in the absence of queuosine nucleoside found in the anticodons of some *E. coli* tRNAs, reassignment of sense codons is slightly enhanced.

Although our Research Topic is dominated by intact cells as platforms for the incorporation of ncAA, cell-free translation systems (CFTS) are also presented. CFTS are particularly useful for expressing so-called "difficult" proteins such as toxins and membrane proteins that would otherwise threaten cell viability. Moreover, the current practice involves encoding the components in plasmids, which may not always be stable. Schloβhauer et al. developed two orthogonal eukaryotic CFTS derived from the Chinese Hamster Ovary (CHO) cell line. The first system is based on transient transfection and expression of aminoacyl-tRNA synthetases (aaRS) prior to cell disruption for extract preparation. The second system is based on stable transfected cells containing aaRS expression cassettes at a defined locus created by the CRISPR/cas9 genomic editing system.

The majority of studies using GCE methods focus on single-point mutations. From a technical perspective, the difficulty of incorporating a particular ncAA at multiple sites is a direct consequence of the relatively low efficiency of most aaRS variants, although the chemical instability of some ncAAs may also play a role. Gueta et al. reported a set of powerful aaRS for the incorporation of 15 different aromatic ncAAs at up to 10 residue positions in the elastin-like polypeptide (ELP), an intrinsically disordered protein. Koch et al. chose a different approach to increase the yield of genetically encoded protein nitration. They prevented the reduction of nitro groups by engineering an *E. coli* strain with reduced nitroreductase activity. The result is an ELP variant carrying up to 60 copies of a nitrobenzyl-containing ncAA, which is the largest number of ncAAs ever introduced in a single polypeptide.

An even greater challenge is the incorporation of two or more distinct ncAAs, partially due to the lack of mutually orthogonal aaRS/tRNA pairs, quintuply orthogonal being the current frontier (Beattie et al., 2023). The two most common aaRS employed in GCE campaigns are the tyrosyl-tRNA synthetase from *Methanocaldococcus jannaschii* (*Mj*TyrRS) and the pyrrolysyl-tRNA synthetase (PylRS) from *Methanosarcinae*. Other PylRS from *Methanomethylophilus alvus* and, more recently, from *Methanococcoides burtonii* (Koch et al., 2023) are also rapidly gaining momentum. Fisher et al. evolved an optimized *Ma*PylRS variant ($MaPylRS_{opt}$) by phage-assisted non-continuous evolution. $MaPylRS_{opt}$ is hyperactive, specifically recognizes Nε-substituted lysines and certain phenylalanine derivatives, but not *para*-substituted ones, and is orthogonal to *Mj*TyrRS, making it an excellent tool for the single and dual incorporation of diverse ncAAs. Another method established by Morosky et al. allows the incorporation of selenocysteine (the 21st proteinogenic amino acid) and Nε-acetyl-lysine (a common post-translational modification) at UGA and UAG codons, respectively. As a result, acetylated selenoproteins can be produced in *E. coli* by dual SCS.

Sometimes, a protein of interest cannot be readily produced in a given host. This is the case with the selenoprotein thioredoxin reductase (trxR1) in mammalian cells. To circumvent this problem, Wright et al. fused a cell-penetrating peptide tag derived from the trans-activator of transcription (TAT) protein of human immunodeficiency virus. Purified TAT-trxR1, recombinantly expressed in *E. coli* by GCE, is efficiently uptaken by human cells, providing a new platform to study such a protein *in situ*.

Finally, another application of ncAAs is in the area of photocontrol to switch/turn protein activity ON and OFF. Pham et al. report the use of a photocaged tyrosine (NBY) to control the binding affinity between two medically relevant proteins, interleukin-24 (a cytokine) and its receptor IL-20R2, by UV light. Cell signaling through the JAK/STAT phosphorylation cascade can, thus, be regulated as needed.

It has been almost three decades since the field of genetic code engineering emerged in the 1990s, and during this time, we have witnessed not only significant advancements in methodology but also the emergence of intriguing concepts. For instance, Szostak and his colleagues proposed that approximately 70% of codons could be reassigned (Herman et al., 2007). Similarly, Söll and his research team estimated that it might be possible to encode the genetic makeup of an organism using only 30 to 40 sense codons (Krishnakumar et al., 2013), leaving over 20 sense codons available for reassignment with ncAAs (Mukai et al., 2015). Against this background, we are pleased that the articles in our Research Topic contribute by presenting various aspects of reprogramming the genetic code, from basic principles to practical examples. These efforts are aimed at fostering further advancements in this technology. For instance, merging orthogonal translation with synthetic metabolism (Völler and Budisa, 2017) would reduce the need for external supplementation of ncAAs (or nucleobases).

Undoubtedly, the unexplored potential of ncAAs will attract researchers from diverse disciplines, including AI, material science, biophysics, biomedicine, and evolutionary biology, among others, to engage in this captivating field. Thanks to these collaborative endeavors, the boundaries of the protein universe and life itself will be pushed, explored, and expanded.

Author contributions

GF: Writing–original draft, Writing–review and editing. KS: Writing–review and editing. NB: Writing–original draft, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. GF acknowledges funding by the Czech Academy of Sciences, grant RVO 86652036.

Conflict of interest

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

Publisher's note

References

Beattie, A. T., Dunkelmann, D. L., and Chin, J. W. (2023). Quintuply orthogonal pyrrolysyl-tRNA synthetase/tRNAPyl pairs. *Nat. Chem.* 15 (7), 948–959. doi:10.1038/s41557-023-01232-y

Fricke, R., Swenson, C. V., Roe, L. T., Hamlish, N. X., Shah, B., Zhang, Z., et al. (2023). Expanding the substrate scope of pyrrolysyl-transfer RNA synthetase enzymes to include non-α-amino acids *in vitro* and *in vivo*. *Nat. Chem.* 15 (7), 960–971. doi:10.1038/s41557-023-01224-y

Herman, C., Hartman, M. C. T., Josephson, K., Lin, C.-W., and Szostak, J. W. (2007). An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS ONE* 2 (10), e972. doi:10.1371/journal.pone.0000972

Koch, N. G., Goettig, P., Rappsilber, J., and Budisa, N. (2023). Cold" orthogonal translation: psychrophilic pyrrolysyl-tRNA synthetase as efficient tool for expanding the genetic code. bioRxiv, Available at: https://doi.org/10.1101/2023.05.23.541947.

Krishnakumar, R., Prat, L., Aerni, H.-R., Ling, J., Merryman, C., Glass, J. I., et al. (2013). Transfer RNA misidentification scrambles sense codon recoding. *ChemBioChem* 14 (15), 1967–1972. doi:10.1002/cbic.201300444

Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., et al. (2015). Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.* 43 (16), 8111–8122. doi:10.1093/nar/gkv787

Völler, J.-S., and Budisa, N. (2017). Coupling genetic code expansion and metabolic engineering for synthetic cells. *Curr. Opin. Biotechnol.* 48, 1–7. doi:10.1016/j.copbio.2017.02.002

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 09 March 2022
doi: 10.3389/fmolb.2022.850613

Directed Evolution of *Methanomethylophilus alvus* Pyrrolysyl-tRNA Synthetase Generates a Hyperactive and Highly Selective Variant

Jonathan T. Fischer[1]*, *Dieter Söll*[1,2] *and Jeffery M. Tharp*[1]

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, United States, [2]Department of Chemistry, Yale University, New Haven, CT, United States

Pyrrolysyl-tRNA synthetase (PylRS) is frequently used for site-specific incorporation of noncanonical amino acids (ncAAs) into proteins. Recently, the active site of *Methanomethylophilus alvus* PylRS (*Ma*PylRS) has been rationally engineered to expand its substrate compatibility, enabling the incorporation of difficult ncAAs. However, mutations beyond the active site that enhance the enzymatic properties of *Ma*PylRS have not been reported. We utilized phage-assisted non-continuous evolution (PANCE) to evolve *Ma*PylRS to efficiently incorporate N^{ε}-Boc-L-lysine (BocK). Directed evolution yielded several mutations outside of the active site that greatly improve the activity of the enzyme. We combined the most effective mutations to generate a new PylRS variant ($PylRS_{opt}$) that is highly active and selective towards several lysine and phenylalanine derivatives. The mutations in $PylRS_{opt}$ can be used to enhance previously engineered PylRS constructs such as $MaPylRS_{N166S}$, and $PylRS_{opt}$ is compatible in applications requiring dual ncAA incorporation and substantially improves the yield of these target proteins.

Keywords: directed evolution, PANCE, PylRS, noncanonical amino acid, tRNA, orthogonal, synthetic biology, pyrrolysyl-tRNA synthetase

OPEN ACCESS

Edited by:
Kensaku Sakamoto, RIKEN, Japan

Reviewed by:
Wenshe Ray Liu, Texas A&M University, United States
Tatsuo Yanagisawa, RIKEN Yokohama, Japan

***Correspondence:**
Jonathan T. Fischer
jonathan.t.fischer@gmail.com

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

Received: *07 January 2022*
Accepted: *15 February 2022*
Published: *09 March 2022*

Citation:
Fischer JT, Söll D and Tharp JM (2022) Directed Evolution of Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase Generates a Hyperactive and Highly Selective Variant. Front. Mol. Biosci. 9:850613. doi: 10.3389/fmolb.2022.850613

INTRODUCTION

The genetic code consists of 61 triplet codons that code for 20 canonical amino acids, as well as three stop codons that function as termination signals to end translation and release the protein from the ribosome. However, exceptions to this rule are apparent, and the plasticity of translation is well-founded. A naturally occurring example is found in several methanogenic species of archaea and bacteria. In these organisms, pyrrolysine (Pyl) is encoded in the active site of methylamine methyltransferase by the amber stop codon UAG (Hao et al., 2002; Srinivasan et al., 2002; Polycarpo et al., 2004). This is accomplished *via* pyrrolysyl-tRNA synthetase (PylRS), a class II aminoacyl-tRNA synthetase that aminoacylates its cognate tRNA, $tRNA^{Pyl}$ (Srinivasan et al., 2002; Polycarpo et al., 2004). This unique tRNA features a CUA anticodon, which enables Pyl-$tRNA^{Pyl}$ to suppress UAG and insert Pyl into the elongating polypeptide through the same translational machinery as canonical tRNAs (Théobald-Dietrich et al., 2004; Zhang et al., 2005; Longstaff et al., 2007).

The PylRS/tRNA^{Pyl} system has been utilized in numerous studies to genetically encode ncAAs at a stop codon (most frequently UAG), altering the target protein's structure and/or function (Budisa, 2005; Wan et al., 2014; Crnković et al., 2016; Tharp et al., 2018). To that end, PylRS has been developed and implemented to incorporate a wide variety of ncAAs, including N^{ε}-substituted lysine (Polycarpo et al., 2006; Neumann et al., 2008; Hancock et al., 2010; Umehara et al., 2012) and, to a lesser extent, *ortho*- and *meta*-substituted phenylalanine (Wang et al., 2011; Ko et al., 2013; Tharp et al., 2021). One of the critical characteristics of the PylRS/tRNA^{Pyl} system is its orthogonality with both bacterial and mammalian host translation machinery (Wan et al., 2014). PylRS/tRNA^{Pyl} can also be used in tandem with mutually orthogonal genetic code expansion tools to incorporate multiple ncAAs into a single protein, such as the *M. jannaschii* TyrRS/tRNA^{Tyr} system. These genetic code expansion tools have been used together to site-specifically incorporate two ncAAs into a single protein in response to two stop codons, most frequently UAG and UAA (Wan et al., 2010; Chatterjee et al., 2013).

Methanomethylophilus alvus PylRS (*Ma*PylRS) is a highly active PylRS homolog. Like its previously characterized counterparts, the ncAA specificity of *Ma*PylRS has been modified through active site engineering (Meineke et al., 2018; Willis and Chin, 2018; Beránek et al., 2019; Seki et al., 2020; Tharp et al., 2021). *Ma*PylRS belongs to the class of PylRS proteins that lacks the poorly soluble N-terminal domain. This confers an advantage to *Ma*PylRS, as *Ma*PylRS is far more soluble than full-length PylRS homologs such as *M. mazei* PylRS (Yanagisawa et al., 2008; Seki et al., 2020). The high activity, solubility, and tunable specificity of *Ma*PylRS position it as a powerful genetic code expansion tool.

Improving the activity or altering the substrate specificity of aminoacyl-tRNA synthetases such as PylRS can be approached in two ways: rational engineering or randomized selection (Vargas-Rodriguez et al., 2018; Krahn et al., 2020). Rational engineering typically relies on structural information to guide the process of mutating key residues in the amino acid binding pocket (Wang et al., 2012; Seki et al., 2020). This approach has been used successfully and is straightforward to execute; however, there are likely many targets for mutation that can have a dramatic impact on activity but are easily overlooked during the rational design process. Conversely, randomized selection through library screening or directed evolution is unbiased and can result in powerful mutations that would likely never have been considered for rational engineering (Bryson et al., 2017; Vargas-Rodriguez et al., 2018; Baumann et al., 2019). One such method of randomized selection is phage-assisted non-continuous evolution (PANCE) (Esvelt et al., 2011; Suzuki et al., 2017; Miller et al., 2020). Unlike library-based evolution methods, PANCE does not require structural information to guide selection. PANCE is also simple to implement, as no specialized equipment is necessary. The technique offers tunable stringency and has been shown to generate highly active mutant variants of enzymes such as PylRS. Indeed, PANCE and its continuous evolution counterpart PACE have been successfully used to evolve highly active variants of chPylRS, an enzyme derived from a fusion construct of *M. mazei* and *M. barkeri* PylRS (Bryson et al., 2017; Suzuki et al., 2017).

In this study, we utilized PANCE to evolve *Ma*PylRS with the goal of developing a hyperactive PylRS variant that recognizes a broad spectrum of ncAAs while still discriminating against the canonical amino acids. Sequencing of evolved PylRS variants revealed a polymorphic population of mutations, almost all of which were located outside of the active site. We screened the activity of the mutants towards a variety of Lys- and Phe-ncAAs using *in vivo* fluorescence and chloramphenicol acetyltransferase assays, and we combined the most active mutations to create a new PylRS variant, PylRS$_{opt}$. Our data indicate that PylRS$_{opt}$ excludes canonical amino acids and recognizes a diverse pool of ncAAs, incorporating them with vastly improved activity compared to wild-type *Ma*PylRS. In combination with the *M. jannaschii* TyrRS, the high activity and selectivity of PylRS$_{opt}$ enables robust incorporation of multiple ncAAs into a single protein, which has applications including bioorthogonal click chemistry (Wan et al., 2010; Meineke et al., 2020), peptide cyclization (Neumann et al., 2010; Hayes et al., 2021), and FRET (Wu et al., 2012).

MATERIALS AND METHODS

PANCE

*Ma*PylRS was cloned into M13 phage, replacing gene III. Phages were initially propagated without selection using the permissive host *E. coli* S1059. Cultures were grown in 2xYT media supplemented with the appropriate antibiotics at 37°C until reaching an OD_{600} of 0.4–0.6. Cells were then infected with a viral load of phage ranging from 10^4–10^6 pfu/ml, and the culture was grown overnight at 37°C. The following day, a 1 ml aliquot of the overnight culture was centrifuged at 14,000 × g for 1 min, and the phage-containing supernatant was decanted and stored at 4°C. Three independent phage lineages were maintained throughout the PANCE process.

Positive selection was performed using *E. coli* S1030 cells transformed with the accessory plasmid pJT017 (gIII$_{1xTAG}$ *Ma*tRNA$^{Pyl}_{(6)}$), as well as mutagenesis plasmid MP4 during rounds of mutagenesis. 30 ml cultures were grown in 2xYT media supplemented with 20 mM glucose and the appropriate antibiotics to an OD_{600} of 0.4–0.6. BocK (5 mM) was then added where indicated, and during rounds of mutagenesis, 5 mM arabinose was also added. Cultures were then infected with 10^4–10^6 pfu/ml M13ΔgIII:*Ma*PylRS. The infected culture grew overnight at 37°C, and phages were harvested the next morning. In later rounds of higher stringency positive selection, plasmids pJT018 (gIII$_{2xTAG}$ *Ma*tRNA$^{Pyl}_{(6)}$) and pJT019 (gIII$_{3xTAG}$ *Ma*tRNA$^{Pyl}_{(6)}$) were used in place of pJT017.

Negative selection was carried out using the negative selection plasmids pJF011 (T7RNAP$_{2xTAG}$ *Ma*tRNA$^{Pyl}_{(6)}$) and pDB016 (carrying gIII under a T7 promoter). 30 ml cultures were grown in 2xYT media, and at an OD_{600} of 0.4–0.6, the cells were infected with 10^4–10^6 pfu/ml of phage. The culture grew overnight at 37°C, and phages were harvested the following morning.

Phage titers were performed at the conclusion of every generation as a checkpoint to ensure that phage propagation remained robust and selective for BocK. Collected phage samples were serially diluted and plated on LB-top agar plates containing *E. coli* S1059 cells. 5 μL of each serial dilution were spotted on the plates, and the pfu/mL for each phage lineage +/− BocK was calculated.

In vivo Fluorescence Assays

Electrocompetent *E. coli* DH10B cells were transformed with a pMW plasmid encoding the PylRS variant, and a pBAD plasmid encoding sfGFP 2TAG and *Mat*RNA^{Pyl}. Variations of the two plasmids were cloned and used where indicated. Colonies of freshly transformed DH10B cells harboring the indicated plasmids were picked and grown at 37°C overnight in 5 ml of LB media supplemented with appropriate antibiotics. 2 μL of the overnight cultures were diluted into 150 μL LB supplemented with 500 μM IPTG, 0.2% arabinose, appropriate antibiotics, and 1 mM of the indicated ncAA. Inoculated cultures were grown shaking at 37°C in a BioTek Synergy plate reader. Fluorescence intensity (λ_{ex} = 485nm, λ_{em} = 528 nm) and OD_{600} were measured every 15 min, and fluorescence/OD_{600} was calculated using the 12-h time point. Fluorescence/OD_{600} measurements and standard deviations were calculated based on data collected from three biological and two technical replicates.

For the dual ncAA incorporation assay, a three-plasmid system was used: 1) pMW AzFRS.2. t1, 2) pULTRA encoding the PylRS variant and the ochre-suppressing mutant tRNA *Mat*RNA$^{Pyl}_{(6)/UUA}$, and 3) pBAD sfGFP 2TAG 149TAA *M. jannaschii* tRNA$^{Tyr}_{CUA}$. The plasmids were co-transformed into electrocompetent *E. coli* DH10B cells. The rest of the procedure follows the above protocol.

Protein Expression and Purification

Electrocompetent DH10B cells were co-transformed with pBAD sfGFP 2TAG *Mat*RNA^{Pyl} and pMW *Ma*PylRS or PylRS$_{opt}$. Fresh colonies were used to inoculate 20 ml LB media supplemented with 0.2% arabinose, 500 μM IPTG, 1 mM BocK, 25 μg/ml spectinomycin, and 100 μg/ml ampicillin. Cultures were grown shaking at 37°C overnight. The following morning, the cultures were centrifuged at 4,000 × g for 10 min, and the pellets were resuspended in 1X BugBuster protein extraction reagent (Millipore-Sigma). The lysate was clarified by centrifugation (15,000 × g for 20 min) and the cleared lysate was incubated with nickel-NTA agarose beads for 20 min. The beads were washed five times with wash buffer (50 mM Tris pH 8.0, 50 mM NaCl, 10 mM imidazole), and eluted with elution buffer (50 mM Tris pH 8.0, 50 mM NaCl, 250 mM imidazole. The purified proteins were concentrated and buffer-exchanged into 25 mM Tris pH 8.0, 25 mM NaCl using an Amicon Ultra Centrifugal Filters (10 kDa MWCO).

Mass Spectroscopy

Mass spectroscopy was performed by Bioinformatics Solutions Inc. in Waterloo, ON, Canada. LC-MS analysis of DTT reduced samples were performed on a Thermo Scientific Orbitrap Exploris 240 mass spectrometer, equipped with a heated electrospray ionization source (H-ESI) in positive ion mode with a Thermo Fisher Ultimate 3000 RSLCnano HPLC System. On the H-ESI source, sheath gas was set to 2 arbitrary units (arb), and auxiliary gas was set to 6 arb. The ion transfer tube was set at 275°C. The vaporizer temp was at 200°C. The sample was analyzed on a MAbPac RP, 4 μM, 3.0 × 50 mm analytical column (ThermoFisher, San Jose, CA, United States), held at 60°C. The protein was eluted at a rate of 500 μL/min for a 10-min gradient, where 0–7 min: 10–70% acetonitrile +0.1% formic acid; 7–8.2 min: 95% acetonitrile +0.1% formic acid, 8.2–10 min: 20% acetonitrile +0.1% formic acid. MS spectra were acquired using full scans at 15,000 resolution in the orbitrap within a range of 700–2,200 m/z. The maximum injection time was set at auto with a standard AGC target. Ten micro scans were employed, and the RF lens was set to 60%. 15 V of insource CID was applied. Thermo BioPharma Finder 4.1 was used for intact mass deconvolution and peak identification.

Chloramphenicol Acetyltransferase Assay

Electrocompetent DH10B cells were transformed with a pMW plasmid encoding the PylRS variant, and a pCAM plasmid encoding chloramphenicol acetyltransferase 112TAG and *Mat*RNA^{Pyl}. Colonies of freshly transformed DH10B cells harboring the indicated plasmids were picked and grown at 37°C overnight in 5 ml of LB media supplemented with appropriate antibiotics. The overnight cultures were serially diluted, and 2 μL of each dilution were spotted onto LB-agar plates supplemented with the appropriate antibiotics for plasmid maintenance, 100 μM IPTG, 1 mM of the ncAA where indicated, and either 100 (PrK plates) or 200 (BocK and ALocK) μg/mL chloramphenicol. The spotted plates were grown at 37°C overnight. Images were taken after 16 h of incubation.

RESULTS

Developing a PANCE Protocol to Evolve *Ma*PylRS

To evolve a hyperactive variant of *Ma*PylRS that also discriminates against canonical amino acids, we adapted a variation of the previously described PANCE system (**Figure 1A**) (Suzuki et al., 2017). For each generation of phage evolution, we performed a three-step selection process: positive selection with mutagenesis, followed by negative selection, and finally positive selection without mutagenesis. In the first step, *E. coli* S1030 cells are transformed with the MP4 mutagenesis plasmid and an accessory plasmid (AP), pJT017 (Badran and Liu, 2015; Bryson et al., 2017). The AP encodes the essential phage protein pIII with one to three TAG codons, as well as *Mat*RNA$^{Pyl}_{(6)}$, an engineered tRNA^{Pyl} variant that is orthogonal to the *M. mazei* PylRS/tRNA^{Pyl} system (**Supplementary Figure S1**) (Willis and Chin, 2018). We utilized *Mat*RNA$^{Pyl}_{(6)}$ in our PANCE system to determine if the tRNA binding domain of *Ma*PylRS would evolve a greater affinity for *Mat*RNA$^{Pyl}_{(6)}$. The transformed *E. coli* S1030 cells are grown to mid-log phase, supplemented with BocK (**Figure 1B**) and arabinose, and

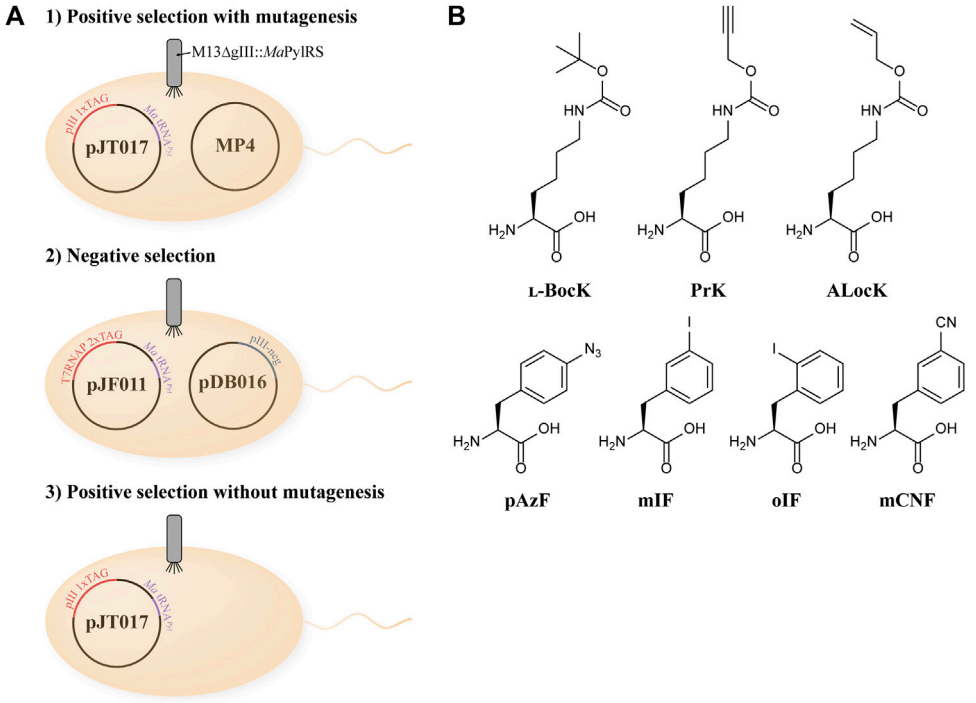


FIGURE 1 | (A) Overview of the PANCE experimental design. **(B)** Structures of ncAAs used in this study.

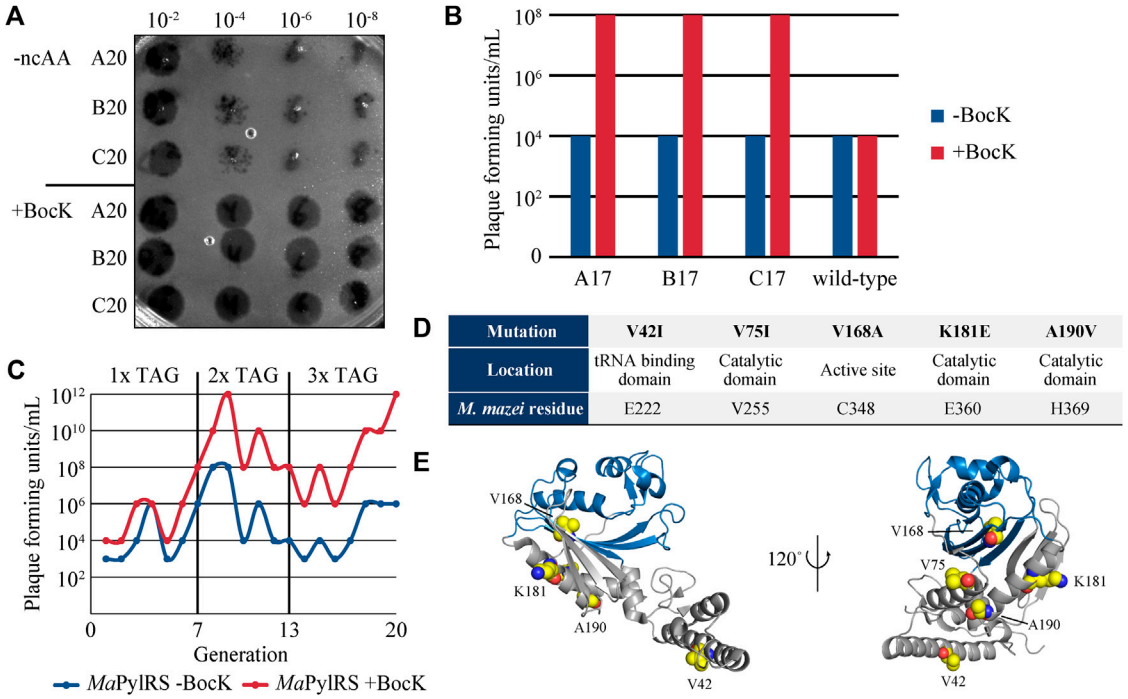
infected by the phage M13 ΔgIII::*Ma*PylRS. Phage propagation is dependent on successful aminoacylation of $tRNA^{Pyl}$ by *Ma*PylRS and subsequent suppression of the TAG codon(s) in gIII (the gene encoding pIII), thus enabling selection for phages carrying the most active PylRS variants.

To avoid the unwanted evolution of a promiscuous synthetase that is active towards canonical amino acids, after each round of positive selection with mutagenesis, we performed negative selection in the absence of BocK. Originally developed as an additional selection mechanism in the continuous directed evolution technique PACE, negative selection is highly effective in maintaining the amino acid selectivity of evolving aminoacyl-tRNA synthetases (Carlson et al., 2014; Bryson et al., 2017). The negative selection system relies on the production of pIII-neg, a dominant-negative version of pIII. Production of pIII-neg inhibits phage propagation, thus depleting the population of phages carrying PylRS mutants that are active towards canonical amino acids and can produce pIII-neg in the absence of BocK. In our pilot experiments to optimize the PANCE system for evolving *Ma*PylRS, we found that without negative selection, significant phage propagation occurs in the absence of BocK after consecutive generations of positive selection with mutagenesis (**Supplementary Figure S2**). This indicates that *Ma*PylRS is sensitive to mutations that enable activity towards a canonical amino acid, most likely Phe. This is predictable, as *M. mazei* PylRS has been shown to readily incorporate Phe when the critical active site residues N346 (N166 in *M. alvus*) and C348 (V168) are mutated (Wang et al., 2011). We therefore adopted the negative selection step to maintain the selectivity of evolving *Ma*PylRS variants.

As a final selection step, the surviving phage population isolated from the prior negative selection was passed through a round of positive selection without mutagenesis. This final step amplifies phages that survived negative selection and are still active towards BocK without the added stringency of mutagenesis. We performed this selection in both the presence and absence of BocK, and used the phages isolated from these two conditions as a checkpoint to verify that evolution was proceeding successfully. We measured the phage titers of both conditions to ensure that phage propagation remained robust in the presence of BocK, and low when BocK was not added to the media (**Figure 2A**). The significant difference between the phage titers indicates that promiscuous variants are not persisting through the negative selection step, and variants that are active towards BocK are indeed being selected. The phage population recovered from this final round of positive selection in the presence of BocK was carried over to the next generation of selection, and the three-step cycle was repeated.

Directed Evolution of *Ma*PylRS

We performed a total of 20 generations of directed evolution and maintained three independent lineages of evolved phages. The stringency of selection was increased as evolution proceeded. Initially, gIII carried a single in-frame TAG codon to be suppressed by *Ma*PylRS/$tRNA^{Pyl}_{(6)}$. After 7 generations of selection, we increased the stringency to two TAG codons in



Mutation	V42I	V75I	V168A	K181E	A190V
Location	tRNA binding domain	Catalytic domain	Active site	Catalytic domain	Catalytic domain
M. mazei residue	E222	V255	C348	E360	H369

FIGURE 2 | (A) Phage titers of the 20th generation of evolved M13 *Ma*PylRS. The numbers above the image indicate the dilution factor. **(B)** Measurements of phage titers comparing the 17th generation of evolved phages and the ancestral generation zero encoding wild-type *Ma*PylRS. Cells were infected and grown in the presence and absence of BocK. **(C)** Summary of phage titers over 20 generations, in the presence and absence of BocK. **(D)** Mutations to *Ma*PylRS evolved through PANCE, as well as their location in the protein and the corresponding *M. mazei* residue as determined by multiple sequence alignment. **(E)** Crystal structure of *Ma*PylRS. Residues that were mutated during PANCE are highlighted as spheres (PDB ID: 6JP2).

gIII. The evolved phage population efficiently propagated in this higher stringency selection when cultures were supplemented with BocK, while maintaining low levels of propagation in the absence of BocK. After 6 rounds of selection with two TAG codons, we again increased the stringency to three TAG codons. Even at this highest level of stringency, phage propagation remained robust and highly selective for BocK. At this stage, we compared the propagation of our evolved phage lineages to that of the ancestral variant (generation zero) that encodes wild-type *Ma*PylRS. We measured propagation under the high-stringency three TAG codon system in the absence of mutagenesis. The phage titer indicated that the evolved phage lineages propagated far more efficiently than generation zero in the presence of BocK, suggesting that more active variants of PylRS had emerged (**Figure 2B**).

Throughout the course of the directed evolution process, phage titers at the end of each generation consistently showed greater phage propagation when BocK was present in the media (**Figure 2C**). After the 20th round of evolution, several plaques from each evolved lineage were sequenced to determine the identity of PylRS mutations. Five mutations were identified, with very little crossover between each of the three independent lineages (**Figure 2D**). The most frequently occurring mutations that arose from our system were V42I, V75I, V168A, K181E, and A190V. V42I and K181E were found as concomitant mutations, while the rest of the variants were point mutations. V168A is the only active site mutation that was observed. This mutation has been identified previously and exploited for its role in substrate specificity (Wang et al., 2011; Wang et al., 2012; Wan et al., 2014; Seki et al., 2020). As such, we hypothesized that this mutation to the smaller alanine residue may confer altered substrate specificity favoring larger ncAAs such as BocK. V75I, K181E, and A190V are located in the catalytic domain just outside of the active site, while V42I is found in the tRNA binding domain (**Figure 2E**). Interestingly, multiple sequence alignment shows that the K181E mutation matches the sequence of several PylRS homologs that also encode Glu at this position (E360 in *M. mazei* PylRS) (**Supplementary Figure S3**) (Yanagisawa et al., 2008). To the best of our knowledge, aside from V168, none of these residues have previously been identified or targeted for engineering in *Ma*PylRS or any other PylRS homologs.

*Ma*PylRS$_{evo}$ Variants Are Highly Active and Selective for ncAAs *in vivo*

To test the activity of our evolved *Ma*PylRS mutants, we cloned the mutant genes into a pMW expression vector and co-transformed them along with a pBAD reporter plasmid carrying sfGFP (2TAG) and tRNA^{Pyl} into *E. coli* DH10B. Read-through of the TAG codon at position 2 is dependent on PylRS aminoacylating its cognate tRNA^{Pyl} with a ncAA.

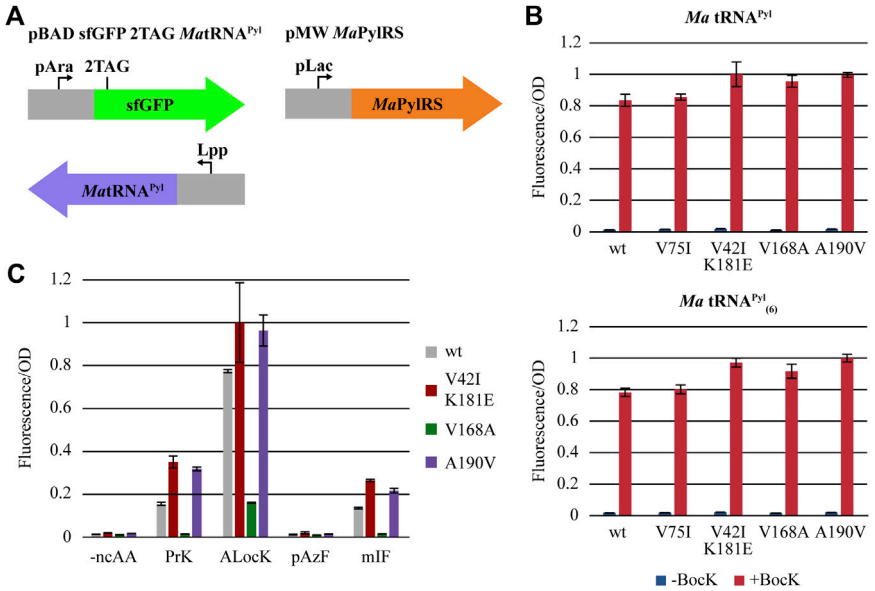


FIGURE 3 | (A) Overview of the *in vivo* fluorescence reporter system. **(B)** Initial fluorescence screening of PylRS mutants, using wild-type *Mat*RNA^{Pyl} (left) and *Mat*RNA$^{Pyl}_{(6)}$ (right). **(C)** Fluorescence measurements of the most active mutants and a selection of ncAAs. Error bars indicate the standard deviation for each condition (n = 6).

Aminoacyl-tRNA^{Pyl} is then utilized in translation to suppress TAG and incorporate the ncAA into the protein (**Figure 3A**). Thus, fluorescence from sfGFP production can be used to measure the activity of PylRS. We measured the fluorescence/OD_{600} of cells expressing our PylRS variants and compared this to those expressing wild-type *Ma*PylRS, both in the presence and absence of BocK (**Figure 3B**). The results show that each of the four mutant constructs as well as wild-type *Ma*PylRS are highly active and selective for BocK. The lack of background activity in the absence of a ncAA for all four mutants substantiates the importance of the negative selection step in our PANCE system. We tested the activity of the variants using both wild-type *Mat*RNA^{Pyl} and *Mat*RNA$^{Pyl}_{(6)}$. The data indicate that both *Mat*RNA^{Pyl} and *Mat*RNA$^{Pyl}_{(6)}$ are efficiently charged with BocK by all four mutant variants and wild-type *Ma*PylRS. The activity of the evolved variants towards *Mat*RNA$^{Pyl}_{(6)}$ is consistent with that of wild-type *Ma*PylRS, suggesting that the variants did not evolve enhanced recognition of *Mat*RNA$^{Pyl}_{(6)}$. However, compared to wild-type *Ma*PylRS, the activity towards BocK is improved for three of the four mutants, with the exception being V75I which is approximately equal to wild-type.

We hypothesized that in addition to having robust activity towards the directed evolution substrate BocK, our mutants may also have improved activity towards other Lys- and Phe-ncAAs. We screened four additional ncAAs: N^{ϵ}-propargyloxycarbonyl-L-lysine (PrK), N^{ϵ}-allyloxycarbonyl-L-lysine (ALocK), 4-azido-L-phenylalanine (pAzF), and 3-iodo-L-phenylalanine (mIF) (**Figure 3C**). The most active variants, V42I/K181E and A190V, have dramatically improved activity towards PrK, ALocK, and mIF. Interestingly, the activity of the active-site mutant V168A is significantly decreased to all the ncAAs we tested except BocK, highlighting the previously established role of this mutation in governing the size and selectivity of the PylRS active site (Wang et al., 2011; Wang et al., 2012; Wan et al., 2014; Seki et al., 2020). The activity of V75I is moderately higher towards PrK compared to the most active variants, and we omitted V75I from further screening. None of the mutants have significant activity towards the *para*-substituted phenylalanine derivative pAzF. As such, we hypothesized that like *Ma*PylRS, the evolved PylRS variants should also be orthogonal to the *M. jannaschii*-derived AzFRS.2. t1, enabling them to be used jointly to incorporate multiple, distinct ncAAs into a single protein (Amiram et al., 2015; Tharp et al., 2021).

Combining the Most Active PANCE Mutations to Create PylRS$_{opt}$

We suspected that combining the mutations that conferred the greatest increases in activity may result in an additive effect, yielding a hyperactive, ncAA-specific PylRS variant. Based on our results from the initial fluorescence experiments, we cloned a *Ma*PylRS variant containing the mutations V42I, K181E, and A190V, and named the construct PylRS$_{opt}$. We tested the activity of *Ma*PylRS, PylRS$_{V42I/K181E}$, PylRS$_{A190V}$, and PylRS$_{opt}$ towards several ncAAs, and found that the activity of PylRS$_{opt}$ was higher than both wild-type and the individual mutants for every ncAA we tested (**Figure 4A**). Notably, background fluorescence in the absence of ncAA remains close to basal levels for PylRS$_{opt}$. Like *Ma*PylRS, PylRS$_{opt}$ is highly selective for ncAAs over canonical amino acids. To support this, we co-transformed DH10B cells with sfGFP 2TAG and either *Ma*PylRS or PylRS$_{opt}$. The cells were supplemented with 1 mM BocK, and sfGFP was purified. The

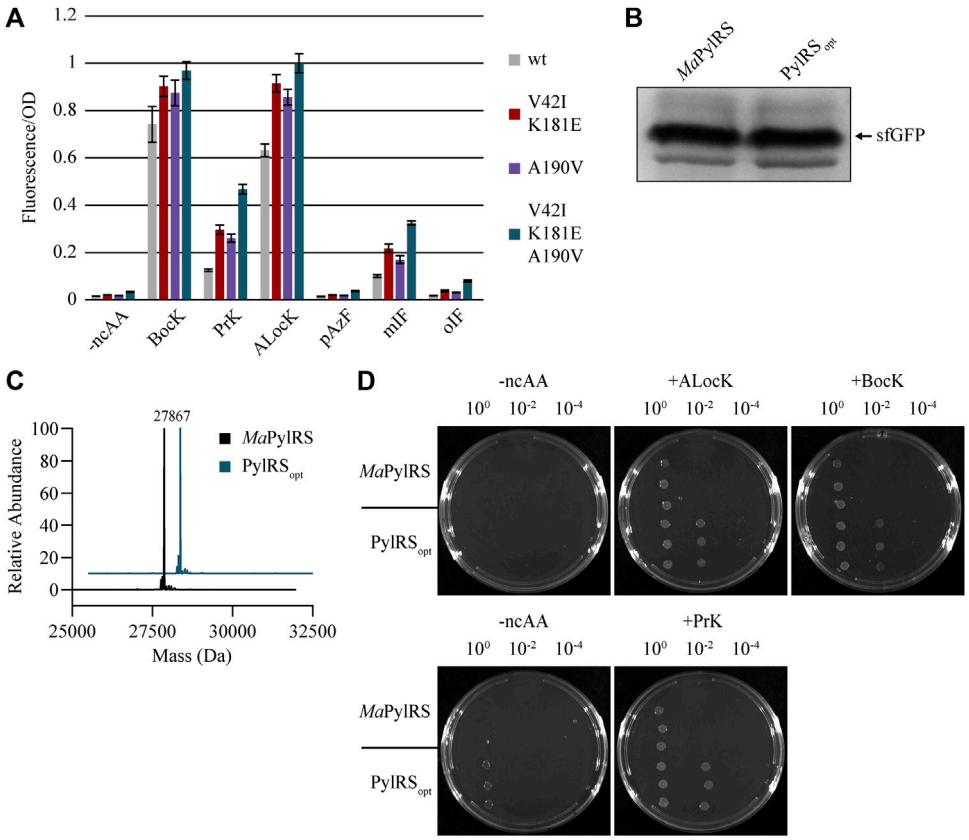


FIGURE 4 | (A) *In vivo* fluorescence assay measuring the activity of wild-type and evolved *Ma*PylRS variants and the combined mutant $PylRS_{opt}$. **(B)** SDS gel showing relative protein yields of sfGFP-2BocK from cells expressing either *Ma*PylRS or $PylRS_{opt}$. **(C)** LC-MS of the samples shown in **Figure 4B**. Peaks were detected for sfGFP-2BocK (theoretical mass = 27,869 Da) purified from cells expressing either *Ma*PylRS or $PylRS_{opt}$. **(D)** Chloramphenicol acetyltransferase assay. The top and bottom row of plates are supplemented with 200 μg/ml and 100 μg/ml chloramphenicol, respectively.

yield of sfGFP was robust at approximately 20 mg/L in cells expressing either *Ma*PylRS or $PylRS_{opt}$ (**Figure 4B**). We analyzed these samples by intact mass spectroscopy, which confirmed that BocK is solely present at position 2 of the protein (**Figure 4C**).

To gain further insight into the robustness of the activity of $PylRS_{opt}$, we titrated BocK, ALocK, and PrK and measured the fluorescence in cells expressing either *Ma*PylRS or $PylRS_{opt}$ (**Supplementary Figure S4**). We tested concentrations as low as 62.5 μM, which is slightly above the K_m that was reported for *M. mazei* PylRS and its natural substrate, pyrrolysine (Guo et al., 2014). We found that the activity of $PylRS_{opt}$ is significantly greater than *Ma*PylRS at all ncAA concentrations tested.

To validate our *in vivo* fluorescence data, we implemented a chloramphenicol acetyltransferase assay as a second reporter system. DH10B cells expressing either *Ma*PylRS or $PylRS_{opt}$, $MatRNA^{Pyl}$, and the chloramphenicol acetyltransferase gene with a single in-frame TAG codon were challenged to grow on plates containing bacteriostatic levels of chloramphenicol in the presence and absence of BocK, PrK, and ALocK (**Figure 4D**). The results corroborated our observations from the fluorescence data. Upon diluting saturated overnight cultures 1/100, cells expressing *Ma*PylRS are unable to grow on chloramphenicol plates regardless of the presence or absence of a ncAA. However, under these same conditions, $PylRS_{opt}$ supports strong colony growth in the presence of BocK, PrK, or ALocK. Neither enzyme supports growth on 200 μg/ml chloramphenicol plates in the absence of a ncAA. We note, however, that on the less stringent 100 μg/ml chloramphenicol plates, spotting undiluted cells expressing $PylRS_{opt}$ enables slow growth in the absence of a ncAA, whereas the same cells expressing *Ma*PylRS cells are unable to grow under these conditions. This phenomenon is not observed at 200 μg/ml chloramphenicol, where growth is only apparent for $PylRS_{opt}$ in the presence of 1 mM BocK or ALocK. These results confirm the previous observation that $PylRS_{opt}$ is more active than *Ma*PylRS towards BocK, ALocK, and PrK, with the caveat that under low stringency conditions, $PylRS_{opt}$ can apparently incorporate a canonical amino acid in the absence of a ncAA and facilitate modest colony growth.

Inserting $PylRS_{opt}$ Mutations Into *Ma*PylRS N166S

Next, we sought to determine whether the mutations found in $PylRS_{opt}$ can improve the activity of a previously engineered variant of *Ma*PylRS that readily incorporates *meta*- and *ortho*-substituted Phe (Tharp et al., 2021). The highly conserved active

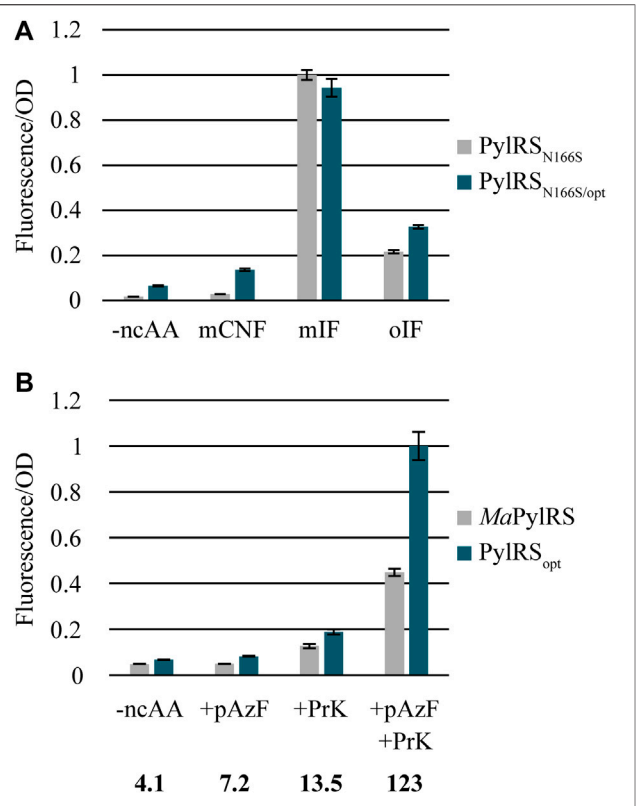


FIGURE 5 | (A) Combining PylRS$_{opt}$ mutations with N166S improves incorporation of mCNF, oIF. **(B)** Dual ncAA incorporation in a reporter protein using AzFRS.2. t1 and either *Ma*PylRS or PylRS$_{opt}$. The numbers below each condition indicate the percent increase of fluorescence/OD in cells expressing PylRS$_{opt}$ relative to those expressing *Ma*PylRS.

site residue N166 (N346 in *M. mazei* PylRS) plays a critical role as the "gatekeeper residue" of the PylRS active site, as the amide nitrogen is involved in the recognition of pyrrolysine and its analogs (Yanagisawa et al., 2008; Wang et al., 2011; Wang et al., 2012). Mutating N166 alters the substrate specificity of *Ma*PylRS, effectively eliminating recognition of lysine-derived ncAAs while greatly increasing its activity towards several phenylalanine derivatives. *Ma*PylRS$_{N166S}$ was recently shown to aminoacylate a variety of *meta*- and *ortho*-substituted phenylalanine derivatives (Tharp et al., 2021). Thus, we created PylRS$_{N166S/opt}$ and compared its activity to *Ma*PylRS$_{N166S}$ to assess whether combining the PylRS$_{opt}$ mutations with N166S can enhance its activity towards Phe derivatives without compromising its selectivity (**Figure 5A**). The results show that PylRS$_{N166S/opt}$ has increased activity towards mCNF and oIF, which are both difficult substrates for *Ma*PylRS$_{N166S}$. However, the PylRS$_{opt}$ mutations do not improve activity towards the well-recognized substrate mIF. As *Ma*PylRS$_{N166S}$ is already highly active towards mIF, the lack of improvement in activity towards this substrate is predictable. It should be noted that the background activity of PylRS$_{N166S/opt}$ is also elevated in the absence of a ncAA. It is plausible that the canonical amino acid may be sufficiently outcompeted when a recognized ncAA is present, as is the case with PylRS$_{opt}$ in the presence of BocK. Overall, the data indicate that the mutations in PylRS$_{opt}$ improve the activity of *Ma*PylRS$_{N166S}$ towards difficult Phe-ncAAs. However, elevated background activity is also apparent for this PylRS variant and may be a drawback for applications that require a homogeneous protein sample. Lower background levels may be attainable when combining the PylRS$_{opt}$ mutations with other active site mutants variants, such as those reported previously for incorporating bulky Lys-ncAAs (Seki et al., 2020).

PylRS$_{opt}$ Mutations Improve the Efficiency of Dual ncAA Incorporation

As a final assessment of the usefulness of PylRS$_{opt}$, we tested whether the enhanced activity and substrate specificity of PylRS$_{opt}$ could be used to improve the incorporation of two distinct ncAAs into a single protein. Translation efficiency drops dramatically when two or more distinct ncAAs are to be inserted into a protein, a problem that is exacerbated when aminoacylation of either of the ncAAs is particularly challenging (Hoesl and Budisa, 2011). Thus, improved activity is crucial to attaining substantial yields of proteins containing multiple ncAAs. To investigate the applicability of PylRS$_{opt}$ for improving dual incorporation of difficult ncAAs, we cloned a system to measure the production of sfGFP featuring two in-frame stop codons: 2TAG and 149TAA. We utilized the *M. jannaschii* TyrRS variant AzFRS.2. t1 and its cognate MjtRNA$^{Tyr}_{CUA}$ to incorporate pAzF, and either *Ma*PylRS or PylRS$_{opt}$ along with *Mat*RNA$^{Pyl}_{(6)/UUA}$ to incorporate PrK. Importantly, PylRS does not discriminate against the anticodon of tRNA^{Pyl} (Ambrogelly et al., 2007), and tRNA$^{Pyl}_{UUA}$ does not suppress UAG codons (Odoi et al., 2013), thus enabling PylRS/tRNA$^{Pyl}_{UUA}$ to function as an orthogonal pair with the *Mj*TyrRS/tRNA$^{Tyr}_{CUA}$ system (Wan et al., 2010). When cultures were supplemented with both pAzF and PrK, we observed over 120% increased fluorescence in cells expressing PylRS$_{opt}$ as opposed to *Ma*PylRS (**Figure 5B**). Background activity in the absence of one or both ncAAs was comparable for PylRS and PylRS$_{opt}$. Interestingly, low levels of expression were seen when PrK was present but pAzF was omitted, consistent with earlier observations of the PylRS/*Mj*TyrRS dual ncAA incorporation system (Wan et al., 2010; Chatterjee et al., 2013). As has been suggested previously, this artifact is likely due to low-level recognition of canonical amino acids by AzFRS.2. t1 in the absence of pAzF. Ultimately, utilizing PylRS$_{opt}$ to incorporate a second ncAA into a protein leads to over a two-fold improvement in the apparent protein yield compared to wild-type *Ma*PylRS. PylRS$_{opt}$ should therefore be quite useful in applications requiring site-specific incorporation of multiple reactive moieties in a single protein, such as bioorthogonal click chemistry (Neumann et al., 2010; Wan et al., 2010).

DISCUSSION

We have shown that the PANCE system can be adjusted to mitigate the emergence of promiscuous PylRS variants by integrating a negative selection step into the procedure. Alternating positive and negative selection while gradually increasing the stringency of positive selection enables the

emergence of highly selective, hyperactive variants of *Ma*PylRS. Screening the activity of *Ma*PylRS mutants that arose from PANCE led us to the identification of $PylRS_{opt}$, a novel PylRS construct. Despite having no modifications to its active site, $PylRS_{opt}$ profiles as a valuable genetic code expansion tool that is highly active towards several Lys- and Phe-ncAAs while maintaining excellent selectivity in discriminating against canonical amino acids. All but one of the mutations that arose from our directed evolution process were previously unidentified, and the mechanism by which these mutations enhance PylRS activity is unclear. Nevertheless, the identification of several residues outside of the active site underscores the power of impartial mutagenesis and selection, as it is unlikely that these residues would have otherwise been targeted for engineering.

Our findings of impactful mutations outside of the active site are consistent with previously reported directed evolution experiments on chPylRS (Bryson et al., 2017; Suzuki et al., 2017). In these studies, increased PylRS activity is proposed to stem from improved binding to the cognate tRNA. In *Ma*PylRS, K181 and A190 are located on an α-helix just outside of the active site (Seki et al., 2020). It is plausible that the mutations we observed may create conformational changes that alter the size, shape, or flexibility of the catalytic core. Conversely, V42 is located in the tRNA binding domain. This mutation is unlikely to impact ncAA binding but may instead facilitate a change in tRNA interaction. In the *D. hafniense* PylRS/$tRNA^{Pyl}$ complex, the corresponding aligned residue Q52 is located in the α-helix of tRNA binding domain 1 that serves as a binding surface for the core of $tRNA^{Pyl}$ (Nozawa et al., 2009). Thus, it is possible that mutation of V42 may alter the binding of *Ma*PylRS to $tRNA^{Pyl}$, although it is unclear how the conserved V42I mutation may impact the overall structure of the domain. It is also possible that V42I is an innocuous mutation and K181E is instead the sole driving force behind the activity of this PylRS variant, as we did not test the activity of each of the two mutants independently. Accordingly, the opposite may hold true as well.

Our directed evolution system led to the emergence of several PylRS variants that are active towards a variety of ncAAs. The lone exception to this observation is the active site mutant V168A, which is capable of efficiently incorporating BocK but discriminates against all the other Lys- and Phe-derivatives we tested. V168 (C348 in the *M. mazei* sequence) is associated with the size of the active site binding pocket and is positioned near the N^{ε}-substituent of pyrrolysine and its analogs (Kavran et al., 2007; Seki et al., 2020). Thus, it stands to reason that substituting the smaller alanine at this position enlarges the binding pocket. This may then allow the bulky BocK to maintain contacts with the active site while simultaneously hindering the recognition of smaller canonical and noncanonical amino acids. Conversely, mutating V168 to a larger residue may improve the recognition of smaller ncAAs as well as canonical phenylalanine, as this was observed in a mutational analysis of *M. mazei* PylRS (Wang et al., 2011).

Combining the most active PANCE mutations to generate $PylRS_{opt}$ resulted in an additive effect: the activity of the combined mutant is demonstrably higher than its predecessors. The enhanced activity of $PylRS_{opt}$ is particularly notable for ncAAs that *Ma*PylRS is relatively weakly active towards, such as PrK. While the greatest activity increases are observed when using weaker substrates, $PylRS_{opt}$ is also more active towards well-recognized ncAAs such as BocK and ALocK. Further, background activity in the absence of a ncAA is not significantly altered by the mutations in $PylRS_{opt}$. We suspect that the activity increase seen in $PylRS_{opt}$ is generalizable and not substrate-specific, as we observed enhanced activity for $PylRS_{opt}$ towards nearly every substrate that the wild-type enzyme is also active towards. Utilizing $PylRS_{opt}$ in future studies to incorporate additional ncAAs should validate this hypothesis. Because the $PylRS_{opt}$ mutations are located outside of the active site, these mutations should theoretically be generalizable like those found in the evolved chPylRS variant (Bryson et al., 2017). Although the results with *Ma*$PylRS_{N166S/opt}$ were somewhat mixed, future studies may reveal other engineered PylRS variants that are improved by the $PylRS_{opt}$ mutations. Finally, we have shown that $PylRS_{opt}$ is orthogonal with the *Mj*TyrRS/$tRNA^{Tyr}$ system and enables a sharp increase in the production of a reporter protein with two distinct ncAAs. It is likely that when used with $tRNA^{Pyl}_{(6)}$, $PylRS_{opt}$ is orthogonal with the *M. mazei* PylRS/$tRNA^{Pyl}$ system, as is the case with *Ma*PylRS/$tRNA^{Pyl}_{(6)}$ (Willis and Chin, 2018). Thus, incorporation of three distinct ncAAs should also be improved by $PylRS_{opt}$ when used in combination with the other two orthogonal systems as previously described (Tharp et al., 2021).

In summary, directed evolution facilitated the development of a hyperactive and highly selective genetic code expansion tool $PylRS_{opt}$, which should be useful for a wide variety of applications moving forward.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JF designed the study, performed the experiments, and drafted the manuscript. DS and JT provided feedback and intellectual support throughout the study and edited the manuscript.

FUNDING

JT was supported by a K99 Pathway to Independence Award from the National Institute of General Medical Sciences (No. K99GM141320). This work was supported by grants from the US

National Institute of General Medical Sciences (No. R35GM122560 to DS), and the US Department of Energy (No. DE-FG02-98ER20311 to DS).

ACKNOWLEDGMENTS

We thank C. Chung, N. Krahn, and O. Vargas-Rodriguez for critical feedback on the project, J. Ho, C. Miller, and E. Chory for insightful conversations, and K. Hoffman for mass spectroscopy data collection and analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.850613/full#supplementary-material

REFERENCES

Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C., and Söll, D. (2007). Pyrrolysine Is Not Hardwired for Cotranslational Insertion at UAG Codons. *Proc. Natl. Acad. Sci.* 104, 3141–3146. doi:10.1073/pnas.0611634104

Amiram, M., Haimovich, A. D., Fan, C., Wang, Y.-S., Aerni, H.-R., Ntai, I., et al. (2015). Evolution of Translation Machinery in Recoded Bacteria Enables Multi-Site Incorporation of Nonstandard Amino Acids. *Nat. Biotechnol.* 33, 1272–1279. doi:10.1038/nbt.3372

Badran, A. H., and Liu, D. R. (2015). Development of Potent *In Vivo* Mutagenesis Plasmids with Broad Mutational Spectra. *Nat. Commun.* 6, 8425. doi:10.1038/ncomms9425

Baumann, T., Hauf, M., Richter, F., Albers, S., Möglich, A., Ignatova, Z., et al. (2019). Computational Aminoacyl-tRNA Synthetase Library Design for Photocaged Tyrosine. *Int. J. Mol. Sci.* 20, 2343. doi:10.3390/ijms20092343

Beránek, V., Willis, J. C. W., and Chin, J. W. (2019). An Evolved *Methanomethylophilus Alvus* Pyrrolysyl-tRNA Synthetase/tRNA Pair Is Highly Active and Orthogonal in Mammalian Cells. *Biochemistry* 58, 387–390. doi:10.1021/acs.biochem.8b00808

Bryson, D. I., Fan, C., Guo, L.-T., Miller, C., Söll, D., and Liu, D. R. (2017). Continuous Directed Evolution of Aminoacyl-tRNA Synthetases. *Nat. Chem. Biol.* 13, 1253–1260. doi:10.1038/nchembio.2474

Budisa, N. (2005). *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins*. Weinheim: Wiley-VCH Verlag GmbH & Co KGaA. doi:10.1002/3527607188

Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., and Liu, D. R. (2014). Negative Selection and Stringency Modulation in Phage-Assisted Continuous Evolution. *Nat. Chem. Biol.* 10, 216–222. doi:10.1038/nchembio.1453

Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013). A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia coli*. *Biochemistry* 52, 1828–1837. doi:10.1021/bi4000244

Crnković, A., Suzuki, T., Söll, D., and Reynolds, N. M. (2016). Pyrrolysyl-tRNA Synthetase, an Aminoacyl-tRNA Synthetase for Genetic Code Expansion. *Croat. Chem. Acta* 89, 163–174. doi:10.5562/cca2825

Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011). A System for the Continuous Directed Evolution of Biomolecules. *Nature* 472, 499–503. doi:10.1038/nature09929

Guo, L.-T., Wang, Y.-S., Nakamura, A., Eiler, D., Kavran, J. M., Wong, M., et al. (2014). Polyspecific Pyrrolysyl-tRNA Synthetases from Directed Evolution. *Proc. Natl. Acad. Sci. USA* 111, 16724–16729. doi:10.1073/pnas.1419737111

Hancock, S. M., Uprety, R., Deiters, A., and Chin, J. W. (2010). Expanding the Genetic Code of Yeast for Incorporation of Diverse Unnatural Amino Acids via a Pyrrolysyl-tRNA Synthetase/tRNA Pair. *J. Am. Chem. Soc.* 132, 14819–14824. doi:10.1021/ja104609m

Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A., and Chan, M. K. (2002). A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase. *Science* 296, 1462–1466. doi:10.1126/science.1069556

Hayes, H. C., Luk, L. Y. P., and Tsai, Y.-H. (2021). Approaches for Peptide and Protein Cyclisation. *Org. Biomol. Chem.* 19, 3983–4001. doi:10.1039/d1ob00411e

Hoesl, M. G., and Budisa, N. (2011). *In Vivo* incorporation of Multiple Noncanonical Amino Acids into Proteins. *Angew. Chem. Int. Ed.* 50, 2896–2902. doi:10.1002/anie.201005680

Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D., and Steitz, T. A. (2007). Structure of Pyrrolysyl-tRNA Synthetase, an Archaeal Enzyme for Genetic Code Innovation. *Proc. Natl. Acad. Sci.* 104, 11268–11273. doi:10.1073/pnas.0704769104

Ko, J.-h., Wang, Y.-S., Nakamura, A., Guo, L.-T., Söll, D., and Umehara, T. (2013). Pyrrolysyl-tRNA Synthetase Variants Reveal Ancestral Aminoacylation Function. *FEBS Lett.* 587, 3243–3248. doi:10.1016/j.febslet.2013.08.018

Krahn, N., Tharp, J. M., Crnković, A., and Söll, D. (2020). Engineering Aminoacyl-tRNA Synthetases for Use in Synthetic Biology. *Enzymes* 48, 351–395. doi:10.1016/bs.enz.2020.06.004

Longstaff, D. G., Blight, S. K., Zhang, L., Green-Church, K. B., and Krzycki, J. A. (2007). *In Vivo* contextual Requirements for UAG Translation as Pyrrolysine. *Mol. Microbiol.* 63, 229–241. doi:10.1111/j.1365-2958.2006.05500.x

Meineke, B., Heimgärtner, J., Lafranchi, L., and Elsässer, S. J. (2018). *Methanomethylophilus Alvus* Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. *ACS Chem. Biol.* 13, 3087–3096. doi:10.1021/acschembio.8b00571

Meineke, B., Heimgärtner, J., Eirich, J., Landreh, M., and Elsässer, S. J. (2020). Site-Specific Incorporation of Two ncAAs for Two-Color Bioorthogonal Labeling and Crosslinking of Proteins on Live Mammalian Cells. *Cel Rep.* 31, 107811. doi:10.1016/j.celrep.2020.107811

Miller, S. M., Wang, T., and Liu, D. R. (2020). Phage-assisted Continuous and Non-continuous Evolution. *Nat. Protoc.* 15, 4101–4127. doi:10.1038/s41596-020-00410-3

Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008). Genetically Encoding Nε-Acetyllysine in Recombinant Proteins. *Nat. Chem. Biol.* 4, 232–234. doi:10.1038/nchembio.73

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010). Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome. *Nature* 464, 441–444. doi:10.1038/nature08817

Nozawa, K., O'Donoghue, P., Gundllapalli, S., Araiso, Y., Ishitani, R., Umehara, T., et al. (2009). Pyrrolysyl-tRNA synthetase-tRNA^{Pyl} Structure Reveals the Molecular Basis of Orthogonality. *Nature* 457, 1163–1167. doi:10.1038/nature07611

Odoi, K. A., Huang, Y., Rezenom, Y. H., and Liu, W. R. (2013). Nonsense and Sense Suppression Abilities of Original and Derivative Methanosarcina Mazei Pyrrolysyl-tRNA Synthetase-tRNA^{Pyl} Pairs in the *Escherichia coli* BL21(DE3) Cell Strain. *PLoS One* 8, e57035. doi:10.1371/journal.pone.0057035

Polycarpo, C., Ambrogelly, A., Bérubé, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., et al. (2004). An Aminoacyl-tRNA Synthetase that Specifically Activates Pyrrolysine. *Proc. Natl. Acad. Sci.* 101, 12450–12454. doi:10.1073/pnas.0405362101

Polycarpo, C. R., Herring, S., Bérubé, A., Wood, J. L., Söll, D., and Ambrogelly, A. (2006). Pyrrolysine Analogues as Substrates for Pyrrolysyl-tRNA Synthetase. *FEBS Lett.* 580, 6695–6700. doi:10.1016/j.febslet.2006.11.028

Seki, E., Yanagisawa, T., Kuratani, M., Sakamoto, K., and Yokoyama, S. (2020). Fully Productive Cell-free Genetic Code Expansion by Structure-Based Engineering of *Methanomethylophilus Alvus* Pyrrolysyl-tRNA Synthetase. *ACS Synth. Biol.* 9, 718–732. doi:10.1021/acssynbio.9b00288

Srinivasan, G., James, C. M., and Krzycki, J. A. (2002). Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science* 296, 1459–1462. doi:10.1126/science.1069588

Suzuki, T., Miller, C., Guo, L.-T., Ho, J. M. L., Bryson, D. I., Wang, Y.-S., et al. (2017). Crystal Structures Reveal an Elusive Functional Domain of Pyrrolysyl-tRNA Synthetase. *Nat. Chem. Biol.* 13, 1261–1266. doi:10.1038/nchembio.2497

Tharp, J. M., Ehnbom, A., and Liu, W. R. (2018). $tRNA^{Pyl}$: Structure, Function, and Applications. *RNA Biol.* 15, 441–452. doi:10.1080/15476286.2017.1356561

Tharp, J. M., Vargas-Rodriguez, O., Schepartz, A., and Söll, D. (2021). Genetic Encoding of Three Distinct Noncanonical Amino Acids Using Reprogrammed Initiator and Nonsense Codons. *ACS Chem. Biol.* 16, 766–774. doi:10.1021/acschembio.1c00120

Théobald-Dietrich, A., Frugier, M., Giegé, R., and Rudinger-Thirion, J. (2004). Atypical Archaeal tRNA Pyrrolysine Transcript Behaves towards EF-Tu as a Typical Elongator tRNA. *Nucleic Acids Res.* 32, 1091–1096. doi:10.1093/nar/gkh266

Umehara, T., Kim, J., Lee, S., Guo, L.-T., Söll, D., and Park, H.-S. (2012). N-Acetyl Lysyl-tRNA Synthetases Evolved by a CcdB-Based Selection possessN-Acetyl Lysine Specificity *In Vitro* and *In Vivo*. *FEBS Lett.* 586, 729–733. doi:10.1016/j.febslet.2012.01.029

Vargas-Rodriguez, O., Sevostyanova, A., Söll, D., and Crnković, A. (2018). Upgrading Aminoacyl-tRNA Synthetases for Genetic Code Expansion. *Curr. Opin. Chem. Biol.* 46, 115–122. doi:10.1016/j.cbpa.2018.07.014

Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P.-J., Russell, D. H., et al. (2010). A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in *Escherichia Coli*. *Angew. Chem. Int. Ed.* 49, 3211–3214. doi:10.1002/anie.201000465

Wan, W., Tharp, J. M., and Liu, W. R. (2014). Pyrrolysyl-tRNA Synthetase: an Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1844, 1059–1070. doi:10.1016/j.bbapap.2014.03.002

Wang, Y.-S., Russell, W. K., Wang, Z., Wan, W., Dodd, L. E., Pai, P.-J., et al. (2011). The *de novo* Engineering of Pyrrolysyl-tRNA Synthetase for Genetic Incorporation of L-Phenylalanine and its Derivatives. *Mol. Biosyst.* 7, 714–717. doi:10.1039/c0mb00217h

Wang, Y.-S., Fang, X., Wallace, A. L., Wu, B., and Liu, W. R. (2012). A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant with a Broad Substrate Spectrum. *J. Am. Chem. Soc.* 134, 2950–2953. doi:10.1021/ja211972x

Willis, J. C. W., and Chin, J. W. (2018). Mutually Orthogonal Pyrrolysyl-tRNA Synthetase/tRNA Pairs. *Nat. Chem* 10, 831–837. doi:10.1038/s41557-018-0052-5

Wu, B., Wang, Z., Huang, Y., and Liu, W. R. (2012). Catalyst-free and Site-specific One-Pot Dual-Labeling of a Protein Directed by Two Genetically Incorporated Noncanonical Amino Acids. *Chembiochem* 13, 1405–1408. doi:10.1002/cbic.201200281

Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008). Crystallographic Studies on Multiple Conformational States of Active-Site Loops in Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* 378, 634–652. doi:10.1016/j.jmb.2008.02.045

Zhang, Y., Baranov, P. V., Atkins, J. F., and Gladyshev, V. N. (2005). Pyrrolysine and Selenocysteine Use Dissimilar Decoding Strategies. *J. Biol. Chem.* 280, 20740–20751. doi:10.1074/jbc.M501458200

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

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

frontiers | Frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fmolb.2022.832379

Cell Engineering and Cultivation of Chinese Hamster Ovary Cells for the Development of Orthogonal Eukaryotic Cell-free Translation Systems

Jeffrey L. Schloßhauer[1,2†], Niño Cavak[1†], Anne Zemella[1], Lena Thoring[1] and Stefan Kubick[1,2,3]*

[1]Branch Bioanalytics and Bioprocesses (IZI-BB), Fraunhofer Institute for Cell Therapy and Immunology (IZI), Potsdam, Germany, [2]Institute of Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany, [3]Faculty of Health Sciences, Joint Faculty of the Brandenburg University of Technology Cottbus –Senftenberg, The Brandenburg Medical School Theodor Fontane and the University of Potsdam, Potsdam, Germany

OPEN ACCESS

Edited by:
Nediljko Budisa,
University of Manitoba, Canada

Reviewed by:
Qiong Wang,
Johns Hopkins University, United States
Cleo Kontoravdi,
Imperial College London, United Kingdom
Allen Liu,
University of Michigan, United States
Lei Kai,
Jiangsu Normal University, China

***Correspondence:**
Stefan Kubick
Stefan.Kubick@izi-bb.fraunhofer.de

[†]These authors have contributed equally to this work

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

Received: 09 December 2021
Accepted: 15 March 2022
Published: 14 April 2022

Citation:
Schloßhauer JL, Cavak N, Zemella A, Thoring L and Kubick S (2022) Cell Engineering and Cultivation of Chinese Hamster Ovary Cells for the Development of Orthogonal Eukaryotic Cell-free Translation Systems. Front. Mol. Biosci. 9:832379. doi: 10.3389/fmolb.2022.832379

The investigation of protein structures, functions and interactions often requires modifications to adapt protein properties to the specific application. Among many possible methods to equip proteins with new chemical groups, the utilization of orthogonal aminoacyl-tRNA synthetase/tRNA pairs enables the site-specific incorporation of non-canonical amino acids at defined positions in the protein. The open nature of cell-free protein synthesis reactions provides an optimal environment, as the orthogonal components do not need to be transported across the cell membrane and the impact on cell viability is negligible. In the present work, it was shown that the expression of orthogonal aminoacyl-tRNA synthetases in CHO cells prior to cell disruption enhanced the modification of the pharmaceutically relevant adenosine A2a receptor. For this purpose, in complement to transient transfection of CHO cells, an approach based on CRISPR/Cas9 technology was selected to generate a translationally active cell lysate harboring endogenous orthogonal aminoacyl-tRNA synthetase.

Keywords: orthogonal translation, cell-free protein synthesis, CRISPR, amber suppression, *E. coli* tyrosyl-tRNA synthetase, *M. mazei* pyrrolysyl-tRNA synthetase, membrane protein, C12orf35

INTRODUCTION

The production of proteins represents an important building block to elucidate the fundamental biochemical signatures of cells and forms the basis for diverse industrial applications. Therefore, further development of appropriate production systems is of particular interest. Cell-based systems are commonly exploited to address questions in basic research and to meet the demand for biotechnologically relevant proteins such as drug targets. Proteins can be equipped with special properties or new reactive groups for broader applications. Although conjugation of proteins *via* cysteine residues with sulfhydryl chemical groups or coupling of lysine residues with N-hydroxysuccinimide esters is widely used, the ligation of the reacting amino acids is not homogeneously distributed and resulted in a heterogeneous protein mixture (Shadish and DeForest, 2020). In addition, there are reports of intein-based protein ligation and conjugation of proteins by cysteine transpeptidases (Debelouchina and Muir, 2017; Vogl et al., 2021). However,

protein sequence motifs located at the N- or C- terminus must be incorporated into the protein of interest to ensure appropriate protein modification. In contrast, genetic code expansion strategies have the advantage that amino acid positions in proteins can be modified site-specifically (Chung et al., 2020). This requires an orthogonal aminoacyl-tRNA synthetase (aaRS)-tRNA pair, which does not cross-react with endogenous aaRS, tRNAs and amino acids, while maintaining high specificity towards the non-canonical amino acids (ncaa). Usually, a desired sense codon in the protein will be changed to an amber stop codon, recognized by the aminoacylated orthogonal tRNA or the release factor, resulting in either amber suppressed or truncated proteins (Beyer et al., 2020). In this process, ncaa with various chemical groups such as azides, alkynes and strained alkenes can be incorporated into the protein of interest, enabling specific reactions based on azide-alkyne cycloadditions, Staudinger ligation as well as tetrazine ligation (Kenry and Liu, 2019; Mushtaq et al., 2019). Moreover, various ncaa could be incorporated into proteins, such as photocrosslinking amino acids, NMR sensitive probes and amino acids, which were equipped with post-translational modifications like phosphorylation (Chin et al., 2002; Jones et al., 2010; Rogerson et al., 2015). Albeit orthogonal systems are utilized for protein modifications *in vivo*, ncaa must be supplemented to the culture medium at high concentrations as uptake into cells is limited, thus imposing a burden on the cells (Takimoto et al., 2010; Lin et al., 2017). Furthermore, overexpression of pharmaceutical proteins including membrane proteins can lead to decreased cell viability and growth, thereby limiting these systems for further applications (Gubellini et al., 2011). On the one hand, an alternative production system based on translationally active cell lysates, offers the possibility of producing proteins such as toxins and membrane proteins that can negatively affect cell growth and viability (Khambhati et al., 2019). Due to their open character, cell-free protein synthesis (CFPS) systems in general can be supplemented by various additives, making it feasible to realize protein modifications based on orthogonal translation systems (Lu, 2017). Previous work on *Sf*21 cell-free protein synthesis systems showed that the modified EGF receptor, which is misregulated in various tumors, was translocated into microsomal structures, originating from the endoplasmic reticulum (Quast et al., 2016; Thoring et al., 2017). In the process described by Quast et al., the ncaa p-azido-L-phenylalanine (AzF) was integrated at different amino acid positions based on the orthogonal mutant *E. coli* tyrosyl-tRNA synthetase (eAzFRS)-tRNAtyr pair. Kapoor et al. demonstrated conjugation of cell-free synthesized malaria surface antigen Pfs25 to a GPI linker derivate and following immunization of mice, leading to higher anti-Pfs25 antibody titers with Pfs25-GPI compared to Pfs25 (Kapoor et al., 2018).

The cell-free system can be further improved by directly overexpressing desired proteins in host cell-lines used for the generation of cell lysate. It has been shown that expression of T7 RNA polymerase in *E. coli* eliminates the need for supplementation of the subsequent cell-free reaction (Des Soye et al., 2019). In a previous study, orthogonal pyrrolysyl-tRNA synthetase (PylRS) was expressed prior to cell disruption of *E. coli* and followed by the incorporation of two unnatural lysine derivatives into EGFP in an orthogonal CFPS reaction (Chemla et al., 2015).

In contrast to frequently used *E. coli* systems, protein expression in eukaryotic systems is often more complex but necessary to provide proteins with post-translational modifications such as glycosylation, disulfide bond formation and lipidation (Gillette et al., 2019; Tripathi and Shrivastava, 2019). Approximately 70% of clinically approved proteins are produced in Chinese hamster ovary (CHO) cells due to their robustness, growth to high cell densities, high protein yields and the ability to synthesize complex mammalian proteins (Jayapal et al., 2007). Nowadays, protein expression is often based on stable transfected cells using sequence-specific recombinases and transposon-based modifications as well as RNA-guided nuclease based integration techniques, rather than random integration methods (Matasci et al., 2011; Zhang et al., 2015; Grav et al., 2017). Of particular importance is the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technology for genome editing in mammalian cells, which now makes it possible, to specifically knock out genes and insert expression cassettes at well-defined positions into the genome with high efficiency (Jinek et al., 2012; Ran et al., 2013). For protein overexpression, transcriptionally active loci in the genome that are less affected by gene silencing are preferentially targeted. To date, various loci including *HPRT*, *H11*, *COSMC*, *Rosa26* and *C12orf35* were addressed to insert desired transgenes into the CHO genome (Lee et al., 2015; Gaidukov et al., 2018; Kawabe et al., 2018; Zhao et al., 2018; Chi et al., 2019).

The objective of this study was to determine whether orthogonal aaRS can be expressed in CHO cells without diminishing the translational activity of the resulting cell lysate for CFPS. To this end, the orthogonal eAzFRS was transiently transfected into CHO cells and convenient cultivation formats were examined to attain a cell lysate optimal for modifying the G-protein-coupled adenosine A2a receptor in cell-free reactions. We also attempted to develop an orthogonal CFPS system based on stable transfection of eAzFRS into the CHO genome by targeting the *HPRT1* and *C12orf35* loci through CRISPR technology. A general workflow of the presented study is shown in **Figure 1**.

MATERIALS AND METHODS

Plasmids

The plasmids A2aRamb and A2aR for cell-free synthesis of adenosine A2a receptor (with and without an amber stop codon at amino acid position 215) were utilized as previously described (Zemella et al., 2019). A Nanoluciferase sequence was linked to the C-terminus of the receptor. The plasmid for cell-free synthesis of the modified *E.coli* tyrosyl-tRNA-synthetase (eAzFRS) in an *E.coli* based cell-free reaction was previously described (Zemella et al., 2019). The utilized PylRS-AF coding sequence contained the mutations Y306A and Y384F as

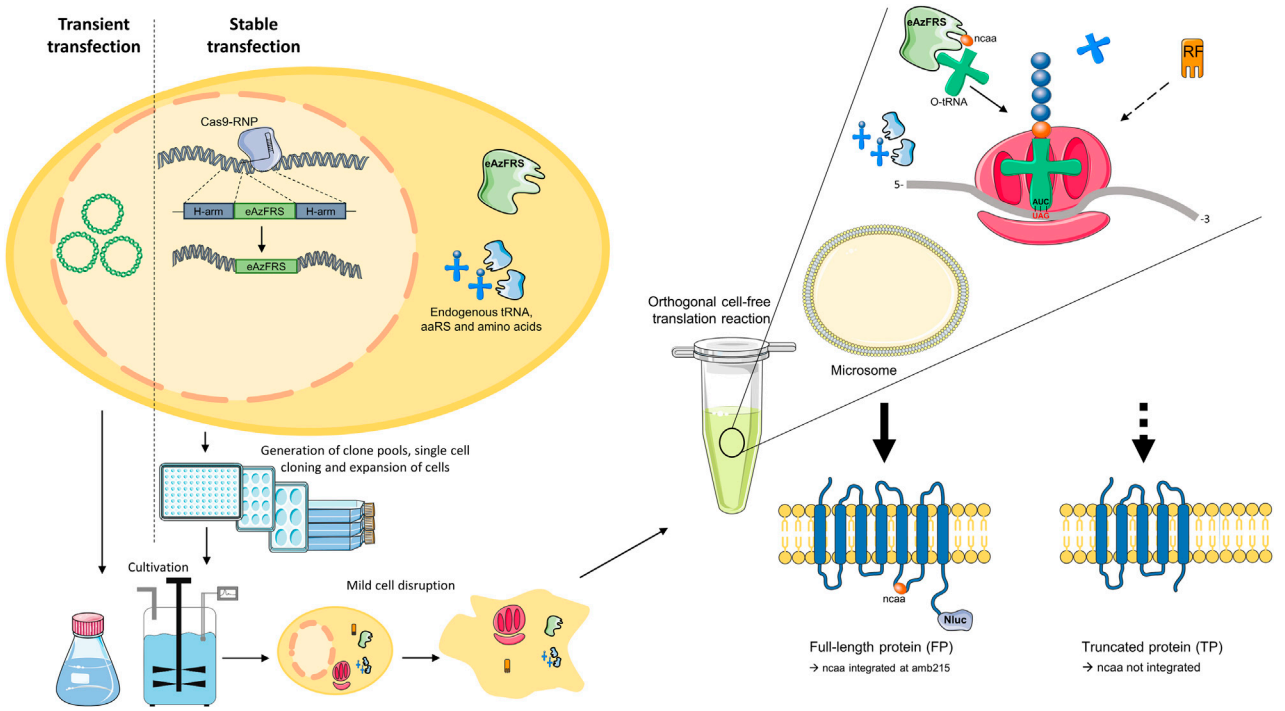


FIGURE 1 | General workflow. CHO cells were either transiently or stably transfected. While transiently transfected CHO cells were cultured in shake flasks and bioreactors for two days, stable transfected cells were subjected to selection pressure to obtain clone pools and clonal cell lines. The adenosine A2a receptor, which is C-terminally coupled to a Nanoluciferase (Nluc), was used to evaluate the developed orthogonal modification system. In the presence of orthogonal eAzFRS and orthogonal tRNA (o-tRNA), as well as the non-canonical amino acid (ncaa) p-azido-L-phenylalanine, the amber stop codon UAG at amino acid position 215 can be addressed to produce amber suppressed full-length protein (FP). Upon binding of the eukaryotic release factor (RF) termination product (TP) will be produced. Arrows indicate the direction of the workflow. The dashed arrow indicates translation termination. Parts of the figure were created by using pictures from Servier Medical Art (http://smart.servier.com/). Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.

previously reported (Yanagisawa et al., 2008). Transient transfection of CHO cells was performed with pcDNA3.1(+)-eAzFRS and pcDNA3.1(+)-PylRS-AF expression vectors. A Strep-Tag was fused to the C-terminus of eAzFRS, while a 6xHis-Tag was fused to the N-terminus of the PylRS-AF. For stable transfection of CHO cells the plasmid pSpCas9(BB)-2A-GFP (Plasmid #48138 from Addgene) was used to express Cas9 enzyme linked to GFP by a 2A self-cleavage peptide. The guide RNA sequences gRNA-HPRT1-T1: 5′-GTAGAATGATCAGTC AACAG-3′ and gRNA-C12orf35-T1: 5′-GCCCCTTACAGC TGTAGATA-3′ targeting the *C12orf35* and *HPRT1* locus in the CHO genome were extracted from the literature (Zhao et al., 2018). Novel gRNA sequences gRNA-HPRT1-T2: 5′-GGG GTTGTACCGCTTGACCA-3′ and gRNA-C12orf35-T2: 5′-GCCGGGACTTAACCACTCGA-3′ were designed with the CRISPOR gRNA design tool. Gibson assembly was used to clone the gRNA sequences into the gRNA_Cloning Vector (Plasmid #41824 from Addgene) according to the protocol previously described (Mali et al., 2013). The donor vector sequence for stable transfection was designed similar to a previously described study on CHO cells (Zhao et al., 2018). Therefore, a human cytomegalovirus promoter was used to drive gene expression of the GFP or eAzFRS gene and a bovine growth hormone (BGH)-polyA sequence for termination of transcription was used. The expression cassette for the resistance gene puromycin was under the control of the phosphoglycerate kinase-1 promoter and BGH-polyA utilized to terminate transcription. The expression cassette was flanked by locus (*HPRT1* or *C12orf35*)-specific homology arms. Additionally, gRNA sequences were located at the ends of the homology arms to create a linear donor in presence of Cas9 enzyme (**Supplementary Figure S1**).

Cell Culture

CHO-K1 cells were provided from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH (DSMZ no: ACC110) and adapted to grow in suspension. CHO-K1 cells were cultured in serum-free ProCHO5 medium (Lonza) at 37°C, 5% CO_2 and 100 rpm. Cells were seeded at densities of 1×10^6 cells/ml and grown up to 5×10^6 cells/ml in shake flasks, unless otherwise noted. Cell growth was monitored by counting cells using trypan blue staining and a Luna-FL cell counter (Logos Biosystems, Gyeonggi-do, Korea).

Transient Transfection

CHO cells were transiently transfected with eAzFRS and PylRS-AF plasmids by polyethylenimine (PEI) reagent. Therefore, cells were grown to 2×10^6 cells/ml and plasmid was added in a ratio of

1.5 μg/10^6 cells, followed by 2 μg PEI reagent/10^6 cells. Afterwards cells were incubated for four hours rotating at 37°C and 5% CO_2 and diluted to obtain a cell density of 1 × 10^6 cells/ml in 1,000 ml final volume in either 2 L shake flasks or 1 L bioreactor vessels (Biostat B-DCUII, Sartorius Stedim Biotech GmbH). CHO cells were cultivated for two days and harvested for cell lysate preparation as described previously (Thoring and Kubick, 2018).

Stable Transfection and Single-Cell Cloning

Plasmid DNA was mixed in a ratio of 2:2:1 (eAzFRS donor vector: gRNA vector: Cas9 vector) to obtain 500 ng plasmid DNA in 5 μL PBS. The DNA was further diluted by addition of 100 μL Opti-MEM serum-free medium (Gibco). Lipid based transfection was conducted by combine the diluted plamsid DNA with 2.5 μL Lipofectamine LTX (Thermo Fisher Scientific) and 0.5 μL Plus reagent. After incubation for 30 min at room temperature the transfection mixture was added to 1 × 10^6 cells/ml in 500 μL serum-free ProCHO5 medium (Lonza) in a 24-well plate. Cells were mixed gently and centrifuged for 15 min at room temperature and 400xg to increase transfection efficiency similar to a previous report (Barbu and Welsh, 2007). Following 48 h at 37°C, 5% CO_2 without shaking, cells were selected by puromycin resistance. Therefore, a concentration of 10 μg/ml puromycin in culture medium was achieved for two weeks. Positively transfected cells were pre-selected by array dilution, while maintaining the puromycin selection pressure. Clone pools were isolated and analyzed by qPCR and submitted to single-cell cloning by limiting dilution.

Fluorescent Microscopy

The GFP fluorescence of stable transfected CHO cells was visualized by a Nikon Eclipse TS2 inverted microscope combined with the NIS-Elements imaging software. A combination of a fluorescent LED and a GFP-B filter cube was used, whereby the filter cube is composed of a 470/40 nm excitation filter, a 500 nm dichroic filter and a 535 nm barrier filter. Images were recorded with a Nikon DS-Fi3 camera system.

Genotyping PCR

Genomic DNA was extracted from 10^6 CHO cells by the Quick-DNA Miniprep Plus Kit (Zymo Research) according to the manufacturer's instruction. Following genotyping primer sequences flanking the integration side of the expression cassette in the *C12orf35* locus were designed: GT-C12-T2-FW 5′-TGCATGCACCACAGAGTCAT-3′ and GT-C12-T2-RV 5′-ACAGGGCGCTTTGATGGTAA-3′. Genotyping PCR was performed by combining 0.5 μM of each primer, 5 ng/μL genomic DNA, 1x HotStar HiFidelity PCR Buffer (Qiagen), 0.05 U/μl HotStar HiFidelity DNA Polymerase (Qiagen) and distilled water in a 20 μL reaction. The following temperature profile was used: 95°C for 5 min; 40 cycles of 94°C for 15 s, 62°C for 60 s, 68°C for 5 min; 72°C for 10 min. PCR-products were run on a 1% agarose gel and product size was analyzed by comparing them to the Quick-Load 2-Log DNA Ladder (0.1–10.0 kbp, New England Biolabs).

Quantitative PCR

Quantitative real-time PCR (qPCR) was used to determine the eAzFRS mRNA levels. Therefore, mRNA was extracted with the High Pure RNA Isolation Kit (Roche) according to the manufacturer's instruction. Afterwards cDNA was prepared by using the Transcriptor First Strand cDNA Synthesis Kit (Roche). qPCR-primer sequences qRT-Gnb1-FW 5′-CCATATGTTTCT TTCCCAATGGC-3′ and qRT-Gnb1-RV 5′-AAGTCGTCGTACCCAGCAAG-3′of the housekeeping gene G protein subunit beta 1 (Gnb1) were extracted from the literature (Brown et al., 2018). The primer sequences qRT-AzFRS-FW 5′-GGATAAGAACAGCGGCAAGG-3′ and qRT-AzFRS-RV 5′-TCCATCTCCACCATAGGCAC-3′ were used for eAzFRS amplification. The qPCR reaction mix consisted of 0.5 μM qPCR primers, 2x FastStart Essential DNA Green master-mix (Roche), 10 ng cDNA and water in a 20 μL reaction. qPCR reactions were analyzed by the LightCycler 96 System (Roche). The temperature profile of the qPCR was 95°C for 5 min, followed by 38 cycles at 95°C for 20 s, 63°C for 20 s and 72°C for 20 s. A melt curve analysis was performed to confirm the specificity of amplification. Therefore, samples were heated for 5 min at 95°C and cooled down to 60°C. Afterwards the samples were heated continuously at 0.1°C/s. Relative quantification was calculated using Gnb1 as reference gene and the sample with the lowest gene expression as a control according to the Pfaffl method (Pfaffl, 2001). Measurements were conducted in triplicate and data were expressed as the mean with standard deviation.

Bioreactor Based Cultivation

Initial cell densities of 0.7 × 10^6 cells/ml, unless otherwise noted, were achieved in a 1 L bioreactor vessel connected to the control unit (Biostat B-DCUII, Sartorius Stedim Biotech GmbH). The cells were cultured while pH and oxygen supply were kept constant until a cell density of approx. 3–7 × 10^6 cells/ml was reached. During fermentation the parameters stirrer speed, pH, pO_2 and cultivation temperature were tracked over time and cell counts were measured regularly. The batch process was repeated up to three cycles, while maintaining a defined initial cell density in each cycle.

Cell-free Protein Synthesis

CHO cell lysate for CFPS was prepared to contain endogenous microsomes, taking advantage of the endoplasmic reticulum, as previously described in detail (Brödel et al., 2014; Thoring and Kubick, 2018). Eukaryotic cell-free reactions were composed of 40% CHO cell lysate, 10 μM PolyG, 30 mM HEPES-KOH (pH 7.5, Carl Roth GmbH), 100 mM sodium acetate (Merck), 3.9 mM magnesium acetate (Merck), 150 mM potassium acetate (Merck), 100 μM amino acids (Merck), 250 μM spermidin (Roche), 2.5 mM Dithiothreitol (Life technologies GmbH), 100 μg/ml creatine phosphokinase (Roche), 20 mM creatine phosphate (Roche), 1.75 mM ATP (Roche) and 0.3 mM GTP (Roche). To enable the qualitative analysis of radio-labeled proteins by electrophoresis 30 μM radioactive ^{14}C-leucine (specific radioactivity 46.15 dpm/pmol, Perkin Elmer) was added to the reaction. Furthermore, 1 U/μl T7 RNA polymerase (Agilent), 0.3 mM of UTP (Roche), 0.3 mM CTP (Roche) and 0.1 mM of

the cap analogue m7G(ppp)G (Prof. Edward Darzynkiewicz, Warsaw University, Poland) were added to the reaction to implement transcription and translation reactions simultaneously. Orthogonal cell-free reactions based on eAzFRS were further supplemented with 4 μM orthogonal $tRNA_{Tyr}$ and 2 mM p-Azido-L-phenylalanine (AzF), while PylRS-AF based reactions were supplemented with 4 μM orthogonal $tRNA_{Pyl}$ and 2 mM strained cyclooctyne (SCO). Reactions containing purified eAzFRS were supplemented with 4 μM eAzFRS unless otherwise noted. Reactions were incubated for three hours at 30°C and 600 rpm. The eAzFRS and the *in vitro* transcribed tRNAtyr were prepared as previously described in detail (Zemella et al., 2019). Briefly, the eAzFRS was synthesized in a cell-free reaction based on *E. coli* lysate using the RTS 500 E.coli HY Kit (biotechrabbit). Purification of eAzFRS was performed using Strep-Tactin Superflow high capacity columns (IBA Life Sciences) due to the presence of a C-terminal Strep tag II within the coding sequence of eAzFRS.

Autoradiography of Radiolabeled Proteins

The crude reaction mixture was precipitated after CFPS by acetone. Therefore, 3 μL of reaction mix was diluted with 47 μL distilled water followed by three times the sample volume of cold acetone. After incubation on ice for 15 min samples were centrifuged 10 min at 16,000 × g. Supernatants were discarded while precipitated proteins were dried for 30 min at 45°C to get rid of remaining acetone. Afterwards, dried protein pellets were resuspended in LDS-sample buffer (Invitrogen) containing 50 mM DTT. Samples were mixed at 1,400 rpm at room temperature. Denaturing polyacrylamide gel electrophoresis (PAGE) was carried out with NuPAGE 10% Bis-Tris Gels. Gels were stained and dried by a vacuum chamber (Unigeldryer 3545D, Uniequip), deposited on phosphor screens (GE Healthcare) for three days and screens were scanned by an imaging system (Typhoon TRIO + Imager, GE Healthcare). The presence of ^{14}C leucine in CFPS allowed the visualization of synthesized proteins based on autoradiography.

Luciferase Assay

The analysis of Nanoluciferase (Promega) activity was performed by utilizing the Nano-Glo Luciferase Assay System (Promega). Therefore, 50 μL of the assay reagent was mixed with 3 μL crude mix of a cell-free reaction and luminescence was detected by the Multimode Microplate Reader Mithras^2 LB 943 (Berthold Technologies GmbH) by using an OD2 filter. Measurements were conducted in triplicate and data were expressed as the mean with standard deviation.

Semi-Quantitative Western Blot

5×10^5 CHO cells were lysed for 30 min on ice using RIPA lysis buffer (10 mM Tris–HCl pH 8, 140 mM NaCl, 1% Triton X-100, 1 mM EDTA, 0.5 mM EGTA, 0.1% SDS, 0.1% sodium deoxycholate), supplemented with cOmplete-ULTRA protease inhibitor cocktail (Roche). The supernatant was isolated after 10 min centrifugation (16,000 × g) at 4°C. The protein concentration was determined by the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) and 2.5 μg proteins were heated at 70°C for 10 min in LDS sample buffer (Invitrogen). Samples were separated by denaturing polyacrylamide gel electrophoresis with NuPAGE 10% Bis-Tris Gels. Protein transfer to a PVDF membrane was carried out by the iBlot Dry Blotting System (Invitrogen). The membrane was blocked with 5% bovine serum albumin (BSA). The C-terminally located Strep-tag II of the eAzFRS coding sequence allowed western blot analysis by using anti-Strep II (ab76949, Abcam) as primary antibody. Furthermore, anti-Actin (I-19) (sc-1616, Santa Cruz Biotechnology) was utilized as primary antibody since β-Actin served as the internal control. Primary antibodies were diluted 1: 1,000 in TBS with 0.1% Tween 20 and 5% BSA. Detection of signals was enabled by using the anti-rabbit IgG, HRP-linked antibody (#7074, Cell Signaling Technology) as secondary antibody (1:3,000 in TBS with 0.1% Tween 20 and 5% BSA). After detection of eAzFRS, removal of antibodies from the blot membrane was achieved by incubating the membrane at 50°C for 30 min in stripping buffer (0.5 M Tris HCl pH 6.8, 10% SDS, 0.8% 2-mercaptoethanol). Following detection of β-Actin was performed on the stripped PVDF membrane. Band intensities of eAzFRS signals were normalized to β-Actin signals.

Data and Statistical Analysis

The amber suppression efficiency was calculated by dividing the relative light unit (RLU) value based on cell-free reactions containing the A2aRamb construct by the RLU value based on cell-free reactions containing the A2aR construct without an amber stop codon. After multiplying the ratio by 100 the amber suppression efficiency was expressed as a percentage. The experimental data were expressed as mean values ± standard error, with "n" representing the number of biological replicates if not otherwise specified. Differences between three independent experiments were analyzed using a Welch's *t*-test. The maximum growth rate of analyzed cell lines was calculated from the slope of log-transformed data. Therefore, non-linear fitting of cell growth data was performed by using the SRichards1 function of Origin (Pro) software, Version Number 2019 (OriginLab Corporation, Northampton, MA, United States).

RESULTS

Expression of eAzFRS in CHO Cells Prior to Cell Lysate Preparation Enables Orthogonal Cell-Free Reactions

A previous study on orthogonal translation based on *E.coli* cell-free systems showed that transformation of *E.coli* with the *Methanosarcina mazei* orthogonal pyrrolysyl-tRNA synthetase prior to cell disruption circumvent time-intensive purification steps, leading to an optimized orthogonal cell-free protein synthesis system (Chemla et al., 2015). Based on this, we aimed to transiently transfect CHO cells with eAzFRS to create an eukaryotic cell-free system with the ability to site-specifically modify difficult-to-express proteins, such as the membrane protein adenosine receptor A2a (A2aR). Additionally, the goal was to identify optimal cultivation

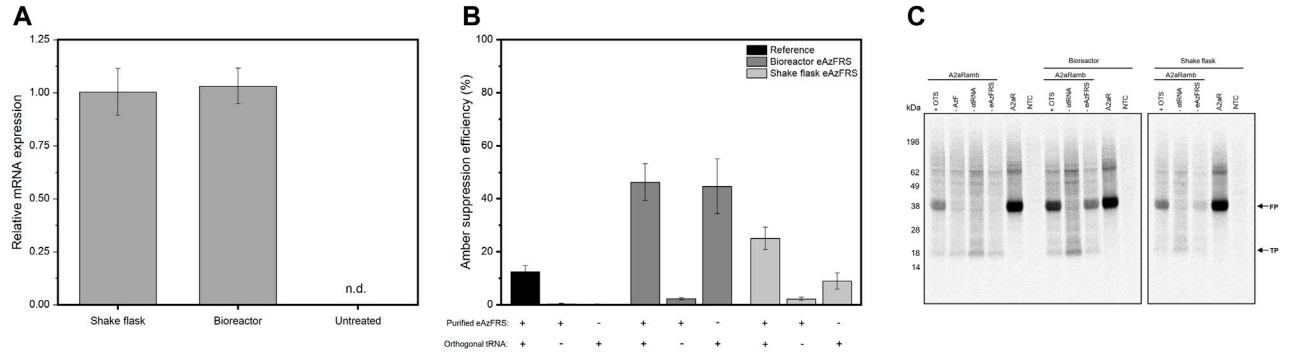


FIGURE 2 | CFPS based on transiently transfected CHO cells expressing eAzFRS. **(A)** Relative eAzFRS mRNA expression levels two days after transfection. Experiments were performed in technical triplicates and bars represent the standard deviation of the mean. **(B)** Amber suppression efficiency based on cell-free synthesized A2aR and A2aRamb constructs using reference cell lysates and cell lysates from transiently transfected CHO cells. Samples with a plus indicate the addition of either 4 μM purified eAzFRS and orthogonal tRNA, while a minus indicates the absence of the particular orthogonal component. Luminescence signals of cell-free reactions without a template (NTC) were subtracted from sample signals prior to calculation of suppression efficiency. Data are shown as mean ± SEM (n = 3). **(C)** Autoradiography of electrophoretically separated proteins based on a cell-free reaction with A2aR, A2aRamb constructs and a no-template control (NTC). n.d.: not detected, FP: Full-length protein, TP: Truncated protein, otRNA: Orthogonal tRNA, AzF: p-Azido-L-phenylalanine, OTS: Orthogonal translation system, TrCHO: Transiently transfected CHO cells.

conditions, while keeping transfection conditions constant, for producing a highly active CHO cell lysate enabling site-specific labeling of proteins in CFPS. Transiently transfected cells were either cultured in a shake flask or in a bioreactor to evaluate the influence of the culture format on cell lysate activity. Transcriptional analysis by quantitative PCR (qPCR) revealed similar eAzFRS mRNA levels for both cultivation formats (**Figure 2A**). As expected, mRNA expression could not be detected in non-treated cells due to the absence of endogenous eAzFRS. CHO cells were harvested two days post-transfection to ensure sufficient protein expression prior to cell lysate preparation. Characterization of orthogonal aaRS activity of the resulting processed cell extracts was achieved by evaluating cell-free synthesis of a reporter gene construct composed of the GPCR coding sequence containing an amber stop codon at position 215 and a C-terminal Nanoluciferase (Nluc) and is referred to as A2aRamb (Zemella et al., 2019). Once, the amber stop codon is suppressed by orthogonal tRNA-ncaa, the C-terminal Nluc activity can be correlated to the aminoacylation activity based on the orthogonal aaRS, since the luminescent output was shown to be linear (**Supplementary Figure S2**). A2aR construct without an amber stop codon was utilized to assess the efficiency of ncaa incorporation into the desired protein, referred to as amber suppression efficiency. A concentration of 4 μM purified eAzFRS was determined to be optimal in cell-free reactions based on CHO lysate (**Supplementary Figure S3**).

To exclude the possible negative effect of PEI reagent on cell lysate and thus on CFPS, a reaction set-up was included that was based on cell lysate derived from CHO cells transfected only with PEI reagent without any DNA template (**Supplementary Figure S4**). Unexpectedly, cell lysate based on CHO cells treated only with PEI resulted in higher protein expression compared to non-treated CHO cells, which may be caused by batch-to-batch variations affecting cell lysate efficiency. Considering the comparable resulting translationally active cell lysate of PEI treated CHO cells, cell lysates processed from transfected cells in a 1 L bioreactor and 2 L shake flask were analyzed for orthogonal translation in CFPS. A reference lysate based on non-treated CHO cells was utilized as a control. CFPS by using the reference lysate shows no translation after reaching the stop codon if omitting either orthogonal tRNA (otRNA) or purified eAzFRS, while combining orthogonal components led to a suppression efficiency of 12% (**Figure 2B**). In contrast, without addition of purified eAzFRS to cell-free reactions based on novel lysates, a higher amber suppression efficiency of 45% for bioreactor and 9% for shake flask based cultivations was detected. When supplementing cell-free reactions with purified eAzFRS, we came across further increase of suppression efficiencies up to 46% (bioreactor) and 25% (shake flask). Qualitative electrophoretic analysis of proteins from cell-free reactions with A2aRamb and A2aR supplemented with ^{14}C-leucine showed expression of full-length proteins (FP) and termination products (TP) as expected (**Figure 2C**). Overall, these results demonstrate the functionality of the orthogonal system by expressing the eAzFRS in CHO cells prior to cell lysate preparation.

Integration of eAzFRS Into the CHO Genome Creates a Robust System for Orthogonal Cell-Free Reactions

One should acknowledge the heterogeneous expression and the fact that eAzFRS expression only persists for a limited time if cells are transiently transfected. Here we conduct an alternative approach to generate an *in vitro* orthogonal translation system by utilizing CHO cells stably transfected with eAzFRS. Several studies have demonstrated the use of CRISPR/Cas technology to overexpress a diverse repertoire of proteins from defined genomic

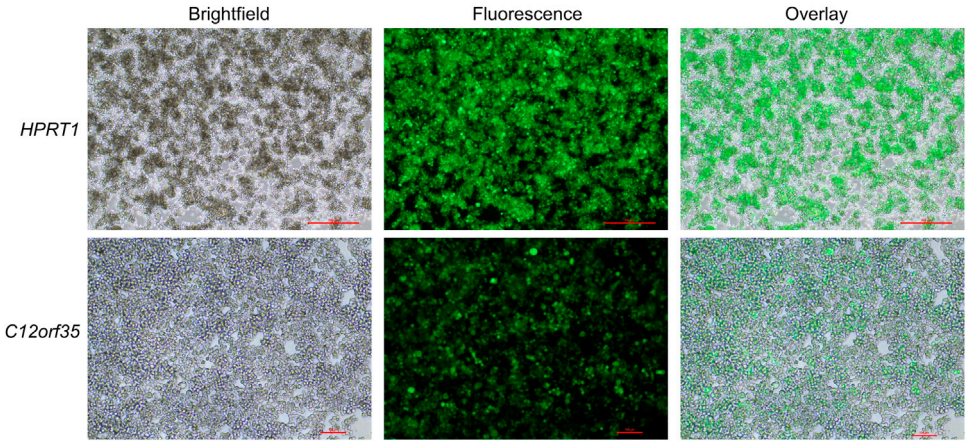


FIGURE 3 | CHO cells stable transfected with GFP. CHO cells were stable transfected by CRISPR/Cas9 to integrate GFP into the *HPRT1* or *C12orf35* locus. Cells were subjected to puromycin selection pressure to enrich positively transfected cells. Microscopic images of GFP clone pools are shown. Scale bars for *HPRT1* measure 500 µm, while scale bars for *C12orf35* measure 100 µm.

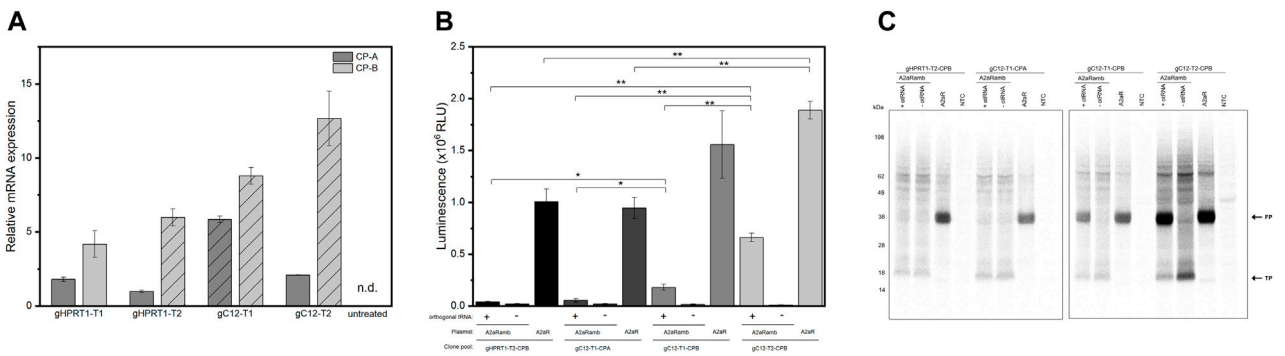


FIGURE 4 | CHO clone pools stable transfected with eAzFRS. **(A)** Relative eAzFRS mRNA expression levels of CHO clone pools. The line pattern indicates selected clone pools for cell-free synthesis. Experiments were performed in technical triplicates and bars represent the standard deviation. **(B)** Luciferase assay based on cell-free synthesized A2aR and A2aRamb constructs using cell lysates based on clone pools gHPRT1-T2-CPB, gC12-T1-CPA, gC12-T1-CPB and gC12-T2-CPB. Samples with a plus or minus indicate the addition or absence of orthogonal tRNA, respectively. Luminescence signals of cell-free reactions without a template (NTC) were subtracted from sample signals. Data are shown as mean ± SEM (n = 3). A Welch's t-test was performed to determine the statistically significant differences between samples (*p-value < 0.05; **p-value < 0.01). **(C)** Exemplary autoradiography of electrophoretically separated proteins of a CFPS based on the different clone pools. CP: Clone pool, n.d: not detected, otRNA: Orthogonal tRNA, FP: Full-length protein, TP: Truncated protein, NTC: No-template control.

TABLE 1 | Utilized guide RNA sequences

gRNA	Locus	Origin
gHPRT1-T1	*HPRT1*	gRNA from Zhao et al. (2018)
gHPRT1-T2	*HPRT1*	Novel gRNA
gC12-T1	*C12orf35*	gRNA from Zhao et al. (2018)
gC12-T2	*C12orf35*	Novel gRNA

loci in CHO cells (Lee et al., 2015; Eisenhut et al., 2018; Iwao et al., 2021). The genomic loci *HPRT1* and *C12orf35* were investigated for the generation of stable CHO cell lines using CRISPR/Cas, as reported in a recent study (Zhao et al., 2018). According to this, we have aimed to target respective loci to integrate an eAzFRS expression cassette into the CHO genome. Therefore, we analyzed four different gRNAs to modify the *HPRT1* and *C12orf35* locus of the CHO genome (**Table 1**). Either locus was targeted with one gRNA (referred to as gHPRT1-T1 and gC12-T1), which was previously identified (Zhao et al., 2018). Additionally, two gRNAs were designed that were predicted to have a higher specificity score value and less off-target effects and were referred to as gHPRT1-T2 and gC12-T2.

First, stable transfection of CHO cells was conducted by using a GFP expression cassette targeting either *HPRT1* or *C12orf35* to examine whether the locations were suitable for expression. Afterwards selection pressure (puromycin) was applied. Qualitative microscopic analysis demonstrated GFP expression of genetically modified cells (**Figure 3**). Both target sites showed a high proportion of enriched fluorescent cells, leading us to integrate an eAzFRS expression cassette into both loci. Selection of stable

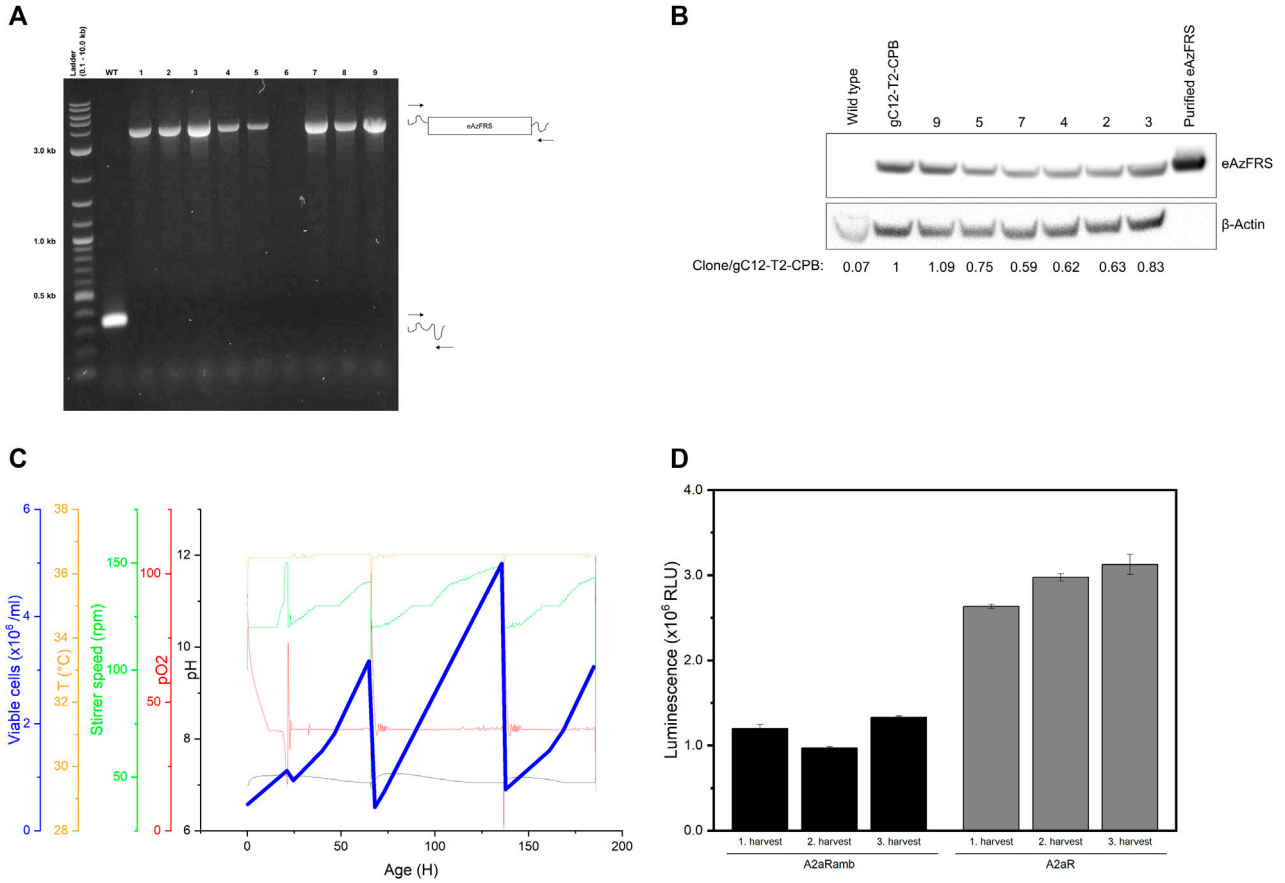


FIGURE 5 | Evaluation of clonal cell lines expressing the eAzFRS. **(A)** Genotyping PCR of clonal cell lines isolated from clone pool gC12-T2-CPB. The expected outcome is indicated by visualize primer binding (arrows) on wilde type (WT) and modified cells. A Quick-Load 2-Log DNA Ladder 0.1–10.0 kb (New England Biolabs) was used. **(B)** Semi-quantitative western blot of cell clones. eAzFRS band intensities were normalized to the internal β-Actin control. The indicated fold change is expressed as ratio of band intensity of the samples compared to clone pool gC12-T2-CPB. **(C)** Exemplary fermentation process of clone RS7. Cell harvesting was performed three times at highest peaks (blue line). **(D)** Luciferase assay of cell-free synthesized A2aR and A2aRamb constructs based on cell lysates prepared from three harvest points of a cultivation process using CHO clone RS7. Experiments were performed in technical triplicates and bars represent the standard deviation.

transfected cells was conducted by supplementation of puromycin. Two promising CHO cell clone pools for each gRNA target were evaluated by qPCR. The relative mRNA expression of eAzFRS is shown in **Figure 4A**. The change in gene expression is relative to the sample with the lowest gene expression. A 8.8-fold and 12.7-fold higher mRNA expression could be observed for the clone pools (CP) B by exploiting gC12-T1 and gC12-T2, respectively, compared to the lowest expression by using gHPRT1-T2 (CP-A). Further investigation on clone pools (marked in **Figure 4A**) by producing cell lysates highlighted their performance in orthogonal cell-free reactions. Therefore, three different cell lysates of gHPRT1-T2-CPB, gC12-T1-CPA, gC12-T1-CPB and gC12-T2-CP-B were generated and the eAzFRS activity was evaluated by the C-terminal Nluc activity. Indeed, significant higher luminescence could be observed for clone pool gC12-T2-CP-B compared to the other examined clone pools in cell-free reactions with the A2aRamb construct (**Figure 4B**). Furthermore, significant higher signals were achieved for clone pool gC12-T2-CP-B compared to gHPRT1-T1-CPB and gC12-T1-CPA using the construct A2aR without an amber position, indicating a generally higher protein production. Thus clone pool gC12-T2-CP-B was identified as the most promising candidate for subsequent isolation of CHO clones. Electrophoretic analysis supported the observed data of the luciferase assay (**Figure 4C**).

Single Clone Selection Circumvents Clonal Variations and Ensures Reproducibility

On the one hand, working with a heterogeneous pool of edited cells provides a time saving method to produce cell lysates containing an endogenous orthogonal aaRS. On the other hand, reproducibility is a main issue. As a consequence, our objective was to create a stable cell line to overcome clonal variability of edited cell pools. Nine single clones were isolated from clone pool gC12-T2-CP-B. Genotyping PCR with primers flanking the site of integration revealed only a short PCR product for WT cells, as expected (**Figure 5A**). Upon successful integration of the eAzFRS expression cassette an increase of PCR product size was anticipated. The clonality of eight produced cell lines could be demonstrated due to the presence of a PCR-

product at the expected size, while clone 6 showed no PCR product. Mono-allelic knock-in of the eAzFRS expression cassette can be excluded, since only one defined PCR product correlating to the exact size of the expression cassette was observed. The eAzFRS CHO clones (RS) 2–5, 7 and 9 were subjected to semi-quantitative western blot analysis, since increased outgrowth was observed compared to RS1 and RS8. A 1.09-fold increase in band intensity was observed for clone RS9 compared to the clone pool gC12-T2-CPB, while clone RS7 showed the lowest band intensity (**Figure 5B**).

Recombinant protein expression is often associated with a metabolic burden on the cells, leading us to compare unmodified CHO cells (wild type) with CHO clones with the highest (RS9) and lowest (RS7) eAzFRS expression by evaluating their maximum growth rates. Therefore, growth curves were plotted (**Supplementary Figure S5**) and resulting growth rates were compared by a Welch's *t*-test. The maximum growth rates of WT ($0.013\ h^{-1} \pm 0.00080$), RS7 ($0.014\ h^{-1} \pm 0.00076$) and RS9 ($0.015\ h^{-1} \pm 0.0047$) were similar and there was no significant difference determined. However, cell-free reactions based on cell lysate from clone RS9 led to a lower suppression efficiency (13%) than previously observed with the cell lysate prepared from the parental pool (**Supplementary Figure S6**). In contrast, higher amber suppression could be observed by using cell lysate based on clone RS7 (**Supplementary Figure S7**). Further examination of RS7 cells by producing cell lysates highlighted their performance during the fermentation process in a 1 L bioreactor and in orthogonal cell-free reactions. In particular, it was analyzed if the absence of puromycin during the fermentation process affects the stability of eAzFRS gene expression. Consequently, we set out to cultivate clone RS7 under controlled conditions for nine days and CHO cells were harvested after three, six and nine days of cultivation as depicted in **Figure 5C**. Therefore, a so-called repeated batch mode was applied, leaving CHO cells in the bioreactor after each harvest point. Cell lysates prepared at three harvest points were analyzed by cell-free synthesis of A2aRamb and A2aR to assess the functionality of the novel orthogonal system over the period of time. Using the A2aRamb construct, it was shown that no reduced enzyme activity was observed after three harvesting times (**Figure 5D**). General protein synthesis was also not negatively affected over the cultivation period, as indicated by using A2aR construct.

Moreover, it was shown that addition of increasing concentrations of purified eAzFRS did not elevate luminescent signals in cell-free reactions with A2aRamb using the RS7 cell lysate (**Supplementary Figure S7**). On the contrary, addition of 3 μM eAzFRS seems to have a negative impact on the reaction.

Transferability of the Approach to Orthogonal PylRS

A promising orthogonal aaRS is the double mutant pyrrolysyl-tRNA synthetase (PylRS-AF), which allows large and bulky ncaa to be incorporated into proteins (Oliveira et al., 2017). A major drawback of the PylRS is the frequently observed protein instability and tendency to aggregate (Wan et al., 2014). Therefore, the intent was to check whether the PylRS can also be expressed to obtain active orthogonal aaRS in the cell lysate, thus avoiding enzyme purification and possibly accompanying enzyme inactivity. For this purpose, we transiently transfected the PylRS-AF into CHO cells and prepared cell lysate based on shake flask and bioreactor based cultivation. We could demonstrate that a suppression efficiency of 24% (bioreactor) and 42% (shake flask) was achieved by incorporating strained cyclooctyne (SCO) into the reporter construct A2aRamb (**Figure 6A**). These findings could be verified by autoradiography due to the presence of ^{14}C leucine during the reaction (**Figure 6B**).

DISCUSSION

Chemical modification of proteins is widely used to study protein function and interactions in biological environments. Applied research can also facilitate medical diagnostics and therapy by linking drugs as well as fluorescent markers to proteins as exemplified by the production of novel antibody-drug conjugates. Suitable antibodies are often coupled to the drug *via* cysteines and lysines (Shadish and DeForest, 2020). Due to many possible coupling partners, depending on the individual amino acid composition, a mixture of conjugates results, displaying different pharmacokinetic properties (Hamblett et al., 2004; Wang et al., 2005). In this context, cell-free protein synthesis can be utilized to modify therapeutically relevant proteins by means of orthogonal systems at precisely defined positions to obtain a homogeneous mixture of antibody-drug conjugates. It has been shown that site-specifically modified Her2-binding IgG antibodies based on *E.coli* cell-free systems have been coupled to the anti-cancer agent monomethyl auristatin F, delivering the antibody-drug conjugate to the antigen (Zimmerman et al., 2014).

The orthogonal aaRS/tRNA pair is usually added to the open cell-free system. On the one hand, the simple addition of a wide variety of reactants to the translation reaction is advantageous. On the other hand, the additive can negatively affect the reaction itself, as the reaction environment is highly dependent on a defined milieu. Proteins such as aaRS are typically located in a solvent that ensures protein stability but does not correspond to the optimum of cell-free reaction conditions in terms of ion concentrations. Various studies have shown that even in the case of a minor change of ion concentrations including magnesium, potassium and chloride ions, a significant effect on protein production was observed (Brigotti et al., 2003; Spice et al., 2020; Garenne et al., 2021). In particular, a reduced amber suppression was detectable with increasing eAzFRS concentrations and thus increasing buffer concentrations. Together with the time consuming purification of aaRS, an alternative approach was achieved by Chemla et al. using cell-free systems based on *E.coli* extracts, where PylRS expression was performed prior to cell disruption (Chemla et al., 2015). Besides, commonly used *E. coli* extracts for CFPS contain the T7 RNA polymerase from T7 phage to bypass the addition of purified enzyme (Köhrer et al., 1996; Des Soye et al., 2019). In contrast to *E.coli* based systems, the present study demonstrates that orthogonal aaRS originating from *E. coli* and archaea were

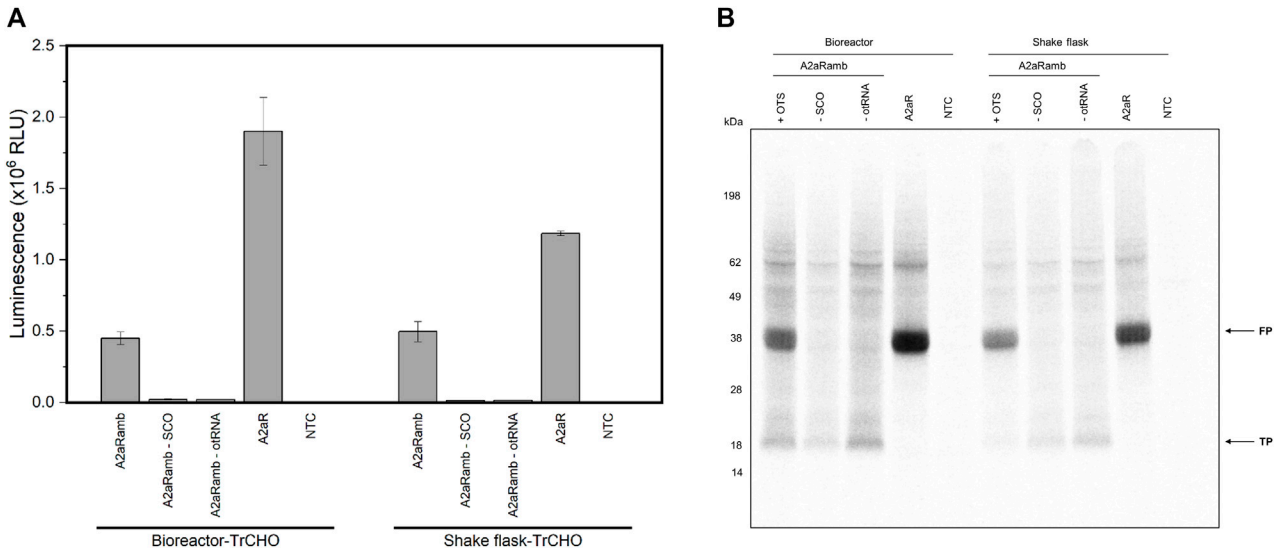


FIGURE 6 | CFPS based on transiently transfected CHO cells expressing PylRS-AF. **(A)** Luciferase assay of cell-free synthesized A2aRamb and A2aR constructs based on transiently transfected CHO cells expressing the PylRS-AF. Samples with a plus indicate the addition of the particular orthogonal component, while a minus indicates the absence of the orthogonal component. Experiments were performed in technical triplicates and bars represent the standard deviation. **(B)** Autoradiography of electrophoretically separated proteins corresponding to CFPS shown in **(A)**. FP: Full-length protein, TP: Truncated protein, NTC: No-template control, SCO: Strained cyclooctyne, otRNA: Orthogonal tRNA, OTS: Orthogonal translation system, TrCHO: Transiently transfected CHO cells.

successfully integrated into eukaryotic cells prior to cell disruption. The resulting orthogonal translation system was used to produce the complex membrane protein A2aR in cell-free systems.

It has already been shown that orthogonal aaRS were transiently transfected to modify proteins *in vivo* (Cohen and Arbely, 2016; Meineke et al., 2020). However, it has been assumed that transfection has a negative impact on the cell lysate for CFPS as cell disruption is typically performed 2–3 days after transfection and transfection reagents can negatively affect cell viability. Contrary to expectations, CFPS based on lysates of CHO cells treated with the commonly used transfection reagent PEI were shown to have similar reporter gene activity as non-treated CHO cells.

Another aspect that was examined in the present study is the cultivation format after transient transfection prior to cell disruption. Higher reporter gene activities based on A2aR were detected with cell lysates based on cultivation using bioreactors. This could be related to the controlled cultivation conditions and thus reduced cell stress in bioreactor based cultivations. Although shake flasks are easy to use, lack of oxygen control and pH control can be challenging throughout the cultivation process (Link and Weuster-Botz, 2011). Nevertheless, satisfactory reporter gene activities could be achieved based on shake flask cultivation. Thus, with the presented orthogonal translation system, cells can be efficiently enabled for orthogonal aaRS production, cost-effective cultivation in shake flasks and subsequent cell lysate preparation for CFPS can be performed in a timesaving procedure. Of particular concern is the role of the novel orthogonal CFPS based on CHO lysate for investigating membrane proteins such as the examined A2aR, since efficient translocation of membrane proteins into microsomal structures can be achieved in certain eukaryotic cell-free systems (Brödel et al., 2014; Sonnabend et al., 2017; Zemella et al., 2017). As GPCRs account for 35% of approved drug targets, the analysis of GPCRs based on interaction studies with potential ligands is of highest interest (Sriram and Insel, 2018). Fluorescence-based interaction studies offer the possibility to identify potential ligands and inhibitors by incorporating ncaa at defined positions in GPCRs subsequently conjugating them with fluorescent molecules, thus eliminating the requirement for large fluorescent fusion proteins that may affect protein interactions and functions (Lee et al., 2019). In this context, we recently demonstrated that site-specifically fluorescently labeled A2aR was excited by the C-terminal localized Nluc after substrate addition and a signal change was measured based on the conformational change of the GPCR after addition of ligand (Zemella et al., 2019). In the present study we demonstrate that transient transfection of PylRS-AF also leads to an intact orthogonal cell-free translational system for the production of site-specifically modified A2aR. Accordingly, purification of the unstable PylRS can be bypassed, which can influence the enzyme`s activity significantly.

Cell lysates based on transient transfection showed variable amber suppression efficiencies, which could be the result of the different transfection efficiencies. Alternatively, to create a cost-effective and reproducible orthogonal cell-free system for long-term use, the eAzFRS was stably transfected into the CHO cells. In particular, stable transfection by CRISPR offers the possibility to target sites in the genome that allow high expression levels and simultaneously reduce silencing (Lo et al., 2017; Li et al., 2020).

While many studies have demonstrated that *HPRT1* is suitable for high levels of antibody production in CHO cells (Wang et al., 2017; Kawabe et al., 2018), our study showed that the *C12orf35* locus is beneficial for incorporating the orthogonal eAzFRS into the CHO genome and subsequently generating active cell lysates for orthogonal translation. Disruption of the *C12orf35* gene, located in the telomeric region of chromosome 8, has been reported to affect recombinant protein production in the resulting cell lines (Ritter et al., 2016a). Using small interfering RNAs, the *C12orf35* gene was silenced and monoclonal antibody production was increased, whereas silencing of the gene and subsequent isolation of clonal cell lines resulted in fast recovery rates during the selection process. Our results are consistent with those reported by Zhao et al. who were able to generate cell lines with highest stability and anti-PD1 monoclonal antibody productivity using the *C12orf35* locus in contrast to other CHO hot spots studied such as *HPRT* and *GRIK1* (Zhao et al., 2018). However, expression of eAzFRS using the *HPRT1* locus led to lower mRNA levels and minor eAzFRS activity in cell lysates. On the one hand reduced eAzFRS transcription and production could be a result of increased epigenetic gene silencing. On the other hand *HPRT1* is a widely used target site for protein production (Wang et al., 2017; Kawabe et al., 2018). Nevertheless, the eAzFRS coding sequence is originated from *E.coli* and might have an impact on the *HPRT1* locus in CHO cells.

Addressing transcriptionally active gene loci by CRISPR ensures controlled overexpression of proteins and overcomes the unpredictable insertion of transgenes into random genomic positions. The on-target efficiency and off-target effects depend mainly on the gRNA sequence used (Wilson et al., 2018). Therefore, we designed a gRNA sequence that addresses the *HPRT1* and *C12orf35* loci and postulated that integration of eAzFRS is improved. Indeed, the novel gRNAs were shown to result in a higher eAzFRS mRNA level in comparison to the recently reported gRNA sequences (Zhao et al., 2018). This may be due to the fact that the sequence context flanking the target site and resulting gene positioning effects are of particular importance. Indeed, it has been shown that deletions in the telomeric region around *C12orf35* resulted in CHO cell lines with increased stability and high protein production rates (Ritter et al., 2016b). Epigenetic regulatory mechanisms such as DNA methylation and histone modification can strongly influence the expression level as well as gene silencing (Gibney and Nolan, 2010; Keller et al., 2019). However, it must be emphasized that off-target effects of the utilized gRNAs have not been investigated and integration of eAzFRS in other transcriptionally active gene loci cannot be excluded. In addition, mRNA expression strongly depends on the inserted gene sequence in the genomic context.

Various orthogonal aaRS would be of interest to incorporate them into CHO cells to expand the repertoire of diverse ncaa for incorporation into proteins in CFPS. In contrast, including orthogonal tRNA in the cell lysate could be challenging, since it was reported that the ratio of expression cassettes of both orthogonal tRNA and corresponding aaRS needs to be adjusted for cell-based production to achieve high suppression efficiency (Ryu and Schultz, 2006; Schmied et al., 2014). This highlights the strength of CFPS, as the tRNA can be transcribed *in vitro* followed by a titration to the cell-free system in optimal ratios.

In summary, the developed system is suitable for the incorporation of diverse orthogonal aaRS as well as other proteins of interest into the cell lysate, thus eliminating the need for the supplementation of enzymes and co-factors. As a result, a novel orthogonal eukaryotic cell-free system speeds up the production of site-specifically modified complex proteins.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

JS was involved in methodology, investigation, formal analysis and writing the original draft. NC was involved in methodology, investigation and formal analysis. AZ was involved in methodology, conceptualization, review and editing the draft. LT was involved in methodology, investigation, conceptualization, review and editing of the draft. SK was involved in methodology, conceptualization, review and editing of the draft, project administration and funding acquisition.

FUNDING

This work was supported by the European Regional Development Fund (EFRE) and the German Ministry of Education and Research (BMBF, Nos. 031B0078A, 031B0831C).

ACKNOWLEDGMENTS

The authors would like to thank Dana Wenzel and Felix Jorde (Fraunhofer IZI-BB, Potsdam-Golm, Germany) for their support in CHO lysate preparation. Moreover, we would like to thank Servier Medical Art (Servier Medical Art—https://smart.servier.com/).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.832379/full#supplementary-material

REFERENCES

Barbu, A., and Welsh, N. (2007). Lipofection of Insulin-Producing RINm5F Cells: Methodological Improvements. *J. Liposome Res.* 17, 49–62. doi:10.1080/01676830701374986

Beyer, J. N., Hosseinzadeh, P., Gottfried-Lee, I., van Fossen, E. M., Zhu, P., Bednar, R. M., et al. (2020). Overcoming Near-Cognate Suppression in a Release Factor 1-Deficient Host with an Improved Nitro-Tyrosine tRNA Synthetase. *J. Mol. Biol.* 432, 4690–4704. doi:10.1016/j.jmb.2020.06.014

Brigotti, M., Petronini, P. G., Carnicelli, D., Alfieri, R. R., Bonelli, M. A., Borghetti, A. F., et al. (2003). Effects of Osmolarity, Ions and Compatible Osmolytes on Cell-free Protein Synthesis. *Biochem. J.* 369, 369–374. doi:10.1042/BJ20021056

Brödel, A. K., Sonnabend, A., and Kubick, S. (2014). Cell-free Protein Expression Based on Extracts from CHO Cells. *Biotechnol. Bioeng.* 111, 25–36. doi:10.1002/bit.25013

Brown, A. J., Gibson, S., Hatton, D., and James, D. C. (2018). Transcriptome-Based Identification of the Optimal Reference CHO Genes for Normalisation of qPCR Data. *Biotechnol. J.* 13, 1700259. doi:10.1002/biot.201700259

Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V., and Alfonta, L. (2015). Genetically Expanded Cell-free Protein Synthesis Using Endogenous Pyrrolysyl Orthogonal Translation System. *Biotechnol. Bioeng.* 112, 1663–1672. doi:10.1002/bit.25587

Chi, X., Zheng, Q., Jiang, R., Chen-Tsai, R. Y., and Kong, L.-J. (2019). A System for Site-specific Integration of Transgenes in Mammalian Cells. *PLoS One* 14, e0219842. doi:10.1371/journal.pone.0219842

Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002). Addition of a Photocrosslinking Amino Acid to the Genetic Code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11020–11024. doi:10.1073/pnas.172226299

Chung, C. Z., Amikura, K., and Söll, D. (2020). Using Genetic Code Expansion for Protein Biochemical Studies. *Front. Bioeng. Biotechnol.* 8, 598577. doi:10.3389/fbioe.2020.598577

Cohen, S., and Arbely, E. (2016). Single-Plasmid-Based System for Efficient Noncanonical Amino Acid Mutagenesis in Cultured Mammalian Cells. *Chembiochem* 17, 1008–1011. doi:10.1002/cbic.201500681

Debelouchina, G. T., and Muir, T. W. (2017). A Molecular Engineering Toolbox for the Structural Biologist. *Quart. Rev. Biophys.* 50, e7. doi:10.1017/S0033583517000051

Des Soye, B. J., Gerbasi, V. R., Thomas, P. M., Kelleher, N. L., and Jewett, M. C. (2019). A Highly Productive, One-Pot Cell-free Protein Synthesis Platform Based on Genomically Recoded *Escherichia coli*. *Cel Chem. Biol.* 26, 1743–1754. doi:10.1016/j.chembiol.2019.10.008

Eisenhut, P., Klanert, G., Weinguny, M., Baier, L., Jadhav, V., Ivansson, D., et al. (2018). A CRISPR/Cas9 Based Engineering Strategy for Overexpression of Multiple Genes in Chinese Hamster Ovary Cells. *Metab. Eng.* 48, 72–81. doi:10.1016/j.ymben.2018.05.017

Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., et al. (2018). A Multi-landing Pad DNA Integration Platform for Mammalian Cell Engineering. *Nucleic Acids Res.* 46, 4072–4086. doi:10.1093/nar/gky216

Garenne, D., Haines, M. C., Romantseva, E. F., Freemont, P., Strychalski, E. A., and Noireaux, V. (2021). Cell-Free Gene Expression. *Nat. Rev. Methods Primers* 1, 49. doi:10.1038/s43586-021-00046-x

Gibney, E. R., and Nolan, C. M. (2010). Epigenetics and Gene Expression. *Heredity* 105, 4–13. doi:10.1038/hdy.2010.54

Gillette, W., Frank, P., Perkins, S., Drew, M., Grose, C., and Esposito, D. (2019). Production of Farnesylated and Methylated Proteins in an Engineered Insect Cell System. *Methods Mol. Biol.*, 259–277. doi:10.1007/978-1-4939-9532-5_20

Grav, L. M., La Cour Karottki, K. J., Lee, J. S., and Kildegaard, H. F. (2017). Application of CRISPR/Cas9 Genome Editing to Improve Recombinant Protein Production in CHO Cells. *Methods Mol. Biol.* 1603, 101–118. doi:10.1007/978-1-4939-6972-2_7

Gubellini, F., Verdon, G., Karpowich, N. K., Luff, J. D., Boël, G., Gauthier, N., et al. (2011). Physiological Response to Membrane Protein Overexpression in *E. coli*. *Mol. Cell Proteomics* 10, M111. doi:10.1074/mcp.M111.007930

Hamblett, K. J., Senter, P. D., Chace, D. F., Sun, M. M. C., Lenox, J., Cerveny, C. G., et al. (2004). Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate. *Clin. Cancer Res.* 10, 7063–7070. doi:10.1158/1078-0432.CCR-04-0789

Iwao, R., Kawabe, Y., Murakami, M., Ito, A., and Kamihira, M. (2021). Targeted Knock-In of Transgenes into the CHO Cell Genome Using CRISPR-Mediated Integration Systems. *MATEC Web Conf.* 333, 07001. doi:10.1051/matecconf/202133307001

Jayapal, K., Wlaschin, K. F., Hu, W. S., and Yap, M. G. S. (2007). Recombinant Protein Therapeutics from CHO Cells - 20 Years and Counting. *Chem. Eng. Prog.* 103, 40

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816–821. doi:10.1126/science.1225829

Jones, D. H., Cellitti, S. E., Hao, X., Zhang, Q., Jahnz, M., Summerer, D., et al. (2010). Site-specific Labeling of Proteins with NMR-Active Unnatural Amino Acids. *J. Biomol. NMR* 46, 89–100. doi:10.1007/s10858-009-9365-4

Kapoor, N., Vanjak, I., Rozzelle, J., Berges, A., Chan, W., Yin, G., et al. (2018). Malaria Derived Glycosylphosphatidylinositol Anchor Enhances Anti-pfs25 Functional Antibodies that Block Malaria Transmission. *Biochemistry* 57, 516–519. doi:10.1021/acs.biochem.7b01099

Kawabe, Y., Komatsu, S., Komatsu, S., Murakami, M., Ito, A., Sakuma, T., et al. (2018). Targeted Knock-In of an scFv-Fc Antibody Gene into the Hprt Locus of Chinese Hamster Ovary Cells Using CRISPR/Cas9 and CRIS-PITCh Systems. *J. Biosci. Bioeng.* 125, 599–605. doi:10.1016/j.jbiosc.2017.12.003

Keller, B.-M., Maier, J., Weldle, M., Segan, S., Traenkle, B., and Rothbauer, U. (2019). A Strategy to Optimize the Generation of Stable Chromobody Cell Lines for Visualization and Quantification of Endogenous Proteins in Living Cells. *Antibodies* 8, 10. doi:10.3390/antib8010010

Kenry and Liu, B. (2019). Bio-orthogonal Click Chemistry for *In Vivo* Bioimaging. *Trends Chem.* 1, 763–778. doi:10.1016/j.trechm.2019.08.003

Khambhati, K., Bhattacharjee, G., Gohil, N., Braddick, D., Kulkarni, V., and Singh, V. (2019). Exploring the Potential of Cell-free Protein Synthesis for Extending the Abilities of Biological Systems. *Front. Bioeng. Biotechnol.* 7, 248. doi:10.3389/fbioe.2019.00248

Köhrer, C., Mayer, C., Gröbner, P., and Piendl, W. (1996). Use of T7 RNA Polymerase in an Optimized *Escherichia coli* Coupled *In Vitro* Transcription-Translation System. Application in Regulatory Studies and Expression of Long Transcription Units. *Eur. J. Biochem.* 236, 234–239. doi:10.1111/j.1432-1033.1996.00234.x

Lee, J. S., Kallehauge, T. B., Pedersen, L. E., and Kildegaard, H. F. (2015). Site-specific Integration in CHO Cells Mediated by CRISPR/Cas9 and Homology-Directed DNA Repair Pathway. *Sci. Rep.* 5, 8572. doi:10.1038/srep08572

Lee, K. J., Kang, D., and Park, H. S. (2019). Site-Specific Labeling of Proteins Using Unnatural Amino Acids. *Mol. Cell* 42, 386–396. doi:10.14348/molcells.2019.0078

Li, G., Zhang, X., Wang, H., Mo, J., Zhong, C., Shi, J., et al. (2020). CRISPR/Cas9-Mediated Integration of Large Transgene into Pig CEP112 Locus. *G3 (Bethesda)* 10, 467–473. doi:10.1534/g3.119.400810

Lin, X., Yu, A. C. S., and Chan, T. F. (2017). Efforts and Challenges in Engineering the Genetic Code. *Life* 7, 12. doi:10.3390/life7010012

Link, H., and Weuster-Botz, D. (2011). "Medium Formulation and Development,". Editor M. Moo-Young (Amsterdam: Elsevier), 1 - 6, 119–134. doi:10.1016/b978-0-08-088504-9.00092-1*Compr. Biotechnol. Principles practices industry, agcriculture, Med. Environ.*

Lo, C.-A., Greben, A. W., and Chen, B. E. (2017). Generating Stable Cell Lines with Quantifiable Protein Production Using CRISPR/Cas9-mediated Knock-In. *Biotechniques* 62, 165–174. doi:10.2144/000114534

Lu, Y. (2017). Cell-free Synthetic Biology: Engineering in an Open World. *Synth. Syst. Biotechnol.* 2, 23–27. doi:10.1016/j.synbio.2017.02.003

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided Human Genome Engineering via Cas9. *Science* 339, 823–826. doi:10.1126/science.1232033

Matasci, M., Baldi, L., Hacker, D. L., and Wurm, F. M. (2011). The PiggyBac Transposon Enhances the Frequency of CHO Stable Cell Line Generation and Yields Recombinant Lines with superior Productivity and Stability. *Biotechnol. Bioeng.* 108, 2141–2150. doi:10.1002/bit.23167

Meineke, B., Heimgärtner, J., Eirich, J., Landreh, M., and Elsässer, S. J. (2020). Site-Specific Incorporation of Two ncAAs for Two-Color Bioorthogonal Labeling and Crosslinking of Proteins on Live Mammalian Cells. *Cel Rep.* 31, 107811. doi:10.1016/j.celrep.2020.107811

Mushtaq, S., Yun, S.-J., and Jeon, J. (2019). Recent Advances in Bioorthogonal Click Chemistry for Efficient Synthesis of Radiotracers and Radiopharmaceuticals. *Molecules* 24, 3567. doi:10.3390/molecules24193567

Oliveira, B. L., Guo, Z., and Bernardes, G. J. L. (2017). Inverse Electron Demand Diels-Alder Reactions in Chemical Biology. *Chem. Soc. Rev.* 46, 4895–4950. doi:10.1039/c7cs00184c

Pfaffl, M. W. (2001). A New Mathematical Model for Relative Quantification in Real-Time RT-PCR. *Nucleic Acids Res.* 29, 45e–45. doi:10.1093/nar/29.9.e45

Quast, R. B., Ballion, B., Stech, M., Sonnabend, A., Varga, B. R., Wüstenhagen, D. A., et al. (2016). Cell-free Synthesis of Functional Human Epidermal Growth Factor Receptor: Investigation of Ligand-independent Dimerization in Sf21 Microsomal Membranes Using Non-canonical Amino Acids. *Sci. Rep.* 6, 34048. doi:10.1038/srep34048

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome Engineering Using the CRISPR-Cas9 System. *Nat. Protoc.* 8, 2281–2308. doi:10.1038/nprot.2013.143

Ritter, A., Rauschert, T., Oertli, M., Piehlmaier, D., Mantas, P., Kuntzelmann, G., et al. (2016a). Disruption of the Gene C12orf35 Leads to Increased Productivities in Recombinant CHO Cell Lines. *Biotechnol. Bioeng.* 113, 2433–2442. doi:10.1002/bit.26009

Ritter, A., Voedisch, B., Wienberg, J., Wilms, B., Geisse, S., Jostock, T., et al. (2016b). Deletion of a Telomeric Region on Chromosome 8 Correlates with Higher Productivity and Stability of CHO Cell Lines. *Biotechnol. Bioeng.* 113, 1084–1093. doi:10.1002/bit.25876

Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S. M., et al. (2015). Efficient Genetic Encoding of Phosphoserine and its Nonhydrolyzable Analog. *Nat. Chem. Biol.* 11, 496–503. doi:10.1038/nchembio.1823

Ryu, Y., and Schultz, P. G. (2006). Efficient Incorporation of Unnatural Amino Acids into Proteins in *Escherichia coli*. *Nat. Methods* 3, 263–265. doi:10.1038/nmeth864

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. doi:10.1021/ja5069728

Shadish, J. A., and DeForest, C. A. (2020). Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials. *Matter* 2, 50–77. doi:10.1016/j.matt.2019.11.011

Sonnabend, A., Spahn, V., Stech, M., Zemella, A., Stein, C., and Kubick, S. (2017). Production of G Protein-Coupled Receptors in an Insect-Based Cell-free System. *Biotechnol. Bioeng.* 114, 2328–2338. doi:10.1002/bit.26346

Spice, A. J., Aw, R., Bracewell, D. G., and Polizzi, K. M. (2020). Improving the Reaction Mix of a Pichia pastoris Cell-free System Using a Design of Experiments Approach to Minimise Experimental Effort. *Synth. Syst. Biotechnol.* 5, 137–144. doi:10.1016/j.synbio.2020.06.003

Sriram, K., and Insel, P. A. (2018). G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol. Pharmacol.* 93, 251–258. doi:10.1124/mol.117.111062

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. *Chem. Eur. J. Chem. Bio.* 11, 2268–2272. doi:10.1002/cbic.201000436

Thoring, L., Dondapati, S. K., Stech, M., Wüstenhagen, D. A., and Kubick, S. (2017). High-yield Production of "Difficult-To-Express" Proteins in a Continuous Exchange Cell-free System Based on CHO Cell Lysates. *Sci. Rep.* 7, 11710. doi:10.1038/s41598-017-12188-8

Thoring, L., and Kubick, S. (2018). Versatile Cell-free Protein Synthesis Systems Based on Chinese Hamster Ovary Cells. *Methods Mol. Biol.* 1850, 289–308. doi:10.1007/978-1-4939-8730-6_19

Tripathi, N. K., and Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Front. Bioeng. Biotechnol.* 7, 420. doi:10.3389/fbioe.2019.00420

Vogl, D. P., Conibear, A. C., and Becker, C. F. W. (2021). Segmental and Site-specific Isotope Labelling Strategies for Structural Analysis of Posttranslationally Modified Proteins. *RSC Chem. Biol.* 2, 1441–1461. doi:10.1039/d1cb00045d

Wan, W., Tharp, J. M., and Liu, W. R. (2014). Pyrrolysyl-tRNA Synthetase: an Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1844, 1059–1070. doi:10.1016/j.bbapap.2014.03.002

Wang, L., Amphlett, G., Blättler, W. A., Lambert, J. M., and Zhang, W. (2005). Structural Characterization of the Maytansinoid-Monoclonal Antibody Immunoconjugate, huN901-DM1, by Mass Spectrometry. *Protein Sci.* 14, 2436–2446. doi:10.1110/ps.051478705

Wang, X., Kawabe, Y., Kato, R., Hada, T., Ito, A., Yamana, Y., et al. (2017). Accumulative scFv-Fc Antibody Gene Integration into the Hprt Chromosomal Locus of Chinese Hamster Ovary Cells. *J. Biosci. Bioeng.* 124, 583–590. doi:10.1016/j.jbiosc.2017.05.017

Wilson, L. O. W., O'Brien, A. R., and Bauer, D. C. (2018). The Current State and Future of CRISPR-Cas9 gRNA Design Tools. *Front. Pharmacol.* 9, 749. doi:10.3389/fphar.2018.00749

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-Azidobenzyloxycarbonyl) Lysine for Site-specific Protein Modification. *Chem. Biol.* 15, 1187–1197. doi:10.1016/j.chembiol.2008.10.004

Zemella, A., Grossmann, S., Sachse, R., Sonnabend, A., Schaefer, M., and Kubick, S. (2017). Qualifying a Eukaryotic Cell-free System for Fluorescence Based GPCR Analyses. *Sci. Rep.* 7, 3740. doi:10.1038/s41598-017-03955-8

Zemella, A., Richter, T., Thoring, L., and Kubick, S. (20191947). A Combined Cell-free Protein Synthesis and Fluorescence-Based Approach to Investigate GPCR Binding Properties. *Methods Mol. Biol.*, 57–77. doi:10.1007/978-1-4939-9121-1_4

Zhang, L., Inniss, M. C., Han, S., Moffat, M., Jones, H., Zhang, B., et al. (2015). Recombinase-mediated Cassette Exchange (RMCE) for Monoclonal Antibody Expression in the Commercially Relevant CHOK1SV Cell Line. *Biotechnol. Prog.* 31, 1645–1656. doi:10.1002/btpr.2175

Zhao, M., Wang, J., Luo, M., Luo, H., Zhao, M., Han, L., et al. (2018). Rapid Development of Stable Transgene CHO Cell Lines by CRISPR/Cas9-mediated Site-specific Integration into C12orf35. *Appl. Microbiol. Biotechnol.* 102, 6105–6117. doi:10.1007/s00253-018-9021-6

Zimmerman, E. S., Heibeck, T. H., Gill, A., Li, X., Murray, C. J., Madlansacay, M. R., et al. (2014). Production of Site-specific Antibody-Drug Conjugates Using Optimized Non-natural Amino Acids in a Cell-free Expression System. *Bioconjug. Chem.* 25, 351–361. doi:10.1021/bc400490z

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

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

frontiers | Frontiers in Molecular Biosciences

REVIEW
published: 24 May 2022
doi: 10.3389/fmolb.2022.851646

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

Michiko Kimoto and Ichiro Hirao**

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

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

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

OPEN ACCESS

Edited by:
Nediljko Budisa,
University of Manitoba, Canada

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

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

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

Received: *10 January 2022*
Accepted: *25 April 2022*
Published: *24 May 2022*

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

INTRODUCTION

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

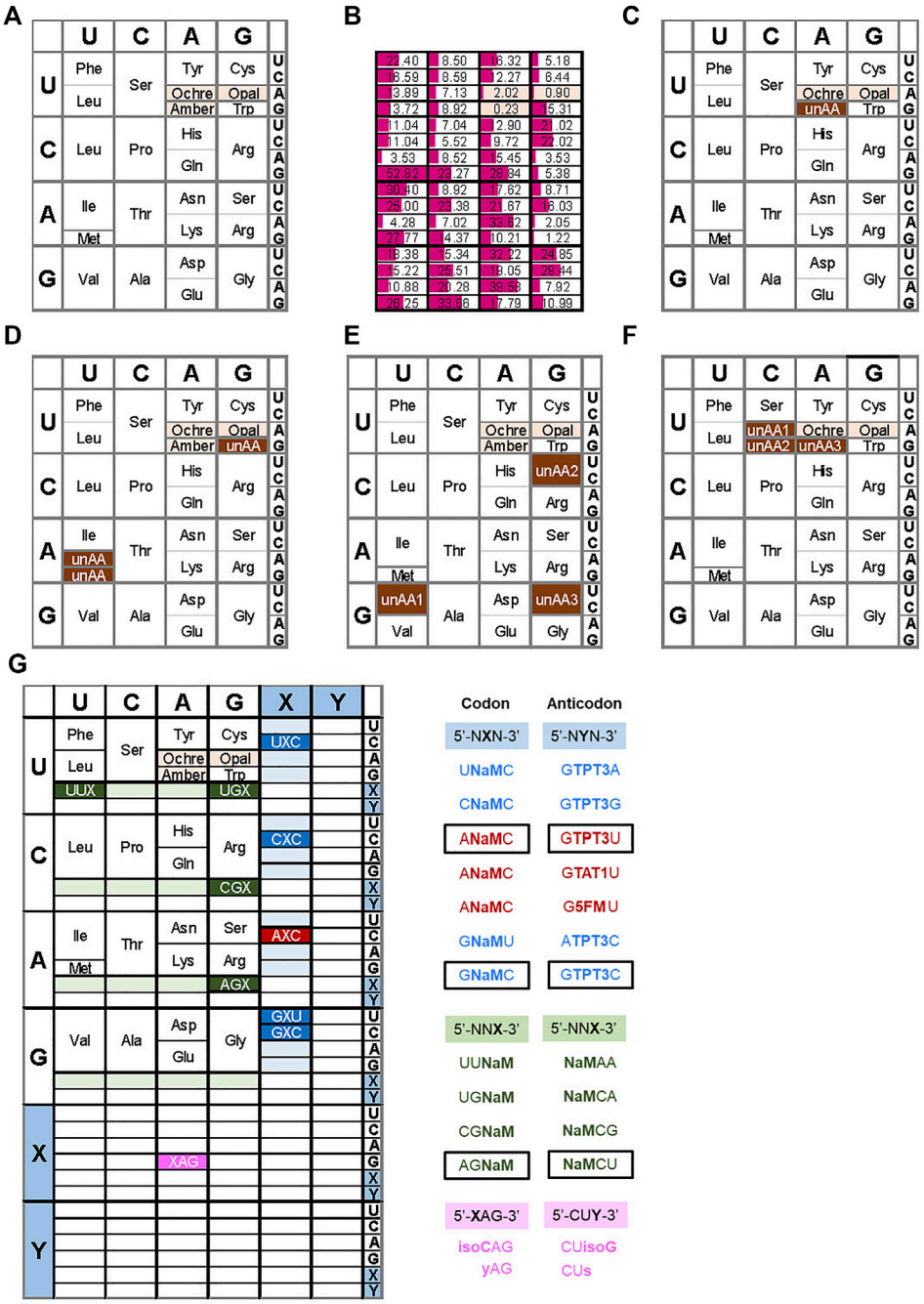


FIGURE 1 | Examples of expanded genetic codon tables **(A)** The original genetic codon table **(B)** Relative frequencies of codon usage in *E. coli* **(C)** Reassignment of the amber codon to an unnatural amino acid (unAA) **(D)** Use of sense codons for unAA, such as Trp codon (UGG) for an unAA (i.e., 4-fluorotryptophan) **(E)** Example of the reprogrammed genetic codon table by Suga's RAPID method **(F)** Reprogrammed genetic codon table in the engineered *E. coli* (Syn61) **(G)** Expanded genetic codon table using UBPs. Examples of the reported UB codon-anticodons are shown on the right.

The artificial incorporation of unAAs into proteins through the natural base pair (NBP) or unnatural base pair (UBP) systems requires an orthogonal "bypassing" system for the pre-existing genetic information flow in the central dogma: replication, transcription, and translation. Living organisms have evolved mechanisms to avoid the misincorporation (non-cognate) events of UBs and unAAs and remove these extra components as errors during nucleic acid and protein biosynthesis. Accordingly, in nature, most of the unnatural components in biopolymers are introduced by post-biosynthesis modifications or other biosynthetic mechanisms. To circumvent the proofreading systems of living organisms, researchers have created several

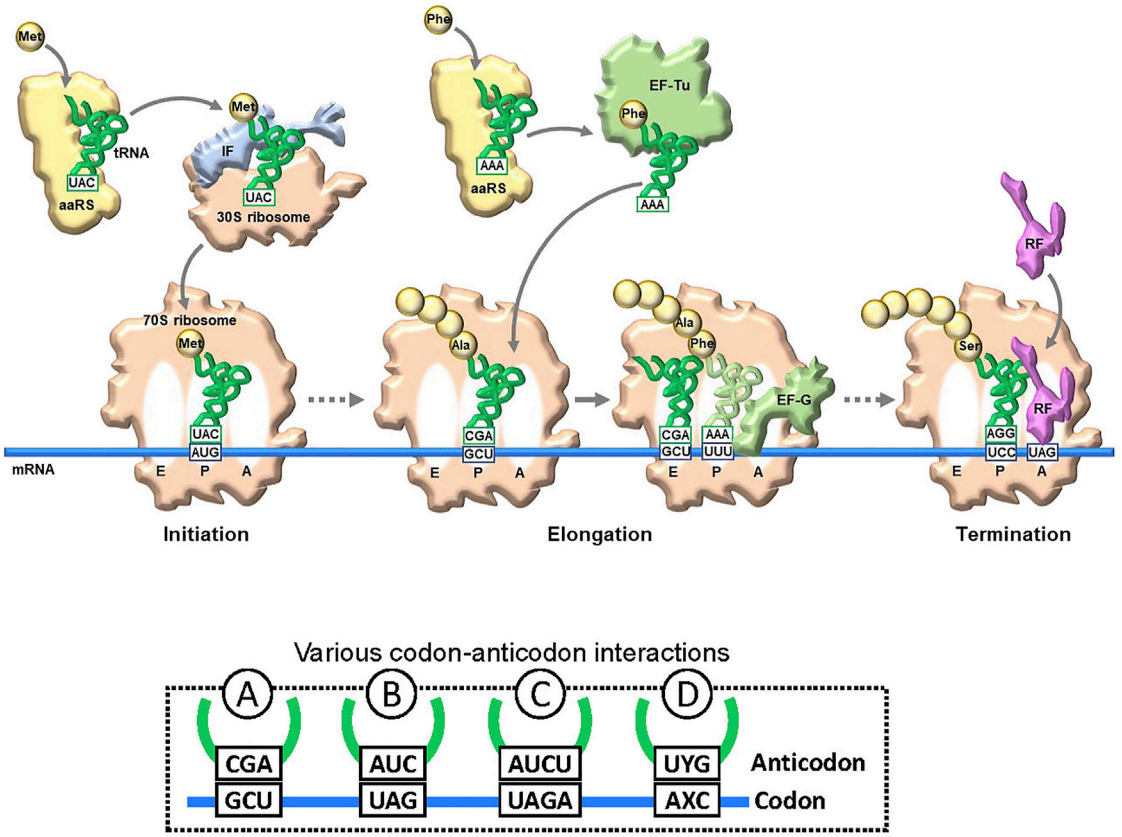


FIGURE 2 | Translation and codon-anticodon interactions. Simplified illustration of translation flow from initiation to termination. The representative key components are schematically illustrated. aaRS: aminoacyl-tRNA synthetase. tRNA and mRNA are shown in blue and green lines, respectively. To decode a specific amino acid, a new codon-anticodon is required. Examples of four different codon-anticodon interactions for (re)assignment of an unnatural amino acid (unAA). A: usual natural codon; B: stop codon (UAG); C: four-base codon (quadruplet codon); D: unnatural-base codon.

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

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

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

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

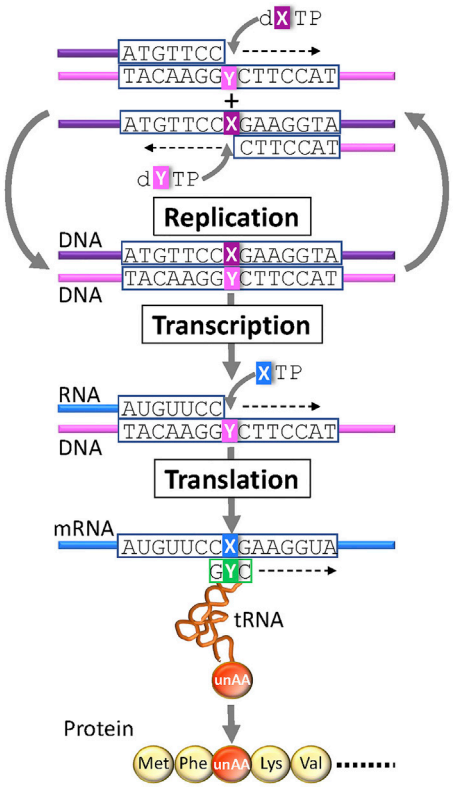


FIGURE 3 | Genetic alphabet expansion using an unnatural base pair (UBP) system for genetic code expansion. A third base pair (X–Y) that functions in replication, transcription, and translation, together with the natural A–T(U) and G–C base pairs, enables the site-specific incorporation of unnatural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

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

GENETIC CODE EXPANSION USING THE NBP SYSTEM

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

Use of stop codons: The most common method for site-specific unAA incorporation is the use of stop codons. As a specific case, archaea and eukaryotes also use stop codons for the incorporation of selenocysteine and pyrrolysine into proteins, using suppressor tRNAs. There are three stop codons, amber (UAG), ochre (UAA), and opal (UGA). Among them, amber is the most popular codon for unAA incorporation (**Figure 1C** and type B in **Figure 2**), because it has the lowest frequency as a stop codon in *Escherichia coli*. The stop codon usage in *E. coli* K12 is 7% UAG, 64% UAA, and 29% UGA (**Figure 1B**) (Openwetware, 2012). Interestingly, the amber codon usage is also low in other organisms, but it is especially low in *E. coli* (Belin and Puigbo, 2022).

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

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

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

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

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

In subsequent experiments, they found that GGGU exhibits the most efficient translation efficiency among the four-base codon contexts (Hohsaka and Sisido, 2002). Using two four-base codons, AGGU and CGGG, they succeeded in the site-specific incorporations of two unAAs into streptavidin (Hohsaka et al., 1999). Hohsaka's team achieved the site-specific labeling of proteins by using dye-conjugated amino acids as unAAs by the four-base codon system (Abe et al., 2010). Schultz's team comprehensively examined the translation efficiency of the four-codon system and found the best four-codon contexts, AGGA, UAGA, CCCU, and CUAG (Magliery et al., 2001).

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

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

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

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

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

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

In the area of codon reprogramming, Chin's team established another GRO system by the total synthesis of the *E. coli* genome with defined synonymous codon compression (Wang et al., 2016; Fredens et al., 2019). They designed and synthesized the 4-Mb *E. coli* genome, in which 18,214 codons for two serine codons (TCA and TCG) and the TAG amber codon in all of the genes were replaced with AGC, AGT, and TAA, respectively. Furthermore, the genes encoding $tRNA^{Ser}_{UGA}$, $tRNA^{Ser}_{CGA}$, and RF1 were also removed from the genome. Therefore, the synthesized *E. coli* (Syn61) uses 61 codons, and the three vacant codons can be used for three unAA codons (**Figure 1F**).

GENETIC CODE EXPANSION USING UBP SYSTEMS

Development of UBP systems *in vitro*: In 1962, Alexander Rich proposed the potential use of a UBP, isoguanine (isoG) and isocytosine (isoC) (**Figure 4**) with different hydrogen-bonding patterns from those of G–C, for a new codon-anticodon system (Rich, 1962). Even though the codon table was still being deciphered at that time, he imagined that two-base genetic codons, instead of three-base codons, using six-letter genetic alphabets could cover the 20 standard amino acids ($20 < 6 \times 6$). Over 2 decades later, in the late 1980s, Benner's team designed several UBPs with alternative hydrogen bonding patterns, including the isoG–isoC pair, and chemically synthesized these UB units. Their biological results opened a new world in which the UBPs could be used for replication and transcription, with orthogonal base pairing to the two natural base pairs (Switzer et al., 1989; Piccirilli et al., 1990). In 1992, they reported an *in vitro* translation system for the site-specific incorporation of 3-iodotyrosine into a peptide, using chemically synthesized mRNA with an (isoC)AG codon and tRNA with a CU(isoG) anticodon (**Figure 1G**) (Bain et al., 1992). Their efforts toward further UBP development and optimization led to the replicable and transcribable P–Z pair, with higher fidelity than those of the isoG–isoC pair (**Figure 4**) (Yang et al., 2010; Yang et al., 2011).

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

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

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

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

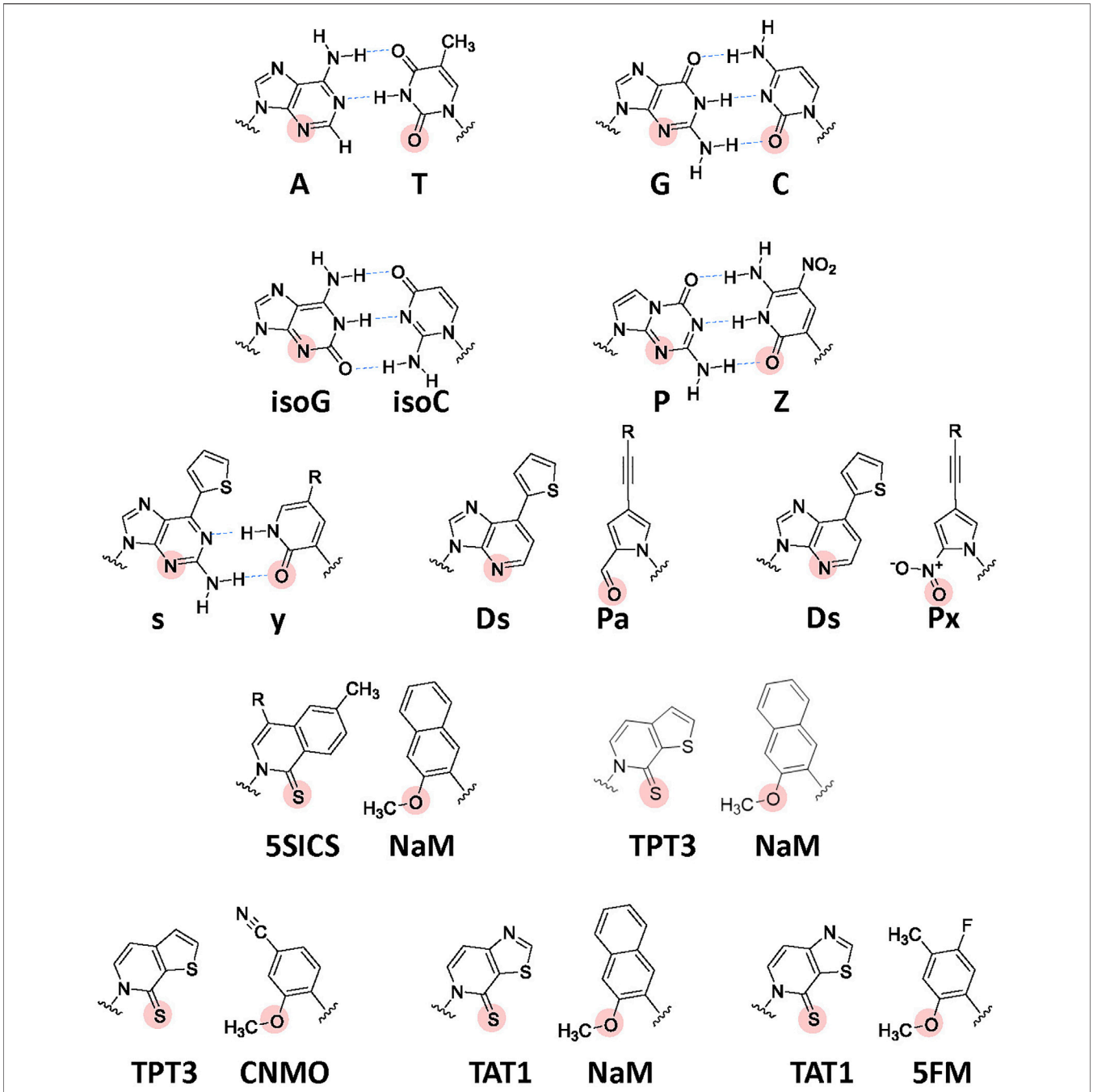


FIGURE 4 | Chemical structures of the natural Watson–Crick base pairs and a variety of UBPs developed to expand the genetic alphabet. Hydrogen-bonding interactions between the cognate base pairs are shown by blue arrows. The important residues (hydrogen acceptors) recognized by DNA and RNA polymerases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

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

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

organism, SSO), which replicates six-letter DNA, using their NaM–5SCIS/TPT3 pairs (Malyshev et al., 2014). To supply the UB substrates for the living cells, they employed media supplemented with UB-nucleoside triphosphates, dNaMTP and d5SCISTP. To facilitate the import of sufficient amounts of the UB substrates within the cells, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (*Pt*NTT2) was expressed in *E. coli* [C41 (DE3) strain]. They prepared a plasmid DNA containing the NaM–TPT3 pair through PCR amplification and transformed it into the engineered *E. coli* expressing *Pt*NTT2. An analysis of the cultured cells revealed the successful replication of the six-letter plasmid DNA with reasonable retention of the NaM–5SICS pair. These results also demonstrated that the UBP was not extensively rejected as a foreign component by the repair system.

They further improved this first generation SSO by switching from the original *E. coli* strain to the BL21 (DE3) strain, which is more suitable for protein expression. The *Pt*NTT2 expression was also optimized because the extremely high expression using T7 RNA polymerase inhibited cell growth. They modified *Pt*NTT2 expression by 1) removing the N-terminal signal peptide (65 aa) of *Pt*NTT2, which is toxic to cell growth, 2) using the optimized codon usage, 3) choosing the best RNA polymerase II promoter sequence, and 4) encoding the truncated *Pt*NTT2 gene within a *lacZYA* locus in the genome, to avoid expression plasmid copy number variations. The resultant second-generation SSO, called the YZ3 strain, greatly increased the retention rates of their UBPs in replicated DNA within various sequence contexts (Zhang et al., 2017a).

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

To evaluate the decoding of the new codon-anticodon interactions involving the NaM–TPT3 pair, they first investigated the incorporation of serine into GFP at position 151 (the TAC codon was replaced by the unnatural codon AXC) through *E. coli* $tRNA^{Ser}$ (*ser*T), where the anticodon was replaced by the unnatural codon GYT. This system can eliminate complicated situations related to unAAs because *E. coli* serine aaRS does not recognize the anticodon for tRNA amino acylation (Shimizu et al., 1992). The efficient full-length sfGFP production with 98.5 ± 0.7% incorporation of serine at position 151 was confirmed in the cells transformed with the plasmid encoding both sfGFP(AXC)^{151} and $tRNA^{Ser}_{GYT}$, cultured in the presence of deoxy- and ribo-nucleoside triphosphates (dNaMTP, dTPT3TP, NaMTP, and TPT3TP). After the validation of the unnatural codon-anticodon interactions, they focused on unAA N^6-[(2-propynyloxy) carbonyl]-L-lysine (PrK) incorporation in sfGFP(AXC)^{151} or sfGFP(GXC)^{151}, utilizing a pair of the *Methanosarcina mazei* $tRNA^{Pyl}_{GYU}$ or RNA^{Pyl}_{GYC} and the *Methanosarcina barkei* pyrrolysyl-tRNA synthetase (PylRS) (Nguyen et al., 2009; Chatterjee et al., 2013). The PylRS was encoded in a separate plasmid with expression controlled by IPTG induction. For another unAA p-azido-phenylalanine (*p*AzF) in sfGFP(AXC)^{151}, they utilized an evolved *Methanococcus jannaschii* TyrRS/$tRNA^{Tyr}$ pair (i.e. *p*AzFRS.$tRNA^{pAzF}_{GYU}$) (Chin et al., 2002b).

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

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

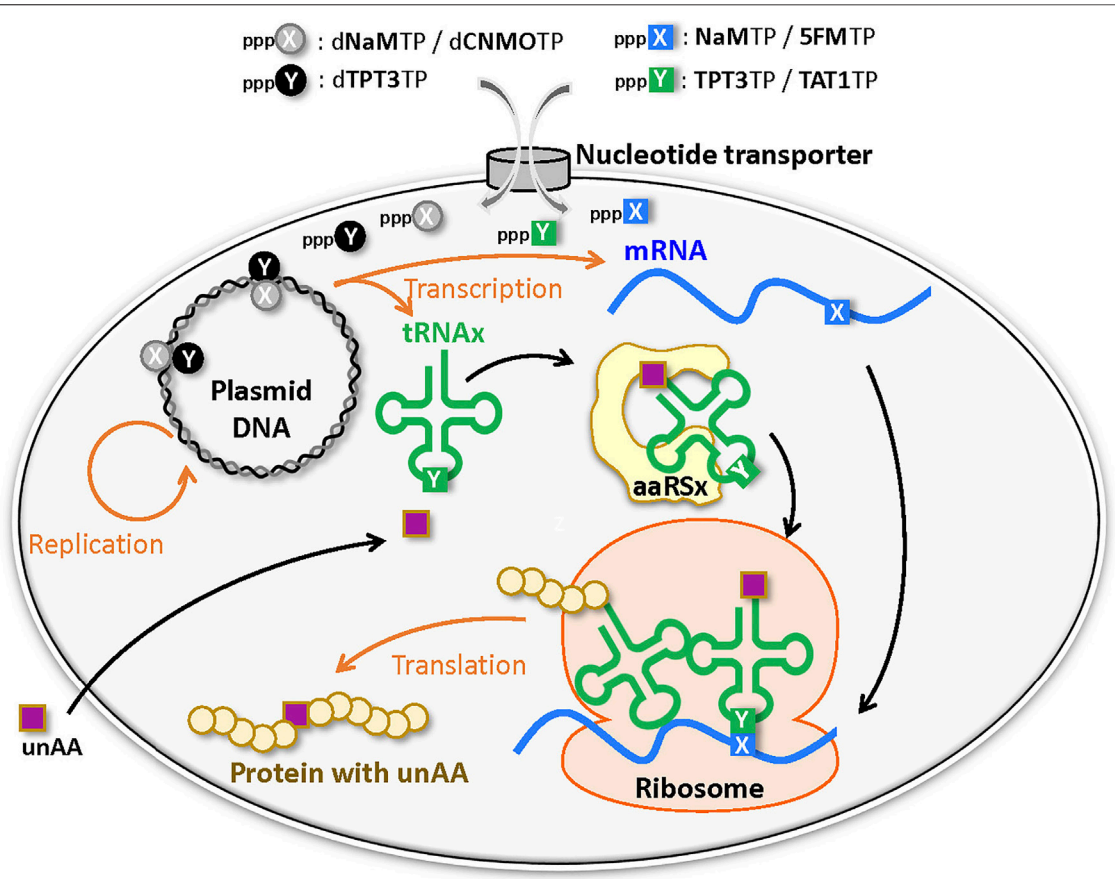


FIGURE 5 | Semi-synthetic organism (SSO) that stores and retrieves the six-letter genetic information expanded by an unnatural base pair (X–Y) system.

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

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

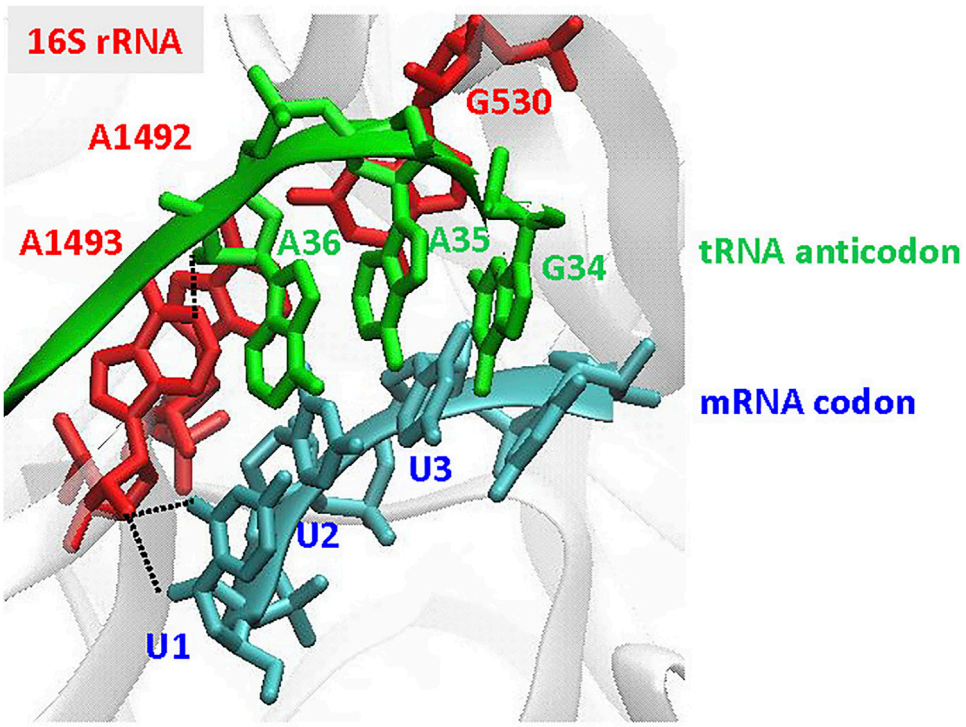


FIGURE 6 | Recognition of the codon-anticodon interaction at the ribosome. The illustration is focused on the decoding site of the 30S subunit, showing the A-site codon (UUU in blue) and the tRNA anticodon (GAA in green), using the coordinates in PDB: 1IBM (Ogle et al., 2001). The important nucleotides,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

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

THERAPEUTIC APPLICATIONS OF ENGINEERED PROTEINS BY GENETIC CODE EXPANSION

These translation systems by genetic code expansion have facilitated the rational design and optimization of proteins suitable for therapeutic applications, by improving the biological functions and pharmacokinetics of biologics in a manner resembling the pursuit of small-molecule therapeutics. Several macrocyclic peptides containing unAAs, including those generated by the ribosomal translation system, are now undergoing clinical tests (Vinogradov et al., 2019). Currently, several engineered proteins generated by these technologies are in pre-clinical and clinical trials as protein therapeutics. Representatives are, but not limited to, PEGylated interleukin-2 (SAR44425, THOR-707) (Manandhar et al., 2021), PEGylated fibroblast growth factor 21 (BMS-986036, pegbelfermin), Anti-HER2 antibody-drug conjugate (a site-specific Herceptin-monomethyl auristatin D (MMAD) conjugate, ARX788) (Sun et al., 2014), and anti-CD3 Folate Bi-Specific (Sun et al., 2014).

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

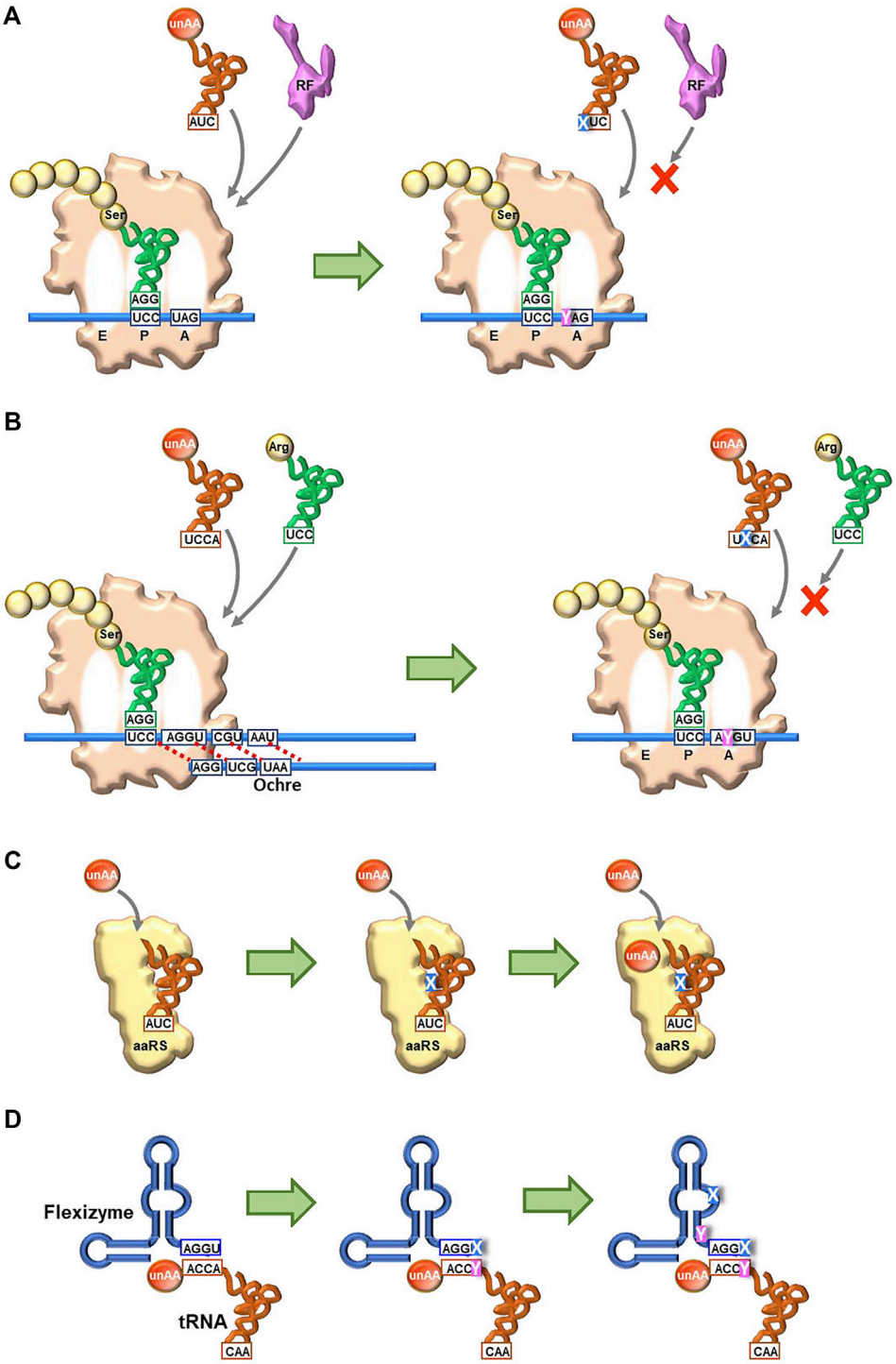


FIGURE 7 | Future perspectives to create new-anticodon interactions using UBP systems for unAA incorporation **(A)** Use of a stop codon (UAG) **(B)** Usage of a four-base codon (quadruplet codon) **(C)** Creation of new orthogonal engineered tRNA and aaRS pairs for specific unAA aminoacylation, using unnatural nucleotides and unAAs **(D)** Creation of new aminoacylation systems using engineered Flexizymes containing unnatural nucleotides.

mPEG, which retained the binding ability to IL-2 Rβγ but lacked that to the undesired IL-2 Rαβγ, and showed an extended half-life (Manandhar et al., 2021; Ptacin et al., 2021; Romesberg, 2021). THOR-707 is currently in a phase I/II study, not only as a monotherapeutic, but also in combination with a checkpoint inhibitor (pembrolizumab or cemiplimab) (Romesberg, 2021).

CONCLUSION AND FUTURE PERSPECTIVE

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

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

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

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

AUTHOR CONTRIBUTIONS

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

FUNDING

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

REFERENCES

Abe, R., Shiraga, K., Ebisu, S., Takagi, H., and Hohsaka, T. (2010). Incorporation of Fluorescent Non-natural Amino Acids into N-Terminal Tag of Proteins in Cell-free Translation and its Dependence on Position and Neighboring Codons. *J. Biosci. Bioeng.* 110, 32–38. doi:10.1016/j.jbiosc.2010.01.003

Agostini, F., Völler, J. S., Koksch, B., Acevedo-Rocha, C. G., Kubyshkin, V., and Budisa, N. (2017). Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chem. Int. Ed.* 56, 9680–9703. doi:10.1002/anie.201610129

Agris, P. F., Vendeix, F. A. P., and Graham, W. D. (2007). tRNA's Wobble Decoding of the Genome: 40 Years of Modification. *J. Mol. Biol.* 366, 1–13. doi:10.1016/j.jmb.2006.11.046

Ambrogelly, A., Palioura, S., and Söll, D. (2007). Natural Expansion of the Genetic Code. *Nat. Chem. Biol.* 3, 29–35. doi:10.1038/nchembio847

Bain, J. D., Diala, E. S., Glabe, C. G., Dix, T. A., and Chamberlin, A. R. (1989). Biosynthetic Site-specific Incorporation of a Non-natural Amino Acid into a Polypeptide. *J. Am. Chem. Soc.* 111, 8013–8014. doi:10.1021/ja00202a052

Bain, J. D., Switzer, C., Chamberlin, R., and Benner, S. A. (1992). Ribosome-mediated Incorporation of a Non-standard Amino Acid into a Peptide through Expansion of the Genetic Code. *Nature* 356, 537–539. doi:10.1038/356537a0

Belin, D., and Puigbò, P. (2022). Why Is the UAG (Amber) Stop Codon Almost Absent in Highly Expressed Bacterial Genes? *Life (Basel)* 12, 431. doi:10.3390/life12030431

Benner, S. A., Karalkar, N. B., Hoshika, S., Laos, R., Shaw, R. W., Matsuura, M., et al. (2016). Alternative Watson-Crick Synthetic Genetic Systems. *Cold Spring Harb. Perspect. Biol.* 8, a023770. doi:10.1101/cshperspect.a023770

Betz, K., Malyshev, D. A., Lavergne, T., Welte, W., Diederichs, K., Dwyer, T. J., et al. (2012). KlenTaq Polymerase Replicates Unnatural Base Pairs by Inducing a Watson-Crick Geometry. *Nat. Chem. Biol.* 8, 612–614. doi:10.1038/nchembio.966

Betz, K., Malyshev, D. A., Lavergne, T., Welte, W., Diederichs, K., Romesberg, F. E., et al. (2013). Structural Insights into DNA Replication without Hydrogen Bonds. *J. Am. Chem. Soc.* 135, 18637–18643. doi:10.1021/ja409609j

Bohlke, N., and Budisa, N. (2014). Sense Codon Emancipation for Proteome-wide Incorporation of Noncanonical Amino Acids: Rare Isoleucine Codon AUA as a Target for Genetic Code Expansion. *FEMS Microbiol. Lett.* 351, 133–144. doi:10.1111/1574-6968.12371

Brustad, E., Bushey, M. L., Brock, A., Chittuluru, J., and Schultz, P. G. (2008). A Promiscuous Aminoacyl-tRNA Synthetase that Incorporates Cysteine, Methionine, and Alanine Homologs into Proteins. *Bioorg. Med. Chem. Lett.* 18, 6004–6006. doi:10.1016/j.bmcl.2008.09.050

Budisa, N. (2013). Expanded Genetic Code for the Engineering of Ribosomally Synthetized and Post-translationally Modified Peptide Natural Products (RiPPs). *Curr. Opin. Biotechnol.* 24, 591–598. doi:10.1016/j.copbio.2013.02.026

Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013). A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. *Biochemistry* 52, 1828–1837. doi:10.1021/bi4000244

Chatterjee, A., Xiao, H., and Schultz, P. G. (2012). Evolution of Multiple, Mutually Orthogonal Prolyl-tRNA Synthetase/tRNA Pairs for Unnatural Amino Acid Mutagenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14841–14846. doi:10.1073/pnas.1212454109

Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003). An Expanded Eukaryotic Genetic Code. *Science* 301, 964–967. doi:10.1126/science.1084772

Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002a). Addition of a Photocrosslinking Amino Acid to the Genetic Code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11020–11024. doi:10.1073/pnas.172226299

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002b). Addition of P-Azido-L-Phenylalanine to the Genetic Code of Escherichia coli. *J. Am. Chem. Soc.* 124, 9026–9027. doi:10.1021/ja027007w

Chow, C. M., and Rajbhandary, U. L. (1993). Saccharomyces cerevisiae Cytoplasmic Tyrosyl-tRNA Synthetase Gene. Isolation by Complementation of a Mutant Escherichia coli Suppressor tRNA Defective in Aminoacylation and Sequence Analysis. *J. Biol. Chem.* 268, 12855–12863. doi:10.1016/s0021-9258(18)31466-2

Chung, C. Z., Amikura, K., and Söll, D. (2020). Using Genetic Code Expansion for Protein Biochemical Studies. *Front. Bioeng. Biotechnol.* 8, 598577. doi:10.3389/fbioe.2020.598577

Cochella, L., Brunelle, J. L., and Green, R. (2007). Mutational Analysis Reveals Two Independent Molecular Requirements during Transfer RNA Selection on the Ribosome. *Nat. Struct. Mol. Biol.* 14, 30–36. doi:10.1038/nsmb1183

Cowie, D. B., and Cohen, G. N. (1957). Biosynthesis by *Escherichia coli* of Active Altered Proteins Containing Selenium Instead of Sulfur. *Biochimica Biophysica Acta* 26, 252–261. doi:10.1016/0006-3002(57)90003-3

De La Torre, D., and Chin, J. W. (2021). Reprogramming the Genetic Code. *Nat. Rev. Genet.* 22, 169–184. doi:10.1038/s41576-020-00307-7

De Simone, A., Acevedo-Rocha, C. G., Hoesl, M. G., and Budisa, N. (2016). Towards Reasignment of the Methionine Codon AUG to Two Different Noncanonical Amino Acids in Bacterial Translation. *Croat. Chem. Acta* 89, 243–253. doi:10.5562/cca2915

Ding, W., Zhao, H., Chen, Y., Zhang, B., Yang, Y., Zang, J., et al. (2020). Chimeric Design of Pyrrolysyl-tRNA Synthetase/tRNA Pairs and Canonical Synthetase/tRNA Pairs for Genetic Code Expansion. *Nat. Commun.* 11, 3154. doi:10.1038/s41467-020-16898-y

Feldman, A. W., Dien, V. T., Karadeema, R. J., Fischer, E. C., You, Y., Anderson, B. A., et al. (2019). Optimization of Replication, Transcription, and Translation in a Semi-synthetic Organism. *J. Am. Chem. Soc.* 141, 10644–10653. doi:10.1021/jacs.9b02075

Fischer, E. C., Hashimoto, K., Zhang, Y., Feldman, A. W., Dien, V. T., Karadeema, R. J., et al. (2020). New Codons for Efficient Production of Unnatural Proteins in a Semisynthetic Organism. *Nat. Chem. Biol.* 16, 570–576. doi:10.1038/s41589-020-0507-z

Fredens, J., Wang, K., De La Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., et al. (2019). Total Synthesis of *Escherichia coli* with a Recoded Genome. *Nature* 569, 514–518. doi:10.1038/s41586-019-1192-5

Fujiwara, T., Kimoto, M., Sugiyama, H., Hirao, I., and Yokoyama, S. (2001). Synthesis of 6-(2-thienyl)purine Nucleoside Derivatives that Form Unnatural Base Pairs with Pyridin-2-One Nucleosides. *Bioorg. Med. Chem. Lett.* 11, 2221–2223. doi:10.1016/s0960-894x(01)00415-2

Futami, K., Kimoto, M., Lim, Y. W. S., and Hirao, I. (2019). Genetic Alphabet Expansion Provides Versatile Specificities and Activities of Unnatural-Base DNA Aptamers Targeting Cancer Cells. *Mol. Ther. Nucleic Acids* 14, 158–170. doi:10.1016/j.omtn.2018.11.011

Goto, Y., and Suga, H. (2009). Translation Initiation with Initiator tRNA Charged with Exotic Peptides. *J. Am. Chem. Soc.* 131, 5040–5041. doi:10.1021/ja900597d

Hashimoto, K., Fischer, E. C., and Romesberg, F. E. (2021). Efforts toward Further Integration of an Unnatural Base Pair into the Biology of a Semisynthetic Organism. *J. Am. Chem. Soc.* 143, 8603–8607. doi:10.1021/jacs.1c03860

Heck, S. D., Siok, C. J., Krapcho, K. J., Kelbaugh, P. R., Thadeio, P. F., Welch, M. J., et al. (1994). Functional Consequences of Posttranslational Isomerization of Ser 46 in a Calcium Channel Toxin. *Science* 266, 1065–1068. doi:10.1126/science.7973665

Heckler, T. G., Chang, L. H., Zama, Y., Naka, T., Chorghade, M. S., and Hecht, S. M. (1984). T4 RNA Ligase Mediated Preparation of Novel "chemically Misacylated" tRNAPheS. *Biochemistry* 23, 1468–1473. doi:10.1021/bi00302a020

Hirao, I., Kimoto, M., Mitsui, T., Fujiwara, T., Kawai, R., Sato, A., et al. (2006). An Unnatural Hydrophobic Base Pair System: Site-specific Incorporation of Nucleotide Analogs into DNA and RNA. *Nat. Methods* 3, 729–735. doi:10.1038/nmeth915

Hirao, I., and Kimoto, M. (2012). Unnatural Base Pair Systems toward the Expansion of the Genetic Alphabet in the Central Dogma. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 88, 345–367. doi:10.2183/pjab.88.345

Hirao, I., Mitsui, T., Kimoto, M., and Yokoyama, S. (2007). An Efficient Unnatural Base Pair for PCR Amplification. *J. Am. Chem. Soc.* 129, 15549–15555. doi:10.1021/ja073830m

Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., et al. (2002). An Unnatural Base Pair for Incorporating Amino Acid Analogs into Proteins. *Nat. Biotechnol.* 20, 177–182. doi:10.1038/nbt0202-177

Hoernes, T. P., Faserl, K., Juen, M. A., Kremser, J., Gasser, C., Fuchs, E., et al. (2018). Translation of Non-standard Codon Nucleotides Reveals Minimal Requirements for Codon-Anticodon Interactions. *Nat. Commun.* 9, 4865. doi:10.1038/s41467-018-07321-8

Hohsaka, T., Ashizuka, Y., Murakami, H., and Sisido, M. (1996). Incorporation of Nonnatural Amino Acids into Streptavidin through *In Vitro* Frame-Shift Suppression. *J. Am. Chem. Soc.* 118, 9778–9779. doi:10.1021/ja9614225

Hohsaka, T., Ashizuka, Y., Sasaki, H., Murakami, H., and Sisido, M. (1999). Incorporation of Two Different Nonnatural Amino Acids Independently into a Single Protein through Extension of the Genetic Code. *J. Am. Chem. Soc.* 121, 12194–12195. doi:10.1021/ja992204p

Hohsaka, T., and Sisido, M. (2002). Incorporation of Non-natural Amino Acids into Proteins. *Curr. Opin. Chem. Biol.* 6, 809–815. doi:10.1016/s1367-5931(02)00376-9

Italia, J. S., Addy, P. S., Wrobel, C. J. J., Crawford, L. A., Lajoie, M. J., Zheng, Y., et al. (2017). An Orthogonalized Platform for Genetic Code Expansion in Both Bacteria and Eukaryotes. *Nat. Chem. Biol.* 13, 446–450. doi:10.1038/nchembio.2312

Iwane, Y., Hitomi, A., Murakami, H., Katoh, T., Goto, Y., and Suga, H. (2016). Expanding the Amino Acid Repertoire of Ribosomal Polypeptide Synthesis via the Artificial Division of Codon Boxes. *Nat. Chem.* 8, 317–325. doi:10.1038/nchem.2446

Iwane, Y., Kimura, H., Katoh, T., and Suga, H. (2021). Uniform Affinity-Tuning of N-Methyl-Aminoacyl-tRNAs to EF-Tu Enhances Their Multiple Incorporation. *Nucleic Acids Res.* 49, 10807–10817. doi:10.1093/nar/gkab288

Katoh, T., Sengoku, T., Hirata, K., Ogata, K., and Suga, H. (2020). Ribosomal Synthesis and De Novo Discovery of Bioactive Foldamer Peptides Containing Cyclic β-amino Acids. *Nat. Chem.* 12, 1081–1088. doi:10.1038/s41557-020-0525-1

Katoh, T., and Suga, H. (2022). *In Vitro* Selection of Foldamer-like Macrocyclic Peptides Containing 2-Aminobenzoic Acid and 3-Aminothiophene-2-Carboxylic Acid. *J. Am. Chem. Soc.* 144, 2069–2072. doi:10.1021/jacs.1c12133

Kimoto, M., and Hirao, I. (2020). Genetic Alphabet Expansion Technology by Creating Unnatural Base Pairs. *Chem. Soc. Rev.* 49, 7602–7626. doi:10.1039/d0cs00457j

Kimoto, M., Kawai, R., Mitsui, T., Yokoyama, S., and Hirao, I. (2009). An Unnatural Base Pair System for Efficient PCR Amplification and Functionalization of DNA Molecules. *Nucleic Acids Res.* 37, e14. e14. doi:10.1093/nar/gkn956

Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S., and Hirao, I. (2013). Generation of High-Affinity DNA Aptamers Using an Expanded Genetic Alphabet. *Nat. Biotechnol.* 31, 453–457. doi:10.1038/nbt.2556

Klapper, J. A., Downey, S. G., Smith, F. O., Yang, J. C., Hughes, M. S., Kammula, U. S., et al. (2008). High-dose Interleukin-2 for the Treatment of Metastatic Renal Cell Carcinoma. *Cancer* 113, 293–301. doi:10.1002/cncr.23552

Koh, C. S., and Sarin, L. P. (2018). Transfer RNA Modification and Infection - Implications for Pathogenicity and Host Responses. *Biochimica Biophysica Acta (BBA) - Gene Regul. Mech.* 1861, 419–432. doi:10.1016/j.bbagrm.2018.01.015

Kreil, G. (1994). Conversion of L- to D-Amino Acids: a Posttranslational Reaction. *Science* 266, 996–997. doi:10.1126/science.7973683

Kreil, G. (1997). D-amino Acids in Animal Peptides. *Annu. Rev. Biochem.* 66, 337–345. doi:10.1146/annurev.biochem.66.1.337

Krieg, C., Létourneau, S., Pantaleo, G., and Boyman, O. (2010). Improved IL-2 Immunotherapy by Selective Stimulation of IL-2 Receptors on Lymphocytes and Endothelial Cells. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11906–11911. doi:10.1073/pnas.1002569107

Kubyshkin, V., Acevedo-Rocha, C. G., and Budisa, N. (2018). On Universal Coding Events in Protein Biogenesis. *Biosystems* 164, 16–25. doi:10.1016/j.biosystems.2017.10.004

Lajoie, M. J., Kosuri, S., Mosberg, J. A., Gregg, C. J., Zhang, D., and Church, G. M. (2013a). Probing the Limits of Genetic Recoding in Essential Genes. *Science* 342, 361–363. doi:10.1126/science.1241460

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., et al. (2013b). Genomically Recoded Organisms Expand Biological Functions. *Science* 342, 357–360. doi:10.1126/science.1241459

Lajoie, M. J., Söll, D., and Church, G. M. (2016). Overcoming Challenges in Engineering the Genetic Code. *J. Mol. Biol.* 428, 1004–1021. doi:10.1016/j.jmb.2015.09.003

Lavergne, T., Degardin, M., Malyshev, D. A., Quach, H. T., Dhami, K., Ordoukhanian, P., et al. (2013). Expanding the Scope of Replicable Unnatural DNA: Stepwise Optimization of a Predominantly Hydrophobic Base Pair. *J. Am. Chem. Soc.* 135, 5408–5419. doi:10.1021/ja312148q

Ledbetter, M. P., Karadeema, R. J., and Romesberg, F. E. (2018). Reprograming the Replisome of a Semisynthetic Organism for the Expansion of the Genetic Alphabet. *J. Am. Chem. Soc.* 140, 758–765. doi:10.1021/jacs.7b11488

Lee, N., Bessho, Y., Wei, K., Szostak, J. W., and Suga, H. (2000). Ribozyme-catalyzed tRNA Aminoacylation. *Nat. Struct. Biol.* 7, 28–33. doi:10.1038/71225

Magliery, T. J., Anderson, J. C., and Schultz, P. G. (2001). Expanding the Genetic Code: Selection of Efficient Suppressors of Four-Base Codons and Identification of "shifty" Four-Base Codons with a Library Approach in Escherichia coli11Edited by M. Gottesman. *J. Mol. Biol.* 307, 755–769. doi:10.1006/jmbi.2001.4518

Malyshev, D. A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J. M., et al. (2014). A Semi-synthetic Organism with an Expanded Genetic Alphabet. *Nature* 509, 385–388. doi:10.1038/nature13314

Malyshev, D. A., Dhami, K., Quach, H. T., Lavergne, T., Ordoukhanian, P., Torkamani, A., et al. (2012). Efficient and Sequence-independent Replication of DNA Containing a Third Base Pair Establishes a Functional Six-Letter Genetic Alphabet. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12005–12010. doi:10.1073/pnas.1205176109

Malyshev, D. A., and Romesberg, F. E. (2015). The Expanded Genetic Alphabet. *Angew. Chem. Int. Ed.* 54, 11930–11944. doi:10.1002/anie.201502890

Malyshev, D. A., Seo, Y. J., Ordoukhanian, P., and Romesberg, F. E. (2009). PCR with an Expanded Genetic Alphabet. *J. Am. Chem. Soc.* 131, 14620–14621. doi:10.1021/ja906186f

Manandhar, M., Chun, E., and Romesberg, F. E. (2021). Genetic Code Expansion: Inception, Development, Commercialization. *J. Am. Chem. Soc.* 143, 4859–4878. doi:10.1021/jacs.0c11938

Matsunaga, K., Kimoto, M., and Hirao, I. (2017). High-affinity DNA aptamer generation targeting von Willebrand factor A1-domain by genetic alphabet expansion for systematic evolution of ligands by exponential enrichment using two types of libraries composed of five different bases. *J. Am. Chem. Soc.* 139, 324–334. doi:10.1021/jacs.6b10767

Matsunaga, K., Kimoto, M., Lim, V. W., Tan, H. P., Wong, Y. Q., Sun, W., et al. (2021). High-affinity Five/six-Letter DNA Aptamers with Superior Specificity Enabling the Detection of Dengue NS1 Protein Variants beyond the Serotype Identification. *Nucleic Acids Res.* 49, 11407–11424. doi:10.1093/nar/gkab515

Melnikov, S. V., and Söll, D. (2019). Aminoacyl-tRNA Synthetases and tRNAs for an Expanded Genetic Code: What Makes Them Orthogonal? *Int. J. Mol. Sci.* 20, 1929. doi:10.3390/ijms20081929

Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., et al. (2010). Codon Reassignment in the *Escherichia coli* Genetic Code. *Nucleic Acids Res.* 38, 8188–8195. doi:10.1093/nar/gkq707

Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., et al. (2015a). Highly Reproductive *Escherichia coli* Cells with No Specific Assignment to the UAG Codon. *Sci. Rep.* 5, 9699. doi:10.1038/srep09699

Mukai, T., Lajoie, M. J., Englert, M., and Söll, D. (2017). Rewriting the Genetic Code. *Annu. Rev. Microbiol.* 71, 557–577. doi:10.1146/annurev-micro-090816-093247

Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., et al. (2015b). Reassignment of a Rare Sense Codon to a Non-canonical Amino Acid inEscherichia Coli. *Nucleic Acids Res.* 43, 8111–8122. doi:10.1093/nar/gkv787

Murakami, H., Hohsaka, T., Ashizuka, Y., and Sisido, M. (1998). Site-directed Incorporation of *P*-Nitrophenylalanine into Streptavidin and Site-To-Site Photoinduced Electron Transfer from a Pyrenyl Group to a Nitrophenyl Group on the Protein Framework. *J. Am. Chem. Soc.* 120, 7520–7529. doi:10.1021/ja971890u

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010). Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome. *Nature* 464, 441–444. doi:10.1038/nature08817

Nguyen, D. P., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A., and Chin, J. W. (2009). Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-tRNA Synthetase/tRNACUA Pair and Click Chemistry. *J. Am. Chem. Soc.* 131, 8720–8721. doi:10.1021/ja900553w

Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001). RNA Tertiary Interactions in the Large Ribosomal Subunit: the A-Minor Motif. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4899–4903. doi:10.1073/pnas.081082398

Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989). A General Method for Site-specific Incorporation of Unnatural Amino Acids into Proteins. *Science* 244, 182–188. doi:10.1126/science.2649980

Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001). Recognition of Cognate Transfer RNA by the 30 S Ribosomal Subunit. *Science* 292, 897–902. doi:10.1126/science.1060612

Ogle, J. M., Murphy, F. V., Tarry, M. J., and Ramakrishnan, V. (2002). Selection of tRNA by the Ribosome Requires a Transition from an Open to a Closed Form. *Cell* 111, 721–732. doi:10.1016/s0092-8674(02)01086-3

Oh, J., Shin, J., Unarta, I. C., Wang, W., Feldman, A. W., Karadeema, R. J., et al. (2021). Transcriptional Processing of an Unnatural Base Pair by Eukaryotic RNA Polymerase II. *Nat. Chem. Biol.* 17, 906–914. doi:10.1038/s41589-021-00817-3

Ohtsuki, T., Kawai, G., Watanabe, Y., Kita, K., Nishikawa, K., and Watanabe, K. (1996). Preparation of Biologically Active Ascaris suum Mitochondrial tRNAMet with a TV-Replacement Loop by Ligation of Chemically Synthesized RNA Fragments. *Nucleic Acids Res.* 24, 662–667. doi:10.1093/nar/24.4.662

Ohuchi, M., Murakami, H., and Suga, H. (2007). The Flexizyme System: a Highly Flexible tRNA Aminoacylation Tool for the Translation Apparatus. *Curr. Opin. Chem. Biol.* 11, 537–542. doi:10.1016/j.cbpa.2007.08.011

Openwetware (2012). Escherichia coli/Codon Usage [Online]. Available at: https://openwetware.org/wiki/Escherichia_coli/Codon_usage (Accessed August 15, 2012).

Passioura, T., and Suga, H. (2014). Flexizymes, Their Evolutionary History and Diverse Utilities. *Top. Curr. Chem.* 344, 331–345. doi:10.1007/128_2013_421

Passioura, T., Katoh, T., Goto, Y., and Suga, H. (2014). Selection-based Discovery of Druglike Macrocyclic Peptides. *Annu. Rev. Biochem.* 83, 727–752. doi:10.1146/annurev-biochem-060713-035456

Piccirilli, J. A., Benner, S. A., Krauch, T., Moroney, S. E., and Benner, S. A. (1990). Enzymatic Incorporation of a New Base Pair into DNA and RNA Extends the Genetic Alphabet. *Nature* 343, 33–37. doi:10.1038/343033a0

Ptacin, J. L., Caffaro, C. E., Ma, L., San Jose Gall, K. M., Aerni, H. R., Acuff, N. V., et al. (2021). An Engineered IL-2 Reprogrammed for Anti-tumor Therapy Using a Semi-synthetic Organism. *Nat. Commun.* 12, 4785. doi:10.1038/s41467-021-24987-9

Ramakrishnan, V. (2002). Ribosome Structure and the Mechanism of Translation. *Cell* 108, 557–572. doi:10.1016/s0092-8674(02)00619-0

Rich, A. (1962). "Problems of Evolution and Biochemical Information Transfer," in *Horizons Biochem.* Editors M. Kasha and B. Pullman (Academic Press), 103–126.

Romesberg, F. E. (2022). Creation, Optimization, and Use of Semi-synthetic Organisms that Store and Retrieve Increased Genetic Information. *J. Mol. Biol.* 434, 167331. doi:10.1016/j.jmb.2021.167331

Saito, H., and Suga, H. (2001). A Ribozyme Exclusively Aminoacylates the 3'-hydroxyl Group of the tRNA Terminal Adenosine. *J. Am. Chem. Soc.* 123, 7178–7179. doi:10.1021/ja015756s

Seo, Y. J., Malyshev, D. A., Lavergne, T., Ordoukhanian, P., and Romesberg, F. E. (2011). Site-specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs. *J. Am. Chem. Soc.* 133, 19878–19888. doi:10.1021/ja207907d

Seo, Y. J., Matsuda, S., and Romesberg, F. E. (2009). Transcription of an Expanded Genetic Alphabet. *J. Am. Chem. Soc.* 131, 5046–5047. doi:10.1021/ja9006996

Shimizu, M., Asahara, H., Tamura, K., Hasegawa, T., and Himeno, H. (1992). The Role of Anticodon Bases and the Discriminator Nucleotide in the Recognition of Some E. coli tRNAs by Their Aminoacyl-tRNA Synthetases. *J. Mol. Evol.* 35, 436–443. doi:10.1007/BF00171822

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free Translation Reconstituted with Purified Components. *Nat. Biotechnol.* 19, 751–755. doi:10.1038/90802

Soma, A., Ikeuchi, Y., Kanemasa, S., Kobayashi, K., Ogasawara, N., Ote, T., et al. (2003). An RNA-Modifying Enzyme that Governs Both the Codon and Amino Acid Specificities of Isoleucine tRNA. *Mol. Cell* 12, 689–698. doi:10.1016/s1097-2765(03)00346-0

Steer, B. A., and Schimmel, P. (1999). Major Anticodon-Binding Region Missing from an Archaebacterial tRNA Synthetase. *J. Biol. Chem.* 274, 35601–35606. doi:10.1074/jbc.274.50.35601

Sun, S. B., Schultz, P. G., and Kim, C. H. (2014). Therapeutic Applications of an Expanded Genetic Code. *Chembiochem* 15, 1721–1729. doi:10.1002/cbic.201402154

Suzuki, T., and Miyauchi, K. (2010). Discovery and Characterization of tRNAIlelysidine Synthetase (TilS). *FEBS Lett.* 584, 272–277. doi:10.1016/j.febslet.2009.11.085

Switzer, C., Moroney, S. E., and Benner, S. A. (1989). Enzymatic Incorporation of a New Base Pair into DNA and RNA. *J. Am. Chem. Soc.* 111, 8322–8323. doi:10.1021/ja00203a067

Syed, J., Palani, S., Clarke, S. T., Asad, Z., Bottrill, A. R., Jones, A. M. E., et al. (2019). Expanding the Zebrafish Genetic Code through Site-specific Introduction of Azido-Lysine, Bicyclononyne-Lysine, and Diazirine-Lysine. *Int. J. Mol. Sci.* 20, 2577. doi:10.3390/ijms20102577

Tanrikulu, I. C., Schmitt, E., Mechulam, Y., Goddard, W. A., 3rd, and Tirrell, D. A. (2009). Discovery of Escherichia coli Methionyl-tRNA Synthetase Mutants for Efficient Labeling of Proteins with Azidonorleucine *In Vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15285–15290. doi:10.1073/pnas.0905735106

Torikai, K., and Suga, H. (2014). Ribosomal Synthesis of an Amphotericin-B Inspired Macrocycle. *J. Am. Chem. Soc.* 136, 17359–17361. doi:10.1021/ja508648s

Tsunoda, M., Kusakabe, Y., Tanaka, N., Ohno, S., Nakamura, M., Senda, T., et al. (2007). Structural Basis for Recognition of Cognate tRNA by Tyrosyl-tRNA Synthetase from Three Kingdoms. *Nucleic Acids Res.* 35, 4289–4300. doi:10.1093/nar/gkm417

Väre, V., Eruysal, E., Narendran, A., Sarachan, K., and Agris, P. (2017). Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. *Biomolecules* 7, 29. doi:10.3390/biom7010029

Vinogradov, A. A., Yin, Y., and Suga, H. (2019). Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *J. Am. Chem. Soc.* 141, 4167–4181. doi:10.1021/jacs.8b13178

Wang, K., Fredens, J., Brunner, S. F., Kim, S. H., Chia, T., and Chin, J. W. (2016). Defining Synonymous Codon Compression Schemes by Genome Recoding. *Nature* 539, 59–64. doi:10.1038/nature20124

Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001). Expanding the Genetic Code of *Escherichia coli*. *Science* 292, 498–500. doi:10.1126/science.1060077

Wang, L., and Schultz, P. G. (2001). A General Approach for the Generation of Orthogonal tRNAs. *Chem. Biol.* 8, 883–890. doi:10.1016/s1074-5521(01)00063-1

Wong, J. T. (1983). Membership Mutation of the Genetic Code: Loss of Fitness by Tryptophan. *Proc. Natl. Acad. Sci. U.S.A.* 80, 6303–6306. doi:10.1073/pnas.80.20.6303

Yamashige, R., Kimoto, M., Takezawa, Y., Sato, A., Mitsui, T., Yokoyama, S., et al. (2012). Highly Specific Unnatural Base Pair Systems as a Third Base Pair for PCR Amplification. *Nucleic Acids Res.* 40, 2793–2806. doi:10.1093/nar/gkr1068

Yang, Z., Chen, F., Alvarado, J. B., and Benner, S. A. (2011). Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J. Am. Chem. Soc.* 133, 15105–15112. doi:10.1021/ja204910n

Yang, Z., Chen, F., Chamberlin, S. G., and Benner, S. A. (2010). Expanded Genetic Alphabets in the Polymerase Chain Reaction. *Angew. Chem. Int. Ed.* 49, 177–180. doi:10.1002/anie.200905173

Young, D. D., and Schultz, P. G. (2018). Playing with the Molecules of Life. *ACS Chem. Biol.* 13, 854–870. doi:10.1021/acschembio.7b00974

Young, T. S., and Schultz, P. G. (2010). Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J. Biol. Chem.* 285, 11039–11044. doi:10.1074/jbc.r109.091306

Zeng, Y., Wang, W., and Liu, W. R. (2014). Towards Reassigning the Rare AGG Codon inEscherichia Coli. *Chembiochem* 15, 1750–1754. doi:10.1002/cbic.201400075

Zhang, Y., Lamb, B. M., Feldman, A. W., Zhou, A. X., Lavergne, T., Li, L., et al. (2017a). A Semisynthetic Organism Engineered for the Stable Expansion of the Genetic Alphabet. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1317–1322. doi:10.1073/pnas.1616443114

Zhang, Y., Ptacin, J. L., Fischer, E. C., Aerni, H. R., Caffaro, C. E., San Jose, K., et al. (2017b). A Semi-synthetic Organism that Stores and Retrieves Increased Genetic Information. *Nature* 551, 644–647. doi:10.1038/nature24659

Zhou, A. X.-Z., Dong, X., and Romesberg, F. E. (2020). Transcription and Reverse Transcription of an Expanded Genetic Alphabet *In Vitro* and in a Semisynthetic Organism. *J. Am. Chem. Soc.* 142, 19029–19032. doi:10.1021/jacs.0c09230

Zhou, A. X.-Z., Sheng, K., Feldman, A. W., and Romesberg, F. E. (2019). Progress toward Eukaryotic Semisynthetic Organisms: Translation of Unnatural Codons. *J. Am. Chem. Soc.* 141, 20166–20170. doi:10.1021/jacs.9b09080

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

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

frontiers | Frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fbioe.2022.913057

Tuning the Properties of Protein-Based Polymers Using High-Performance Orthogonal Translation Systems for the Incorporation of Aromatic Non-Canonical Amino Acids

*Osher Gueta[1†], Ortal Sheinenzon[1†], Rotem Azulay[1], Hadas Shalit[2], Daniela S. Strugach[1], Dagan Hadar[1], Sigal Gelkop[1], Anat Milo[2] and Miriam Amiram[1]**

[1]Avram and Stella Goldstein-Goren Department of Biotechnology Engineering, Ben-Gurion University of the Negev, Beersheba, Israel, [2]Department of Chemistry, Ben-Gurion University of the Negev, Beersheba, Israel

OPEN ACCESS

Edited by:
Gustavo Fuertes,
Institute of Biotechnology (ASCR), Czechia

Reviewed by:
Huibin Zou,
Qingdao University of Science and Technology, China
Jeffery M. Tharp,
Yale University, United States

***Correspondence:**
Miriam Amiram
mamiram@bgu.ac.il

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Nanobiotechnology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 05 April 2022
Accepted: 27 April 2022
Published: 30 May 2022

Citation:
Gueta O, Sheinenzon O, Azulay R, Shalit H, Strugach DS, Hadar D, Gelkop S, Milo A and Amiram M (2022) Tuning the Properties of Protein-Based Polymers Using High-Performance Orthogonal Translation Systems for the Incorporation of Aromatic Non-Canonical Amino Acids. Front. Bioeng. Biotechnol. 10:913057. doi: 10.3389/fbioe.2022.913057

The incorporation of non-canonical amino acids (ncAAs) using engineered aminoacyl-tRNA synthetases (aaRSs) has emerged as a powerful methodology to expand the chemical repertoire of proteins. However, the low efficiencies of typical aaRS variants limit the incorporation of ncAAs to only one or a few sites within a protein chain, hindering the design of protein-based polymers (PBPs) in which multi-site ncAA incorporation can be used to impart new properties and functions. Here, we determined the substrate specificities of 11 recently developed high-performance aaRS variants and identified those that enable an efficient multi-site incorporation of 15 different aromatic ncAAs. We used these aaRS variants to produce libraries of two temperature-responsive PBPs—elastin- and resilin-like polypeptides (ELPs and RLPs, respectively)—that bear multiple instances of each ncAA. We show that incorporating such aromatic ncAAs into the protein structure of ELPs and RLPs can affect their temperature responsiveness, secondary structure, and self-assembly propensity, yielding new and diverse families of ELPs and RLPs, each from a single DNA template. Finally, using a molecular model, we demonstrate that the temperature-responsive behavior of RLPs is strongly affected by both the hydrophobicity and the size of the unnatural aromatic side-chain. The ability to efficiently incorporate multiple instances of diverse ncAAs alongside the 20 natural amino acids can help to elucidate the effect of ncAA incorporation on these and many other PBPs, with the aim of designing additional precise and chemically diverse polymers with new or improved properties.

Keywords: non-canonical amino acids (ncAAs), elastin-like polypeptides (ELP), resilin-like polypeptides (RLP), smart biomaterials, genetic code expansion

INTRODUCTION

Incorporating non-canonical amino acids (ncAAs) into proteins has emerged as a powerful methodology to improve, alter, or introduce new functions into proteins. The incorporation of ncAAs bearing a variety of chemical groups can facilitate the elucidation of protein structure–function relationship (Debelouchina and Muir, 2017; Chen et al., 2018) or protein–protein interactions (Nguyen et al., 2018), or the production of protein-based therapeutics (Huang and Liu, 2018) and biomaterials with novel functions (Connor and Tirrell, 2007; Israeli et al., 2019), among many other applications (Johnson et al., 2010; Chin, 2017; Voller and Budisa, 2017; Young and Schultz, 2018; Lee et al., 2019; Zhou and Deiters, 2021). There are two common strategies for expressing recombinant proteins containing ncAAs. The first entails substituting one of the natural AAs with an ncAA. This technique permits the incorporation of multiple instances of a ncAA into the target protein and in entire proteomes (Johnson et al., 2010). However, substitution allows the incorporation of only 19 natural AAs alongside the ncAA, and the ncAA must be a close analog of the natural AA that it replaces, therefore using this approach limits the flexibility of protein engineering. Alternatively, ncAAs can be incorporated by codon reassignment or frameshift codons using orthogonal translation systems consisting of an aminoacyl-tRNA synthetase (aaRS) and a tRNA pair that do not interact with the native aaRSs and tRNAs in the expression host. This technique has enabled the incorporation of >200 different ncAAs, using aaRS·tRNA pairs such as the $tRNA^{Tyr}$·TyrRS pair from *Methanocaldococcus jannaschii* (Kobayashi et al., 2003) or the $tRNA^{Pyl}$·PylRS pair from various methanogenic archaea and bacteria, such as from the *Methanosarcina* or *Methanomethylophilus* spp. (Wan et al., 2014; Willis and Chin, 2018). Typically, a TAG stop codon is assigned to the ncAA and the orthogonal tRNA anticodon is mutated to CUA (if needed). In addition, the evolution of the aaRS by mutagenesis of its AA-binding site is required to accommodate the ncAA. However, traditional evolution methodologies typically yield aaRS variants capable of introducing the ncAA at only one or a few instances within a polypeptide chain, and often with reduced yields compared to wild-type proteins (O'Donoghue et al., 2013; Vargas-Rodriguez et al., 2018).

We recently developed an aaRS evolution platform that utilizes multiplex automated genome engineering (MAGE) to create a library of chromosomal MjTyrRS variants. High-performance aaRS variants were selected from these libraries, enabling multi-site incorporation (10–30 instances per protein) of the aromatic ncAAs p-acetyl-phenylalanine (pAcF) (Amiram et al., 2015), p-azido-phenylalanine (pAzF) (Amiram et al., 2015), 4-propargyloxy-L-phenylalanine (pPR) (Hadar et al., 2021), and phenylalanine-4'-azobenzene (AzoPhe) (Israeli et al., 2021). Although these aaRS variants were selected based on their ability to incorporate the ncAAs mentioned above, it had been previously demonstrated by us and others that most aaRS variants exhibit some degree of poly-specificity, that is, they can recognize and charge their cognate tRNA with ncAAs bearing various chemical groups (Young et al., 2011; Amiram et al., 2015). The elucidation of the substrate specificities of high-performance aaRS variants can be a facile means for generating highly efficient aaRS–ncAA pairs for the multi-site incorporation of additional ncAAs. Such an expanded set of aaRS–ncAA pairs for the efficient incorporation of multiple instances of ncAAs within a single protein chain can enable the engineering of proteins with the desired new or improved functions and their high-yield production (Ohtake et al., 2015; Rezhdo et al., 2019; Hayashi et al., 2021). In addition, analyzing the relationship between the mutations in the aaRS AA-binding site and the corresponding ncAA binding specificities can inform computational models and facilitate the selection of smaller, focused libraries for future aaRS evolution experiments (Hauf et al., 2017; Baumann et al., 2019).

Multi-site ncAA incorporation is particularly desirable for the design and production of protein-based polymers (PBPs), which consist of tandem repeats of either natural or artificial short peptide motifs, and whose properties are determined by the sequence of the AAs and ncAAs in these motifs (Connor and Tirrell, 2007; Israeli et al., 2019; Varanko et al., 2020; Chang et al., 2021). Although the incorporation of ncAAs in numerous PBPs can be used to expand their range of properties and functions (Connor and Tirrell, 2007; Israeli et al., 2019), we focus here on the effect of ncAA incorporation on the properties of two families of bio-inspired, thermo-responsive PBPs: elastin-like polypeptides (ELPs) and resilin-like polypeptides (RLPs). ELPs are arguably the most well-studied family of artificial PBPs that comprise multiple repeats (typically 5–200 repeats) of the VPGXG motif (variations of the tropoelastin-derived VPGVP motif), where X is permissive to any amino acid (AA) except proline (MacEwan and Chilkoti, 2010). ELPs undergo a reversible soluble-to-insoluble phase transition at their lower critical solution temperature (LCST), often termed the transition temperature, which depends on the ELP composition and is maintained in ELP fusion proteins and conjugates (MacEwan and Chilkoti, 2010; Amiram et al., 2013; MacEwan and Chilkoti, 2014; Despanie et al., 2016). Extensive characterization of the mechanism underlying this phase transition permits the prediction and tuning of the LCST of the ELP, primarily by varying the hydrophobicity of the X-guest residue and the molecular weight (MW) of the ELP (Meyer and Chilkoti, 2004; McDaniel et al., 2013). RLPs are artificial, repetitive PBPs, comprising sequences identical to, or inspired by, motifs found in natural resilin proteins (Balu et al., 2021). RLPs exhibit a temperature-responsive behavior characterized by an upper critical solution temperature (UCST), with some variants also showing an additional LCST-type phase transition. The role of aromatic AAs in the UCST- and LCST-type transition behaviors appears to have particular significance (Martin and Mittag, 2018). The LCST of ELPs is affected mainly by hydrophobic interactions (Pak et al., 2016; Ruff et al., 2018) but may also be affected by hydrogen bonding and π-π stacking of the aromatic side-chains (Taylor et al., 2020). In contrast, several studies have demonstrated that the UCST transition appears to be strongly influenced by the fraction, position, and identity of aromatic AAs and their interactions with other AAs (e.g., cation-π interactions), as well as by the overall hydrophobicity (Joseph et al., 2021).

In this work, we have sought to characterize the substrate specificities of 11 of our previously described, MAGE-evolved aaRS variants using 38 different aromatic ncAAs. We identified efficient aaRS variants that are capable of incorporating 15 of these ncAAs in up to 30 instances in a single ELP chain. We applied this information to produce ELP and RLP libraries bearing multiple instances of each ncAA and characterized the effect of the newly introduced chemical group on the properties of these PBPs.

MATERIALS AND METHODS

Materials

ncAAs were purchased from various vendors, as indicated in **Supplementary Table S1**. Plasmid purification was conducted with Plasmid HiYield mini-prep (RBC Bioscience). SDS solution was purchased from Bio-Rad. Anhydrotetracycline hydrochloride was purchased from Sigma-Aldrich. Arabinose and 2xYT media were purchased from Tivan Biotech, and IPTG from Biolab-Biology. C321. ΔA (Isaacs lab) was a gift from Farren Isaacs (Addgene plasmids # 73,581). Isomerization experiments were performed with a 365 nm UV lamp (VL-6.LC, 12W, Vilber Lourmat) and blue mounted LEDs: 405 nm, 1,000 mW, 800 mA (M405L4, Thor-Labs). Plasmids encoding the pAcFRS.1.t1, pAcFRS.2.t1, pAzFRS.1.t1, pAzFRS.2.t1, Mut1-RS, Mut2-RS, and AzoRS-4 can be obtained from Addgene.

Analysis of GFP Expression by Intact-Cell Fluorescence Measurements

For 96-well plate-based assays, strains harboring plasmids encoding the orthogonal translation systems and ELP-GFP reporter plasmids were inoculated with transformed cells from either a fresh agar plate or from stocks stored at −80°C, and grown to confluence overnight. Cultures were then inoculated at a 1:30 dilution in 2xYT medium supplemented with kanamycin (30 μg ml^{-1}), chloramphenicol (25 μg ml^{-1}), and the ncAA (0.25 mM of ncAAs 1–5 or 1 mM of all other ncAAs). The expression of the aaRS was induced by adding arabinose (0.2%). Cells were allowed to grow at 34°C to an OD_{600} of 0.5–0.8 in a shaking plate incubator at 550 rpm (~5 h) and GFP expression was induced by adding anhydrotetracycline (60 ng ml^{-1}). Following overnight expression, the cells were centrifuged at 4,000 rpm for 3 min, the supernatant medium was removed, and the cells were resuspended in PBS. GFP fluorescence was measured on a Biotek spectrophotometric plate reader by using excitation and emission wavelengths of 485 nm and 528 nm, respectively. Fluorescence signals were normalized by dividing the fluorescence counts by the OD_{600} reading. Shown are representative results of at least three different assays conducted for each reporter protein and expression condition.

ELP Expression and Purification

Before batch expression, starter cultures (1:40 v/v of final expression volume) of 2xYT media, supplemented with kanamycin (30 μg ml^{-1}) and chloramphenicol (25 μg ml^{-1}), were inoculated with transformed cells from either a fresh agar plate or from stocks stored at −80°C, incubated overnight at 34°C while shaking at 220 rpm, and transferred to expression flasks containing 2xYT media, antibiotics, arabinose (0.2%), and the ncAA (0.25 mM of ncAAs 5 or 1 mM of all other ncAAs). Cells were allowed to grow to an OD_{600} of 0.5–0.8 and protein expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM). The cells were harvested 24 h after inoculation by centrifugation at 4,000 g for 30 min at 4°C. The cell pellet was then resuspended by vortex in milli-Q water (~4 ml) and either stored at −80°C or purified immediately. For purification, resuspended pellets were lysed by ultrasonic disruption (18 cycles of 10 s sonication, separated by 40 s intervals of rest). Poly (ethyleneimine) was added (0.2 ml of a 10% solution) to each lysed suspension before centrifugation at 4,000 rpm for 15 min at 4°C to separate cell debris from the soluble cell lysate. All ELP constructs were purified by a modified inverse transition cycling (ITC) protocol (Hassouneh et al., 2010) consisting of multiple "hot" and "cold" spins by using sodium chloride to trigger the phase transition. Before purification, the soluble cell lysate was incubated for up to 2 min at 42–65°C to denature the native *E. coli* proteins. The cell lysate was then cooled on ice, centrifuged for 2 min at ~14,000 rpm, and the pellet was discarded. For "hot" spins, the ELP phase transition was triggered by adding sodium chloride to the cell lysate or to the product of a previous cycle of ITC at a final concentration of up to ~5 M. The solutions were then centrifuged at ~14,000 rpm for 10 min and the pellets were resuspended in milli-Q water, after which a 2 min "cold" spin was performed without sodium chloride to remove denatured contaminants. Additional rounds of ITC were conducted as needed using a saturated solution of sodium chloride until sufficient purification was achieved. Purified proteins were visualized on SDS-PAGE.

Protein concentrations were calculated by measuring the OD_{280} of the purified protein according to the extinction coefficients in **Supplementary Table S2**. Predicted LogD values of the ncAAs were determined using ChemAxon logD predictor (Kubyshkin, 2021).

RLP Expression and Purification

Batch expression was performed as described above for ELP production. The cells were harvested 24 h after inoculation by centrifugation at 4,000 g for 30 min at 4°C. Purification was performed according to a previously described protocol (Dzuricky et al., 2020). Purified proteins were visualized on SDS-PAGE. Protein concentrations were calculated by measuring the OD_{280} of the purified protein in 8M urea according to the extinction coefficients in **Supplementary Table S2**.

Intact Mass Measurements

The intact mass of ELP and RLP variants was measured using MALDI-TOF/TOF autoflex speed at the Ilse Katz Institute for Nanoscale Science and Technology (Ben-Gurion University of the Negev). Spectrum analysis was performed by the Flexanalysis software.

TABLE 1 | Annotations of specific mutations in evolved aaRS variants compared with the WT *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (MjTyrRS) sequence. In addition to the indicated mutations, all mutants harbored the R257G and D286R mutations, which have been shown to improve tRNA binding.

Position	32	65	107	108	109	158	159	162	167
aaRS									
Tyr-RS	Y	L	E	F	Q	D	I	L	A
PacFRS.1.t1	L	L	E	F	Q	G	C	R	D
PacFRS.2.t1	L	V	E	F	Q	G	C	R	D
PazFRS.2.t1	L	L	T	Y	M	G	C	R	A
Mut1-RS	L	L	E	F	Q	S	M	K	H
Mut2-RS	L	V	E	F	Q	G	A	E	H
Mut3-RS	T	V	A	Y	M	G	C	R	D
Mut4-RS	L	V	E	F	Q	G	M	S	H
AzoRS-1	G	V	E	F	Q	G	Y	S	F
AzoRS-2	L	V	S	V	S	G	Y	S	F
AzoRS-3	L	V	N	V	L	G	Y	S	F
AzoRS-4	G	V	E	F	Q	G	Y	R	A

Phase Transition Analysis

To characterize the inverse transition temperature of the ELP and RLP variants, the OD_{600} of the ELP solution was monitored as a function of temperature, with heating and cooling performed at a rate of 1°C min^{-1} on a UV-vis spectrophotometer equipped with a multicell thermoelectric temperature controller (Thermo Scientific).

Dynamic Light Scattering (DLS) Analysis

ELP and RLP self-assembly was analyzed using a Zetasizer Nano ZS (Malvern Pananalytical). For each sample, 11–17 acquisitions (determined automatically by the instrument) were obtained at 5°C or 60°C for ELPs or RLPs, respectively. Populations comprising less than 1% of the total mass (by volume) were excluded from the analysis.

Circular Dichroism (CD) Analysis

The secondary structure of ELPs was studied using an Jasco J-715 spectropolarimeter (Tokyo) equipped with a PTC-348WI temperature controller, using a 1-mm quartz cuvette by scanning from 280 to 180 nm at 5°C. Purified constructs were diluted to 5 μM in water. Data were considered for analysis whenever the Dynode voltage was below 800 V.

RESULTS AND DISCUSSION

Profiling the Substrate Specificity of High-Performance aaRSs

We first profiled the substrate specificities of our previously selected aaRS variants: pAcFRS.1.t1, pAcFRS.2.t1, and pAzFRS.2.t1 were originally selected for pAcF and pAzF incorporation (Amiram et al., 2015), MutRS-1—MutRS-4 were selected for pPR incorporation (Hadar et al., 2021), and AzoRS-1-AzoRS-4 were selected for AzoPhe incorporation (Israeli et al., 2021). A summary of the mutations in the AA binding site of these aaRS variants, as compared with the parent MjTyrRS, which natively recognizes tyrosine, is provided in **Table 1**. To identify high-performance aaRSs capable of efficient incorporation of multiple instances of the ncAAs per protein, we assessed the ability of the aaRSs to support the incorporation of either 10 or 30 instances of each ncAA per protein via the expression of two reporter proteins: ELP(10TAG)-GFP and ELP(30TAG)-GFP (**Supplementary Table S3**). These reporter proteins were produced in the presence of each aaRS·tRNA pair using the *Escherichia coli* strain C321.ΔRF1, which lacks all the native TAG codons and the associated release factors (RF-1) (Lajoie et al., 2013).

Notably, some high-performance aaRS variants, particularly pAcFRS.1.t1 and pAzFRS.2.t1, also exhibit a relatively high incorporation of natural AAs in the absence of their cognate ncAA. This property results in high levels of "background" expression of the ELP-GFP reporters in the absence of a ncAA, although, when present, they accurately incorporate their cognate ncAA (Amiram et al., 2015; Hadar et al., 2021; Israeli et al., 2021). Such background expression can be reduced in conditions that highlight the differences in aaRS efficiencies, such as by increasing the number of TAG codons in the reporter protein, since the aaRSs incorporate the ncAAs more efficiently than the natural AAs. Specifically, relative protein production in the presence of the cognate ncAA, as compared with the absence of any ncAA, is lower for ELP(30TAG)-GFP than for ELP(10TAG)-GFP (Amiram et al., 2015; Hadar et al., 2021). Consequently, we conducted our initial evaluation of multi-site ncAA incorporation using the ELP(30TAG)-GFP reporter to eliminate the possibility of masking the production of ncAA-containing reporter proteins by background expression in the absence of the ncAA.

The expression of the ELP(30TAG)-GFP reporter protein in the presence of 38 different ncAAs revealed that 15 of the ncAAs (including four of the cognate ncAAs for which the aaRS variants were initially selected; **Supplementary Figure S1**) are efficiently incorporated by one or more of the aaRS variants (**Figures 1A–C**, **Supplementary Figure S2**). Notably, all but one (ncAA 3) of the 15 ncAAs that were successfully incorporated are phenylalanine derivatives that harbor 4' (para) substituents. The expression of the ELP(10TAG)-GFP reporter protein in the presence of the same 15 ncAAs confirms the high background incorporation previously observed for some of the variants and reveals that some aaRS variants appear to have increased efficiency in the expression of this reporter, as compared with ELP(30TAG)-GFP (**Supplementary Figure S3**). This finding is in agreement with previous studies demonstrating that differences in aaRS efficiencies are revealed by increasing the number of TAG codons per protein (Amiram et al., 2015; Hadar et al., 2021; Israeli et al., 2021).

An examination of the specificities of the aaRS variants (indicated as heat maps in **Figures 1A–C**) reveals several distinctions in the substrate profiles of some of the aaRSs. First, generally, the larger ncAAs (1–5) were suitable substrates for AzoRS-1-4, which were selected for the incorporation of 5 (AzoPhe). In contrast, smaller ncAAs were more efficiently recognized by the remaining aaRSs, which were selected for the incorporation of pAcF (13), pAzF (11), and pPR

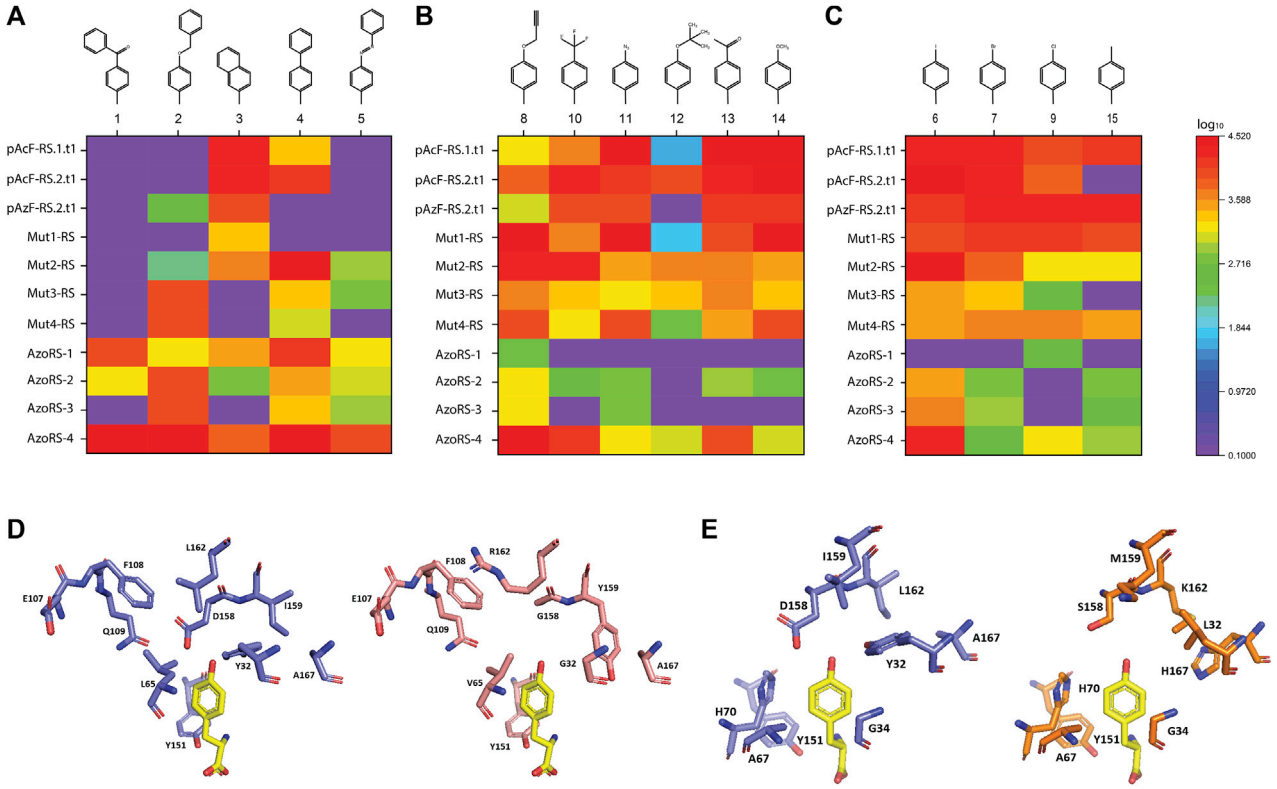


FIGURE 1 | (A–C) Specificities of high-performance aaRS variants for the 15 ncAAs depicted in **Supplementary Figure S1**, assessed by their incorporation into the reporter protein ELP(30TAG)-GFP. **(D–E)** A comparison of the AA-binding pocket, shown in complex with the native cognate AA tyrosine (yellow), of **(D)** native MjTyrRS (purple) compared with AzoRS-4 (pink), and **(E)** native MjTyrRS (purple, different angle) and Mut1-RS (orange). The structural models were generated (based on PDB ID: 1J1U) using the PyMOL program (Schrödinger, LLC).

(8). In addition, while some aaRSs, such as pAcFRS.1.t1, pAcFRS.2.t1, and AzoRS-4, exhibited a broad substrate spectrum and could efficiently accommodate >9 different ncAAs, other aaRSs, such as Mut3-RS, Mut4-RS, AzoRS-1, and AzoRS-3, were more selective and can efficiently aminoacylate only 1-3 ncAAs. We also profiled the substrate specificities of an additional aaRS, namely pAzFRS.1.t1, using our panel of 38 ncAAs. We found that it is exceptionally specific for pAzF and excluded all other ncAAs in our panel (**Supplementary Figure S4**), in accordance with our previous analysis of this aaRS using another panel of ncAAs (Amiram et al., 2015). Finally, mutually orthogonal aaRS–ncAA pairs, such as AzoRS-1:ncAA1 and Mut2-RS:ncAA6, could be identified from these analyses. Such orthogonal pairs may assist in constructing multiple orthogonal translation systems for the incorporation of two or more ncAAs within a single protein.

A comparison of the mutations in each of the aaRS variants (**Table 1**) provides information that may aid in the construction of future libraries for aaRS evolution. First, the differences and similarities in the mutations found in the aaRS variants revealed patterns in the types of AAs substituted at certain positions (**Figures 1D,E**). For example, Y32 and D158, which originally form hydrogen bonds with the tyrosine hydroxyl group, were mutated in all variants to a smaller AA—leucine, threonine, or glycine (for Y32) or glycine and serine (for D158), enlarging the ncAA binding pocket. Similarly, I159 was mutated in all variants to either smaller AAs, such as cysteine or alanine in aaRSs selected for the incorporation of 8, 11, or 13, or to tyrosine in variants selected for the incorporation of 5, where π-π interactions between the side chain of tyrosine 159 and the additional phenyl ring in ncAAs 1–5 may facilitate their binding. Finally, A167 was frequently replaced with a charged AA or with phenylalanine in aaRSs that are compatible with ncAAs bearing the relatively small, or phenyl-based substituents, respectively. Second, although 12 AAs in the AA-binding pocket were targeted for diversification in the selections of all aaRS variants, mutations in only nine of these AAs were evident in the selected aaRSs. It is possible that any changes in the remaining three AAs—G34, A67 and Y151—which remained unaffected in all of our variants, are not beneficial to this group of ncAAs or deleterious to the binding of any AA. A mutagenesis of G34 and A67, which are relatively small AAs, may increase the size of the AA-binding pocket and, thereby, interfere with the binding of aromatic AAs and ncAAs. The mutagenesis of Y151, which forms a hydrogen bond with the AA/ncAA backbone amino group (Kobayashi et al., 2003), may prevent AA binding. Finally, it may be that other AAs in

the AA-binding pocket, such as H70, must be mutagenized to enable the efficient incorporation of aromatic ncAAs bearing meta or ortho substituents (Neumann et al., 2008; Sakamoto et al., 2009; Hauf et al., 2017).

Our results suggest that the aaRS variants analyzed in this study may also efficiently incorporate other phenylalanine derivatives that harbor 4' (para) substituents, which were not included in the current panel. It is also possible that some of these aaRS variants can efficiently incorporate one or a few instances of other aromatic ncAAs, including the remaining 23 ncAAs that we examined in this study and for which we did not identify an efficient aaRS for the incorporation of 10 or 30 instances per protein. However, the experimental conditions for the identification of such low-to-medium efficiency aaRS:ncAA pairs must be carefully selected, since the background expression of reporter protein in the absence of the ncAA can be high, and it increases with decreasing number of TAG codons per protein. Such analyses may be facilitated by decreasing aaRS production, for example by expression of a chromosomally integrated aaRS, or by reducing the copy number of the plasmid encoding the aaRS, or the amount of the respective inducer of aaRS expression (Amiram et al., 2015; Hadar et al., 2021). Finally, the high-performance aaRS:ncAA pairs identified in this study may also be utilized to improve the expression of ncAA-containing proteins and PBPs in other *E. coli* strains, such as derivatives of the commonly utilized BL21 strain, although competition with RF-1 is expected to reduce the yield of the full-length protein in these strains (Hadar et al., 2021).

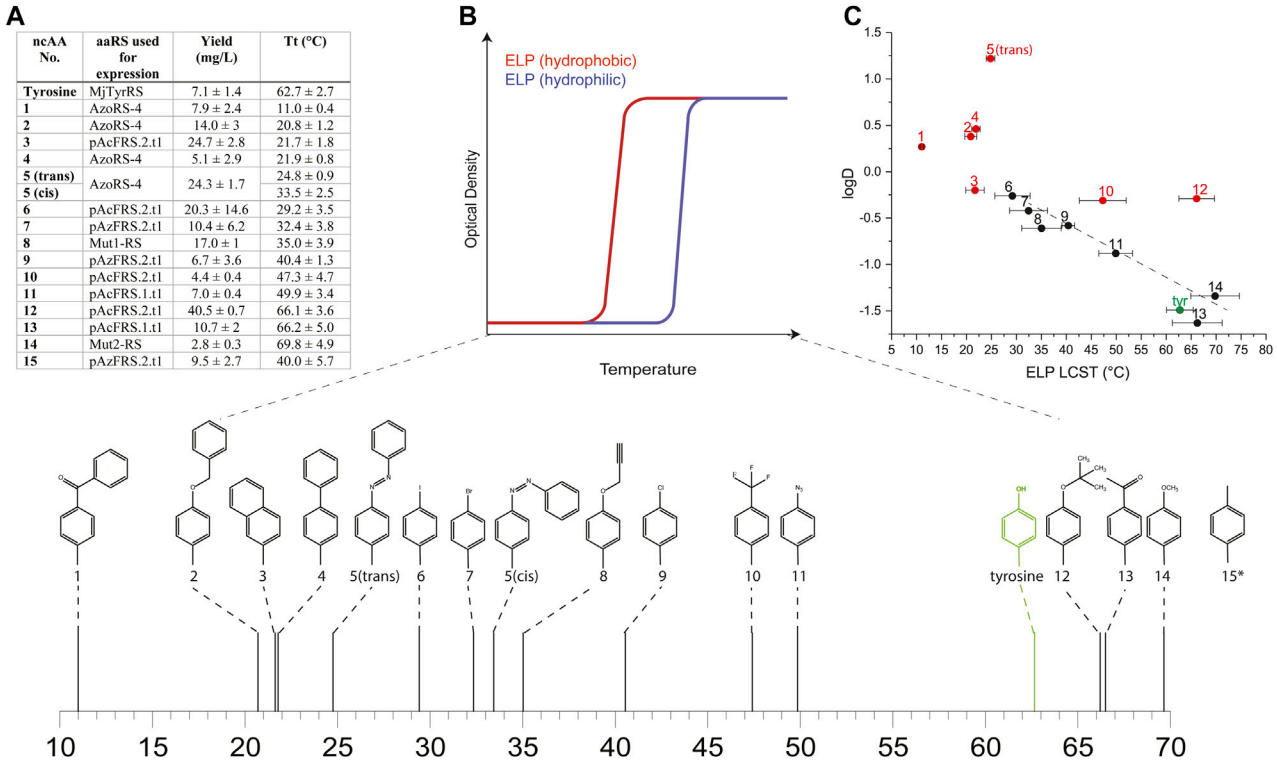
Multi-Site ncAA Incorporation in ELPs

The characterization of ncAA-containing ELPs requires a suitable ELP variant, such that ncAA incorporation will result in ELPs with measurable LCSTs (i.e., ~10–90°C). However, the LCST of the above-mentioned ELP-GFP reporters is predicted to be either above (in GFP-fused proteins) or below (in unfused proteins) the measurable temperature range used for LCST analysis and characterization. Therefore, we selected a gene that encodes another ELP variant, termed ELP_{60}(10TAG), which we previously utilized to determine the effect of the incorporation and isomerization of 5 on the ELP LCST (Israeli et al., 2021). The protein encoded by the ELP_{60}(10TAG) gene is based on previously described hydrophilic ELPs, which have a high LCST (>90°C for a 25 μM solution (Meyer and Chilkoti, 2002)) and are composed of glycine and alanine, alternating in the X-guest residue position (**Supplementary Table S3**). We chose this variant since we hypothesized that the multi-site incorporation of aromatic ncAAs is expected to dramatically reduce the LCST of the ELP, based on the observation that the incorporation of aromatic residues, such as tyrosine, phenylalanine, tryptophan (Urry et al., 1992; Seifried et al., 2018), and 5 (Israeli et al., 2021), reduce the LCST, as compared with the incorporation of more hydrophilic AAs. Below, we name proteins according to the identity of the ncAA incorporated in the TAG codons. For example, ELP_{60}(1 × 10) is the protein product of the ELP_{60}(10TAG) gene, wherein 1 is incorporated in 10 encoded TAG codons.

We expressed proteins from the ELP_{60}(10TAG) gene in the C321. ΔRF1 *E. coli* strain by using either aaRSs suitable for each of the 15 ncAAs, or the native MjTyrRS (for the incorporation of tyrosine), as designated in **Figure 2A**. We quantified protein yields by using small-batch expression and found that the yields of ncAA-containing ELP_{60}(10TAG) proteins ranged from ~3 to 40 mg/L, as compared with the yield of tyrosine-containing ELP_{60}(tyrosine×10), which was ~7 mg/L (**Figure 2A**). The yield of purified proteins is likely affected both by the purification process, in which more hydrophobic ELPs generally tend to precipitate more easily, and by the efficiency of the aaRS. Protein purity and MW were determined by SDS-PAGE and intact mass-spectrometry (**Supplementary Table S4** and **Supplementary Figure S5**).

To determine the effect of ncAA incorporation on ELP properties, we first examined the effect of ncAA identity on the LCST of the ELPs in ddH_2O at a fixed concentration of 25 μM (**Figures 2A**,**B**, **Supplementary Figure S6**). The incorporation of 10 instances of each ncAA in the ELP_{60}(10TAG) protein profoundly affected the LCST, which ranged from ~11 to 70°C for ELP_{60}(1 × 10)- ELP_{60}(14 × 10), while the LCST of ELP_{60}(15 × 10) was above 90°C in these conditions. We next examined the effect of ncAA incorporation on the self-assembly of this ELP family, motivated by an earlier finding that the incorporation of 5 engendered the self-assembly of ELP_{60}(5 × 10) into thin 2-dimensional sheets. An assessment of self-assembly was conducted by a dynamic light scattering (DLS) analysis of a solution of each protein (25 μM in ddH_2O) at a temperature below the LCST of all ELPs (5°C). The DLS analysis indicated that ncAAs 1–5, 10, 12, and 15 appeared to engender ELP self-assembly (**Supplementary Figure S7**). We note that previously reported ELP_{60}(1 × 10) produced using a first-generation aaRS did not self-assemble, perhaps due to differences in mis-incorporation tendencies or in the purification protocol (Israeli et al., 2021). The effect of ncAA incorporation on the LCST of the ELPs can be viewed as an extension to Urry's AA hydrophobicity scale, which is based on the effect of natural AAs on the ELP LCST (Urry et al., 1992). Therefore, we examined the relationship between the ELP LCST and predicted logD values of the ncAAs at pH 7 (computed using ChemAxon, which was previously shown to generate values that correlated well with the experimental logD measurements of AAs and similar ncAAs (Kubyshkin, 2021)). While a poor correlation was observed overall between the LCST of the ELPs and the computed logD values, a strong correlation (R^2 = 0.92) was observed for those ELPs that did not self-assemble in the solution (**Figure 2C**). This finding could be explained by the fact that self-assembly effectively alters the local concentration of the ELPs—and, therefore, its apparent LCST—as was suggested in other self-assembled ELP systems (Dreher et al., 2008; Prhashanna et al., 2019).

Next, we examined the secondary structure of ncAA-containing ELPs by using circular dichroism (CD) spectroscopy at a temperature below the LCST of all ELPs (5°C). All ELPs, including the control ELP_{60}(Tyrosine×10), showed negative peaks at around ~190 nm, and most ELPs exhibited an additional negative peak at ~220 nm; these peaks



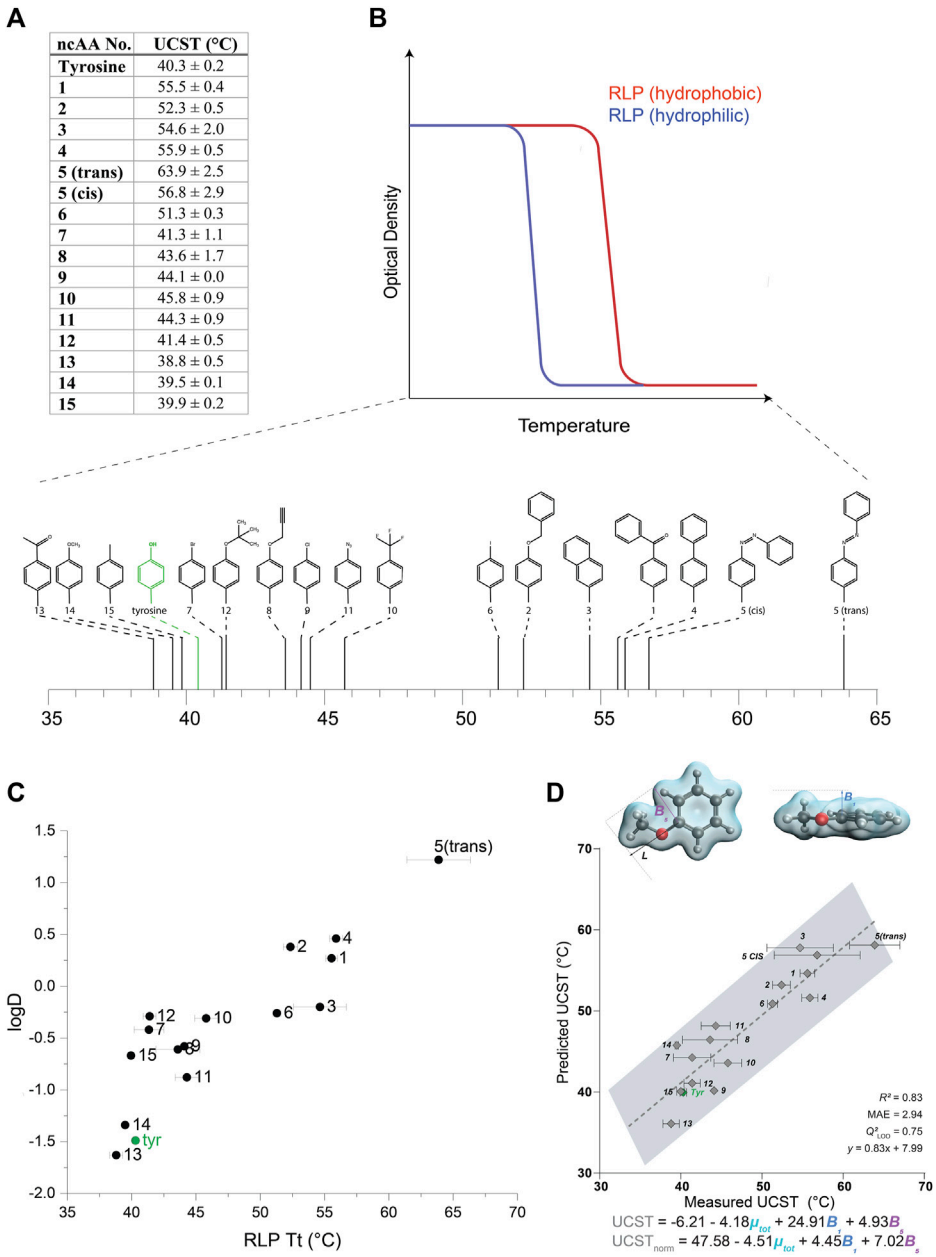
ncAA No.	aaRS used for expression	Yield (mg/L)	Tt (°C)
Tyrosine	MjTyrRS	7.1 ± 1.4	62.7 ± 2.7
1	AzoRS-4	7.9 ± 2.4	11.0 ± 0.4
2	AzoRS-4	14.0 ± 3	20.8 ± 1.2
3	pAcFRS.2.t1	24.7 ± 2.8	21.7 ± 1.8
4	AzoRS-4	5.1 ± 2.9	21.9 ± 0.8
5 (trans)	AzoRS-4	24.3 ± 1.7	24.8 ± 0.9
5 (cis)			33.5 ± 2.5
6	pAcFRS.2.t1	20.3 ± 14.6	29.2 ± 3.5
7	pAzFRS.2.t1	10.4 ± 6.2	32.4 ± 3.8
8	Mut1-RS	17.0 ± 1	35.0 ± 3.9
9	pAzFRS.2.t1	6.7 ± 3.6	40.4 ± 1.3
10	pAcFRS.2.t1	4.4 ± 0.4	47.3 ± 4.7
11	pAcFRS.1.t1	7.0 ± 0.4	49.9 ± 3.4
12	pAcFRS.2.t1	40.5 ± 0.7	66.1 ± 3.6
13	pAcFRS.1.t1	10.7 ± 2	66.2 ± 5.0
14	Mut2-RS	2.8 ± 0.3	69.8 ± 4.9
15	pAzFRS.2.t1	9.5 ± 2.7	40.0 ± 5.7

FIGURE 2 | (A) Yields and LCST values calculated for ELPs containing each of the 15 ncAAs depicted in **Supplementary Figure S1** and expressed using the indicated aaRS. **(B)** A schematic illustration of the effect of the relative hydrophobicity of the AA or ncAA incorporated into the ELP sequence on its LCST, and the measured LCST values for ELPs bearing ncAAs 1–14 (*the LCST of ELP_{60} (15 × 10) was too high to be measured in ddH_2O, and was therefore determined in ddH_2O supplemented with 1M NaCl). **(C)** The relationship between the LCST of ELPs bearing ncAAs 1–14 and the predicted logD values of the ncAAs. Red and black data points indicate ELPs that were found to self-assemble or that remained as monomers, respectively.

are characteristic of random-coil and β-turn structures, respectively (**Supplementary Figure S8**) (Janib et al., 2014). The magnitude and ratio of these peaks varied for each ELP, indicating that ncAA incorporation can also affect the secondary structure of these proteins. However, no correlation was found between these structural properties, the LCST, or the self-assembly propensity of the ELPs, as determined in these conditions.

Precise tuning of the properties, such as the LCST, of ELPs (and other PBPs), either as single- or multi-block polymers or as fusions with other proteins and PBPs (e.g., ELP–RLP fusions), is required in many applications (McDaniel et al., 2013), such as for drug delivery (Bhattacharyya et al., 2016; Varanko et al., 2020; Balu et al., 2021), tissue engineering (Bhattacharyya et al., 2016; Balu et al., 2021), sensing (Dhandhukia et al., 2013; Dhandhukia et al., 2017), metabolic engineering (Dzuricky et al., 2020), and protein purification (Hassouneh et al., 2010; Bhattacharyya et al., 2016). Although factors such as salinity and MW can also be used to tune the LCST or UCST of PBPs, such adjustments are often limited by the specific application (McDaniel et al., 2013). Previous studies have demonstrated that either installing or incorporating unnatural chemical groups can modulate ELP properties, such as LCST. For example, a global replacement of proline residues with various proline analogs in ELPs resulted in changes to the secondary structure of the ELPs and altered their LCST (Kim et al., 2005; Kim et al., 2006; Catherine et al., 2015). Similarly, the chemical modification of methionine residues encoded in the guest-residue position in ELPs also enabled the tuning of the LCST (Kramer et al., 2015; Petitdemange et al., 2017; Rosselin et al., 2019) and triggered the self-assembly of di-block ELPs (Dai et al., 2021). However, until recently, the moderate efficiency of orthogonal translation systems has limited the production of PBPs that contain multiple instances of ncAAs by genetic code expansion, and thereby hindered the analysis of the resulting properties (Wang et al., 2021).

Here, we demonstrate that high-performance aaRSs enable the incorporation of 15 different aromatic ncAAs in ELPs—and, in turn, the production of a family of ELPs with LCSTs that vary within a wide temperature range—from a single DNA template. In addition, the analysis of these ELPs revealed that ncAA incorporation affects the secondary structure and self-assembly of the ELPs. Importantly, as compared with global replacement or chemical modification, which modify every instance of a particular AA, incorporating ncAAs using high-performance orthogonal translation components enables the multi-site



ncAA No.	UCST (°C)
Tyrosine	40.3 ± 0.2
1	55.5 ± 0.4
2	52.3 ± 0.5
3	54.6 ± 2.0
4	55.9 ± 0.5
5 (trans)	63.9 ± 2.5
5 (cis)	56.8 ± 2.9
6	51.3 ± 0.3
7	41.3 ± 1.1
8	43.6 ± 1.7
9	44.1 ± 0.0
10	45.8 ± 0.9
11	44.3 ± 0.9
12	41.4 ± 0.5
13	38.8 ± 0.5
14	39.5 ± 0.1
15	39.9 ± 0.2

FIGURE 3 | (A) UCST values calculated for RLPs containing each of the 15 ncAAs depicted in **Supplementary Figure S1**. **(B)** A schematic illustration of the effect of the relative hydrophobicity of the AA or ncAA incorporated in the RLP sequence on its UCST, and a schematic indicating the measured UCST values for the RLPs bearing ncAAs 1–15. **(C)** The relationship between the UCST of RLPs bearing ncAAs 1–15, and predicted logD values of the ncAAs. **(D)** Multivariate linear regression model that includes B_1, B_5, and μ-total as predictive molecular descriptors for the aggregation temperature of different RLP analogs. The goodness-of-fit is indicated by R^2 and Q^2 for a leave-one-one (LOO) cross validation. The equations predicting the transition temperature, UCST and $UCST_{norm}$, are indicated for raw and normalized parameters (by subtracting the respective mean and dividing by the standard deviation), respectively.

incorporation of the selected unnatural side-chain alongside the entire set of the 20 natural AAs. The distinct advantage of this approach is particularly valuable in the design of ELP-fusion proteins wherein unintended AA modifications can result in suboptimal PBP properties or reduce or abolish the activity of the fused protein (Costa et al., 2019; Vanderschuren et al., 2022).

Multi-Site ncAA Incorporation in RLPs

While a few previous studies have reported the effects of ncAA incorporation or side-chain modifications on the properties and behaviors of ELPs (Kim et al., 2005; Kim et al., 2006; Catherine et al., 2015; Kramer et al., 2015; Petitdemange et al., 2017; Rosselin et al., 2019), we are not aware of any similar analyses

of such alterations in other PBPs. To examine the effect of ncAA incorporation on the UCST-type phase transition behavior, we turned our attention to a recently described family of RLPs constructed from tandem repeats of the GRGDSPYS peptide (Dzuricky et al., 2020). We produced a series of proteins from the $GRGDSPYS_{40}$(6TAG) gene (**Supplementary Table S3**), with either tyrosine (which results in the original "wild type" sequence) or ncAAs 1–15 encoded in the TAG codons. Protein purity and MW were determined by SDS-PAGE and intact mass-spectrometry (**Supplementary Table S5** and **Supplementary Figure S9**).

We then examined the effect of the different ncAAs on the UCST of this set of RLPs (**Figures 3A**,**B**, **Supplementary Figure S10**). The incorporation of more hydrophilic ncAAs, as estimated by predicted logD values, typically resulted in a reduction of the UCST, such that the solubilization of the protein occurred at lower temperatures (**Figures 3A**,**B**) (we note that all the RLP variants appeared to self-assemble above their UCST, **Supplementary Figure S11**). As expected, this effect was generally opposite to that displayed by the ELPs, wherein the more hydrophilic ncAAs increased the LCST, such that the aggregation of the protein occurred at higher temperatures. Indeed, the correlation ($R^2 = 0.77$) between the predicted logD values and the UCST of the RLPs indicated that increasing ncAA hydrophobicity generally increases the UCST of the RLP (**Figure 3C**).

Although hydrophobic interactions broadly explain the aggregation of different RLP analogs at various temperatures, it is well established that other interactions (e.g., π-π, cation-π, hydrogen bonds) also strongly affect the UCST. Therefore, we were interested in utilizing the data collected on this large set of aromatic substitutions to gain a more in-depth understanding of their effect on the phase-transition behavior at the molecular level. To this end, a multivariate model was identified to elucidate the specific molecular features of the RLP derivatives that may influence their UCST. Based on the versatile modifications on the aromatic ring of the tyrosine side chain, we evaluated the steric, electronic, and stereo-electronic molecular descriptors, such as Sterimol parameters (Verloop et al., 1976), vibration frequencies and intensities (Milo et al., 2014), dipole moment and NBO charges (Glendening et al., 2012) (**Supplementary Appendix 1**). The best-fitted model included the total dipole moment of the side-chain and two steric Sterimol parameters, B1 and B5, representing the minimal and maximal width of the substituent, respectively (**Figure 3D**). Overall, wider analogs with decreased polarity, reflected as larger B1 and B5 values and a smaller total dipole variable, led to an increased transition temperature. We speculate that the decreased polarity could reflect hydrophobicity, whereas the width parameters could represent steric hindrance. These results suggest a decreased transition temperature for both more hydrophilic residues and for less sterically hindered substituents that do not impede hydrogen bonding with the aqueous solution.

Several studies demonstrated that the number, position, and identity of aromatic AAs have a significant effect on the UCST of RLPs and other natural and artificial intrinsically disordered proteins\regions (IDP\Rs) known to exhibit UCST-type behavior. Specifically, an extensive characterization of RLPs composed of the "GRGDSPYS" motif, and the effect of substitutions of tyrosine residues to other aromatic or non-aromatic AAs, revealed that the UCST was substantially effected by the identity of the aromatic AA in the repeat sequence, and by an arginine-to-lysine substitution, which affects cation-π interactions with the aromatic AAs (Dzuricky et al., 2020). Here, we show that substituting only six of the 40 tyrosine residues in the repeating $GRGDSPYS_{40}$ protein sequence can generate a difference of ~25°C in the UCST of these PBPs. Likewise, it is expected that ncAA incorporation will enable the tuning of the UCST of other PBPs and IDP/Rs whose properties can be altered by aromatic AAs. For example, mutations of aromatic and positively charged AAs have been shown to modulate the UCST of several natural IDP/Rs, such as proteins derived from the Fused in Sarcoma (FUS) family (Lin et al., 2017; Wang et al., 2018), IDR regions of Ddx4 (Nott et al., 2015; Brady et al., 2017), and the intrinsically disordered RGG domain of the LAF-1 protein (a component of P granules) (Schuster et al., 2020). We expect that the ability to incorporate multiple instances of various aromatic ncAAs in these and many other PBPs and IDPs will continue to inform the molecular-level behavior of their phase transition and enable the rational tuning of their UCSTs and, thereby, also of their self-assembly propensities.

CONCLUSION

We report a suit of high-performance aaRSs for the multi-site incorporation of 15 different aromatic ncAAs. Furthermore, our comparative analysis of 11 aaRS variants that evolved from a single ancestor (MjTyrRS), using an identical mutagenesis strategy, enabled the analysis of the mutational landscape at the ncAA-binding site which can inform future aaRS evolution efforts. This analysis indicates a distinct pattern of mutations found in the ncAA-binding site and shows that the mutagenesis of only nine of the 12 residues targeted in our library designs may suffice for the selection of efficient aaRS variants for the incorporation of para-substituted aromatic ncAAs. The application of this knowledge toward the creation of more focused libraries will enable the exploration of a larger subset of the relevant theoretical library space, and thus the selection of more efficient variants (Hauf et al., 2017; Baumann et al., 2019). Employing these capabilities, we were able to provide the first investigation of the effect of various unnatural aromatic groups on the ELP LCST- and RLP UCST-type phase-transition behaviors. We show that the incorporation of aromatic ncAAs can generate, from a single DNA template, a family of ELP proteins that have a wide range (spanning >60°C) of LCSTs. Further analysis of the ELP properties reveals that ncAA incorporation affects both the secondary structure of the ELPs and, for some ncAAs, the propensity to self-assemble. Moreover, we show that aromatic ncAA incorporation can also be utilized to tune the phase-transition behavior of RLPs, and that this behavior is strongly affected by both the hydrophobicity and the size of the ncAA side-chain. Given the recognized importance of aromatic residues for the phase transition behavior—primarily of the UCST-type, but also of the LCST-type—the ability to efficiently encode aromatic ncAAs alongside the entire set of 20 natural AAs will allow access to new families of precise and chemically diverse PBPs and PBP-protein fusions. Expanding the chemical diversity of aromatic side chains incorporated in ELP- and RLP-based PBPs will permit further investigation of the effect of aromatic character on the different

phase-transition behaviors and advance our understanding of the functions and sequence determinants of LCST- and UCST-type phase transitions in natural and artificial IDPs and PBPs (Ruff et al., 2018). Beyond the specific families explored in this work, these capabilities will empower future studies of the effect of ncAA incorporation on the self-assembly propensities and morphologies of multi-block polymers composed of domains derived from ELPs, RLPs, or other PBPs. We expect that further elucidation of the sequence and molecular determinants of these and many other PBPs, IDP/Rs, and ncAAs will enable the rational design of semi-synthetic polymers with bespoke new or improved properties.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

OG and OS contributed equally. OG, OS, RA, DSS, DH, and SG carried out the experiments, HS and AM conducted the molecular modeling analysis, MA supervised the study and wrote the manuscript with feedback from all authors.

FUNDING

This work was funded by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation program, grant agreement 756996 and by the Israel Science Foundation (grant number 939\21) (to MA). MA gratefully acknowledges support from the Elaine S. and Alvin W. Wene Career Development Chair in Biotechnology Engineering. HS acknowledges support from the Emergency Postdoctoral Fellowships for Israeli Researchers in Israel.

ACKNOWLEDGMENTS

We thank Dr. Mark Karpasas from the Ilse Katz Institute for Nanoscale Science and Technology for the professional help with the mass spectrometry experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.913057/full#supplementary-material

REFERENCES

Amiram, M., Haimovich, A. D., Fan, C., Wang, Y.-S., Aerni, H.-R., Ntai, I., et al. (2015). Evolution of Translation Machinery in Recoded Bacteria Enables Multi-Site Incorporation of Nonstandard Amino Acids. *Nat. Biotechnol.* 33 (12), 1272–1279. doi:10.1038/nbt.3372

Amiram, M., Luginbuhl, K. M., Li, X., Feinglos, M. N., and Chilkoti, A. (2013). A Depot-Forming Glucagon-like Peptide-1 Fusion Protein Reduces Blood Glucose for Five Days with a Single Injection. *J. Control. Release* 172 (1), 144–151. doi:10.1016/j.jconrel.2013.07.021

Balu, R., Dutta, N. K., Dutta, A. K., and Choudhury, N. R. (2021). Resilin-mimetics as a Smart Biomaterial Platform for Biomedical Applications. *Nat. Commun.* 12 (1), 149. doi:10.1038/s41467-020-20375-x

Baumann, T., Hauf, M., Richter, F., Albers, S., Möglich, A., Ignatova, Z., et al. (2019). Computational Aminoacyl-tRNA Synthetase Library Design for Photocaged Tyrosine. *Ijms* 20 (9), 2343. doi:10.3390/ijms20092343

Bhattacharyya, J., Bellucci, J. J., and Chilkoti, A. (2016). "Elastin-Like Polypeptides: Bio-Inspired Smart Polymers for Protein Purification, Drug Delivery and Tissue Engineering," in *Biomaterials from Nature for Advanced Devices and Therapies* (Hoboken, NJ: John Wiley & Sons, Inc.), 106–126. doi:10.1002/9781119126218.ch7

Brady, J. P., Farber, P. J., Sekhar, A., Lin, Y.-H., Huang, R., Bah, A., et al. (2017). Structural and Hydrodynamic Properties of an Intrinsically Disordered Region of a Germ Cell-specific Protein on Phase Separation. *Proc. Natl. Acad. Sci. U.S.A.* 114 (39), E8194–E8203. doi:10.1073/pnas.1706197114

Catherine, C., Oh, S. J., Lee, K.-H., Min, S.-E., Won, J.-I., Yun, H., et al. (2015). Engineering Thermal Properties of Elastin-like Polypeptides by Incorporation of Unnatural Amino Acids in a Cell-free Protein Synthesis System. *Biotechnol. Bioproc E* 20 (3), 417–422. doi:10.1007/s12257-015-0190-1

Chang, M. P., Huang, W., and Mai, D. J. (2021). Monomer-scale Design of Functional Protein Polymers Using Consensus Repeat Sequences. *J. Polym. Sci.* 59 (22), 2644–2664. doi:10.1002/pol.20210506

Chen, H., Venkat, S., McGuire, P., Gan, Q., and Fan, C. (2018). Recent Development of Genetic Code Expansion for Posttranslational Modification Studies. *Molecules* 23 (7), 1662. doi:10.3390/molecules23071662

Chin, J. W. (2017). Expanding and Reprogramming the Genetic Code. *Nature* 550 (7674), 53–60. doi:10.1038/nature24031

Connor, R. E., and Tirrell, D. A. (2007). Non-Canonical Amino Acids in Protein Polymer Design. *Polym. Rev.* 47 (1), 9–28. doi:10.1080/15583720601109552

Costa, S. A., Mozhdehi, D., Dzuricky, M. J., Isaacs, F. J., Brustad, E. M., and Chilkoti, A. (2019). Active Targeting of Cancer Cells by Nanobody Decorated Polypeptide Micelle with Bio-Orthogonally Conjugated Drug. *Nano Lett.* 19 (1), 247–254. doi:10.1021/acs.nanolett.8b03837

Dai, M., Georgilis, E., Goudounet, G., Garbay, B., Pille, J., van Hest, J. C. M., et al. (2021). Refining the Design of Diblock Elastin-Like Polypeptides for Self-Assembly into Nanoparticles. *Polymers* 13 (9), 1470. doi:10.3390/polym13091470

Debelouchina, G. T., and Muir, T. W. (2017). A Molecular Engineering Toolbox for the Structural Biologist. *Q. Rev. Biophysics* 50, e7. doi:10.1017/s0033583517000051

Despanie, J., Dhandhukia, J. P., Hamm-Alvarez, S. F., and MacKay, J. A. (2016). Elastin-like Polypeptides: Therapeutic Applications for an Emerging Class of Nanomedicines. *J. Control. Release* 240, 93–108. doi:10.1016/j.jconrel.2015.11.010

Dhandhukia, J. P., Brill, D. A., Kouhi, A., Pastuszka, M. K., and MacKay, J. A. (2017). Elastin-like Polypeptide Switches: A Design Strategy to Detect Multimeric Proteins. *Protein Sci.* 26 (9), 1785–1795. doi:10.1002/pro.3215

Dhandhukia, J., Weitzhandler, I., Wang, W., and MacKay, J. A. (2013). Switchable Elastin-Like Polypeptides that Respond to Chemical Inducers of Dimerization. *Biomacromolecules* 14 (4), 976–985. doi:10.1021/bm301558q

Dreher, M. R., Simnick, A. J., Fischer, K., Smith, R. J., Patel, A., Schmidt, M., and Chilkoti, A. (2008). Temperature Triggered Self-Assembly of Polypeptides into Multivalent Spherical Micelles. *J. Am. Chem. Soc.* 130 (2), 687–694. doi:10.1021/ja0764862

Dzuricky, M., Rogers, B. A., Shahid, A., Cremer, P. S., and Chilkoti, A. (2020). De Novo engineering of Intracellular Condensates Using Artificial Disordered Proteins. *Nat. Chem.* 12 (9), 814–825. doi:10.1038/s41557-020-0511-7

Glendening, E. D., Landis, C. R., and Weinhold, F. (2012). Natural Bond Orbital Methods. *WIREs Comput. Mol. Sci.* 2 (1), 1–42. doi:10.1002/wcms.51

Hadar, D., Gelkop, S., Vaserman, L., and Amiram, M. (2021). Efficient Incorporation of Clickable Unnatural Amino Acids Enables Rapid and Biocompatible Labeling of Proteins *In Vitro* and in Bacteria. *Chembiochem* 22, 1379–1384. doi:10.1002/cbic.202000663

Hassouneh, W., Christensen, T., and Chilkoti, A. (2010). Elastin-Like Polypeptides as a Purification Tag for Recombinant Proteins. *Curr. Protoc. Protein Sci.* 6 (61), 1–20. Unit 6 11. doi:10.1002/0471140864.ps0611s61

Hauf, M., Richter, F., Schneider, T., Faidt, T., Martins, B. M., Baumann, T., et al. (2017). Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code. *Chembiochem* 18 (18), 1819–1823. doi:10.1002/cbic.201700327

Hayashi, A., Haruna, K. i., Sato, H., Ito, K., Makino, C., Ito, T., et al. (2021). Incorporation of Halogenated Amino Acids into Antibody Fragments at Multiple Specific Sites Enhances Antigen Binding. *Chembiochem* 22 (1), 120–123. doi:10.1002/cbic.202000429

Huang, Y., and Liu, T. (2018). Therapeutic Applications of Genetic Code Expansion. *Synthetic Syst. Biotechnol.* 3 (3), 150–158. doi:10.1016/j.synbio.2018.09.003

Israeli, B., Strugach, D. S., Gelkop, S., Weber, S., Gozlan, D. S., and Amiram, M. (2021). Genetically Encoding Light-Responsive Protein-Polymers Using Translation Machinery for the Multi-Site Incorporation of Photo-Switchable Unnatural Amino Acids. *Adv. Funct. Mater.* 31, 1–14. doi:10.1002/adfm.202011276

Israeli, B., Vaserman, L., and Amiram, M. (2019). Multi-Site Incorporation of Nonstandard Amino Acids into Protein-Based Biomaterials. *Isr. J. Chem.* 60, 1118–1128. doi:10.1002/ijch.201900043

Janib, S. M., Pastuszka, M. F., Aluri, S., Folchman-Wagner, Z., Hsueh, P. Y., Shi, P., et al. (2014). A Quantitative Recipe for Engineering Protein Polymer Nanoparticles. *Polym. Chem.* 5 (5), 1614–1625. doi:10.1039/c3py00537b

Johnson, J. A., Lu, Y. Y., Van Deventer, J. A., and Tirrell, D. A. (2010). Residue-specific Incorporation of Non-canonical Amino Acids into Proteins: Recent Developments and Applications. *Curr. Opin. Chem. Biol.* 14 (6), 774–780. doi:10.1016/j.cbpa.2010.09.013

Joseph, J. A., Reinhardt, A., Aguirre, A., Chew, P. Y., Russell, K. O., Espinosa, J. R., et al. (2021). Physics-driven Coarse-Grained Model for Biomolecular Phase Separation with Near-Quantitative Accuracy. *Nat. Comput. Sci.* 1 (11), 732–743. doi:10.1038/s43588-021-00155-3

Kim, W., Hardcastle, K. I., and Conticello, V. P. (2006). Fluoroproline Flip-Flop: Regiochemical Reversal of a Stereoelectronic Effect on Peptide and Protein Structures. *Angew. Chem. Int. Ed.* 45 (48), 8141–8145. doi:10.1002/anie.200603227

Kim, W., McMillan, R. A., Snyder, J. P., and Conticello, V. P. (2005). A Stereoelectronic Effect on Turn Formation Due to Proline Substitution in Elastin-Mimetic Polypeptides. *J. Am. Chem. Soc.* 127 (51), 18121–18132. doi:10.1021/ja054105j

Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., et al. (2003). Structural Basis for Orthogonal tRNA Specificities of Tyrosyl-tRNA Synthetases for Genetic Code Expansion. *Nat. Struct. Mol. Biol.* 10 (6), 425–432. doi:10.1038/nsb934

Kramer, J. R., Petitdemange, R., Bataille, L., Bathany, K., Wirotius, A.-L., Garbay, B., et al. (2015). Quantitative Side-Chain Modifications of Methionine-Containing Elastin-Like Polypeptides as a Versatile Tool to Tune Their Properties. *ACS Macro Lett.* 4 (11), 1283–1286. doi:10.1021/acsmacrolett.5b00651

Kubyshkin, V. (2021). Experimental Lipophilicity Scale for Coded and Noncoded Amino Acid Residues. *Org. Biomol. Chem.* 19 (32), 7031–7040. doi:10.1039/d1ob01213d

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., et al. (2013). Genomically Recoded Organisms Expand Biological Functions. *Science* 342 (6156), 357–360. doi:10.1126/science.1241459

Lee, K. J., Kang, D., and Park, H. S. (2019). Site-Specific Labeling of Proteins Using Unnatural Amino Acids. *Mol. Cells* 42 (5), 386–396. doi:10.14348/molcells.2019.0078

Lin, Y., Currie, S. L., and Rosen, M. K. (2017). Intrinsically Disordered Sequences Enable Modulation of Protein Phase Separation through Distributed Tyrosine Motifs. *J. Biol. Chem.* 292 (46), 19110–19120. doi:10.1074/jbc.M117.800466

MacEwan, S. R., and Chilkoti, A. (2014). Applications of Elastin-like Polypeptides in Drug Delivery. *J. Control. Release* 190, 314–330. doi:10.1016/j.jconrel.2014.06.028

MacEwan, S. R., and Chilkoti, A. (2010). Elastin-like Polypeptides: Biomedical Applications of Tunable Biopolymers. *Biopolymers* 94 (1), 60–77. doi:10.1002/bip.21327

Martin, E. W., and Mittag, T. (2018). Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* 57 (17), 2478–2487. doi:10.1021/acs.biochem.8b00008

McDaniel, J. R., Radford, D. C., and Chilkoti, A. (2013). A Unified Model for De Novo Design of Elastin-like Polypeptides with Tunable Inverse Transition Temperatures. *Biomacromolecules* 14 (8), 2866–2872. doi:10.1021/bm4007166

Meyer, D. E., and Chilkoti, A. (2002). Genetically Encoded Synthesis of Protein-Based Polymers with Precisely Specified Molecular Weight and Sequence by Recursive Directional Ligation: Examples from the Elastin-like Polypeptide System. *Biomacromolecules* 3 (2), 357–367. doi:10.1021/bm015630n

Meyer, D. E., and Chilkoti, A. (2004). Quantification of the Effects of Chain Length and Concentration on the Thermal Behavior of Elastin-like Polypeptides. *Biomacromolecules* 5 (3), 846–851. doi:10.1021/bm034215n

Milo, A., Bess, E. N., and Sigman, M. S. (2014). Interrogating Selectivity in Catalysis Using Molecular Vibrations. *Nature* 507 (7491), 210, 214. doi:10.1038/nature13019

Neumann, H., Hazen, J. L., Weinstein, J., Mehl, R. A., and Chin, J. W. (2008). Genetically Encoding Protein Oxidative Damage. *J. Am. Chem. Soc.* 130 (12), 4028–4033. doi:10.1021/ja710100d

Nguyen, T.-A., Cigler, M., and Lang, K. (2018). Expanding the Genetic Code to Study Protein-Protein Interactions. *Angew. Chem. Int. Ed.* 57 (44), 14350–14361. doi:10.1002/anie.201805869

Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., et al. (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* 57 (5), 936–947. doi:10.1016/j.molcel.2015.01.013

O'Donoghue, P., Ling, J., Wang, Y.-S., and Söll, D. (2013). Upgrading Protein Synthesis for Synthetic Biology. *Nat. Chem. Biol.* 9 (10), 594–598. doi:10.1038/nchembio.1339

Ohtake, K., Yamaguchi, A., Mukai, T., Kashimura, H., Hirano, N., Haruki, M., et al. (2015). Protein Stabilization Utilizing a Redefined Codon. *Sci. Rep.* 5, 9762. doi:10.1038/srep09762

Pak, C. W., Kosno, M., Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., et al. (2016). Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol. Cell* 63 (1), 72–85. doi:10.1016/j.molcel.2016.05.042

Petitdemange, R., Garanger, E., Bataille, L., Dieryck, W., Bathany, K., Garbay, B., et al. (2017). Selective Tuning of Elastin-like Polypeptide Properties via Methionine Oxidation. *Biomacromolecules* 18 (2), 544–550. doi:10.1021/acs.biomac.6b01696

Prhashanna, A., Taylor, P. A., Qin, J., Kiick, K. L., and Jayaraman, A. (2019). Effect of Peptide Sequence on the LCST-Like Transition of Elastin-Like Peptides and Elastin-Like Peptide-collagen-Like Peptide Conjugates: Simulations and Experiments. *Biomacromolecules* 20 (3), 1178–1189. doi:10.1021/acs.biomac.8b01503

Rezhdo, A., Islam, M., Huang, M., and Van Deventer, J. A. (2019). Future Prospects for Noncanonical Amino Acids in Biological Therapeutics. *Curr. Opin. Biotechnol.* 60, 168–178. doi:10.1016/j.copbio.2019.02.020

Rosselin, M., Xiao, Y., Belhomme, L., Lecommandoux, S., and Garanger, E. (2019). Expanding the Toolbox of Chemoselective Modifications of Protein-Like Polymers at Methionine Residues. *ACS Macro Lett.* 8 (12), 1648–1653. doi:10.1021/acsmacrolett.9b00862

Ruff, K. M., Roberts, S., Chilkoti, A., and Pappu, R. V. (2018). Advances in Understanding Stimulus-Responsive Phase Behavior of Intrinsically Disordered Protein Polymers. *J. Mol. Biol.* 430 (23), 4619–4635. doi:10.1016/j.jmb.2018.06.031

Sakamoto, K., Murayama, K., Oki, K., Iraha, F., Kato-Murayama, M., Takahashi, M., et al. (2009). Genetic Encoding of 3-Iodo-L-Tyrosine in *Escherichia coli* for Single-Wavelength Anomalous Dispersion Phasing in Protein Crystallography. *Structure* 17 (3), 335–344. doi:10.1016/j.str.2009.01.008

Schuster, B. S., Dignon, G. L., Tang, W. S., Kelley, F. M., Ranganath, A. K., Jahnke, C. N., et al. (2020). Identifying Sequence Perturbations to an

Intrinsically Disordered Protein that Determine its Phase-Separation Behavior. *Proc. Natl. Acad. Sci. U.S.A.* 117 (21), 11421–11431. doi:10.1073/pnas.2000223117

Seifried, B. M., Cao, J., and Olsen, B. D. (2018). Multifunctional, High Molecular Weight, Post-Translationally Modified Proteins through Oxidative Cysteine Coupling and Tyrosine Modification. *Bioconjugate Chem.* 29 (6), 1876–1884. doi:10.1021/acs.bioconjchem.7b00834

Taylor, P. A., Huang, H., Kiick, K. L., and Jayaraman, A. (2020). Placement of Tyrosine Residues as a Design Element for Tuning the Phase Transition of Elastin-Peptide-Containing Conjugates: Experiments and Simulations. *Mol. Syst. Des. Eng.* 5 (7), 1239–1254. doi:10.1039/d0me00051e

Urry, D. W., Gowda, D. C., Parker, T. M., Luan, C.-H., Reid, M. C., Harris, C. M., et al. (1992). Hydrophobicity Scale for Proteins Based on Inverse Temperature Transitions. *Biopolymers* 32 (9), 1243–1250. doi:10.1002/bip.360320913

Vanderschuren, K., Arranz-Gibert, P., Khang, M., Hadar, D., Gaudin, A., Yang, F., et al. (2022). Tuning Protein Half-Life in Mouse Using Sequence-Defined Biopolymers Functionalized with Lipids. *Proc. Natl. Acad. Sci. U.S.A.* 119 (4). doi:10.1073/pnas.2103099119

Varanko, A., Saha, S., and Chilkoti, A. (2020). Recent Trends in Protein and Peptide-Based Biomaterials for Advanced Drug Delivery. *Adv. Drug Deliv. Rev.* 156, 133–187. doi:10.1016/j.addr.2020.08.008

Vargas-Rodriguez, O., Sevostyanova, A., Söll, D., and Crnković, A. (2018). Upgrading Aminoacyl-tRNA Synthetases for Genetic Code Expansion. *Curr. Opin. Chem. Biol.* 46, 115–122. doi:10.1016/j.cbpa.2018.07.014

Verloop, A., Hoogenstraaten, W., and Tipker, J. (1976). "Chapter 4 - Development and Application of New Steric Substituent Parameters in Drug Design," in *Drug Design*. Editor E. J. Ariëns (Amsterdam: Academic Press), 165–207. doi:10.1016/b978-0-12-060307-7.50010-9

Völler, J.-S., and Budisa, N. (2017). Coupling Genetic Code Expansion and Metabolic Engineering for Synthetic Cells. *Curr. Opin. Biotechnol.* 48, 1–7. doi:10.1016/j.copbio.2017.02.002

Wan, W., Tharp, J. M., and Liu, W. R. (2014). Pyrrolysyl-tRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochimica Biophysica Acta (BBA) - Proteins Proteomics* 1844 (6), 1059–1070. doi:10.1016/j.bbapap.2014.03.002

Wang, B., Patkar, S. S., and Kiick, K. L. (2021). Application of Thermoresponsive Intrinsically Disordered Protein Polymers in Nanostructured and Microstructured Materials. *Macromol. Biosci.* 21 (9), 1–20. doi:10.1002/mabi.202100129

Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., et al. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 174 (3), 688–699. e616. doi:10.1016/j.cell.2018.06.006

Willis, J. C. W., and Chin, J. W. (2018). Mutually Orthogonal Pyrrolysyl-tRNA Synthetase/tRNA Pairs. *Nat. Chem.* 10 (8), 831–837. doi:10.1038/s41557-018-0052-5

Young, D. D., and Schultz, P. G. (2018). Playing with the Molecules of Life. *ACS Chem. Biol.* 13 (4), 854–870. doi:10.1021/acschembio.7b00974

Young, D. D., Young, T. S., Jahnz, M., Ahmad, I., Spraggon, G., and Schultz, P. G. (2011). An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity,. *Biochemistry* 50 (11), 1894–1900. doi:10.1021/bi101929e

Zhou, W., and Deiters, A. (2021). Chemogenetic and Optogenetic Control of Post-translational Modifications through Genetic Code Expansion. *Curr. Opin. Chem. Biol.* 63, 123–131. doi:10.1016/j.cbpa.2021.02.016

Conflict of Interest: MA, DH, DSS, and SG filed a patent related to this technology.

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

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

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 31 August 2022
DOI 10.3389/fmolb.2022.938114

OPEN ACCESS

EDITED BY
Nediljko Budisa,
University of Manitoba, Canada

REVIEWED BY
Lluís Ribas De Pouplana,
Institute for Research in Biomedicine, Spain
Ute Kothe,
University of Manitoba, Canada

*CORRESPONDENCE
John D. Fisk,
john.fisk@ucdenver.edu

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

RECEIVED 07 May 2022
ACCEPTED 27 July 2022
PUBLISHED 31 August 2022

CITATION
Tittle JM, Schwark DG, Biddle W, Schmitt MA and Fisk JD (2022), Impact of queuosine modification of endogenous E. coli tRNAs on sense codon reassignment.
Front. Mol. Biosci. 9:938114.
doi: 10.3389/fmolb.2022.938114

Impact of queuosine modification of endogenous *E. coli* tRNAs on sense codon reassignment

Jillyn M. Tittle, David G. Schwark, Wil Biddle, Margaret A. Schmitt and John D. Fisk*

Department of Chemistry, University of Colorado Denver, Denver, CO, United States

The extent to which alteration of endogenous tRNA modifications may be exploited to improve genetic code expansion efforts has not been broadly investigated. Modifications of tRNAs are strongly conserved evolutionarily, but the vast majority of *E. coli* tRNA modifications are not essential. We identified queuosine (Q), a non-essential, hypermodified guanosine nucleoside found in position 34 of the anticodons of four *E. coli* tRNAs as a modification that could potentially be utilized to improve sense codon reassignment. One suggested purpose of queuosine modification is to reduce the preference of tRNAs with guanosine (G) at position 34 of the anticodon for decoding cytosine (C) ending codons over uridine (U) ending codons. We hypothesized that introduced orthogonal translation machinery with adenine (A) at position 34 would reassign U-ending codons more effectively in queuosine-deficient *E. coli*. We evaluated the ability of introduced orthogonal tRNAs with AUN anticodons to reassign three of the four U-ending codons normally decoded by Q34 endogenous tRNAs: histidine CAU, asparagine AAU, and aspartic acid GAU in the presence and absence of queuosine modification. We found that sense codon reassignment efficiencies in queuosine-deficient strains are slightly improved at Asn AAU, equivalent at His CAU, and less efficient at Asp GAU codons. Utilization of orthogonal pair-directed sense codon reassignment to evaluate competition events that do not occur in the standard genetic code suggests that tRNAs with inosine (I, 6-deaminated A) at position 34 compete much more favorably against G34 tRNAs than Q34 tRNAs. Continued evaluation of sense codon reassignment following targeted alterations to endogenous tRNA modifications has the potential to shed new light on the web of interactions that combine to preserve the fidelity of the genetic code as well as identify opportunities for exploitation in systems with expanded genetic codes.

KEYWORDS

genetic code expansion, sense codon reassignment, tRNA modification, queuosine modification, synthetic biology, orthogonal translation machinery

1 Introduction

The incorporation of an amino acid into a growing peptide chain is the ultimate outcome of a series of competition events. Amino acids, tRNAs, and aminoacyl tRNA synthetases (aaRSs) identify their cognate partners to appropriately charge each tRNA. In the ribosome, each mRNA codon is sampled by the complement of aminoacylated tRNAs in the cell. Following recognition between the codon-anticodon pair, the amino acid may be incorporated into the growing protein chain. Thermodynamic interactions interpreted through kinetic decisions by the ribosome and elongation factors enable a cognate tRNA to progress through the successive steps of translation more efficiently than a non-cognate tRNA (Rodnina and Wintermeyer, 2001; Fluitt et al., 2007).

Expansion of the genetic code via incorporation of noncanonical amino acids (ncAAs) in response to sense or stop codons is the outcome of the same set of competition events, complicated by the addition of orthogonal translation machinery (tRNA/aaRS pair). Increasing the efficiency of codon reassignment by an orthogonal pair may be achieved by altering one or more of the steps in the progression from tRNA aminoacylation to peptide bond formation in the ribosome. A significant barrier to broadly expanding the genetic code is the incomplete quantitative understanding of the factors that control the fidelity of the genetic code, e.g. codon-anticodon interaction energies, aminoacylation efficiency, relative tRNA concentrations, and the influence of tRNA modifications on each of these components. As the genetic code is engineered to allow codon-specific incorporation of multiple copies of multiple ncAAs into a single protein, the ability to improve orthogonal pair competition by modulating precise functions of the endogenous translation machinery will become increasingly important.

The fidelity of protein translation is the extent to which the "correct" tRNA successfully incorporates its charged amino acid and "incorrect" tRNAs are rejected at some point along the path. Misincorporation of amino acids into proteins occurs at a rate of approximately 1 error in 1,000 to 10,000 peptide bond forming events (Parker, 1989; Kramer and Farabaugh, 2007). While missense errors are thought to be generally destabilizing, missense mutations induced by antibiotic treatment, defective aaRS editing, or sense codon reassignment are broadly tolerated by cells (Min et al., 2003; Ruan et al., 2008; Schmitt et al., 2018). The missense error rate is a function of the complement of translational machinery present in the cell. Missense errors differ across codons in the same cell. Missense errors at individual codons differ with growth rate and across organisms (Parker, 1989; Kramer and Farabaugh, 2007).

In *E. coli*, tRNA species with 40 anticodon sequences decode the 61 sense codons. Most tRNA species are tasked with decoding more than a single sense codon. The pairing of tRNA anticodon to mRNA codon is often mediated by modified nucleotides in the tRNA. *E. coli* is one of the few organisms for which the complete

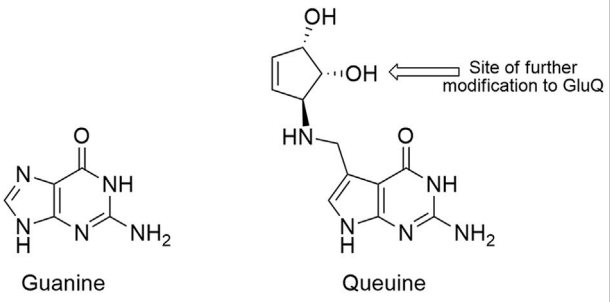


FIGURE 1
Structures of guanine (G, left) and the modified guanine, queuine (Q, right). The arrow on queuine indicates the position of glutamic acid esterification, a further modification found in the *E. coli* aspartic acid tRNA.

set of tRNA modifications and modifying enzymes have been identified. *E. coli* tRNA species contain, on average, 8 modified bases each, nearly 12% of the entire tRNA molecule (El Yacoubi et al., 2012; Björk et al., 2014). The anticodon stem loop is the most highly and most diversely-modified region of tRNAs. Within the anticodon, position 34, which interacts with the third/wobble position of the mRNA codon, may be modified to one of 14 different non-UCAG nucleotides. tRNA modifications serve multiple functions in translation, including modulating the tRNA-protein, tRNA-mRNA, and tRNA-ribosome interactions that combine to determine the fidelity of translation.

Certain position 34 tRNA anticodon modifications serve to expand the space of recognition to codons not typically decoded, while other modified anticodons restrict decoding to prevent missense errors (reviewed in: El Yacoubi et al., 2012; Björk et al., 2014). *E. coli* tRNAs with uridine at position 34 (U34) feature two types of modifications, largely determined by whether the tRNA is part of a four-codon box (one amino acid for all codons in the box) or a two-codon box (U3/C3 codons encode one amino acid; A3/G3 codons encode another). In most four boxes, uridine-5-oxyacetic acid at position 34 allows a single tRNA to decode the expected A3 and G3 codons as well as the U3 codon. In two boxes, 5-methylaminomethyl-2-thiouridine at position 34 restricts decoding to the A3 and G3 codons only because the U3 codon encodes a different amino acid. A single *E. coli* tRNA, $tRNA^{Arg2}_{ICG}$, includes modification of A34 to inosine. The inosine modification expands anticodon-codon recognition beyond the expected A34/U3 base pairing and allows $tRNA^{Arg2}$ to decode the U3, C3, and A3 codons. A single *E. coli* tRNA, $tRNA^{Ile2}_{LAU}$, is transcribed as C34 and subsequently modified to lysidine. The lysidine modification restricts decoding to the Ile AUA codon only (Lysidine 34/A3 pairing). A U34 tRNA with a UAU anticodon would be expected to also decode the Met AUG codon via a U34/G3 wobble pairing. L34 does not pair with G3 (El Yacoubi et al., 2012; Björk et al., 2014).

The extent to which alteration of endogenous tRNA modifications may be exploited to improve genetic code expansion efforts has not been broadly investigated. Despite having important evolutionarily conserved functions, the vast majority of *E. coli* tRNA modifications are not essential. Of the 29 different modifications encoded by 50 genes, only 4 are essential (6 of 50 genes) (Datsenko and Wanner, 2000; Kitagawa et al., 2005). We identified queuosine (Q, Figure 1), a non-essential, hypermodified guanosine nucleoside found in position 34 of the anticodons of four *E. coli* tRNAs as an anticodon modification that could potentially be exploited to improve sense codon reassignment (Bienz and Kubli, 1981; Dineshkumar et al., 2002). The four NAU/NAC codon pairs [tyrosine (UAU/UAC), histidine (CAU/CAC), asparagine (AAU/AAC), and aspartic acid (GAU/GAC)] are each decoded by a single tRNA species $tRNA_{QUN}$ in *E. coli*. The encoded guanosine at position 34 of each of these tRNAs is replaced by queuosine, one of the larger and more complex nucleotide modifications (Björk et al., 2014).

Eight gene products are involved in queuosine biosynthesis and tRNA modification (Iwata-Reuyl, 2003; Björk et al., 2014). Five enzymes are responsible for the synthesis of the 7-methylaminodeazaguanine precursor base, queuine (queuine is the modified base, queuosine is the modified nucleotide). An enzyme catalyzes the exchange of the queuine base for the genetically-encoded guanine base in appropriate tRNAs. Two additional enzymes function to further decorate the 7-methylamino group of queuine with a ribose-derived cyclopentenediol ring (Figure 1). In aspartic acid tRNAs (QUC anticodon), queuosine is further modified to GluQ by addition of a glutamic acid residue (Salazar et al., 2004). Queuosine modification has been identified in tRNAs having QUN anticodons across most organisms, save yeast, archaea, and *Thermus thermophilus*. Prokaryotes have the ability to synthesize queuine from guanosine. Eukaryotes strip queuine from the anticodons of bacterial tRNAs acquired either through their diet or, in higher organisms, from gut microbiota.

Early studies on queuosine modifications in translation suggest that a primary function may be to reduce the C3 over U3 codon bias of G34 tRNAs. Comparisons of tRNA-tRNA anticodon parings suggested that Q34-C3 pairings were less energetically favorable that G34-C3 pairings while Q34-U3 pairings were more stable than G34-U3 pairings (Grosjean et al., 1978). In *E. coli*, queuosine-modified tyrosine tRNAs appear to sample the ribosome twice as fast as unmodified tRNAs (Noguchi et al., 1982). Evaluation of the *in vivo* decoding properties of Drosophila histidine tRNAs in Xenopus ooctyes suggested that unmodified G34 tRNAs prefer C3 codons while Q34 tRNAs equally decode C- and U-ending codons (Meier et al., 1985). tRNA identity, reading frame maintenance, and prevention of missense errors are additional functions in protein translation ascribed to queuosine modification (Giege et al., 1998; Urbonavicius et al., 2001; Manickam et al., 2014; Manickam et al., 2016). Beyond the direct evaluations of tRNA modifications affecting decoding properties in natural genetic codes, differential tRNA modifications have been proposed as a general cellular response to stress and as a component of the development of cancer and other disease states (Chen and Wu, 1994; Nagaraja et al., 2021). Queuosine modification of tRNAs has been shown to vary across developmental stages of Drosophila (Zaborske et al., 2014).

We hypothesized that the inherent bias of G34 tRNAs for C-ending codons could be exploited to improve the reassignment of U-ending codons (Figure 2). The introduction of an orthogonal tRNA capable of Watson-Crick base pairing to the U-ending codon in queuosine-deficient strains could lead to increased U-ending codon reassignment in two ways: first, by taking advantage of the energetic favorability of orthogonal tRNA A34/U3 base pairing relative to the endogenous G/Q34/U3 wobble, and second, by reducing the efficiency of competition from the endogenous tRNA G34/U3 wobble relative to the Q34/U3 wobble. We evaluated the effects of queuosine knockout on the reassignment efficiency of the *Methanocaldococcus jannaschii* (*M. jannaschii*) tyrosyl tRNA/aaRS orthogonal pair at three of the four U-ending codons read by Q34 *E. coli* tRNAs using our previously-described fluorescence-based screen (Biddle et al., 2015). As our screen relies on differentiating tyrosine incorporation from the naturally encoded amino acid, we are unable to evaluate the tyrosine (UAU/UAC) codon pair.

Prior studies on the effect of queuosine modification in protein translation reveal divergent effects across organisms and among the subset of queuosine-modified tRNAs within a single organism. Our findings bear out these observations. Orthogonal tRNAs with AUN anticodons in queuosine modification-deficient cells exhibit sense codon reassignment efficiencies that are slightly improved at Asn AAU, equivalent at His CAU, and less efficient at Asp GAU codons relative to the parent, non-queuosine-deficient strain. Utilization of an introduced orthogonal pair in our screen also allowed evaluation of base pairing competition events that are not present within standard genetic codes. We found that an orthogonal tRNA whose A34 anticodon is partially modified to inosine competed much more effectively against the queuosine-deficient *E. coli* $tRNA^{His}_{GUG}$ than $tRNA^{His}_{QUG}$ for decoding the His CAU codon.

2 Methods and materials

The supplementary materials file includes detailed experimental protocols for the previously-published fluorescence-based screen. The file also includes general reagents and materials, cell strain details, and a procedure for the preparation of electrocompetent cells (Sambrook and Russell, 2001).

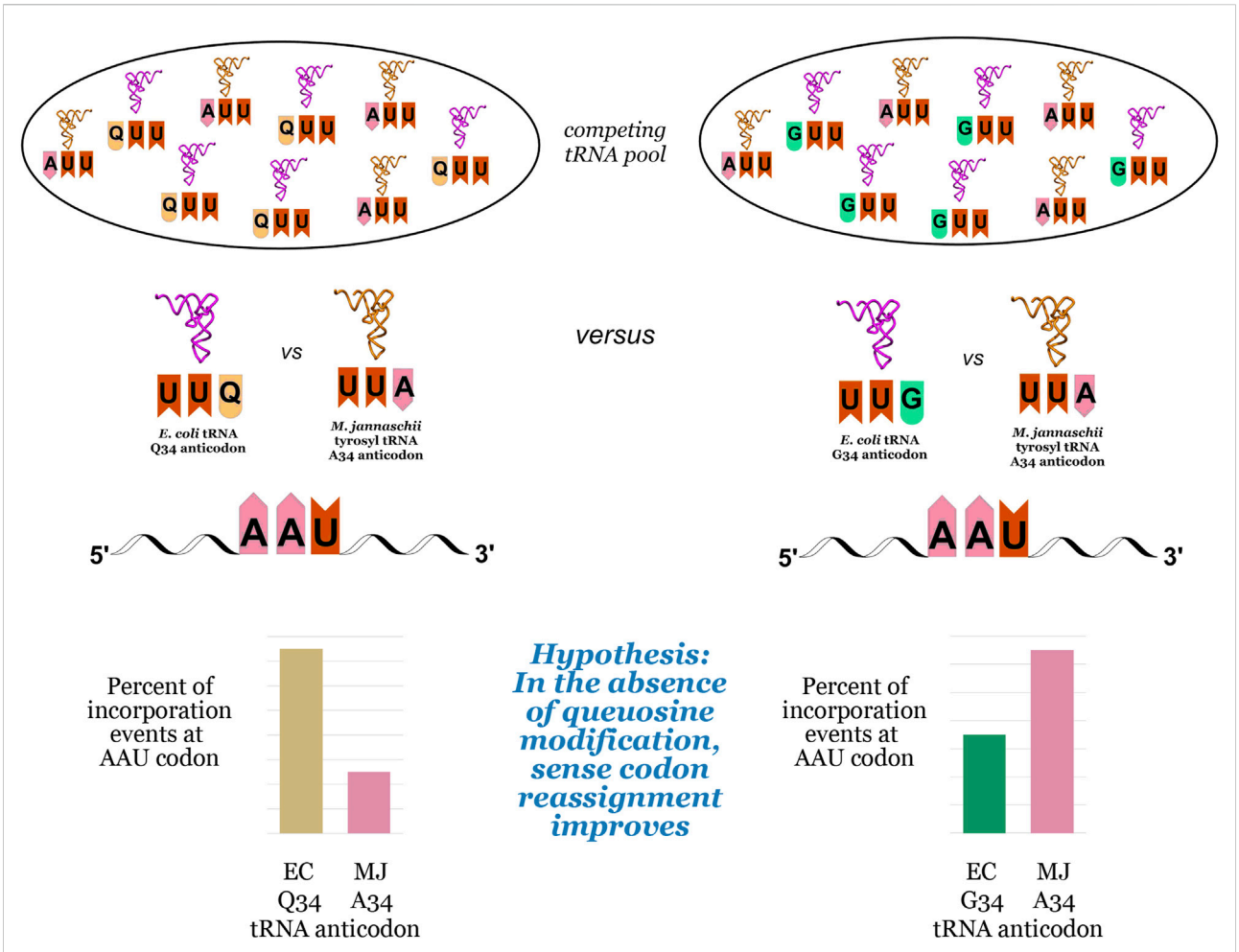


FIGURE 2
Competition events for decoding NAU codons in *E. coli*, using the asparagine AAU codon as an example. The pool of competing tRNAs includes the orthogonal *M. jannaschii* tRNA aminoacylated with tyrosine and the endogenous tRNAs with either queuosine (Q) at position 34 in the anticodon (left) or guanosine (G) at position 34 in the anticodon (right). The pair of bar graphs at the bottom of the figure illustrate the hypothesis that an orthogonal tRNA would more efficiently compete against G34 endogenous tRNAs than Q34 endogenous tRNAs, leading to increased efficiency of sense codon reassignment (SCR).

2.1 Principle of the fluorescence-based screen

Unlike stop codon suppression in which missed incorporations by an orthogonal pair lead to truncated proteins, sense codon reassignment (SCR) results in a heterogeneous mixture of full length proteins. As such, SCR efficiency cannot be readily measured by the yield of full length protein. Rather, the efficiency of incorporation of the amino acid on the orthogonal tRNA relative to that by endogenous tRNAs must be determined via another metric. Our in cell, fluorescence-based screen exploits the absolute requirement of tyrosine at position 66 in green fluorescent protein (GFP) for fluorophore formation. The assay evaluates the extent to which an introduced orthogonal aaRS capable of charging tyrosine to its cognate tRNA is able to reassign the meaning of a test codon placed at the essential fluorophore Tyr position in GFP (Figure 3). The observed per cell fluorescence represents a direct measurement of the orthogonal tRNA's ability to compete against endogenous tRNAs to decode the targeted codon. Sense codon reassignment efficiencies distill down to a single data point the complex interplay of levels of orthogonal and endogenous tRNAs, the interactions between the orthogonal aaRS and its cognate tRNA, and differences in codon-anticodon interaction energies between the introduced and endogenous tRNAs.

2.2 Considerations for screen deployment and reporter gene sequence design

Codon reassignment efficiency is quantified by bracketing the measured per cell fluorescence of each system under evaluation between the measured per cell fluorescence of 0%

10.3389/fmolb.2022.938114

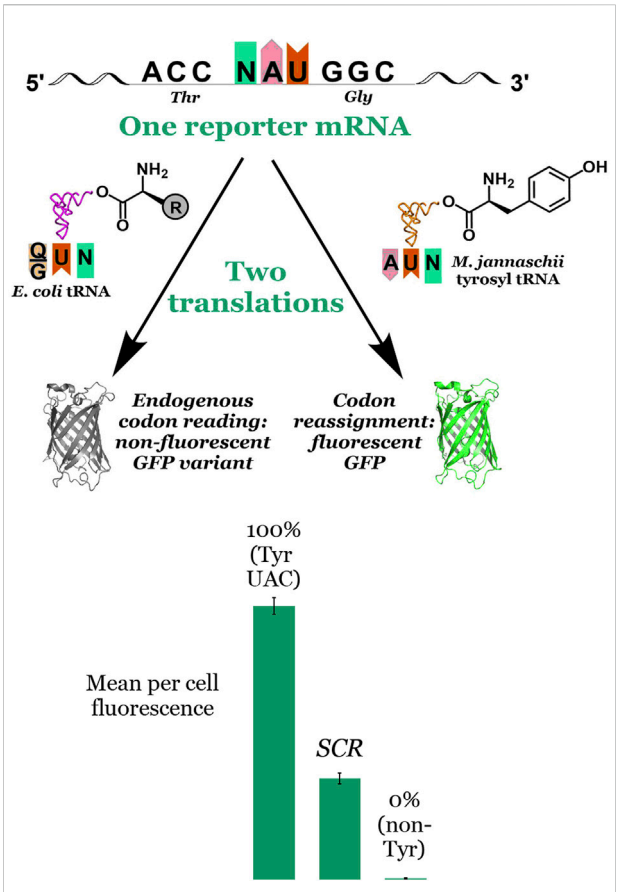


FIGURE 3
Principle of and data visualization for the fluorescence-based screen for codon reassignment. The fluorescence-based screen exploits the absolute requirement of tyrosine at position 66 in superfolder GFP for fluorophore formation. Reporter vectors with another codon at that position are co-transformed with the *M. jannaschii* tyrosyl aminoacyl tRNA synthetase and a variant of its cognate tRNA. Both the engineered orthogonal tRNA and endogenous tRNAs (with either Q or G at position 34) compete to decode the evaluated codon. Incorporation of tyrosine by the orthogonal tRNA in response to the codon at position 66 in the GFP reporter leads to production of fluorescent protein. Incorporation of a different canonical amino acid by the endogenous translation machinery leads to production of non-fluorescent protein. The mean per cell fluorescence of the system under evaluation is bracketed between a 100% fluorescence reference system in which the codon at position 66 encodes tyrosine and a 0% fluorescence reference system in which the codon at position 66 encodes another amino acid. The middle bar labelled "SCR" represents the intermediate per cell fluorescence observed for a hypothetical system under evaluation. In this report, SCR efficiencies range from 2.1% to 13.4% depending on the system. More broadly, we have reported SCR efficiencies between 1.0% and 70% using the *M. jannaschii* orthogonal pair.

and 100% reassigning control systems. The 100% fluorescence reference is established in each experiment by monitoring the per cell fluorescence of a reporter vector with a tyrosine codon specifying the fluorophore. The 0% fluorescence reference is established in each experiment by monitoring the per cell fluorescence of a reporter vector with a non-tyrosine codon specifying the fluorophore. Both control systems also include a vector expressing a variant of the orthogonal translation machinery to maintain a metabolic burden on the cells similar to that of the systems under evaluation for reassignment efficiency.

To reduce variation across reassignment systems, the DNA sequences for the reporter vectors for each targeted codon are identical, save the codon at the critical tyrosine position (test position). The gene sequence of the GFP reporter utilizes a reduced codon set to minimize the possibility that reassignment of a particular sense codon at a non-permissive position would interfere with protein folding and fluorophore formation, resulting in a lower apparent SCR efficiency for that codon. The reporter vectors utilized in this evaluation are those used in our broad evaluation of the reassignment potential of more than 30 rarely-used and/or *E. coli* wobble codons (Schmitt et al., 2018; Schwark et al., 2020). Supplementary Figure S1 depicts the codon usage frequency for the wild type superfolder GFP gene used in the 100% fluorescence reference system. The nearly identical sequences of the two vectors used across all evaluations should minimize apparent reassignment differences as a result of different rates or levels of mRNA transcription/codon context effects.

2.3 Chromosomal knockout of a key gene in the queuosine biosynthetic pathway, QueC

QueC knockout strains were prepared in house according to the method of Datsenko and Wanner (Datsenko and Wanner, 2000). The method employs a two-step process in which the lambda red recombination system (beta, gam, and exo gene products) facilitates the recombination of an introduced linear DNA sequence in place of a targeted gene, followed by flp-mediated recombination to remove the initially-introduced selective marker. The linear DNA segment is prepared by amplification of a cassette containing a selectivity marker flanked by frp recombinase sites. The amplification primers introduce sequences identical to the sequence flanking the chromosomal region to be removed. The linear DNA is introduced into cells containing a temperature-sensitive plasmid expressing the lambda red system. After selection for recombination and plasmid loss, a second temperature-sensitive plasmid containing the flp recombinase is introduced into the cells, and the initial selective marker is removed. Chromosomal knockout of specified genes was confirmed by sequencing purified PCR products which amplified relevant regions of the chromosome. Sequences of the oligonucleotides used to affect and confirm the genomic knockouts are listed in the supplementary material (Supplementary Table S1).

3 Results

Our fluorescence-based screen (Figure 3) rapidly generates single point measurements of the outcome of the multiple competition events that are involved in the incorporation of a given amino acid in response to a given codon. An orthogonal tRNA/aaRS pair, in this case the *M. jannaschii* tyrosyl tRNA/aaRS pair, is introduced into a cell harboring a vector encoding a GFP reporter in which the codon for the essential Tyr residue at position 66 has been replaced by a codon under evaluation. The orthogonal tRNA is aminoacylated with tyrosine by the orthogonal aaRS. Fluorescent GFP is produced when the orthogonal tRNA decodes the codon at position 66 in the GFP reporter vector. Non-fluorescent protein is produced when the endogenous translation machinery decodes the test codon. Bracketing the observed per cell (i.e., normalized to optical density) fluorescence in the system under interrogation with systems in which no fluorescent GFP is produced (0% reassignment control) and one in which wild type superfolder GFP is produced (100% reassignment control) provides a quantitative measurement of the efficiency of reassignment of the particular codon by the orthogonal tRNA variant. In the lower section of Figure 3, the middle bar labelled "SCR" represents the intermediate per cell fluorescence observed for a hypothetical system under evaluation. In this report, SCR efficiencies range from 2.1% to 13.4%. Other anticodon variants of the *M. jannaschii* orthogonal pair facilitate sense codon reassignment to tyrosine with efficiencies between 1.0% and 70% (Schmitt et al., 2018; Schwark et al., 2020).

Most sense codon reassignment evaluations reported by our laboratory have been performed in *E. coli* SB3930, a histidine auxotroph (Biddle et al., 2015; Biddle et al., 2016; Schmitt et al., 2018; Schwark et al., 2020). Our first evaluation of the effect of queuosine modification of endogenous tRNAs on codon reassignment by an orthogonal tRNA looked at reassignment of His CAU in SB3930 and SB3930/QueC-. We chose to employ knockouts of the QueC gene because spontaneous QueC deletions have been shown to eliminate queuosine modification of tRNAs without inducing growth defects (Dineshkumar et al., 2002; Gaur and Varshney, 2005). The QueC gene is involved in the initial steps of the queuosine biosynthetic pathway in *E. coli* (Gaur and Varshney, 2005; El Yacoubi et al., 2012). Removal of queuosine modification does not affect other tRNA modifications (Manickam et al., 2016). Reassignment efficiency of the His CAU codon improved more than 50% from 6.3 ± 0.4% in SB3930 to 10.1 ± 0.7% in SB3930/QueC-.

We proceeded with further evaluations of the effects of queuosine modification on sense codon reassignment in the more frequently-used laboratory strain *E. coli* Top10 and derivatives. The observed reassignment efficiency of His CAU was 7.4 ± 0.2% in Top10 to 11.5 ± 0.3% in Top10/HisB-/QueC-. Differences in incorporation efficiency between strains of *E. coli* despite using identical vectors and system conditions (e.g., media, antibiotics) have been widely observed though not holistically evaluated. In order to confirm that any observed changes in sense codon reassignment efficiency between Top10 *E. coli* and a Top10 strain incapable of modifying tRNAs with queuosine were the result of the change to the queuosine pathway as opposed to the histidine biosynthetic pathway (a notable difference between SB3930 and Top10), we evaluated the sense codon reassigning systems in 3 *E. coli* strains: Top10, Top10/HisB-, and Top10/HisB-/QueC-. Every codon evaluated showed identical reassignment efficiencies in both Top10 and Top10/HisB-, suggesting that reassignment efficiency in rich media is independent of the strain's ability to synthesize its own histidine. Reassignment efficiencies in Top10/HisB- are provided in the supporting information.

We evaluated the ability of introduced orthogonal tRNAs with AUN anticodons to reassign three of the four U-ending codons normally decoded by Q34 endogenous tRNAs: His CAU, Asn AAU, and Asp GAU in the presence and absence of queuosine modification (Table 1). The reassignment efficiency of the asparagine AAU codon improved modestly (1.2-fold) in the absence of queuosine. In the case of aspartic acid, however, removal of the queuosine modification resulted in a 1.6-fold *decrease* in reassignment of the GAU codon. *E. coli* $tRNA^{Asp}_{QUC}$ is hypermodified at position 34 beyond the addition of queuosine; a glutamic acid residue is esterified to the dihydroxycyclopentene ring to form a GluQ nucleoside. Very little is understood about the function of GluQ in translation. Previous evaluations of the effects of queuosine modification on frameshift maintenance did not evaluate aspartic acid codons (Urbonavicius et al., 2001). Evaluations of the effect of queuosine modification on missense incorporation indicated that the absence of queuosine manifested different effects for different tRNAs (Manickam et al., 2014; Manickam et al., 2016).

Sense codon reassignment of histidine CAU codons utilized two variants of the orthogonal tRNA. We previously demonstrated that the *M. jannaschii* $tRNA_{AUG}$ is partially (~50%) modified to inosine in *E. coli* and that anticodon loop modifications outside the anticodon could eliminate the partial inosine modification (Biddle et al., 2016). Both the A34/I34 mixed population tRNA as well as *M. jannaschii* $tRNA_{AUG\text{-}G37}$ in which no inosine-modification is detected were evaluated. The partially inosine-modified orthogonal tRNA competed much more effectively in the queuosine-deficient strain, displaying a 1.6-fold improvement in reassignment efficiency. The non-inosine-modified orthogonal tRNA reassigned the CAU codon equally well in both strains.

Importantly, neither sense codon reassignment nor the removal of queuosine modification from GUN anticodon endogenous tRNAs is fatal to cells. QueC chromosomal knockouts have previously been reported to not show large growth defects (Dineshkumar et al., 2002; Gaur and Varshney, 2005). The limited effect of QueC knockout on cell growth is

TABLE 1 Sense codon reassignment efficiency (SCR) by variants of the *M. jannaschii* orthogonal translation machinery with AUN anticodons in queuosine-containing and queuosine-deficient *E. coli*.

Orthogonal anticodon	Codon evaluated	Reassignment efficiency in Top10	Biological replicates Top10	Reassignment efficiency in Top10/HisB-/QueC-	Biological replicates Top10/HisB-/QueC-	Fold change in queuosine-deficient strain
AUU	AAU (Asn)	7.5 ± 0.5%	18	8.9 ± 0.4%	11	1.2
	AAC (Asn)	B.D[a]	18	B.D.	16	—
AUC	GAU (Asp)	3.0 ± 0.2%	18	2.1 ± 0.2%	12	0.7
	GAC (Asp)	B.D.	18	B.D.	17	—
AUG/IUG	CAU (His)	7.4 ± 0.2%	24	11.5 ± 0.3%	12	1.6
	CAC (His)	2.8 ± 0.2%	22	2.6 ± 0.1%	11	1.0
AUG G37	CAU (His)	13.2 ± 0.4%	24	13.4 ± 0.8%	15	1.0
	CAC (His)	B.D.	24	B.D.	17	—

[a]B.D., indicates that the codon was evaluated with the specified tRNA, and the measurement was below the detection limit of the in cell assay (0.15%).

consistent with our observation of similar optical density vs. time profiles of the same codon reassignment system in Top10 and Top10/HisB-/QueC- strains. Our previous evaluations of sense codon reassignments have revealed that *E. coli* are largely tolerant of widespread missense incorporation of tyrosine across their proteomes (Schmitt et al., 2018; Schwark et al., 2020). While the carrying capacity of sense codon reassigning systems is slightly reduced relative to the nonreassigning controls, none of the systems exhibits large growth defects. Representative OD600 vs. time plots for systems expressing all four evaluated *M. jannaschii* tRNAs and their targeted GFP reporters are provided in the supplementary material (Supplementary Figure S2). The supplementary material also includes representative fluorescence versus time plots for the controls and codon reassigning systems in both the queuosine-containing and queuosine-deficient strains (Supplementary Figure S3). The lack of queuosine does not have a significant impact on protein production. Similar fluorescence for the 100% control (GFP with a tyrosine codon at the fluorophore) is observed in both *E. coli* strains.

4 Discussion

4.1 Could engineering endogenous tRNA modifications serve as a handle for improving the function of expanded genetic code systems?

Increasing the efficiency of codon reassignment by an orthogonal pair may be achieved either by *improving* the function of the orthogonal pair in the system into which it is transplanted or by *decreasing* the efficacy of the endogenous translation components competing for the targeted codon. Improving the function of the orthogonal pair may be as simple as increasing the effective concentrations of orthogonal components by expressing both tRNAs and aaRSs from cassettes with stronger promoters in plasmids with increased copy numbers (Ryu and Schultz, 2006; Young et al., 2010; Chatterjee et al., 2013). Adjustments to media composition, including adjusting amino acid concentrations, have been utilized to improve codon reassignment (Dougherty et al., 1993; Budisa et al., 1995; Kwon et al., 2003; Pal and Budisa, 2010). Increasing the assimilation of the orthogonal machinery into the endogenous translation system may also increase the efficiency of codon reassignment. Amber stop codon suppression has been improved by altering nucleotides on the tRNA that modulate interactions with elongation factor Tu (EF-Tu) (Guo et al., 2009; Schrader et al., 2011; Mittelstaet et al., 2013; Fan et al., 2015; Maranhao and Ellington, 2017).

Improving the interactions between the orthogonal tRNA and its cognate synthetase is another means by which the function of the orthogonal tRNA may be increased. Orthogonal tRNA/aaRS recognition becomes particularly relevant as the anticodon of the orthogonal tRNA is changed to target different sense codons. The anticodon sequence is often an important identity element for proper recognition and aminoacylation of tRNAs (Giege et al., 1998). Poor charging of an amino acid onto its cognate tRNA lowers the effective concentration of aminoacylated tRNA available to compete against endogenous components for a targeted codon. The impacts of changes to anticodon sequences on the efficiency of aminoacylation have been evaluated for the *M. jannaschii* Tyr tRNA/aaRS pair, one of the two orthogonal pairs most commonly used for incorporation of ncAAs in response to amber stop codons (Fechter et al., 2001). Changing the anticodon of the *M. jannaschii* tRNA from GAU (Tyr) to Watson-Crick base pair with other codons decreases aminoacylation efficiency over a range of three orders of magnitude. Several selection strategies have been developed to improve the function of stop codon suppressing orthogonal pair systems, although

the maturation of initially-identified functional aaRSs utilized for ncAA incorporation are not commonly performed (Pott et al., 2014; Wang et al., 2015; Rauch et al., 2016; Maranhao and Ellington, 2017; Owens et al., 2017). Directed evolution of *M. jannaschii* tRNA/aaRS pair variants for improved reassignment of the amber stop codon as well as Lys AAG, His CAU, and Arg AGG sense codons have been described (Takimoto et al., 2009; Kuhn et al., 2010; Mittelstaet et al., 2013; Biddle et al., 2015; Wang et al., 2015; Biddle et al., 2016; Biddle et al., 2022).

The alternative strategy of *decreasing* competition from the endogenous translation components that compete with the orthogonal machinery has been applied to improve amber stop codon reassignment and rare arginine codon reassignment. Three different strategies of genomic modifications have been described that enable the elimination of the normally essential peptide release factor, RF1, that recognizes the amber UAG codon as a stop signal. Systems from which RF1 has been eliminated show improved incorporation of ncAAs in response to UAG codons (Johnson et al., 2011; Lajoie et al., 2013; Mukai et al., 2015). Reassignment of the low frequency Arg AGG codon in *E. coli* has been demonstrated after knocking out one of the two competing tRNA genes (Mukai et al., 2015; Lee et al., 2016). Additional genome rewriting projects seek to generate multiple "free" codons for ncAA incorporation (Ostrov et al., 2016; Wang et al., 2016; Robertson et al., 2021). A free codon is one that is not employed in any genes in the organism and does not have a tRNA to decode it. Total genome synthesis and transplantaion are costly and experimentally demanding and may have potential unintended consequences on organismal viability as a result of inadvertant disruptions to regulatory sequences or RNA folding.

An alternative and not widely-explored approach is to modulate the modifications of tRNA species that affect their decoding properties. Functions for tRNA modifications based on their locations in tRNAs and roles that modified tRNAs play in translation have been broadly explored but many questions still remain (Agris, 1991; Lim and Curran, 2001; El Yacoubi et al., 2012; Björk et al., 2014; de Crecy-Lagard and Jaroch, 2021). We sought to evaluate the extent that directed changes in queuosine modification of endogenous anticodons at position 34 could be exploited to alter the reading of the genetic code in *E. coli*. Queuosine is thought to bias the codon reading preferences of the endogenous tRNAs, although data is sparse, and the effects appear to differ across organisms and even between modified tRNAs within the same organism (Noguchi et al., 1982; Manickam et al., 2014; Manickam et al., 2016). Unmodified G34 tRNAs efficiently decode C-ending codons through traditional Watson-Crick interactions and wobble to U-ending codons with reduced efficiency. Q34 tRNAs appear to read U- and C-ending codons equally well (Meier et al., 1985).

4.2 Do A34 tRNAs compete better against G34 than Q34 tRNAs for U-ending codons?

Evaluation of A34 tRNA competition is complicated by the limited number of Q34 modified tRNAs and the inability of the fluorescence-based screen to evaluate natural tyrosine incorporation. The fluorescence-based screen quantifies incorporation of tyrosine in response to non-tyrosine codons, which excludes evaluation of the impact of queuosine modification on sense codon reassignment of the Tyr UAU codon. Interpreting the quantification of reassignment of the aspartic acid codon GAU is fraught, as the modified nucleotide on aspartic acid tRNAs is the hypermodified variant of Q, GluQ. Finally, we previously identified partial inosine modification of the *M. jannaschii* $tRNA_{AUG}$ used to reassign the His CAU codon. A single nucleotide change at position 37 of the anticodon loop abrogated inosine modification.

Four codon pairs whose decoding is potentially influenced by queuosine modification exist. We evaluated two systems that directly address the competition of A34 orthogonal vs. Q34 or G34 endogenous tRNAs (*M. jannaschii* $tRNA_{AUU}$ for Asn AAU codons and *M. jannaschii* $tRNA_{AUG\text{-}G37}$ for His CAU codons). We found that *M. jannaschii* $tRNA_{AUU}$ reassigned the asparagine AAU codon 1.2 times better in the queuosine-deficient strain (7.5 ± 0.5% vs. 8.9 ± 0.4%). We found that *M. jannaschii* $tRNA_{AUG\text{-}G37}$ reassigned the histidine CAU codon equally well in either strain (13.2 ± 0.4% and 13.4 ± 0.8%). Sections 4.3, 4.4 elaborate on the evaluation of two additional systems that offer insight into the impact of modified A34 orthogonal or hypermodified Q34 endogenous tRNAs on competition for the targeted codons.

Meier and co-workers demonstrated that Drosophila histidine tRNA isoforms differing only at the identity of the nucleotide at position 34 showed distinct decoding preferences for histidine CAU versus CAC codons (Meier et al., 1985). Tritiated histidine was aminoacylated onto either the G34 isoform or the Q34 isoform. The amount of label incorporated in response to CAU and CAC codons while in competition with the other isoform aminoacylated with unlabeled histidine provided a window on the decoding preferences of the two tRNAs. When the G34 tRNA was labeled, the labeled amino acid was incorporated in response to the CAC codon twice as often as in response to the CAU codon. In contrast, when the Q34 tRNA was labeled, the labeled amino acid was incorporated in response to the CAC and CAU codons nearly equally, with even a slight preference for CAU over CAC.

Urbonavicius and co-workers evaluated the ways in which Q34 and G34 tRNAs differentially interact with the His CAU and CAC codons (Urbonavicius et al., 2001). Employing a frameshifting reporter, they measured the kinetics of entry of the two tRNA isoforms into the ribosomal a site. The likelihood of frameshifting by a P site tRNA increases as the amount of time

required to accept a tRNA into the A site increases. Urbonavicius et al. found that when U-ending codons for tyrosine (UAU) and histidine (CAU) were in the ribosomal A site, queuosine-deficient G34 tRNAs had increased sampling time in the ribosomal A site leading to measurable P site frameshifting. In contrast, when the C-ending codons for tyrosine (UAC) and histidine (CAC) were in the A site, queuosine-deficient G34 tRNAs did not increase P site frameshifting. The absence of queuosine modification did not increase frameshifting when the asparagine U-ending codon (AAU) was in the A site.

Both reports concluded that queuosine is more important for efficient translation of the U-ending codons than C-ending codons, particularly for histidine and tyrosine. Additionally, the presence or absence of queuosine modification influences global cellular protein translation and has been implicated in organismal development and several disease states, including cancer (Chen and Wu, 1994; Zaborske et al., 2014; Nagaraja et al., 2021). Genes containing more NAC codons are expressed when queuosine modification is at a low level. Genes containing more NAU codons are expressed as queuosine modification of relevant tRNAs increases (Zaborske et al., 2014). These observations suggest that queuosine deficiency would disfavor reading U-ending codons by endogenous tRNAs and improve codon reassignment by orthogonal tRNAs targeted to U-ending codons. Our quantification of the reassignment efficiencies of asparagine and histidine U3 codons by *M. jannaschii* tRNAs with A34 anticodons, $tRNA_{AUU}$ and $tRNA_{AUG\text{-}G37}$ (Table 1), is in line with these observations and suggests that removal of queuosine modification results in a modest improvement in sense codon reassignment.

4.3 Evaluation of competition between A34 vs. G34 or A34 vs. GluQ34 in translation

The *E. coli* tRNA that translates aspartic acid codons is further modified at position 34 relative to the other *E. coli* tRNAs with QUN anticodons by esterification of a glutamic acid to the queuosine nucleotide (Figure 1) (El Yacoubi et al., 2012; Björk et al., 2014). Very little is known about the function of the hypermodified GluQ base. Unlike A34 orthogonal tRNA competition that was either modestly improved or unchanged in the absence of queueosine modification, we found that U-ending codon reassignment decreased in efficiency in the absence of GluQ modification. *M. jannaschii* $tRNA_{AUC}$ reassigned the aspartic acid GAU codon only 0.7 times as well in the GluQ-deficient strain (3.0 ± 0.2% vs. 2.1 ± 0.2%).

Neither Meier nor Urbonavicius examined the impact of GluQ modification on translation or frameshifting at aspartic acid U- and C-ending codons. In a series of studies of the effects of tRNA modifications on missense incorporation, Manickam et al. observed different degrees of misreading by *E. coli* tyrosine and aspartic acid tRNAs in the presence and absence of queuosine. The authors employed a series of gain of function enzymatic reporters to evaluate the extent that aspartic acid and tyrosine tRNAs could read alternative codons to restore enzyme function. Their observations suggest that queuosine modification is important for controlling missense errors related to second position mismatches, e.g., $tRNA^{Asp}_{GluQUC}$ decoding the glycine GGC codon using a U35/G2 interaction. In queuosine-deficient strains, missense incorporation of aspartic acid in response to Gly GGC increased. In contrast, for another potential U35/G2 interaction, $tRNA^{Tyr}_{QUA}$ decoding Cys UGU and UGC codons, missense incorporation of tyrosine decreased in queuosine-deficient strains (Manickam et al., 2016).

The experiments of Manickam et al. evaluated the impact of tRNA modifications on the propensity of tRNAs to incorrectly decode codons for other amino acids as opposed to the effects of modifications on "cognate" codons. Our fluorescence-based gain of function screen evaluates competition between tRNAs that are both (although differently) cognate for the evaluated codon. While not directly comparable, our experiments and those described by Manickam and co-workers suggest that tRNAs modified with Q vs. GluQ impact translation differently. However, the outcome of removing Q vs. GluQ modification on codon recognition at the wobble position appears to be different than the outcome at codons with second position mismatches. At the second position of the codon, removing GluQ modification increases missense incorporation. At the wobble position, removing GluQ modification reduces the efficiency of directed sense codon reassignment.

4.4 Evaluation of competition between inosine and queuosine-modified tRNAs

The fluorescence-based screen and other gain of function enzyme systems can be employed to evaluate base pairing competition events that are not present within standard genetic codes in the context of wholly-functional translation systems in living cells. For example, the base pairing preferences of unmodified A34 tRNAs are relatively unknown because in nearly every naturally encoded tRNA, the A34 is enzymatically-modified to inosine. Similarly, the relative strengths of I34/U3 and Q34/U3 base pairing interactions are largely uninvestigated because natural systems do not contain codon boxes in which such competition would exist.

In natural systems, Q34 and I34 tRNAs are only in direct competition in missense situations. One of the two tRNAs would fully base pair with a certain codon while the other tRNA would have a mismatch at position 35 of the anticodon and incorporate an incorrect amino acid. In *E. coli*, this outside-the-box competition could happen between His tRNA QUG and Arg tRNA ICG. In eukaryotes, which utilize I34 tRNAs more extensively in several 4 codon boxes, the instances of

Q34 tRNAs potentially competing against I34 tRNAs increases. Manickam and co-workers did not examine the case of *E. coli* $tRNA^{His}_{QUG}$, the only one of the four queuosine-containing tRNAs in which a second position misreading error would place a Q34 tRNA (anticodon QUG) in competition with an I34 tRNA (anticodon ICG) for decoding the CGU and CGC arginine codons.

We found that *M. jannaschii* $tRNA_{AUG}$ with partial inosine modification reassigned the histidine CAU codon 1.6 times better in the queuosine-deficient strain (7.4 ± 0.2% vs. 11.5 ± 0.3%). Partial inosine modification of this orthogonal tRNA was identified following observation of unexpected decoding of the histidine CAC codon (Biddle et al., 2016). A34 tRNAs would not be expected to decode C3 codons, whereas I34 tRNAs typically take the place of G34 and U34 tRNAs in their respective boxes. I34 decoding the CAC codon was unaffected by the presence of queuosine in the endogenous *E. coli* tRNA (2.7 ± 0.2% vs. 2.6 ± 0.1% in QueC + vs. QueC- strains). The observation of similar behavior in in QueC + vs. QueC- strains is consistent with the studies suggesting that G34 and Q34 show similar efficiencies for C3 codons. Decoding of the C-ending codons by all other orthogonal tRNAs described in this manuscript is below the limit of detection in both QueC+ and QueC- Top10 *E. coli* (less than 0.15%).

5 Conclusion

The relative quantitative contribution of various factors affecting codon-specific translational efficiency are not well understood. Factors including tRNA modifications, tRNA abundance, codon-anticodon interaction energy, codon context effects, interactions between the nascent peptide chain and the ribosome, and interactions between different amino acids in the peptidyl transferase site, among others, all contribute to decoding. The relative importance of the individual factors is codon- and tRNA-dependent. Competition between endogenous and introduced orthogonal tRNAs, integrating all of the above interactions, ultimately determines the efficiency of sense codon reassignment. Our previous quantification of the reassignment efficiency of more than 30 sense codons using the fluorescence-based screen adds to the dissection of the relative quantitative importance of tRNA abundance, aminoacylation efficiency, and codon-anticodon interaction energy as contributors to translational fidelity (Biddle et al., 2015; Schmitt et al., 2018; Schwark et al., 2020).

Our system for evaluating orthogonal pair-directed codon reassignment allows evaluation of the impact of tRNA modifications on the fidelity of translation. Using the fluorescence-based screen and two related *E. coli* strains, one with queuosine-modified tRNAs and one without, the relative efficiencies of five distinct base pairing combinations: A34/U3, G34/U3, Q34/U3, GluQ34/U3, and I34/U3 have been evaluated. These five types of base pairing interactions are analyzed through evaluation of five distinct groups of competition events: A34 orthogonal tRNAs vs. Q34, GluQ34, or G34 endogenous tRNAs or I34 orthogonal tRNAs vs. Q34 or G34 endogenous tRNAs. Quantifying the effect of tRNA modification on decoding specific codons is critical for understanding the overall fidelity of protein translation and for engineering systems with expanded genetic codes.

Removal of queuosine modification of endogenous tRNAs leads to a 0.7 to 1.6-fold improvement of sense codon reassignment by *M. jannaschii* orthogonal tRNAs, depending on both the orthogonal tRNA and the codon targeted for reassignment. These observed effects of queuosine modification on sense codon reassignment may be contextualized by examining the extent to which the efficiency of stop codon suppression is modulated through system modifications and directed evolution. Efforts to improve system function by targeting expression levels, tRNA interactions with EF-Tu and aaRSs, as well as the kinetics of aminoacylation typically result in 1 to 10-fold improvement. Improved orthogonal tRNAs for amber codon suppression were evolved based on optimization of interactions with *E. coli* EF-Tu (Guo et al., 2009). The evolved tRNA sequences typically improved incorporation of noncanonical amino acids 1 to 3-fold, although a handful of improvements up to 20-fold were reported. Maturation of orthogonal tRNA/aaRS pairs initially-evolved for specific noncanonical amino acids yields improvements of 1 to 10-fold (Kuhn et al., 2010; Maranhao and Ellington, 2017; Owens et al., 2017; Thyer et al., 2021). The functional improvements in the efficiency of noncanonical amino acid incorporation are the results of smaller, 1 to 5-fold improvements in the kinetic efficiency of orthogonal pairs (Amiram et al., 2015; Rauch et al., 2016). The context of the targeted stop codon also contributes to the efficiency of suppression (Pott et al., 2014; Schwark et al., 2018).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization: WB, DS, and JF; funding acquisition: MS and JF; investigation: JT and DS; methodology: JT, DS, WB, and MS; supervision: JF; visualization: JT and MS; writing—original draft: JT and MS; writing—review and editing: MS and JF.

Funding

This research was funded by the National Science Foundation, grant number CHE-1507055 to J.D.F. The APC was funded by University of Colorado Denver.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.938114/full#supplementary-material

References

Agris, P. F. (1991). Wobble position modified nucleosides evolved to select transfer-rna codon recognition - a modified-wobble hypothesis. *Biochimie* 73 (11), 1345–1349. doi:10.1016/0300-9084(91)90163-u

Amiram, M., Haimovich, A. D., Fan, C. G., Wang, Y. S., Aerni, H. R., Ntai, I., et al. (2015). Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* 33 (12), 1272–1279. doi:10.1038/nbt.3372

Biddle, W., Schmitt, M. A., and Fisk, J. D. (2015). Evaluating sense codon reassignment with a simple fluorescence screen. *Biochemistry* 54 (50), 7355–7364. doi:10.1021/acs.biochem.5b00870

Biddle, W., Schmitt, M. A., and Fisk, J. D. (2016). Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment. *Nucleic Acids Res.* 44 (21), 10042–10050. doi:10.1093/nar/gkw948

Biddle, W., Schwark, D. G., Schmitt, M. A., and Fisk, J. D. (2022). Directed evolution pipeline for the improvement of orthogonal translation machinery for genetic code expansion at sense codons. *Front. Chem.* 10, 815788. doi:10.3389/fchem.2022.815788

Bienz, M., and Kubli, E. (1981). Wild-type transfer-RNA gTyr reads the tmv RNA stop codon, but Q-base-modified transfer-RNA gTyr does not. *Nature* 294 (5837), 188–190. doi:10.1038/294188a0

Björk, G. R., Hagervall, T. G., and Lovett, S. T. (2014). Transfer RNA modification: presence, synthesis, and function. *EcoSal Plus* 6 (1). DOI: doi:doi:10.1128/ecosalplus.ESP-0007-2013

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (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), 788–796. doi:10.1111/j.1432-1033.1995.tb20622.x

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 (10), 1828–1837. doi:10.1021/bi4000244

Chen, Y. L., and Wu, R. T. (1994). Altered queuine modification of transfer-RNA involved in the differentiation of human K562 erythroleukemia-cells in the presence of distinct differentiation inducers. *Cancer Res.* 54 (8), 2192–2198.

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* 97 (12), 6640–6645. doi:10.1073/pnas.120163297

de Crecy-Lagard, V., and Jaroch, M. (2021). Functions of bacterial tRNA modifications: from ubiquity to diversity. *Trends Microbiol.* 29 (1), 41–53. doi:10.1016/j.tim.2020.06.010

Dineshkumar, T. K., Thanedar, S., Subbulakshmi, C., and Varshney, U. (2002). An unexpected absence of queuosine modification in the tRNAs of an *Escherichia coli* B strain. *Microbiology* 148, 3779–3787. doi:10.1099/00221287-148-12-3779

Dougherty, M. J., Kothakota, S., Mason, T. L., Tirrell, D. A., and Fournier, M. J. (1993). Synthesis of a genetically engineered repetitive polypeptide containing periodic selenomethionine residues. *Macromolecules* 26 (7), 1779–1781. doi:10.1021/ma00059a045

El Yacoubi, B., Bailly, M., and de Crecy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* 46, 69–95. doi:10.1146/annurev-genet-110711-155641

Fan, C. G., Xiong, H., Reynolds, N. M., and Söll, D. (2015). Rationally evolving tRNA(Pyl) for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.* 43 (22), e156. doi:10.1093/nar/gkv800

Fechter, P., Rudinger-Thirion, J., Tukalo, M., and Giege, R. (2001). Major tyrosine identity determinants in *Methanococcus jannaschii* and *Saccharomyces cerevisiae* tRNA(Tyr) conserved but expressed differently. *Eur. J. Biochem.* 268 (3), 761–767. doi:10.1046/j.1432-1327.2001.01931.x

Fluitt, A., Pienaar, E., and Viljoen, H. (2007). Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.* 31 (5-6), 335–346. doi:10.1016/j.compbiolchem.2007.07.003

Gaur, R., and Varshney, U. (2005). Genetic analysis identifies a function for the queC (ybaX) gene product at an initial step in the queuosine Biosynthetic pathway in *Escherichia coli*. *J. Bacteriol.* 187 (20), 6893–6901. doi:10.1128/jb.187.20.6893-6901.2005

Giege, R., Sissler, M., and Florentz, C. (1998). Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* 26 (22), 5017–5035. doi:10.1093/nar/26.22.5017

Grosjean, H. J., Dehenau, S., and Crothers, D. M. (1978). Physical basis for ambiguity in genetic coding interactions. *Proc. Natl. Acad. Sci. U. S. A.* 75 (2), 610–614. doi:10.1073/pnas.75.2.610

Guo, J. T., Melancon, 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. Engl.* 48 (48), 9148–9151. doi:10.1002/anie.200904035

Iwata-Reuyl, D. (2003). Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. *Bioorg. Chem.* 31 (1), 24–43. doi:10.1016/s0045-2068(02)00513-8

Johnson, D. B. F., Xu, J. F., Shen, Z. X., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., et al. (2011). RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* 7 (11), 779–786. doi:10.1038/nchembio.657

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., et al. (2005). Complete set of ORF clones of *Escherichia coli* ASKA library (A complete Set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* 12 (5), 291–299. doi:10.1093/dnares/dsi012

Kramer, E. B., and Farabaugh, P. J. (2007). The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* 13 (1), 87–96. doi:10.1261/rna.294907

Kuhn, S. M., Rubini, M., Fuhrmann, M., Theobald, I., and Skerra, A. (2010). Engineering of an orthogonal aminoacyl-tRNA synthetase for efficient incorporation of the non-natural amino acid O-Methyl-L-tyrosine using fluorescence-based bacterial cell sorting. *J. Mol. Biol.* 404 (1), 70–87. doi:10.1016/j.jmb.2010.09.001

Kwon, I., Kirshenbaum, K., and Tirrell, D. A. (2003). Breaking the degeneracy of the genetic code. *J. Am. Chem. Soc.* 125 (25), 7512–7513. doi:10.1021/ja0350076

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., et al. (2013). Genomically recoded organisms expand biological functions. *Science* 342 (6156), 357–360. doi:10.1126/science.1241459

Lee, K. B., Hou, C. Y., Kim, C. E., Kim, D. M., Suga, H., and Kang, T. J. (2016). Genetic code expansion by degeneracy reprogramming of arginyl codons. *Chembiochem* 17 (13), 1198–1201. doi:10.1002/cbic.201600111

Lim, V. I., and Curran, J. F. (2001). Analysis of codon : anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. *RNA* 7 (7), 942–957. doi:10.1017/s135583820100214x

Manickam, N., Joshi, K., Bhatt, M. J., and Farabaugh, P. J. (2016). Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res.* 44 (4), 1871–1881. doi:10.1093/nar/gkv1506

Manickam, N., Nag, N., Abbasi, A., Patel, K., and Farabaugh, P. J. (2014). Studies of translational misreading *in vivo* show that the ribosome very efficiently discriminates against most potential errors. *RNA* 20 (1), 9–15. doi:10.1261/rna.039792.113

Maranhao, A. C., and Ellington, A. D. (2017). Evolving orthogonal suppressor tRNAs to incorporate modified amino acids. *ACS Synth. Biol.* 6 (1), 108–119. doi:10.1021/acssynbio.6b00145

Meier, F., Suter, B., Grosjean, H., Keith, G., and Kubli, E. (1985). Queuosine modification of the wobble base in transfer rnahis influences invivo decoding properties. *EMBO J.* 4 (3), 823–827. doi:10.1002/j.1460-2075.1985.tb03704.x

Min, B., Kitabatake, M., Polycarpo, C., Pelaschier, J., Rraczniak, G., Ruan, B., et al. (2003). Protein synthesis in *Escherichia coli* with mischarged tRNA. *J. Bacteriol.* 185 (12), 3524–3526. doi:10.1128/jb.185.12.3524-3526.2003

Mittelstaet, J., Konevega, A. L., and Rodnina, M. V. (2013). A kinetic safety gate controlling the delivery of unnatural amino acids to the ribosome. *J. Am. Chem. Soc.* 135 (45), 17031–17038. doi:10.1021/ja407511q

Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., et al. (2015). Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.* 43 (16), 8111–8122. doi:10.1093/nar/gkv787

Nagaraja, S., Cai, M. W., Sun, J. J., Varet, H., Sarid, L., Trebicz-Geffen, M., et al. (2021). Queuine is a nutritional regulator of entamoeba histolytica response to oxidative stress and a virulence attenuator. *Mbio* 12 (2), e03549–20. doi:10.1128/mBio.03549-20

Noguchi, S., Nishimura, Y., Hirota, Y., and Nishimura, S. (1982). Isolation and characterization of an escherichia-coli mutant lacking transfer rna-guanine transglycosylase - function and biosynthesis of queuosine in transfer-RNA. *J. Biol. Chem.* 257 (11), 6544–6550. doi:10.1016/s0021-9258(20)65176-6

Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., et al. (2016). Design, synthesis, and testing toward a 57-codon genome. *Science* 353 (6301), 819–822. doi:10.1126/science.aaf3639

Owens, A. E., Grasso, K. T., Ziegler, C. A., and Fasan, R. (2017). Two-tier screening platform for directed evolution of aminoacyl-tRNA synthetases with enhanced stop codon suppression efficiency. *Chembiochem* 18 (12), 1109–1116. doi:10.1002/cbic.201700039

Pal, P. P., and Budisa, N. (2010). "Engineering green fluorescent proteins using an expanded genetic code," in *Reviews in fluorescence 2008*. Editor C. D. Geddes (New York: Kluwer Academic/Plenum Publ), 5, 359–386.

Parker, J. (1989). Errors and alternatives in reading the universal genetic-code. *Microbiol. Rev.* 53 (3), 273–298. doi:10.1128/mr.53.3.273-298.1989

Pott, M., Schmidt, M. J., and Summerer, D. (2014). Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids. *ACS Chem. Biol.* 9 (12), 2815–2822. doi:10.1021/cb5006273

Rauch, B. J., Porter, J. J., Mehl, R. A., and Perona, J. J. (2016). Improved incorporation of noncanonical amino acids by an engineered tRNA(TYr) suppressor. *Biochemistry* 55 (3), 618–628. doi:10.1021/acs.biochem.5b01185

Robertson, W. E., Funke, L. F. H., de la Torre, D., Fredens, J., Elliott, T. S., Spinck, M., et al. (2021). Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* 372 (6546), 1057–1062. doi:10.1126/science.abg3029

Rodnina, M. V., and Wintermeyer, W. (2001). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu. Rev. Biochem.* 70, 415–435. doi:10.1146/annurev.biochem.70.1.415

Ruan, B. F., Palioura, S., Sabina, J., Marvin-Guy, L., Kochhar, S., LaRossa, R. A., et al. (2008). Quality control despite mistranslation caused by an ambiguous genetic code. *Proc. Natl. Acad. Sci. U. S. A.* 105 (43), 16502–16507. doi:10.1073/pnas.0809179105

Ryu, Y. H., and Schultz, P. G. (2006). Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*. *Nat. Methods* 3 (4), 263–265. doi:10.1038/nmeth864

Salazar, J. C., Ambrogelly, A., Crain, P. F., McCloskey, J. A., and Söll, D. (2004). A truncated aminoacyl-tRNA synthetase modifies RNA. *Proc. Natl. Acad. Sci. U. S. A.* 101 (20), 7536–7541. doi:10.1073/pnas.0401982101

Sambrook, J., and Russell, D. W. (2001). *Molecular cloning : a laboratory manual. Cold spring harbor*. N.Y.: Cold Spring Harbor Laboratory Press.

Schmitt, M. A., Biddle, W., and Fisk, J. D. (2018). Mapping the plasticity of the *Escherichia coli* genetic code with orthogonal pair-directed sense codon reassignment. *Biochemistry* 57 (19), 2762–2774. doi:10.1021/acs.biochem.8b00177

Schrader, J. M., Chapman, S. J., and Uhlenbeck, O. C. (2011). Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. *Proc. Natl. Acad. Sci. U. S. A.* 108 (13), 5215–5220. doi:10.1073/pnas.1102128108

Schwark, D. G., Schmitt, M. A., Biddle, W., and Fisk, J. D. (2020). The influence of competing tRNA abundance on translation: Quantifying the efficiency of sense codon reassignment at rarely used codons. *ChemBioChem* 21, 2274–2286. doi:10.1002/cbic.202000052

Schwark, D. G., Schmitt, M. A., and Fisk, J. D. (2018). Dissecting the contribution of release factor interactions to amber stop codon reassignment efficiencies of the *Methanocaldococcus jannaschii* orthogonal pair. *Genes* 9 (11), E546. doi:10.3390/genes9110546

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 (9), 931–934. doi:10.1039/b904228h

Thyer, R., d'Oelsnitz, S., Blevins, M. S., Klein, D. R., Brodbelt, J. S., and Ellington, A. D. (2021). Directed evolution of an improved aminoacyl-tRNA synthetase for incorporation of L-3, 4-dihydroxyphenylalanine (L-DOPA). *Angew. Chem. Int. Ed. Engl.* 60 (27), 14811–14816. doi:10.1002/anie.202100579

Urbonavicius, J., Qian, O., Durand, J. M. B., Hagervall, T. G., and Bjork, G. R. (2001). Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* 20 (17), 4863–4873. doi:10.1093/emboj/20.17.4863

Wang, K. H., Fredens, J., Brunner, S. F., Kim, S. H., Chia, T. S., and Chin, J. W. (2016). Defining synonymous codon compression schemes by genome recoding. *Nature* 539 (7627), 59–64. doi:10.1038/nature20124

Wang, N. X., Ju, T., Niu, W., and Guo, J. T. (2015). Fine-tuning interaction between aminoacyl-tRNA synthetase and tRNA for efficient synthesis of proteins containing unnatural amino acids. *ACS Synth. Biol.* 4 (3), 207–212. doi:10.1021/sb500195w

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 (2), 361–374. doi:10.1016/j.jmb.2009.10.030

Zaborske, J. M., DuMont, V. L. B., Wallace, E. W. J., Pan, T., Aquadro, C. F., and Drummond, D. A. (2014). A nutrient-driven tRNA modification alters translational fidelity and genome-wide protein coding across an animal genus. *PLoS Biol.* 12 (12), e1002015. doi:10.1371/journal.pbio.1002015

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 11 October 2022
DOI 10.3389/fmolb.2022.1031756

OPEN ACCESS

EDITED BY
Gustavo Fuertes,
Institute of Biotechnology (ASCR), Czechia

REVIEWED BY
Natalie Krahn,
Yale University, United States
Weimin Xuan,
Tianjin University, China

*CORRESPONDENCE
Patrick O'Donoghue,
patrick.odonoghue@uwo.ca

†These authors have contributed equally to this work

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

RECEIVED 30 August 2022
ACCEPTED 20 September 2022
PUBLISHED 11 October 2022

CITATION
Wright DE, Siddika T, Heinemann IU and O'Donoghue P (2022), Delivery of the selenoprotein thioredoxin reductase 1 to mammalian cells.
Front. Mol. Biosci. 9:1031756.
doi: 10.3389/fmolb.2022.1031756

Delivery of the selenoprotein thioredoxin reductase 1 to mammalian cells

David E. Wright†, Tarana Siddika†, Ilka U. Heinemann and Patrick O'Donoghue*

Departments of Biochemistry and Chemistry, The University of Western Ontario, London, ON, Canada

Over-expression of genetically encoded thioredoxin reductase 1 (TrxR1) TrxR1 can be toxic to cells due to the formation of a truncated version of the enzyme. We developed a new mammalian cell-based model to investigate TrxR1 activity. Fusion of the HIV-derived cell penetrating peptide (TAT) enabled efficient cellular uptake of purified TrxR1 containing 21 genetically encoded amino acids, including selenocysteine. The TAT peptide did not significantly alter the catalytic activity of TrxR1 *in vitro*. We monitored TrxR1-dependent redox activity in human cells using a TrxR1-specific red fluorescent live-cell reporter. Using programmed selenocysteine incorporation in *Escherichia coli*, our approach allowed efficient production of active recombinant human selenoprotein TrxR1 for delivery to the homologous context of the mammalian cell. The delivered TAT-TrxR1 showed robust activity in live cells and provided a novel platform to study TrxR1 biology in human cells.

KEYWORDS

cell biology, cell penetrating peptide, enzymology, genetic code expansion, redox biology, selenocysteine, thioredoxin reductase, trans-activator of transcription peptide

Introduction

Mammalian thioredoxin reductases (TrxR) are an important enzyme family that helps to protect cells from damage caused by reactive oxygen species (ROS) (Mustacich and Powis, 2000) and serves critical roles in redox signaling (Ren et al., 2017). TrxR1 is the most abundant member of the TrxR family and is expressed in the cytosol (Mustacich and Powis, 2000). TrxR1 is also a validated cancer target, and several efforts are underway to generate new inhibitors of TrxR1 that can compromise the rapid growth of cancer cells (Arnér, 2017; Hasan et al., 2022). For example, recent work on compounds with a 4,5-dichloropyridazinone core structure were identified as irreversible TrxR1 inhibitors that displayed potent toxicity in a series of cancer cells (Busker et al., 2020).

TrxR1 is a selenoprotein (Tamura and Stadtman, 1996), and its activity relies on the efficient incorporation of selenocysteine (Sec, U) in response to the UGA codon at position 550. TrxR1 is a complex enzyme that catalyzes the transfer of electrons from a bound NADPH cofactor, with the aid of a second FAD cofactor, to a disulfide N-terminal active site. From there, TrxR1 transfers electrons to the C-terminal selenylsulfide active

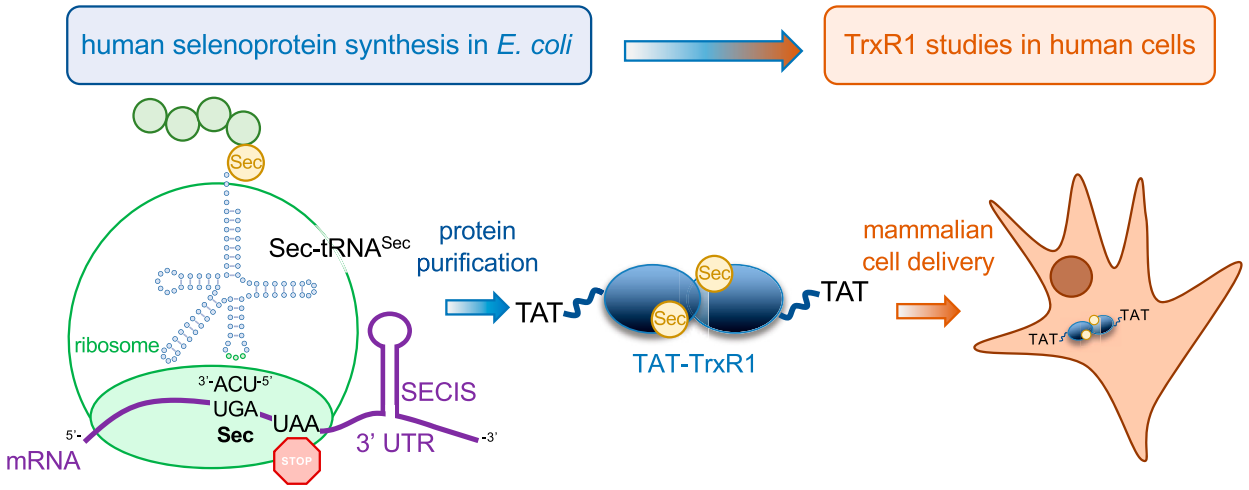


FIGURE 1
Production of recombinant selenocysteine containing TrxR1 in *E. coli* and delivery to mammalian cells. Recombinant production of human TrxR1 containing Sec550 was accomplished by fusing a bacterial Sec insertion sequence (SECIS) hairpin to the 3′ untranslated region of the human TrxR1 gene (Bröcker et al., 2014). We have previously confirmed that the approach leads to the stoichiometric installation of Sec and the production of abundant and active human TrxR1 from bacterial cells (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021). The construct used here also contains an N-terminal TAT peptide fusion (YGRKKRRQRRRPQ) that enables efficient delivery of purified recombinant TAT-TrxR1 to mammalian cells.

site (GCUG). The enzyme then reduces the selenylsulfide to selenolthiol that subsequently acts to reduce oxidized substrates in the cell. Thus, the Sec residue forms a critical part of the TrxR1 active site and is necessary for its enzymatic reduction of the oxidized thioredoxin (Trx), a major TrxR1 substrate (Wright et al., 2018). The reduced Trx1 is then able to resolve oxidative damage in cellular proteins (Lee et al., 2013). TrxR1 itself is also able to directly reduce protein disulfide substrates (Lundstrom-Ljung et al., 1995) and small molecule ROS, including hydrogen peroxide (Zhong and Holmgren, 2000), lipoic acid (May et al., 2007), selenite (Kumar et al., 1992), and quinones (Xia et al., 2003).

Because of the vital role of TrxR1 in human biology and applications in medicine, there is a continuing need to generate mammalian cells with precise levels of TrxR1 activity. Applications of these cell lines will provide insight into how TrxR1 activity regulates redox biology in human cells and provide a critical platform to screen new inhibitors of TrxR1. One roadblock in this area involves the challenges of generating mammalian cell lines that over-express TrxR1. Unfortunately, normal approaches to genetically over-express TrxR1 in mammalian cells lead to increased cell death. Over-expression of TrxR1 by stable transfection produced a > 2-fold increase in cell death in MCF-7 cells compared to empty vector control (Ma et al., 2001). A form of TrxR1 that stops at the UGA550 codon, and lacks the C-terminal Sec-Gly dipeptide, is toxic and induces cell death in human A549 cells (Anestal and Arnér, 2003). In HEK 293T cells, one recent report generated a stable cell line over-expressing TrxR1, yet no experiments were provided to demonstrate incorporation of Sec in TrxR1 or to evaluate apoptosis (Lu et al., 2022). Direct lipid-mediated transfection of TrxR1 protein lacking the two C-terminal residues caused a significant increase in cell death, while transfection of the full-length and Sec-containing TrxR1 showed no change in cell toxicity (Anestal and Arnér, 2003). These observations may result from selenium compromised thioredoxin reductase-derived apoptotic proteins (SecTRAPS) (Anestal et al., 2008). Indeed, inhibition of genetically encoded TrxR1 in mammalian cells can also produce SecTRAPS and induce apoptosis (Zhang et al., 2022).

We previously developed a genetic code expansion approach with Sec to generate recombinant human TrxR1 from *Escherichia coli* containing stoichiometric incorporation of Sec at position 550 (Figure 1) (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021). Here we used the same approach and produced a novel form of recombinant human TrxR1 that also contains an N-terminal cell penetrating peptide (CPP) tag derived from the human immunodeficiency virus (HIV) Trans-Activator of Transcription (TAT) protein (Han et al., 2000; Vives et al., 2003; Lichtenstein et al., 2021). CPPs, such as the TAT peptide, include a diverse and growing catalog of small peptide tags that can be used to deliver proteins (Kurrikoff et al., 2021), mRNAs (Yokoo et al., 2021), and small molecules (Tian and Zhou, 2021), including pharmaceutical compounds, directly to mammalian cells (Shoari et al., 2021). We used a CPP tag to circumvent the need to genetically over-

express TrxR1 due to the pitfalls associated with expressing TrxR1 in mammalian cells. Using a genetically encoded red fluorescent TrxR1 activity reporter (TrxRFP1) (Fan et al., 2017), we found that TAT-dependent cellular uptake was rapid and led to robust TrxR1 activity in mammalian cells. The TAT-tag provides a greatly needed platform to study TrxR1 biology and our technology will enable future studies to identify novel TrxR1 inhibitors for applications in cancer and other human diseases.

Materials and methods

Plasmids and molecular cloning

Wild-type TrxR1 was expressed in *E. coli* from the plasmid pET-TrxR1, previously described (Wright et al., 2018), that contains N-terminally His_6-tagged human TrxR1 (isoform 4) with a UGA codon at residue Sec550. The TrxR1 gene is followed by selenocysteine insertion sequence (SECIS) RNA hairpin loop, which is derived from the *E. coli fdhF* gene, in the 3′ untranslated region (3′ UTR) (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021). Site-specific insertion of Sec at the UGA550 codon depends on recruitment of the native *E. coli* Sec insertion machinery that decodes the UGA550 codon as Sec due to the adjacent SECIS in the 3' UTR, as previously (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021).

The bacterial expression vector pTAT-HA, a kind gift from Steven Dowdy (Addgene plasmid #35612), was used to produce TAT fused TrxR1 protein. The pTAT-HA vector includes an N-terminal His_6-tag upstream of the adjacent TAT peptide sequence (YGRKKRRQRRR) (Nagahara et al., 1998). The TrxR1 gene including the 3′UTR SECIS was polymerase chain reaction (PCR) amplified using primers with *Nco*I and *Eco*RI restriction sites on the 5′ and 3′ ends, respectively. Following digestion of the pTAT-HA vector and TrxR1 insert, T4 DNA ligase (New England Biolabs, Ipswich, MA, USA) was used to ligate TrxR1 into the *Nco*I and *Eco*RI sites in the similarly digested pTAT-HA vector. The resulting construct was verified by DNA sequencing (Azenta Life Sciences, New Jersey, USA). Sequences of the TrxR1 and TAT-TrxR1 gene constructs are available in the Supplementary Data.

TrxR1 protein purification

E. coli strain BL21 (DE3) was transformed with pET-TrxR1 or pTAT-TrxR1 plasmid. Following selection on lysogeny broth (LB) agar plates containing 100 μg/mL ampicillin, at least 3 independent single colonies were inoculated into separate overnight 10 mL pre-cultures in LB with 100 μg/mL ampicillin. The pre-cultures were then used to inoculate preparative 1 L cultures in LB containing 100 μg/mL ampicillin and 10 μM sodium selenite and were incubated with shaking at 37°C. We followed a previously optimized protocol for the production of selenoproteins in *E. coli* (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021). At A_{600} = 1.2, the temperature was decreased to 20°C. At A_{600} = 1.5, 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to the culture with shaking for 18 h to induce protein production. Cell pellets were harvested by centrifugation and stored at -80°C until purification. TrxR1 proteins were purified as previously described (Wright et al., 2018). Briefly, cell pellets were resuspended in 30 ml phosphate buffer (100 mM potassium phosphate, pH 7.2, 10% glycerol), supplemented with lysozyme (1 mg/mL), and disrupted by sonication with 70% amplification. After centrifugation at 6,250 × *g* for 1 h, clear cell lysates were purified by affinity chromatography using Ni^{2+}-Nitrilotriacetic acid (NTA) resin (HisPur Ni-NTA Resin, PI88222, ThermoFisher Scientific, Mississauga, ON, Canada) as detailed before (Wright et al., 2018). Purified TrxR1 proteins were kept in storage buffer containing 100 mM potassium phosphate pH 7.2 and 50% glycerol at -80°C until use.

TrxR1 *in vitro* activity assays

TrxR1 activity was monitored, as we did before (Wright et al., 2018), using 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB) to detect the level of reductive activity from TrxR1 and TAT-TrxR1. Active TrxR1 reduces DTNB to 2-nitro-5-thiobenzoate (TNB^{2-}), which absorbs at wavelength 414 nm (A_{414}). Each reaction mixture contained 50 nM or 100 nM TrxR1, 300 μM Nicotinamide adenine dinucleotide phosphate (NADPH) and 5 mM DTNB in buffer containing 100 mM potassium phosphate, 1 mM Ethylenediaminetetraacetic acid (EDTA) pH 7.0. Protein concentration was assessed by the Bradford protein assay (Catalog #5000006; Bio-Rad) at 595 nm according to the manufacturer's instructions. Reactions were initiated in a 96 well micro-plate by the addition of DTNB to the reaction mixture containing TrxR1 (at 50 nM or 100 nM as indicated) and NADPH in a final volume of 100 μL. Measurements were taken in a Biotek Synergy H1 microplate reader every 1 min for an 80 min time course. All assays were performed in triplicate using three independent enzyme reactions for each condition tested.

Endogenous TrxR1 activity in live cells and fluorescence microscopy

A TrxR1 activity biosensor (TrxRFP1) was used to monitor the TrxR1 activity in live cells by measuring the changes in red fluorescence of the reporter (Fan et al., 2017). The pcDNA3-TrxRFP1 was a kind gift from Huiwang Ai (Addgene plasmid # 98996). HEK 293T cells were cultured in Dulbecco's Modified Eagle Medium (Cellgro DMEM, ThermoFisher Scientific)

containing 10% fetal bovine serum and 1% penicillin/streptomycin in 5% CO_2 at 37°C. Equal numbers of HEK 293T cells were seeded onto a 6 well culture plates containing media and transfected with Lipofectamine 2000 (Invitrogen) and 2 μg/mL plasmid DNA bearing the TrxRFP1 reporter according to manufacturer's instructions. In three biological and three technical replicates, un-transfected cells and transfected cells were monitored by fluorescent microscopy (excitation 542 ± 20 nm, emission 593 ± 40 nm) using an EVOS FL Cell Imaging System (ThermoFisher Scientific). At 24 h after transfection, cell images were acquired every 1 min for a 60 min time course. At the 10 min time point, TrxR1 inhibitor aurothioglucose (ATG) was added to the media at a concentration of 100 μM.

TAT-TrxR1 delivery and activity in live cells

HEK 293T cells were transfected with the plasmid bearing the TrxRFP1 reporter as above. At 24 h after transfection, cells were incubated with no protein or with 0.5 μM of TrxR1 or TAT-TrxR1 protein. Immediately before addition of the protein and at 30 min and at 24 h after incubation with protein or no protein, bright field and fluorescent (excitation 542 ± 20 nm, emission 593 ± 40 nm) cell images for each condition were acquired EVOS FL Cell Imaging System. Fluorescence intensities were analyzed by image J software and statistical analysis was based on 3 biological replicates with 3 technical replicates each using pairwise single factor analysis of variance (ANOVA).

Western blotting

HEK 293T cells were cultured in DMEM (Cellgro) media containing 10% fetal bovine serum and 1% penicillin/streptomycin in 5% CO_2 at 37°C. Equal numbers of cells were plated onto 6 well culture plates and incubated with buffer only or with 1 μM TrxR1 or TAT-TrxR1 protein. After a 1-h incubation, cells were washed with phosphate buffered saline (PBS) and cell pellets were collected by centrifugation. Cells were resuspended in lysis buffer containing (50 mM Na_2HPO_4, 1 mM $Na_4P_2O_7$, 20 mM NaF, 2 mM EDTA, 2 mM EGTA, 1 mM dithiothreitol, 300 μM phenylmethylsulfonyl fluoride (PMSF), and protease inhibitor tablet (Catalog number 04 693 159 001, Roche Canada, Mississauga, ON, Canada). Pellets were kept on ice for 10 min and vortexed every 2 min. Supernatant was collected from cells representing three biological replicates by centrifugation at 15,000 × *g* for 15 min at 4°C.

Protein concentration was measured by Bradford assay and extracted proteins were diluted to 5 μg/μL, and 50 μg of protein was separated using 10% sodium dodecyl sulfate-poly acrylamide gel electrophoresis (SDS-PAGE). The protein was transferred to a polyvinylidene difluoride (PVDF) membrane using Turbo-Blot Turbo transfer system (Bio-Rad). The membrane was blocked in PBS with 0.1% (v/v) Tween 20 (PBST) and 5% (w/v) skim milk powder for 1 h at room temperature. The membrane was immunoblotted with specific anti-TrxR1 (B-2) mouse monoclonal primary antibody (sc-28321, Santa Cruz Biotechnology, Santa Cruz, CA, USA) overnight at 4°C followed by 3 × 10 min washes with agitation in PBST with 1% (w/v) skim milk powder. The membrane was then incubated in fluorescent secondary antibody (Goat-anti-mouse #AQ127, Sigma Aldrich, Oakville, Canada) for 1 h at room temperature with agitation and washed 3 × 10 min in PBST. The membrane was stored in PBS for further analysis and bands were visualized using a LiCor (LiCor Biosciences, Lincoln, Nebraska USA) fluorescence imager (693 nm emission, 700 nm excitation). A glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibody (#ab8245, abcam, Toronto, Canada) and secondary fluorescent antibody (Goat-anti-mouse #AQ127, Sigma Aldrich, Oakville, Canada) were used to establish loading controls.

In identical and independent experiments, additional western blots were performed by chemiluminescence imaging with the His_6 and Vinculin antibodies. Briefly, the membrane was immunoblotted with specific mouse anti-His antibody (#H1029; Sigma, Oakville, Canada) overnight at 4°C followed by 3 × 10 min washes with agitation in PBST with 1% (w/v) skim milk powder. The membrane was then incubated in sheep anti-mouse immunoglobulin G horseradish peroxidase linked antibody (#GENA931, GE Healthcare, Oakville, Canada) for 1 h at room temperature with agitation and washed 3 × 10 min in PBST. The membrane was stored in PBS for further analysis and chemiluminescent signal detection was performed on a ChemiDoc MP system (BioRad). Mouse anti-vinculin antibody (#sc25336, Santa Cruz Biotechnology, Dallas, Texas, USA) and secondary antibody sheep anti-mouse immunoglobulin G horseradish peroxidase linked antibody (#GENA931 GE Healthcare, Oakville, Canada) were used to establish loading controls.

Results

Purification and activity of TrxR1 and TAT-TrxR1 variants

In previous work (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021), we developed a facile approach to produce human TrxR1 in *E. coli*. The method involves recombinant expression of the human TrxR1 gene with a bacterial SECIS derived from the *E. coli fdhF* gene that was appended to the 3' UTR (Figure 1). As we have demonstrated using mass spectrometry (MS) (Bröcker et al., 2014; Wright et al., 2018), elemental analysis with inductively coupled plasma mass spectrometry (ICP-MS) (Bröcker et al., 2014; Wright et al., 2018), and biochemical activity (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021), our method leads to a yield of

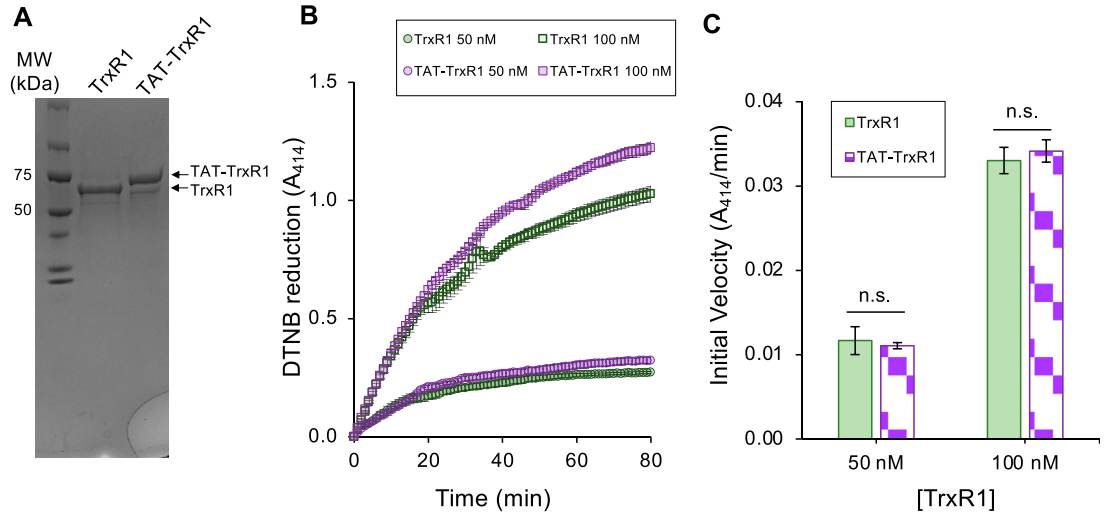


FIGURE 2
Purity and biochemical activity of TrxR1 and TAT-TrxR1 variants. The wild-type TrxR1 and TAT-tagged TrxR1 (TAT-TrxR1) were produced in *E. coli* according to our established approach (Figure 1) (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021). Both the wild-type and TAT-TrxR1 variants were produced at high yield (~2 mg/L *E coli* culture). **(A)** Coomassie stained gel shows the purity and molecular weights of TrxR1 and TAT-TrxR1. We used the reduction of Ellman's reagent (DTNB) monitored by absorbance at 414 nm (A_{414}) to assay **(B)** the activity of the wild-type (green curves) or TAT-TrxR1 (purple curves) at a concentration of 50 nM or 100 nM enzyme. Based on linear regression to the linear phase of the enzyme reactions, we derived **(C)** the initial velocity (A_{414}/min) of each reaction and determined that there was no significant different between TrxR1 and TAT-TrxR1 activity at 50 nM or 100 nM concentrations. Error bars represent 1 standard deviation of 3 independent enzyme reactions and statistical analysis was based on pairwise single factor ANOVA (ns, not significant).

~2 mg/L *E. coli* culture of active TrxR1 with stoichiometrically incorporated Sec at the UGA550 codon (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021). We previously characterized the incorporation of Sec in recombinant TrxR1 using tandem MS/MS analysis to demonstrate Sec incorporation, and we also conducted ICP-MS, which demonstrated quantitative incorporation at a level of 98.5% Sec in the recombinant TrxR1 protein (Bröcker et al., 2014). In independent work, we again provided tandem MS/MS analysis to demonstrate Sec incorporation in TrxR1, and we also performed matrix-assisted laser desorption/ionization (MALDI)-MS analysis to demonstrate Sec incorporation at UGA550 (Wright et al., 2018). Finally, we again employed ICP-MS to demonstrate quantitative incorporation of Sec in the recombinant TrxR1 protein (Wright et al., 2018).

Here, we modified this original TrxR1 construct to include a TAT peptide at the N-terminus of the TrxR1 construct that produces a TAT-TrxR1 protein. Following purification, we confirm similar purity and yield of the TAT-TrxR1 protein compared to the wild-type TrxR1 enzyme (Figure 2A). We previously characterized recombinant TrxR1 activity using Ellman's reagent (5,5′-Disulfanediylbis (2-nitrobenzoic acid, DTNB), 9,10-phenanthrequinone, as well as an insulin linked Trx1 assay (Wright et al., 2018). We found that recombinant TrxR1 showed similar and robust activity with each of these substrates (Wright et al., 2018). Because of this fact and because reduction of DTNB is a straightforward assay for TrxR1 activity that is monitored at A_{414} by the yellow color of the reduced form of DTNB (2-nitro-5-thiobenzoate anion, TNB^{2-}), we chose to characterize the TAT-TrxR1 activity using DTNB.

Thus, to compare the activity of the wild-type and TAT-tagged TrxR1 enzymes, we measured the DTNB reduction activity of TrxR1 and TAT-TrxR1 at 50 nM and 100 nM concentrations over a time course (Figure 2B). In comparing the initial velocity of the reactions at 50 nM or 100 nM TrxR1 or TAT-TrxR1 concentrations, we recorded no significant difference between the wild-type or TAT-tagged TrxR1 proteins (Figure 2C). The TAT-TrxR1 has activity that is robust and indistinguishable from the recombinant TrxR1 protein and because the activity of TrxR1 relies critically on Sec550 incorporation, we can confirm the TAT-TrxR1 protein contains the same level of Sec as in TrxR1, which we documented before using multiple mass spectrometry methods (Bröcker et al., 2014; Wright et al., 2018). Thus, the TAT-tag does not significantly affect that activity of the TrxR1 enzyme, and the TAT-tagged version of the protein will provide an appropriate model of TrxR1 activity that can be readily delivered to mammalian cells.

Assessing the TrxRFP1 reporter in mammalian cells

Before attempting to deliver the TrxR1 or TAT-TrxR1 proteins to mammalian cells, we characterized the response of a live-cell reporter for TrxR1 activity. The TrxRFP1 reporter was

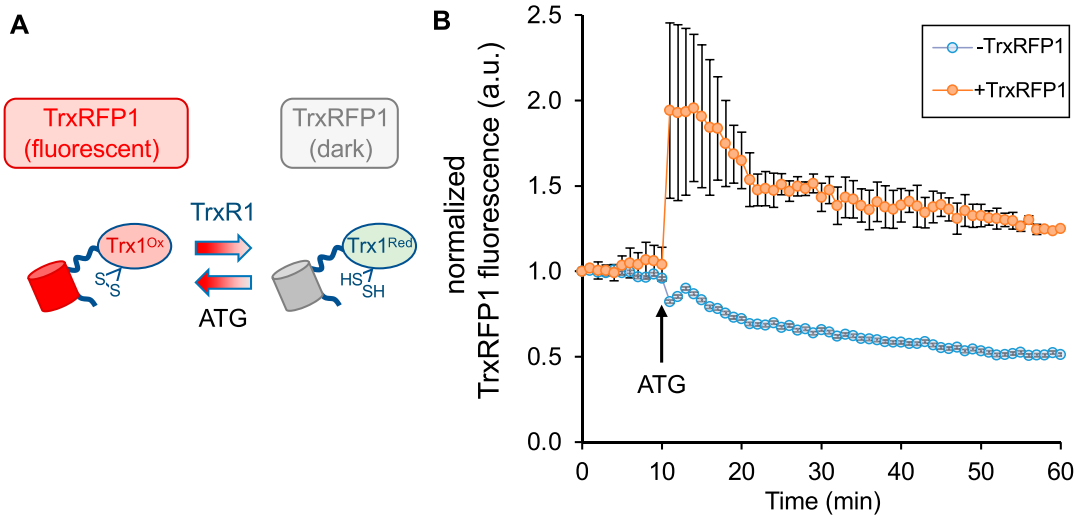


FIGURE 3
Measurement of endogenous TrxR1 activity in live cells. HEK 293T cells were transfected with a plasmid expressing a TrxR1 specific activity reporter. **(A)** The reporter (TrxRFP1) is a fusion of RFP with the TrxR1 substrate protein Trx1 (Fan et al., 2017). In the oxidized state ($Trx1^{Ox}$) the fused RFP (red cylinder) is fluorescent, while in the reduced state ($Trx1^{Red}$) the RFP fluorescence (gray cylinder) is lost in proportion to the amount of TrxR1 activity. **(B)** The time course begins 24 h after transfection and records TrxRFP1 fluorescence (excitation 542 ± 20 nm, emission 593 ± 40 nm) in HEK 293T cells transfected with (orange circles) or without (blue circles) the plasmid encoding TrxRFP1. Following application of TrxR1 inhibitor aurothioglucose (ATG) at 10 min after we began recording red fluorescence, cells with the TrxRFP1 reporter show a strong induction of red fluorescence that is proportional to the level of TrxR1 inhibition. Error bars represent ±1 standard deviation of three biological replicates.

engineered and developed previously (Fan et al., 2017). We obtained a plasmid that genetically encodes the TrxRFP1 reporter, which is a fusion protein of RFP with the natural substrate of TrxR1, Trx. In TrxRFP1, Trx1 was fused to the N-terminal of an engineered RFP protein where native cysteine residues (Cys) were replaced in the protein and a new redox coupled pair of Cys residues was introduced near the N- and C-terminus of the RFP domain. Thus, following reduction of the Trx1 domain by TrxR1, the Trx1 moiety can then reduce the engineered Cys redox couple in the RFP domain and eliminate red fluorescence of RFP. Once expressed in mammalian cells, TrxRFP1 fluorescence is inversely correlated with TrxR1 activity. In the reduced form, the Trx1 domain of the reporter disrupts the normal red fluorescence of the fused RFP, while in the oxidized form of Trx1 the red fluoresce of RFP is restored in proportion to the amount of oxidized Trx1. Because Trx1 is a selective substrate of TrxR1 (Fan et al., 2017), the system provides a specific and robust biosensor to monitor TrxR1 activity in live cells (Figure 3A).

To test the reporter in our hands, we transiently transfected the plasmid bearing TrxRFP1 into HEK 293T cells, and we used cells transfected without plasmid as a control. At 24 h after transfection, we began monitoring red fluorescence in the control cells and in cells transfected with the TrxRFP1 plasmid. After 10 min, we then added the well-established TrxR1 inhibitor aurothioglucose (ATG) (Gromer et al., 1998) to the media at a concentration of 100 μM. Using fluorescence microscopy, we observed a sharp increase in red fluorescence across 3 biological replicates, only in cells expressing the TrxRFP1 reporter (Figure 3B). The data demonstrate inhibition of endogenous TrxR1 in live HEK 293T cells, and in agreement with previous observations (Fan et al., 2017), confirm the proper function of the TrxRFP1 reporter.

Delivery of TAT-TrxR1 to mammalian cells

As noted above, genetic approaches to over-express TrxR1 in mammalian cells can lead to substantial cytotoxicity, and lipid-based protein transfection approaches with TrxR1 (Anestal and Arnér, 2003) or other proteins (Suhorutsenko et al., 2011) are of limited efficiency. Furthermore, lipid-based transfection reagents also induce toxicity and immunogenic reactions in mammalian cells (Suhorutsenko et al., 2011), demonstrating the need for an improved approach to deliver active TrxR1 to cells.

In the absence of any transfection reagents, we incubated HEK 293T cells with buffer only, with wild-type TrxR1, or with TAT-tagged TrxR1. At 1 h after incubation with the proteins, cells were washed to remove any extracellular protein and analyzed by western blotting with anti-TrxR1 and with anti-GAPDH as a loading control (Figure 4A). In the case of TAT-TrxR1, we observed a consistent level of successfully delivered TAT-TrxR1 to cells, as evident by the larger molecular weight of TAT-TrxR1 compared to the endogenous protein.

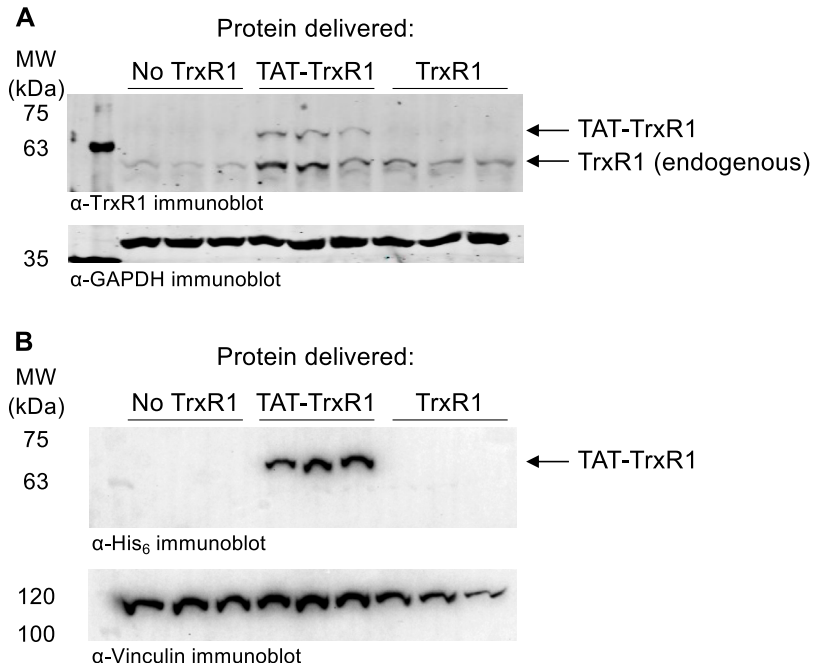


FIGURE 4
Immunoblotting confirms delivery of TAT-TrxR1 to cells. **(A)** Biological triplicates of HEK 293T cells were incubated for 1 h with buffer only (No TrxR1), with TAT-tagged TrxR1 (TAT-TrxR1), or with TrxR1 lacking the TAT-tag (TrxR1). Following incubation, cells were washed to remove any protein remaining outside of the cells. The location of endogenous TrxR1 and TAT-tagged TrxR1 are indicated in a western blot with anti-TrxR1 antibody (above), and GAPDH was used as a loading control (below). **(B)** In an independent set of biological triplicates an identical experiment was performed. HEK 293T cell were incubated for 1 h with buffer only (No TrxR1), with TAT-tagged TrxR1 (TAT-TrxR1), or with TrxR1 lacking the TAT-tag (TrxR1). Following washing of the cells, a western blot was performed with the anti-His_6 antibody as well as the anti-Vinculin antibody as a loading control. Both TAT-TrxR1 and TrxR1 protein contain the His_6 tag, yet only the TAT-TrxR1 protein was imported into the cells.

To confirm selective uptake of the TAT-TrxR1, we also performed western blotting of an identical and independent set of biological triplicate experiments using the anti-His_6 antibody and anti-Vinculin antibody as a loading control. As above, at 1 h after incubation of HEK 293T cells with buffer only, with wild-type TrxR1, or with TAT-tagged TrxR1, cells were washed to remove any protein remaining outside of the cells. Both the TAT-TrxR1 and TrxR1 recombinant proteins contain the His_6 tag. We detected only the TAT-TrxR1 as successfully delivered to the interior of the cells (Figure 4B).

TAT-tagged dependent uptake and activity of TAT-TrxR1 in live cells

To determine if the successfully delivered TAT-TrxR1 was active in HEK 293T cells, we conducted a series of experiments using the TrxRFP1 reporter examined above. First, HEK 293T cells were transfected with the plasmid encoding the TrxRFP1 reporter. At 24 h after transfection, cells were incubated with buffer only, with wild-type TrxR1, or with TAT-TrxR1. We then monitored the changes in red fluorescence of the TrxRFP1 reporter at 0 min, 30 min, and 24 h after incubation with the TrxR1 protein variants or with no protein added. At each time point, both brightfield and fluorescent images of the cells were captured and quantitated (Figure 5). As anticipated, in the case of the wild-type TrxR1 and the no protein control, we found no significant change in the TrxRFP1 fluorescence over the 24-h time course. In contrast, only the cells incubated with TAT-TrxR1 showed significant and marked decreases in red fluorescence (Figure 5).

These data agree precisely with our findings based on western blotting (Figure 4). Because only the TAT-TrxR1 was delivered successfully to the interior of the cells, only cells treated with TAT-TrxR1 and not those treated with TrxR1 showed a significant increase in TrxR1-specific activity. Thus, together the data demonstrate that only incubation with TAT-TrxR1 led to robust TrxR1-specific reduction activity inside the live human cells. The TAT-tag, therefore, represents a facile strategy to deliver active human TrxR1 selenoprotein to mammalian cells.

As a first demonstration of the TAT-tagging approach to delivery selenoprotein to human cells, we chose to work with HEK 293T cells. These cells are well established cell biological model systems that display robust growth and high transfection efficiencies (Thomas and Smart, 2005). This aspect was important since we relied on the

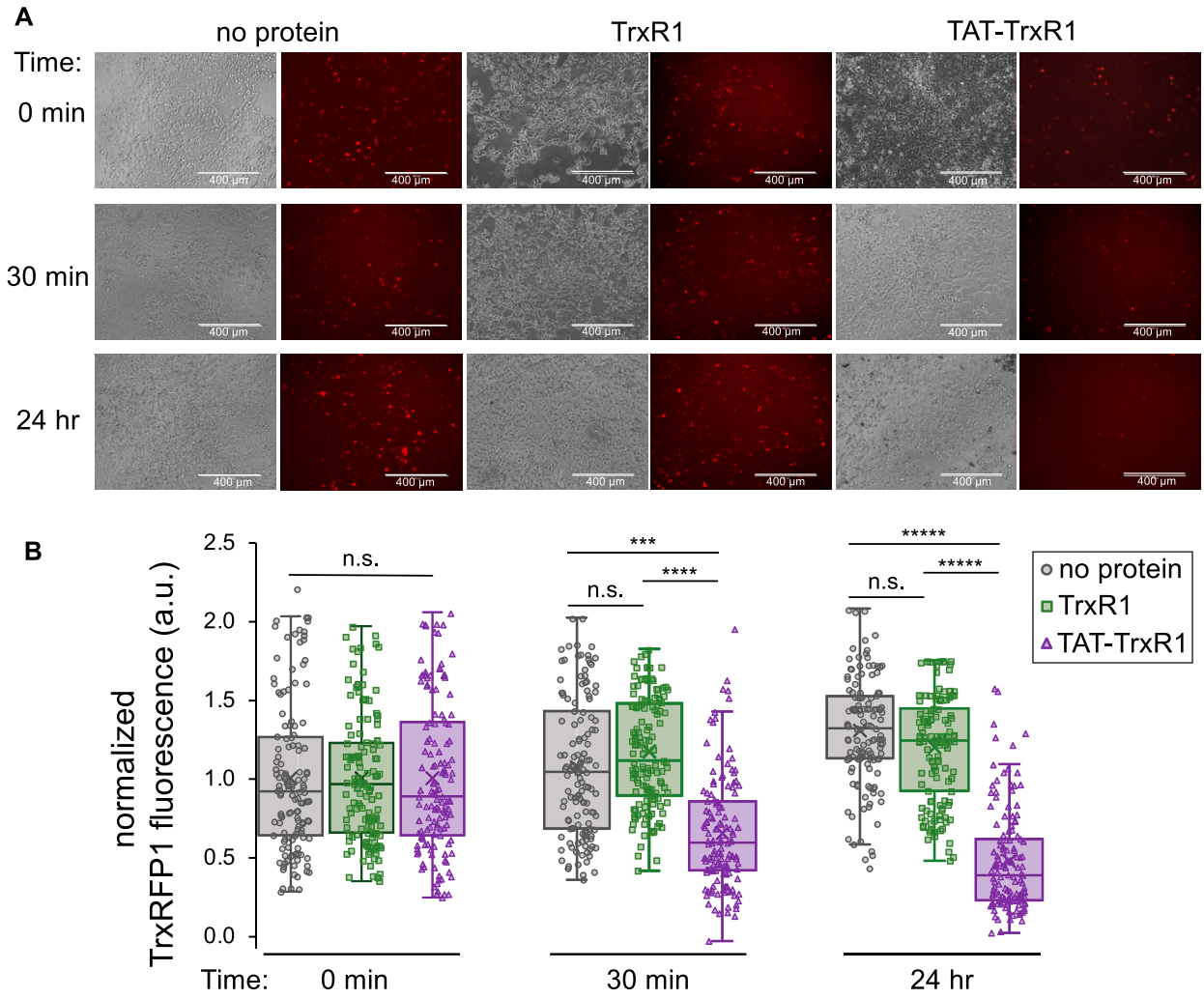


FIGURE 5
TAT-dependent delivery of active TrxR1 to mammalian cells. At 24 h after transfection with a plasmid bearing the TrxRFP1 activity reporter, cells were incubated with buffer only (no protein), with wild-type TrxR1 lacking the TAT-tag (TrxR1), or with TAT-tagged TrxR1 (TAT-TrxR1). **(A)** Following addition of protein to the media, brightfield images of the cells and fluorescence images of the TrxRFP1 reporter (excitation 542 ± 20 nm, emission 593 ± 40 nm) were recorded before addition of protein (0 min), at 30 min, and at 24 h after protein delivery. Decreasing red fluorescence demonstrates increased TrxR1 activity inside the cells. **(B)** At the 0 min time point, the average value of the fluorescence was normalized to 1.0 across each condition. Only the cells incubated with TAT-TrxR1 displayed a time dependent and statistically significant decrease in TrxRFP1 fluorescence during the time course. Data points taken from individual cells treated with no protein (gray circles), TrxR1 (green squares), or TAT-TrxR1 (purple triangles) are shown overlaid with box and whisker plots indicating the mean (x), median (line), range (bars), and lower and upper quartiles (boxed) of the data. Statistical analysis was based on 3 biological replicates and 3 technical replicates each and pairwise single factor ANOVA (ns, not significant; ***$p < 0.005$; ****$p < 0.0001$; *****$p < 10^{-5}$).

TrxRFP1 live cell reporter to demonstrate the activity from the TAT-TrxR1 protein delivered to the human cells. Furthermore, since HEK 293T cells are derived from kidney cells, where endogenous TrxR1 is abundant and active (see Figures 3, 4), these particular cells are also appropriate models systems to study TrxR1 biology (Zhu et al., 2022), including in the context of chemotherapeutics where TrxR1 has been implicated as a major source of adverse effects of cisplatin compounds (Cheng et al., 2020). In the future, we will apply these approaches to other human and mammalian cell lines.

Discussion

Genetic code expansion and codon reassignment

Genetic code expansion is a cornerstone application in synthetic biology that enables protein production with amino acid building blocks beyond the canonical 20 amino acids that are normally used to make proteins (O'Donoghue et al., 2013). Here

we used a genetic code expansion approach to incorporate the 21st amino acid, Sec, into a recombinant human protein produced in *E. coli*. The genetic code expansion method, which we established previously (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021), involves expression of a synthetic gene construct where we fused SECIS from the 3′UTR of an *E. coli fdhF* gene to a human TrxR1 gene that we codon optimized for *E. coli*. The resulting protein product is an active TrxR1 enzyme containing 21 different amino acids. As a novel extension of our approach, here we employed the CPP TAT-tag to enable our studies of the recombinant human TrxR1 enzyme in the homologous context of human cells without any engineering steps to otherwise modify the human cells themselves.

Genetic code expansion systems enable site-specific incorporation of non-canonical amino acids (ncAAs) into proteins to introduce 21 (Bain et al., 1992; Bröcker et al., 2014; Haruna et al., 2014), 22 (Köhrer et al., 2003; Venkat et al., 2018; Wright et al., 2018; Wright et al., 2021), or recently, in a few cases, 23 different amino acids into proteins (Dunkelmann et al., 2020; Tharp et al., 2021). These approaches, which normally require engineering a new tRNA synthetase and tRNA pair, are powerful in their ability to introduce new chemistry into proteins and functionalities that enable bio-orthogonal protein labeling (Ngo and Tirrell, 2011). Applications include creating new probes for protein function (O'Donoghue et al., 2021; Courtney and Deiters, 2018) and protein structure (Mihaila et al., 2022) as well as protein and peptide pharmaceuticals to provide novel antibiotics (Vinogradov and Suga, 2020), vaccines (Si et al., 2016), and to treat major human diseases, including cancers (Huang and Liu, 2018) and auto-immune disorders (Gauba et al., 2011).

Despite the success of genetic code expansion and its many applications in biology as well as studies of health and disease, the methods often involve re-assignment of a particular codon, normally one or more of the stop codons (UGA, UAA, and UAG), not only in a particular protein of interest but also in every instance of that codon throughout the proteome. Indeed, this situation leads to the extension of many natural proteins beyond their normal termination point (Heinemann et al., 2012). When used in a cell free context (Ranji Charna et al., 2022), or simply in a production host to make proteins for downstream applications, this defect is of less importance. Genetic code expansion approaches, however, were also deployed in the very same cells that are the exact object of study. In one example, expanding the genetic code by reassigning the UAG codon to N_{ε}-acetyl-lysine in mammalian cells to study the impact of a specific histone modification also caused transcriptome-wide changes in gene expression ranging from 2 to 100-fold (Elsasser et al., 2016). Thus, the engineered cells were substantially more affected by the codon reassignment than by the insertion of a specific acetylation into histone proteins. The combination of CPPs with genetic code expansion that we demonstrated here allowed engineering in one kind of cell and then subsequent delivery of a protein product of genetic code expansion into otherwise naïve mammalian cells, thus minimizing perturbations associated with expanding the genetic code.

Delivery of ncAA-containing proteins with cell penetrating peptides

Since genetic code expansion can have such a dramatic impact on the nature of the cell under investigation, we sought to devise a novel approach to deliver proteins containing ncAAs to mammalian cells, eliminating the need to engineer the mammalian cells themselves. CPPs represent a diverse and growing catalog of small peptides that have already demonstrated great utility in facilitating the delivery of proteins (Kurrikoff et al., 2021), mRNAs (Yokoo et al., 2021), and small molecules (Tian and Zhou, 2021) to mammalian cells (Shoari et al., 2021). The TAT peptide is derived from the HIV TAT protein, which is essential for HIV replication (Arya et al., 1985). A chemically synthetized 86 amino acid full-length TAT protein was rapidly taken up by HeLa cells, and the protein delivery activity was localized to the region including residues 37-72 of the synthetic TAT protein (Green and Loewenstein, 1988). Shortly thereafter, this same region of the TAT protein was chemically cross-linked to several different larger proteins, including beta-galactosidase, horseradish peroxidase, RNase A, and domain III of *Pseudomonas* exotoxin A that were all shown to be rapidly delivered to several different mammalian cell types in culture (Fawell et al., 1994).

Our work demonstrated a novel approach to deliver an active and full-length human selenoprotein to mammalian cells *via* fusion with a genetically encoded CPP for the first time. We found that the selenoprotein TAT-TrxR1 was efficiently delivered to cells and already displayed significant activity 30 min after incubation of the protein with HEK 293T cells. We found the TAT-TrxR1 continued to display robust and TrxR1-specific reduction activity in the cells for at least 24 h after protein delivery. To our knowledge, there are no other examples in the literature using cell penetrating peptides to deliver Sec-containing full-length proteins to mammalian cells. A recent study successfully engineered a Sec-containing small peptide, called PSELT (FQICVSUGYR), and demonstrated efficient delivery of the peptide to human neuroblastoma SH-SY5Y cells (Alsharif et al., 2021). PSELT protected the cells against oxidative damage associated with Parkinson's disease, indicating an important avenue for therapeutic application of Sec-containing peptides.

Indeed, there is just one example of delivering ncAA-containing protein to mammalian cells using a CPP tag. The report demonstrated production of the murine dihydrofolate reductase (DHFR) in *E. coli* containing the ncAA N_{ε}-

propargyloxycarbonyl-lysine (Bi et al., 2018). The authors then used click chemistry along with an azide-functionalized TAT peptide to chemically link the TAT-peptide to the DHFR protein, achieving a 90% yield. The DHFR protein was also chemically cyclized to help prevent degradation in HeLa cells. Although the engineered DHFR protein showed relatively efficient delivery to the cells at 1 h after incubation with the protein, by the 2-h time point most of the protein was degraded. In the case of the linear or non-cyclized TAT functionalized DHFR, nearly all of the protein was degraded after 2 h (Bi et al., 2018). Thus, our work demonstrates additional novelty in that the TAT-tag was genetically encoded in *E. coli* in the same protein as the ncAA Sec, and no downstream chemical synthetic steps were required. Our simpler approach also showed a longer lasting effect as we observed robust activity of the TAT-TrxR1 protein up to the 24-h time point.

A new system to study TrxR1-dependent activity in mammalian cells

Our approach to engineering selenoproteins production in *E. coli* has allowed us to develop a new system to study TrxR1 redox biology in a homologous context of the mammalian cell without engineering the mammalian cells. Our approach also requires no addition of costly and toxic transfection reagents. Previous work found that lipid-mediated transfection of TrxR1 was achieved but with limited efficiency (Anestal and Arnér, 2003). Indeed, compared to standard protein transfection reagents, including lipofectamine derivates, cell penetrating peptides, including TAT are well tolerated in mammalian cells. While lipofectamine can strongly depress cell viability, induced apoptosis, and generate significant immune responses in human acute monocytic leukemia cells (THP-1), cell penetrating peptides were found to be neither toxic nor immunogenic (Suhorutsenko et al., 2011).

Conclusion

We demonstrated a simple and efficient approach to deliver the active human selenoprotein TrxR1 to mammalian cells. We showed that a genetically encoded TAT-tagged TrxR1 was produced in *E. coli* with similar efficiency and activity to an untagged TrxR1 in *E. coli* that we already demonstrated had stoichiometric incorporation of Sec550 by multiple methods (Bröcker et al., 2014; Wright et al., 2018; Wright et al., 2021). The TAT-tag enabled rapid delivery of the active selenoprotein to mammalian cells that provided robust Trx1-specific reductive activity for at least 24 h. Given the complex nature of selenoprotein synthesis in mammalian cells, our approach represents an important step forward to aid studies in the native and homologous context of mammalian cells of other human selenoproteins as well as ncAA-containing proteins produced in engineered cells with expanded genetic codes. Recent breakthroughs also enable site-specific incorporation of Sec at any location in a recombinant protein in *E. coli* (Haruna et al., 2014; Liu et al., 2018; Thyer et al., 2018; Welegedara et al., 2018; Wang et al., 2021), thus indicating that our approach with cell penetrating peptides can be applied to other selenoproteins and in combination with other genetic code expansion systems.

Finally, because TrxR1 is an important and validated drug target for cancers (Arnér, 2017; Busker et al., 2020; Hasan et al., 2022) and other human diseases, including Rheumatoid arthritis (Duan et al., 2022) and neurodegeneration (Wang et al., 2022), the system we developed here will not only be a valuable tool for studying TrxR1 biology in human cells, but also represents a novel platform to screen inhibitors of TrxR1. Since TrxR1 inhibitors have already found application in clinical settings (Sachweh et al., 2015; Zhang et al., 2021), our approach will facilitate applications in developing the next generation of therapeutics to target TrxR1 activity in human diseases.

Data availability statement

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

Author contributions

The authors are recognized for the following contributions: designed experiments (DW and PO), performed experiments (DW and TS), analysed data (DW, TS, and PO), wrote the paper (TS, IH, and PO), edited the paper (TS, IH, and PO).

Funding

This work was supported from the Natural Sciences and Engineering Research Council of Canada (04282 to PO; 04776 to IH); Canada Foundation for Innovation (229917 to PO); the Ontario Research Fund (229917 to PO); Canada Research Chairs (232341 to PO); and the Canadian Institutes of Health Research (165985 to PO); Ontario Ministry of Research and Innovation (ER-18-14-183 to IH).

Acknowledgments

We are grateful to Paul Walton for critical discussions and suggestions.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.1031756/full#supplementary-material

References

Alsharif, I., Boukhzar, L., Lefranc, B., Godefroy, D., Aury-Landas, J., Rego, J. D., et al. (2021). Cell-penetrating, antioxidant SELENOT mimetic protects dopaminergic neurons and ameliorates motor dysfunction in Parkinson's disease animal models. *Redox Biol.* 40, 101839. doi:10.1016/j.redox.2020.101839

Anestal, K., and Arnér, E. S. (2003). Rapid induction of cell death by selenium-compromised thioredoxin reductase 1 but not by the fully active enzyme containing selenocysteine. *J. Biol. Chem.* 278 (18), 15966–15972. doi:10.1074/jbc.M210733200

Anestal, K., Prast-Nielsen, S., Cenas, N., and Arnér, E. S. (2008). Cell death by SecTRAPs: Thioredoxin reductase as a prooxidant killer of cells. *PLoS One* 3 (4), e1846. doi:10.1371/journal.pone.0001846

Arnér, E. S. J. (2017). Targeting the selenoprotein thioredoxin reductase 1 for anticancer therapy. *Adv. Cancer Res.* 136, 139–151. doi:10.1016/bs.acr.2017.07.005

Arya, S. K., Guo, C., Josephs, S. F., and Wong-Staal, F. (1985). Trans-activator gene of human T-lymphotropic virus type III (HTLV-III). *Science* 229 (4708), 69–73. doi:10.1126/science.2990040

Bain, J. D., Switzer, C., Chamberlin, A. R., and Benner, S. A. (1992). Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature* 356 (6369), 537–539. doi:10.1038/356537a0

Bi, X., Yin, J., Hemu, X., Rao, C., Tam, J. P., and Liu, C. F. (2018). Immobilization and intracellular delivery of circular proteins by modifying a genetically incorporated unnatural amino acid. *Bioconjug. Chem.* 29 (7), 2170–2175. doi:10.1021/acs.bioconjchem.8b00244

Bröcker, M. J., Ho, J. M., Church, G. M., Söll, D., and O'Donoghue, P. (2014). Recoding the genetic code with selenocysteine. *Angew. Chem. Int. Ed. Engl.* 53 (1), 319–323. doi:10.1002/anie.201308584

Busker, S., Qian, W., Haraldsson, M., Espinosa, B., Johansson, L., Attarha, S., et al. (2020). Irreversible TrxR1 inhibitors block STAT3 activity and induce cancer cell death. *Sci. Adv.* 6 (12), eaax7945. doi:10.1126/sciadv.aax7945

Cheng, P., Liu, H., Li, Y., Pi, P., Jiang, Y., Zang, S., et al. (2020). Inhibition of thioredoxin reductase 1 correlates with platinum-based chemotherapeutic induced tissue injury. *Biochem. Pharmacol.* 175, 113873. doi:10.1016/j.bcp.2020.113873

Courtney, T., and Deiters, A. (2018). Recent advances in the optical control of protein function through genetic code expansion. *Curr. Opin. Chem. Biol.* 46, 99–107. doi:10.1016/j.cbpa.2018.07.011

Duan, D., Wang, Y., Pan, D., Jin, X., Yan, Y., Song, P., et al. (2022). Rheumatoid arthritis drug sinomenine induces apoptosis of cervical tumor cells by targeting thioredoxin reductase *in vitro* and *in vivo*. *Bioorg. Chem.* 122, 105711. doi:10.1016/j.bioorg.2022.105711

Dunkelmann, D. L., Willis, J. C. W., Beattie, A. T., and Chin, J. W. (2020). Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* 12 (6), 535–544. doi:10.1038/s41557-020-0472-x

Elsasser, S. J., Ernst, R. J., Walker, O. S., and Chin, J. W. (2016). Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nat. Methods* 13 (2), 158–164. doi:10.1038/nmeth.3701

Fan, Y., Makar, M., Wang, M. X., and Ai, H. W. (2017). Monitoring thioredoxin redox with a genetically encoded red fluorescent biosensor. *Nat. Chem. Biol.* 13 (9), 1045–1052. doi:10.1038/nchembio.2417

Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L. L., Pepinsky, B., et al. (1994). Tat-mediated delivery of heterologous proteins into cells. *Proc. Natl. Acad. Sci. U. S. A.* 91 (2), 664–668. doi:10.1073/pnas.91.2.664

Gauba, V., Grunewald, J., Gorney, V., Deaton, L. M., Kang, M., Bursulaya, B., et al. (2011). Loss of CD4 T-cell-dependent tolerance to proteins with modified amino acids. *Proc. Natl. Acad. Sci. U. S. A.* 108 (31), 12821–12826. doi:10.1073/pnas.1110042108

Green, M., and Loewenstein, P. M. (1988). Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. *Cell* 55 (6), 1179–1188. doi:10.1016/0092-8674(88)90262-0

Gromer, S., Arscott, L. D., Williams, C. H., Jr., Schirmer, R. H., and Becker, K. (1998). Human placenta thioredoxin reductase. Isolation of the selenoenzyme, steady state kinetics, and inhibition by therapeutic gold compounds. *J. Biol. Chem.* 273 (32), 20096–20101. doi:10.1074/jbc.273.32.20096

Han, K., Jeon, M. J., Kim, K. A., Park, J., and Choi, S. Y. (2000). Efficient intracellular delivery of GFP by homeodomains of Drosophila Fushi-tarazu and Engrailed proteins. *Mol. Cells* 10 (6), 728–732. doi:10.1007/s10059-000-0728-7

Haruna, K., Alkazemi, M. H., Liu, Y., Söll, D., and Englert, M. (2014). Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic Acids Res.* 42 (15), 9976–9983. doi:10.1093/nar/gku691

Hasan, A. A., Kalinina, E., Tatarskiy, V., and Shtil, A. (2022). The thioredoxin system of mammalian cells and its modulators. *Biomedicines* 10 (7), 1757. doi:10.3390/biomedicines10071757

Heinemann, I. U., Rovner, A. J., Aerni, H. R., Rogulina, S., Cheng, L., Olds, W., et al. (2012). Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS Lett.* 586 (20), 3716–3722. doi:10.1016/j.febslet.2012.08.031

Huang, Y., and Liu, T. (2018). Therapeutic applications of genetic code expansion. *Synth. Syst. Biotechnol.* 3 (3), 150–158. doi:10.1016/j.synbio.2018.09.003

Köhrer, C., Yoo, J. H., Bennett, M., Schaack, J., and RajBhandary, U. L. (2003). A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression. *Chem. Biol.* 10 (11), 1095–1102. doi:10.1016/j.chembiol.2003.10.013

Kumar, S., Bjornstedt, M., and Holmgren, A. (1992). Selenite is a substrate for calf thymus thioredoxin reductase and thioredoxin and elicits a large non-stoichiometric oxidation of NADPH in the presence of oxygen. *Eur. J. Biochem.* 207 (2), 435–439. doi:10.1111/j.1432-1033.1992.tb17068.x

Kurrikoff, K., Vunk, B., and Langel, U. (2021). Status update in the use of cell-penetrating peptides for the delivery of macromolecular therapeutics. *Expert Opin. Biol. Ther.* 21 (3), 361–370. doi:10.1080/14712598.2021.1823368

Lee, S., Kim, S. M., and Lee, R. T. (2013). Thioredoxin and thioredoxin target proteins: From molecular mechanisms to functional significance. *Antioxid. Redox Signal.* 18 (10), 1165–1207. doi:10.1089/ars.2011.4322

Lichtenstein, M., Zabit, S., Hauser, N., Farouz, S., Melloul, O., Hirbawi, J., et al. (2021). TAT for enzyme/protein delivery to restore or destroy cell activity in human diseases. *Life (Basel)* 11 (9), 924. doi:10.3390/life11090924

Liu, J., Zheng, F., Cheng, R., Li, S., Rozovsky, S., Wang, Q., et al. (2018). Site-specific incorporation of selenocysteine using an expanded genetic code and palladium-mediated chemical deprotection. *J. Am. Chem. Soc.* 140 (28), 8807–8816. doi:10.1021/jacs.8b04603

Lu, X., Zhou, Z., Zhu, L., Zhou, J., Huang, H., Zhang, C., et al. (2022). Construction and identification of a HEK293 cell line with stable TrxR1 overexpression. *Nan Fang. Yi Ke Da Xue Xue Bao* 42 (4), 554–560. doi:10.12122/j.issn.1673-4254.2022.04.11

Lundstrom-Ljung, J., Birnbach, U., Rupp, K., Soling, H. D., and Holmgren, A. (1995). Two resident ER-proteins, CaBP1 and CaBP2, with thioredoxin domains, are substrates for thioredoxin reductase: Comparison with protein disulfide isomerase. *FEBS Lett.* 357 (3), 305–308. doi:10.1016/0014-5793(94)01386-f

Ma, X., Karra, S., Guo, W., Lindner, D. J., Hu, J., Angell, J. E., et al. (2001). Regulation of interferon and retinoic acid-induced cell death activation through thioredoxin reductase. *J. Biol. Chem.* 276 (27), 24843–24854. doi:10.1074/jbc.M100380200

May, J. M., Qu, Z. C., and Nelson, D. J. (2007). Uptake and reduction of alpha-lipoic acid by human erythrocytes. *Clin. Biochem.* 40 (15), 1135–1142. doi:10.1016/j.clinbiochem.2007.06.009

Mihaila, T. S., Bate, C., Ostersehlt, L. M., Pape, J. K., Keller-Findeisen, J., Sahl, S. J., et al. (2022). Enhanced incorporation of subnanometer tags into cellular proteins for fluorescence nanoscopy via optimized genetic code expansion. *Proc. Natl. Acad. Sci. U. S. A.* 119 (29), e2201861119. doi:10.1073/pnas.2201861119

Mustacich, D., and Powis, G. (2000). Thioredoxin reductase. *Biochem. J.* 346 (1), 1–8. doi:10.1042/bj3460001

Nagahara, H., Vocero-Akbani, A. M., Snyder, E. L., Ho, A., Latham, D. G., Lissy, N. A., et al. (1998). Transduction of full-length TAT fusion proteins into mammalian cells: TAT-p27Kip1 induces cell migration. *Nat. Med.* 4 (12), 1449–1452. doi:10.1038/4042

Ngo, J. T., and Tirrell, D. A. (2011). Noncanonical amino acids in the interrogation of cellular protein synthesis. *Acc. Chem. Res.* 44 (9), 677–685. doi:10.1021/ar200144y

O'Donoghue, P., Heinemann, I. U., and Fan, C. (2021). Editorial: Synthetic nucleic acids for expanding genetic codes and probing living cells. *Front. Bioeng. Biotechnol.* 9, 720534. doi:10.3389/fbioe.2021.720534

O'Donoghue, P., Ling, J., Wang, Y. S., and Söll, D. (2013). Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol.* 9 (10), 594–598. doi:10.1038/nchembio.1339

Ranji Charna, A., Des Soye, B. J., Ntai, I., Kelleher, N. L., and Jewett, M. C. (2022). An efficient cell-free protein synthesis platform for producing proteins with pyrrolysine-based noncanonical amino acids. *Biotechnol. J.* 17, e2200096. doi:10.1002/biot.202200096

Ren, X., Zou, L., Zhang, X., Branco, V., Wang, J., Carvalho, C., et al. (2017). Redox signaling mediated by thioredoxin and glutathione systems in the central nervous system. *Antioxid. Redox Signal.* 27 (13), 989–1010. doi:10.1089/ars.2016.6925

Sachweh, M. C., Stafford, W. C., Drummond, C. J., McCarthy, A. R., Higgins, M., Campbell, J., et al. (2015). Redox effects and cytotoxic profiles of MJ25 and auranofin towards malignant melanoma cells. *Oncotarget* 6 (18), 16488–16506. doi:10.18632/oncotarget.4108

Shoari, A., Tooyserkani, R., Tahmasebi, M., and Lowik, D. (2021). Delivery of various cargos into cancer cells and tissues via cell-penetrating peptides: A review of the last decade. *Pharmaceutics* 13 (9), 1391. doi:10.3390/pharmaceutics13091391

Si, L., Xu, H., Zhou, X., Zhang, Z., Tian, Z., Wang, Y., et al. (2016). Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science* 354 (6316), 1170–1173. doi:10.1126/science.aah5869

Suhorutsenko, J., Oskolkov, N., Arukuusk, P., Kurrikoff, K., Eriste, E., Copolovici, D. M., et al. (2011). Cell-penetrating peptides, PepFects, show no evidence of toxicity and immunogenicity *in vitro* and *in vivo*. *Bioconjug. Chem.* 22 (11), 2255–2262. doi:10.1021/bc200293d

Tamura, T., and Stadtman, T. C. (1996). A new selenoprotein from human lung adenocarcinoma cells: Purification, properties, and thioredoxin reductase activity. *Proc. Natl. Acad. Sci. U. S. A.* 93 (3), 1006–1011. doi:10.1073/pnas.93.3.1006

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

Thomas, P., and Smart, T. G. (2005). HEK293 cell line: A vehicle for the expression of recombinant proteins. *J. Pharmacol. Toxicol. Methods* 51 (3), 187–200. doi:10.1016/j.vascn.2004.08.014

Thyer, R., Shroff, R., Klein, D. R., d'Oelsnitz, S., Cotham, V. C., Byrom, M., et al. (2018). Custom selenoprotein production enabled by laboratory evolution of recoded bacterial strains. *Nat. Biotechnol.* 36 (7), 624–631. doi:10.1038/nbt.4154

Tian, Y., and Zhou, S. (2021). Advances in cell penetrating peptides and their functionalization of polymeric nanoplatforms for drug delivery. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* 13 (2), e1668. doi:10.1002/wnan.1668

Venkat, S., Sturges, J., Stahman, A., Gregory, C., Gan, Q., and Fan, C. (2018). Genetically incorporating two distinct post-translational modifications into one protein simultaneously. *ACS Synth. Biol.* 7 (2), 689–695. doi:10.1021/acssynbio.7b00408

Vinogradov, A. A., and Suga, H. (2020). Introduction to thiopeptides: Biological activity, biosynthesis, and strategies for functional reprogramming. *Cell Chem. Biol.* 27 (8), 1032–1051. doi:10.1016/j.chembiol.2020.07.003

Vives, E., Richard, J. P., Rispal, C., and Lebleu, B. (2003). TAT peptide internalization: Seeking the mechanism of entry. *Curr. Protein Pept. Sci.* 4 (2), 125–132. doi:10.2174/1389203033487306

Wang, H., Sun, S., Ren, Y., Yang, R., Guo, J., Zong, Y., et al. (2022). Selenite ameliorates cadmium-induced cytotoxicity through downregulation of ROS levels and upregulation of selenoprotein thioredoxin reductase 1 in SH-SY5Y cells. *Biol. Trace Elem. Res.*, 1–10. doi:10.1007/s12011-022-03117-6

Wang, Y., Liu, P., Chang, J., Xu, Y., and Wang, J. (2021). Site-specific selenocysteine incorporation into proteins by genetic engineering. *Chembiochem.* 22 (20), 2918–2924. doi:10.1002/cbic.202100124

Welegedara, A. P., Adams, L. A., Huber, T., Graham, B., and Otting, G. (2018). Site-specific incorporation of selenocysteine by genetic encoding as a photocaged unnatural amino acid. *Bioconjug. Chem.* 29 (7), 2257–2264. doi:10.1021/acs.bioconjchem.8b00254

Wright, D. E., Altaany, Z., Bi, Y., Alperstein, Z., and O'Donoghue, P. (2018). Acetylation regulates thioredoxin reductase oligomerization and activity. *Antioxid. Redox Signal.* 29 (4), 377–388. doi:10.1089/ars.2017.7082

Wright, D. E., Panaseiko, N., and O'Donoghue, P. (2021). Acetylated thioredoxin reductase 1 resists oxidative inactivation. *Front. Chem.* 9, 747236. doi:10.3389/fchem.2021.747236

Xia, L., Nordman, T., Olsson, J. M., Damdimopoulos, A., Bjorkhem-Bergman, L., Nalvarte, I., et al. (2003). The mammalian cytosolic selenoenzyme thioredoxin reductase reduces ubiquinone. A novel mechanism for defense against oxidative stress. *J. Biol. Chem.* 278 (4), 2141–2146. doi:10.1074/jbc.M210456200

Yokoo, H., Oba, M., and Uchida, S. (2021). Cell-penetrating peptides: Emerging tools for mRNA delivery. *Pharmaceutics* 14 (1), 78. doi:10.3390/pharmaceutics14010078

Zhang, H., Wu, J., Yuan, J., Li, H., Zhang, Y., Wu, W., et al. (2021). Ethaselen synergizes with oxaliplatin in tumor growth inhibition by inducing ROS production and inhibiting TrxR1 activity in gastric cancer. *J. Exp. Clin. Cancer Res.* 40 (1), 260. doi:10.1186/s13046-021-02052-z

Zhang, Y., Sun, S., Xu, W., Yang, R., Yang, Y., Guo, J., et al. (2022). Thioredoxin reductase 1 inhibitor shikonin promotes cell necroptosis via SecTRAPs generation and oxygen-coupled redox cycling. *Free Radic. Biol. Med.* 180, 52–62. doi:10.1016/j.freeradbiomed.2021.12.314

Zhong, L., and Holmgren, A. (2000). Essential role of selenium in the catalytic activities of mammalian thioredoxin reductase revealed by characterization of recombinant enzymes with selenocysteine mutations. *J. Biol. Chem.* 275 (24), 18121–18128. doi:10.1074/jbc.M000690200

Zhu, J., Fu, M., Gao, J., Dai, G., Guan, Q., and Du, C. (2022). Upregulation of thioredoxin reductase 1 expression by flavan-3-ols protects human kidney proximal tubular cells from hypoxia-induced cell death. *Antioxidants (Basel)* 11 (7), 1399. doi:10.3390/antiox11071399

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fmolb.2022.992748

OPEN ACCESS

EDITED BY
Kerstin G Blank,
Johannes Kepler University of Linz, Austria

REVIEWED BY
Miriam Amiram,
Ben-Gurion University of the Negev, Israel
Ryan Mehl,
Oregon State University, United States

*CORRESPONDENCE
Nediljko Budisa,
nediljko.budisa@tu-berlin.de

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

RECEIVED 13 July 2022
ACCEPTED 16 September 2022
PUBLISHED 24 October 2022

CITATION
Koch NG, Baumann T, Nickling JH, Dziegielewski A and Budisa N (2022), Engineered bacterial host for genetic encoding of physiologically stable protein nitration.
Front. Mol. Biosci. 9:992748.
doi: 10.3389/fmolb.2022.992748

Engineered bacterial host for genetic encoding of physiologically stable protein nitration

Nikolaj G. Koch[1,2], Tobias Baumann[2], Jessica H. Nickling[2], Anna Dziegielewski[2] and Nediljko Budisa[2,3]*

[1]Bioanalytics Group, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany, [2]Biocatalysis Group, Institute of Chemistry, Technische Universität Berlin, Berlin, Germany, [3]Chemical Synthetic Biology Group, Department of Chemistry, University of Manitoba, Winnipeg, MB, Canada

Across scales, many biological phenomena, such as protein folding or bioadhesion and cohesion, rely on synergistic effects of different amino acid side chains at multiple positions in the protein sequence. These are often fine-tuned by post-translational modifications that introduce additional chemical properties. Several PTMs can now be genetically encoded and precisely installed at single and multiple sites by genetic code expansion. Protein nitration is a PTM of particular interest because it has been associated with several diseases. However, even when these nitro groups are directly incorporated into proteins, they are often physiologically reduced during or shortly after protein production. We have solved this problem by using an engineered *Escherichia coli* host strain. Six genes that are associated with nitroreductase activity were removed from the genome in a simple and robust manner. The result is a bacterial expression host that can stably produce proteins and peptides containing nitro groups, especially when these are amenable to modification. To demonstrate the applicability of this strain, we used this host for several applications. One of these was the multisite incorporation of a photocaged 3,4-dihydroxyphenylalanine derivative into Elastin-Like Polypeptides. For this non-canonical amino acid and several other photocaged ncAAs, the nitro group is critical for photocleavability. Accordingly, our approach also enhances the production of biomolecules containing photocaged tyrosine in the form of ortho-nitrobenzyl-tyrosine. We envision our engineered host as an efficient tool for the production of custom designed proteins, peptides or biomaterials for various applications ranging from research in cell biology to large-scale production in biotechnology.

KEYWORDS

elastin-like polypeptides, MjTyrRS, ONBY, lambda red recombineering, Keio collection, nitroreduction, mussel-inspired adhesives, genetic code expansion

1 Introduction

Over the past two decades, genetic code expansion has evolved into a powerful technique to genetically introduce non-canonical amino acids (ncAAs) into proteins and peptides (Liu and Schultz, 2010; Mukai et al., 2017; de la Torre and Chin, 2021). More than 200 ncAAs with a broad repertoire of chemistries and new-to-nature functionalities have been incorporated by various routes (Wan et al., 2014; Dumas et al., 2015; Koch et al., 2021; Pagar et al., 2021). The most commonly used method is site-specific stop codon suppression (SCS) *in vivo*, using bacterial hosts. A key application is to genetically encode post-translational modifications (PTMs) to elucidate their effects on protein activity (Hoppmann et al., 2017; Luo et al., 2017; Young and Schultz, 2018; Zhu et al., 2019; Italia et al., 2020). In particular, mimicking protein nitration is of great interest since this PTM has been linked to several diseases (Neumann et al., 2008; Porter et al., 2019; Jang et al., 2020). However, a major challenge in obtaining these recombinant proteins remains: Especially in bacterial hosts, molecules containing nitro groups are often enzymatically reduced to hydroxylamino and/or amino derivatives (Copp et al., 2017). This side chain modification also occurs in a variety of ncAAs and was also observed in our previous study in which a photo-caged derivative of 3,4-dihydroxyphenylalanine (Dopa) was genetically encoded (Nguyen et al., 2014; Ren et al., 2015; Hauf et al., 2017; Baumann et al., 2019; Böcker et al., 2019). Dopa moieties play a crucial role in wet surface adhesion of marine mussels mediated by foot proteins (Ahn et al., 2015; Maier et al., 2015; Seo et al., 2015; Waite, 2017; Budisa and Schneider, 2019).

A long-standing goal of our laboratory is to harness these properties and to engineer biomaterial-based wet-adhesion agents that can be used, for example, for *in vivo* tissue or wound healing and bone regeneration (Haller et al., 2012; Kaushik et al., 2015; Kim et al., 2021). Unfortunately, the tendency of Dopa to oxidize to its quinone moiety under normal (oxidative) protein production and purification conditions makes handling Dopa containing proteins challenging. One solution was to genetically encode a Dopa analog, *meta*-(*ortho*-(2-nitrobenzyl))-3,4-dihydroxyphenylalanine (*m*-oNB-Dopa (**1**), Figure 1) which is resistant to oxidization. Moreover, the protecting group is photocleavable, conferring spatiotemporal control over the adhesion mechanism (Hauf et al., 2017). Unfortunately, as mentioned above, the nitro group is reduced in bacterial production hosts, which drastically reduces photocleavability.

Despite the fascinating properties of mussel adhesion proteins, their recombinant production in high yields is a general obstacle. The main reasons for this are probably the low protein solubility and the low in-frame SCS efficiency during ribosomal synthesis (Hauf et al., 2017). Considering their natural purpose and function, it makes sense that mussels produce adhesion proteins that are insoluble in water-based solutions. Moreover, the incorporation efficiency of ncAAs depends on the protein scaffold and sequence context (Pott et al., 2014; Chemla et al., 2018; Schwark et al., 2018). We thus concluded that these problems could be circumvented in one step by switching the target protein scaffold from mussel proteins to elastin-like polypeptides (ELPs). ELPs are biopolymers composed of the repeating pentapeptide sequence $(VPGXG)_n$ (see Supplementary Material for details). Hallmark of the ELP structure are tandem repeats with repeating proline (conferring stiffness) and glycine (conferring flexibility), while the "host" position X tolerates variations in hydrophobicity and charge well (Roberts et al., 2015). Interestingly, the ELP structure also provides a sequence context for high readthrough in orthogonal translation, allowing suppression of up to 30 in-frame stop codons in a single polypeptide chain (Amiram et al., 2015). Thus, when position X is chosen as the site of ncAA installation *via* genetic code expansion, ELPs appear to provide an excellent context compared to other protein scaffolds. This is consistent with empirical observations suggesting the existence of so-called mRNA "context effects" in SCS. The insertion of a stop codon into a given genetic sequence can lead to large deviations from natural mRNA folding and subsequently affect the strength of binding site interactions within the ribosome (Gorochowski et al., 2015).

ELPs are a class of self-assembling peptides derived from the water-soluble portion of the tropoelastin protein (Urry, 1997; Vrhovski and Weiss, 1998; Bellingham et al., 2003). Elastin is an

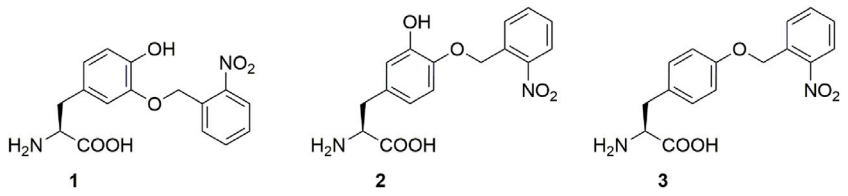


FIGURE 1
Photocaged non-canonical amino acids used in this study. *meta*-(*ortho*-(2-nitrobenzyl))-3,4-dihydroxyphenylalanine (*m*-oNB-Dopa, **1**), *para*-(*ortho*-(2-nitrobenzyl))-3,4-dihydroxyphenylalanine (*p*-oNB-Dopa, **2**) and *ortho*-nitro-benzyl-tyrosine (ONBY, **3**).

extracellular matrix protein found in multicellular organisms as part of connective, skin, lung, and tendon tissues. Due to its high biological similarity to natural tissues, elastin and ELPs are well tolerated and result in a very mild immune response when used for human applications (Hale et al., 2005; Vanderschuren et al., 2022). In addition, ELPs possess a lower critical solution temperature (LCST), which is useful for the development of smart adhesives. The LCST depends on the sequence length of ELPs and the ionic strength of the solution they are located in. Since the length of the ELPs can be precisely genetically encoded, the LCST can be easily adjusted to the desired temperature. A desirable project end goal is to create temperature-controlled ELP-based wet adhesives. By obtaining soluble protein in a biotechnological production setting, purification and application would be facilitated, but upon application to a desired body part (where the temperature is above LCST), the smart adhesive reaches the coacervation point to induce robust adhesion. The combined properties of adhesion (induced by *m*-oNB-Dopa (**1**)) and cohesion (induced by ELP above LCST) are essential for this endeavor.

In this context, we have set the goal to further push the limits of orthogonal translation by rational engineering of the complex system to enable genetic encoding of *m*-oNB-Dopa (**1**) at the scale of up to 60 in-frame stop codons in an ELP gene. To address nitro group reduction *in vivo*, we modified the *Escherichia coli* host by deleting six non-essential nitroreductase genes from the genome. The resulting bacterial strain has drastically reduced nitroreductase activity which enables the intracellular and chemically stable introduction of *m*-oNB-Dopa (**1**) and other nitro group containing ncAAs into peptides and proteins. The overall performance of our combined approach is demonstrated by the production of additional recombinant biomolecules such as the antimicrobial peptide nisin equipped with side chains bearing intact aromatic nitro groups. In this way, we succeeded in establishing a robust, cost-effective and rapid method for the production of nitrated polypeptides through expansion of the genetic code.

2 Materials and methods

2.1 Plasmid vectors

2.1.1 Orthogonal pairs

All plasmids containing the orthogonal translation system (OTS) were pUltra backbones (Chatterjee et al., 2013). *Mj*ONB-DOPARS and *Mj*ONBYRS were constructed previously (Hauf et al., 2017; Baumann et al., 2019). *Mj*PCNFRS was a gift from Abhishek Chatterjee (Boston College, MA, United States).

2.1.2 Recursive Directional Ligation

Monomer gene fragments encoding the different ELP scaffolds were synthesized by GeneArt (ThermoFisher Scientific, Waltham, MA, United States). The overall cloning strategy is shown in Supplementary Figure S1. To generate the ELP constructs, the monomers were transferred into the high copy pSB1C3 vector (Registry of Standard Biological Parts, 2022). Monomer elongation was performed by digesting the desired plasmid with restriction enzyme BglI followed by dephosphorylation with alkaline phosphatase (FastAP, ThermoFisher Scientific, Waltham, MA, United States). Subsequently, the desired insert fragment was obtained by double digestion with BglI and PflMI and ligation, yielding an elongated ELP scaffold. To verify that the correct insert was formed, an analytical digest of isolated plasmids with XbaI and PstI was performed (*cf.* Supplementary Figure S9). This was necessary because colony PCR with ELP constructs bigger than the ELP(10x amber) resulted in unsuccessful amplifications, likely due to the high GC content and the repetitive nature of the ELP gene. ELP scaffolds >2000 bp (up to the ELP(40x amber) construct) were also analyzed by Sanger sequencing from both directions. When the ELP scaffold reached the desired repeat length, the construct was double digested (with BglI and PflMI) and ligated to a suitable fragment of pET-28a expression vector. Successful DNA assembly was verified by double digestion with BglI and XbaI (Supplementary Figure S9). ELP scaffolds >1,000 bp (up to the ELP(20x amber) construct) were also verified by Sanger sequencing (forward direction).

2.2 Analysis of sfGFP expression by intact cell fluorescence

Electrocompetent *E. coli* cells were transformed with the orthogonal translation system and reporter plasmids. LB agar plates contained 1% glucose and corresponding antibiotics. Single colonies were used for inoculation of 2 ml LB (in 14 ml tubes) containing 1% glucose and appropriate antibiotics and grown to saturation overnight. Assays were performed in 96-well plate format. Cultures were added to each well at a dilution of 1:100 in ZYP-5052 auto-induction medium to a final volume of 100 µL supplemented with antibiotics and ncAAs (various concentrations as indicated). Cells were grown in black 96-well flat bottom plates (Greiner Bio-One, Kremsmünster, Austria) covered with a gas permeable foil (Breathe-Easy®, Diversified Biotech, Doylestown, PA, United States) and shaken at 37°C for 18 h. For endpoint measurements, the plate foil was removed and fluorescence was measured on a M200 plate reader (Tecan, Männedorf, Switzerland). For OD_{600} measurements, 50 µL of ZYP-5052 medium was added to clear 96-well µ-plates and 50 µL of culture was added. The excitation and emission wavelengths for fluorescence measurements were set to 481 and 511 nm, respectively. Fluorescence values were normalized to the corresponding OD_{600}. Biological triplicates (three

independent replicates) were used for the measurements of each aaRS construct. Relative fluorescence was normalized to the highest value. The data (incl. Standard deviation) represent the mean of three biological replicates. The experiment which led to Figure 3 was repeated once, with similar results (data not shown).

2.3 Recombinant protein expression

For expression of target protein variants, *E. coli* strains were used in 10 ml ZYP-5052 medium supplemented with 1 mM ncAA and appropriate antibiotics. The expression medium was inoculated with a starter culture (1:100). Shake flasks were incubated at 37°C for 20 h while shaking at 200 rpm. Cells were harvested by centrifugation and stored at −80°C or used directly for protein purification.

2.4 Recombinant protein purification

Harvested cell pellets were resuspended (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0) and lysed with B-PER Bacterial Protein Extraction Reagent (ThermoFisher Scientific, Waltham, MA, United States) according to protocol, with the addition of phenylmethanesulfonyl fluoride (PMSF, 1 mM final concentration), DNAse and RNAse. The purified lysates were loaded onto an equilibrated Protino Ni-NTA column (Macherey-Nagel, Düren, Germany) and purified using a peristaltic pump (Pharmacia Biotech, now: Cytiva, Marlborough, MA, United States). After washing with 10 column volumes of resuspension buffer, elution buffer (50 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0) was used to elute the his-tagged target proteins. The first 2 ml (covering the dead volume) were discarded. Subsequently, the eluate (1.5 ml) was collected and dialyzed in cellulose film tubes against 5 L buffer (50 mM sodium phosphate, 300 mM NaCl, pH 8.0) for at least 2 h with three changes into fresh buffer each. The concentrations of purified proteins were determined by measuring the absorbance of the sfGFP chromophore at 488 nm or for amilCP by measuring the absorption at 280 nm.

2.5 Decaging of ortho-nitrobenzyl-tyrosine

For decaging of the photocleavable ONBY (**3**) analog in solutions of purified protein or whole cell lysates, samples were placed in a HPLC glass vial. A homemade 365 nm LED lamp (radiant flux ~720 mW) was used at a distance of ~3 mm for the indicated time periods.

2.6 Time-resolved chromophore maturation of amilCP(Y63ONBY)

Proteins were expressed and purified according to the protocol described above. After purification, 100 µL protein solution at a concentration of 1 mg/ml was transferred to HPLC glass vials (due to strong light absorption, plastic vials were avoided). Irradiation with UV light was performed according to the protocol described above. After irradiation, the samples were transferred to a 96-well clear bottom plate (Greiner Bio-One, Kremsmünster, Austria), and monitored for 24 h. Absorption was measured static at 589 nm using a M200 plate reader (Tecan, Männedorf, Switzerland). This experiment was performed once.

2.7 Antimicrobial activity assay

To determine antimicrobial activity of recombinant peptides, an overnight culture of the nisin-sensitive indicator strain *L. lactis* NZ9000 pNZ_nisPT was incubated in M17 medium containing 1% (w/v) glucose (GM17) and 5 µg/ml chloramphenicol at 30°C without agitation (Khusainov et al., 2011; Khusainov and Kuipers, 2013). Fresh GM17 medium was inoculated and cells were grown at 30°C until OD_{600} of 0.4–0.6 was reached. 1 ml of the culture was added to 50 ml of molten GM17-agar supplemented with chloramphenicol and poured into a large Petri dish. Holes were poked into the solidified agar using the wide end of a glass Pasteur pipette (Nickling et al., 2018). As source of the antimicrobial peptides, 1 ml of the *E. coli* expression cultures was centrifuged (3 min, 7,000 *g*) and the cell pellet resuspended in 500 µL buffer (50 mM sodium phosphate, pH 7.4). Cells were lysed by sonification (Sonoplus HD3200, MS72 microtip, Bandelin, Berlin, Germany) at 30% amplitude with pulse cycles of 1 s on and 5 s off for 3 min. Cell debris was removed by centrifugation (4°C, 10 min, 13,000 g). The obtained supernatants were diluted and normalized to 1 ml OD_{600} = 0.6 or 1.0 relative to the initially harvested cell density. 40 µL of each normalized sample was added into the holes of the indicator plate and incubated overnight at 30°C. Corresponding *E. coli* lysates with recombinant nisin containing non-decaged ONBY served as a negative control. Chloramphenicol at a concentration of 400 µg/ml served as a positive control compound. This experiment was performed once.

2.8 Intact protein ESI-MS

An Agilent 6530 Q-TOF LC/MS system was used. Samples were infused at a flow rate of 0.3–0.5 ml min^{-1} in a gradient from 5% acetonitrile 0.1% formic acid in water to 80% acetonitrile 0.1% formic acid in water over a Discovery Bio Wide Pore C5 column, 2.1 × 100, 3 micron (Supelco analytical, Sigma-Aldrich, St. Lois, United States) for 20 min. Deconvolution of spectra was performed with Agilent MassHunter Qualitative

Analysis software version B.06.00 Bioconfirm Intact Mass Module employing the Maximum Entropy Deconvolution Algorithm. Raw data were plotted using Origin.

2.9 Chromosomal gene deletion by homologous recombination

Chromosomal gene deletions were performed according to a modified method of Datsenko and Wanner (Jensen et al., 2015). *Via* PCR, the kanamycin resistance cassette was amplified with the flippase recognition sites (FRT) from genomic DNA of *E. coli* clones obtained from the Keio Single Gene Knockout Collection (Baba et al., 2006). These knockout cassettes can be conveniently PCR-amplified by using either bacterial colonies or bacterial cultures as source of the genomic DNA. Primers were designed to amplify the cassette and, in addition, up to 200 bp of homologous regions up- and downstream of the target genes. Taq polymerase was used since amplification using high-fidelity DNA polymerases (Q5, Phusion) failed. PCR products were purified by gel extraction. Between 100 and 250 ng DNA was used to transform B-95.ΔA cells carrying the pSIJ8 plasmid (Jensen et al., 2015; Mukai et al., 2015). Expression of the pSIJ8-encoded λ-Red recombination system was induced *via* 0.2% arabinose (w/v) at an OD_{600} of 0.3 for 30–45 min during preparation of these electrocompetent cells. After transformation and 2 h recovery (30°C, 220 rpm), cells were plated on LB-Kan agar plates and incubated overnight at 30°C. Homologous recombination of the FRT-*kan*-FRT cassette and the *E. coli* genome was verified by colony PCR. Antibiotic markers were removed from the genome by growing the desired strain at 30°C in LB, inducing the FLP gene with 50 mM L-rhamnose (final concentration) at an OD_{600} between 0.1 and 0.4 for 4–6 h and plating the strain afterwards on LB agar containing 100 μg/mL Ampicillin. Antibiotic resistance cassette removal was again verified by PCR. Then, either the next round of gene deletion was performed as described above, or at the final stage the strain was cured of the pSIJ8 plasmid. The final clone was incubated at 42 °C overnight to cure the strain of the temperature-sensitive pSIJ8 plasmid and verified by PCR. All culturing steps were made in a 14 mL culture tube and 2 mL LB with apropriate antibiotics and additives described above. Loss of the plasmid was confirmed by PCR and by growing corresponding clones on LB agar supplied with and without ampicillin.

3 Results and discussion

3.1 Optimizing *Mj*ONB-DopaRS activity and fidelity

Tyrosyl-tRNA synthetase from *Methanocaldococcus janaschii* (*Mj*TyrRS) forms the basis for the ribosomal incorporation of all ncAAs used in this study. Previous studies hypothesized that the native enzyme recognizes the *para*-OH group of its natural substrate tyrosine *via* two critical residues, Y32 and D158, in the amino acid binding pocket analogous to *Bacillus stearothermophilus* TyrRS (*Bst*TyrRS). To shift the amino acid substrate recognition from tyrosine to novel ncAA substrates, it was assumed that this natural mode of substrate recognition must be abolished and that D158 is functionally more important than Y32 (Antonczak et al., 2011). In engineered *Mj*TyrRS enzymes, these two residues are usually mutated to smaller residues, as it is the case in the first aaRS reported to activate ONBY (**3**) (Deiters et al., 2006). The same strategy was followed in the design of the *o*NB-Dopa aaRS gene library in our group, with the original goal to genetically encode *p*-*o*NB-Dopa (**2**) (Hauf et al., 2017). Interestingly, the obtained *Mj*TyrRS enzyme recognizes *m*-*o*NB-Dopa (**1**) and not *p*-*o*NB-Dopa (**2**). Therefore, the question arose whether the intentionally fixed mutations Y32A and D158A are indeed required for genetic encoding of *m*-*o*NB-Dopa (**1**) and how they affect the orthogonality and activity of the aaRS. Review of the structural data suggested that there may be an interaction of the Y32 OH group with the *para* OH group of *m*-*o*NB-Dopa (**1**) that could enhance ncAA recognition. Because this had not been explored in the previous work, we reversed these mutations.

This confirmed the hypothesis that residue D158 is more important for Tyr recognition, because the D158A mutation is necessary individually and in combination to maintain aaRS orthogonality, whereas Y32A is not (Figure 2A). Although the A32Y reversion did not improve *m*-*o*NB-Dopa (**1**) incorporation, our results suggest an increased aaRS fidelity, and thus was used throughout this study. The level of background suppression (stop codon readthrough with canonical amino acid incorporation) decreased by around 30% (Figure 2B). This behavior highlights that this amino acid substrate recognition is quite different compared to other *Mj*TyrRS mutants (Dumas et al., 2015; Gueta et al., 2022). Reversion of Y32A generally resulted in complete inactivation of ncAA recognition or a drastic reduction in catalytic activity and fidelity for the corresponding ncAAs (Antonczak et al., 2011). Certain, albeit low, levels of background suppression are often observed for OTSs due to either near cognate suppression or noncomplete orthogonality of aaRS in the absence of the ncAA, which then lead to the incorporation of canonical amino acids (AA). In engineered *Mj*TyrRS variants, tRNA mischarging most likely occurs with Tyr (O'Donoghue et al., 2012; Aerni et al., 2015).

3.2 Multisite incorporation of *m*-*o*NB-Dopa into elastin-like-polypeptides at up to 60 positions

Approaching more and more ncAA incorporation sites within the repeat-based ELP scaffold required a reliable modular assembly method for the genetic constructs. We decided to prepare them from two monomeric initial sequences using Recursive Directional Ligation (RDL, see

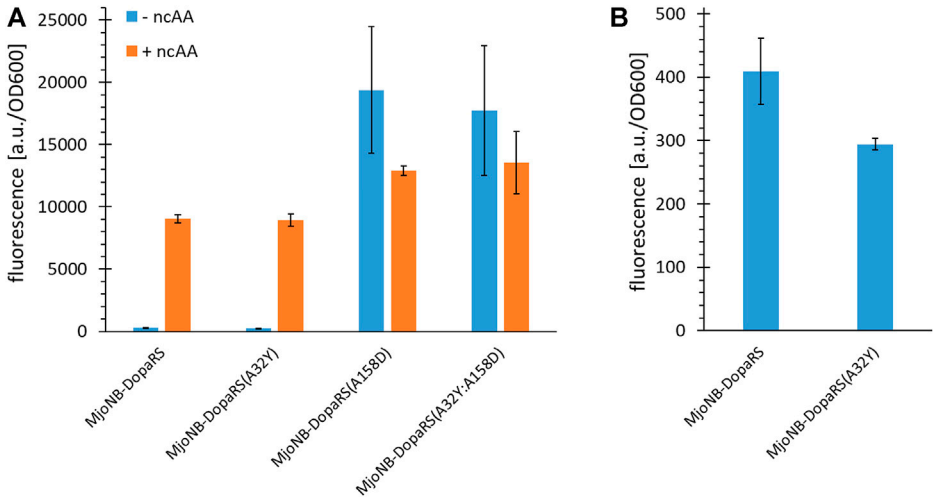


FIGURE 2
Ribosomal incorporation efficiency of *m*-oNB-Dopa upon mutation of the ncAA binding pocket of *Mj*TyrRS. **(A)** OTS efficiency, see main text for a detailed description of the reporter constructs. Ribosomal incorporation with ncAA (+ ncAA = 1 mM *m*-oNB-Dopa) and corresponding controls without ncAA supplementation. **(B)** Zoomed in part of (**A**) to visualize differences in background stop codon suppression of the first two setups. Intact cell fluorescence of *E. coli* strain BL21(DE3) with SUMO-sfGFP(R2amber) as reporter protein, endpoint measurements after 18 h of incubation. Fluorescence values were normalized with corresponding OD_{600} values to correct for differences in cell density. Data (incl. Standard deviation) represent the mean of three biological replicates.

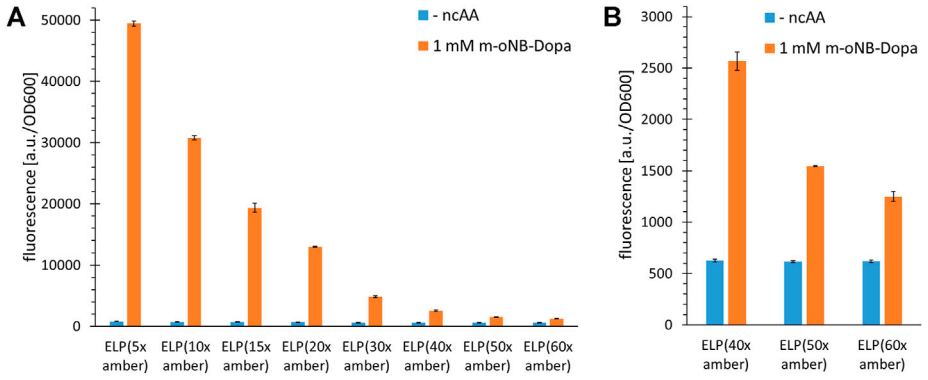


FIGURE 3
Incorporation efficiency of *m*-oNB-Dopa for increasing numbers of installation sites in ELP constructs. Stop codon suppression and OTS constructs as indicated in the main text, see Supplementary Material for sequence details. **(A)** Ribosomal incorporation of *m*-oNB-Dopa (1 mM) with ELP constructs containing a varying number (5–60) of in-frame amber stop codons. Controls lack ncAA supplementation (- ncAA). **(B)** Data extract with focus on the suppression of 40, 50 and 60 in-frame stop codons for ncAA incorporation. Intact cell fluorescence of *E. coli* B-95.ΔA (RF1 deficient strain), endpoint measurements after 18 h of incubation. Fluorescence values were normalized with corresponding OD_{600} values to correct for differences in cell density. Data (incl. Standard deviation) represent the mean of three biological replicates.

Supplementary Figure S1), which is well suited to construct the genes of block/repeat based polymer sequences (Meyer and Chilkoti, 2002). After determining the most efficient *Mj*TyrRS enzyme variant for ribosomal incorporation of *m*-oNB-Dopa (**1**), the production of ELP constructs with up to 60 in-frame amber stop codons was attempted (Figure 3). To facilitate readout, ELP variants were fused to the N-terminus of superfolder Green Fluorescent Protein (sfGFP) (see Supplementary Material for sequence information). In this way, translation of the full length ELP gene also leads to translation of sfGFP. As previously noted, translation efficiency can be directly observed with fluorescence measurements (Amiram et al., 2015).

For the efficient incorporation of ncAAs at multiple sites, protein production was performed in a recoded and release factor

one (RF1) deficient *E. coli* host derived from BL21(DE3), as in our original study (Mukai et al., 2015; Hauf et al., 2017). In this host (strain B-95.ΔA), the competition between the desired in-frame amber SCS and the endogenous translation termination is abolished. To be consistent with our previous studies and because we did not intend to use the strain in minimal media, we used B-95.ΔA rather than its derivative B-95.ΔAΔfabR. Nevertheless, we would like to note that the latter could improve production in minimal media and at low temperatures (*cf.* (Mukai et al., 2015)). Figure 3 shows that high full length production yields can be achieved for ELP(5x ONB-Dopa). This is consistent with previous reports and also with quantification of small scale protein production yields (for the latter, see paragraph 3.5) (Amiram et al., 2015; Gueta et al., 2022). Reporter fluorescence signals show that constructs with up to 60 stop codons can produce full length protein (Figure 3B). To date, at most 30 stop codons have been suppressed *in vivo* (Amiram et al., 2015; Gueta et al., 2022) and 40 *in vitro*, respectively (Martin et al., 2018). This is intriguing since it is well known that increasing the number of in-frame stop codons normally dramatically decreases the efficiency of translation. We would like to emphasize that the fluorescence signals are an indirect estimate of protein yield. The experimentally determined sfGFP signal depends on a variety of factors including folding efficiency and solubility of the fusion protein. For this screening of increasing incorporation sites, we chose not to include indirect control constructs without stop codons because the physico-chemical properties of natural amino acids (e.g., Tyr, which can be ribosomally incorporated *via* wild-type *Mj*TyrRS) differ significantly from those of ncAAs. While fluorescence-based protein yield estimates for sfGFP constructs with a single site of ncAA incorporation are commonly accurate, fusion protein ELP-sfGFP constructs may behave differently (Martin et al., 2018). It has also been observed that ELP repeat length can correlate with decreasing fluorescence signals, independent of ncAA incorporation efficiency (Amiram et al., 2015; Martin et al., 2018; Hadar et al., 2021). Depending on ELP and ncAA properties, cumulative effects could arise from differences in the strength of gene expression or from perturbations that ELP parts of increasing length exert on sfGFP folding and fusion protein solubility.

Based on our experience with various reporters and MS-based detection of ncAA incorporation into proteins and peptides, we expect a fluorescence signal that is twice as high as the background suppression level to indicate a robust incorporation system. It is worth noting that it remains unknown why such a large number of in-frame stop codons can be suppressed in the ELP scaffold. For other scaffolds such as sfGFP, protein yields decrease dramatically even with only five in-frame stop codons in the gene sequence. To start off with a highly efficient ncAA incorporation system, we initially conducted a prescreening with *Mj*ONBYRS and a second high performing aaRS enzyme with a panel of ncAA substrates (Supplementary Figure S2). This confirmed that the orthogonal translation system (OTS) with *Mj*ONBYRS is highly active and allows the ribosomal incorporation of several ncAAs. Also for O-propargyl- and O-allyl-tyrosine supplied at low ncAA concentrations, it became our preferred aaRS enzyme (Supplementary Figure S4). But even with these efficient OTSs, sfGFP production yields decrease far more upon multisite ncAA installation compared to those of the ELP scaffolds (Supplementary Figure S5). A decrease in translation efficiency of up to 75% was observable when the number of in-frame amber stop codons in sfGFP was increased from one to five. This should be investigated in the future to uncover the underlying higher-level context effects, which could be very useful for orthogonal translation systems in general. It should be noted that the decrease in fluorescence signal could be attributed to misfolding or aggregation of sfGFP caused by incorporation of ncAA, as previously reported for an *in vitro* protein production system (Martin et al., 2018). ELPs are unusual, non-globular polypeptide structures that represent a kind of linear "string" of tandem motifs. We hypothesize that these glycine- and proline-rich repeat structures are predominantly intrinsically disordered. This could facilitate their ribosomal translation, since no fine regulation or balance between translation rate and co-translational folding is required. The result is an excellent context for orthogonal translation with an expanded genetic code.

The fact that the fluorescence signal decreases in response to increasing the number of stop codons (most likely because truncation products are formed) despite the absence of endogenous RF1 suggests that the OTS pair cannot deliver enough charged tRNA to prevent ribosome stalling. For a given translation rate, increasing the distance between stop codons could provide more time for loading and delivery of the orthogonal tRNA, which in turn could lead to improved protein production. To better understand the correlation between the number and spacing of in-frame stop codons within the ELP sequence, constructs with alternating stop codon replacements were used, combined with the same fluorescence readout as before. Notably, we cloned and produced construct pairs with identical ELP length to control for the potential impact on the fluorescence reporter. ELP constructs with multiples of five stop codons were built using ELP(5x amber) as origin, while those of with multiples of eight where built using the ELP(8x amber) construct. Details of these constructs can be found in Supplementary Table S1. Using these pairwise comparisons, we found that the amount of full-length protein produced was predominantly correlated with the total number of in-frame stop codons; the distance between these and the overall ELP length had little to no effect (Figure 4). This is best observed by comparing equidistant ELP constructs with their counterparts showing alternating placement of amber codon (e.g. ELP(10x amber) vs ELP(16x amber)). Increasing the number of stop codons and ncAA incorporation sites by 60%

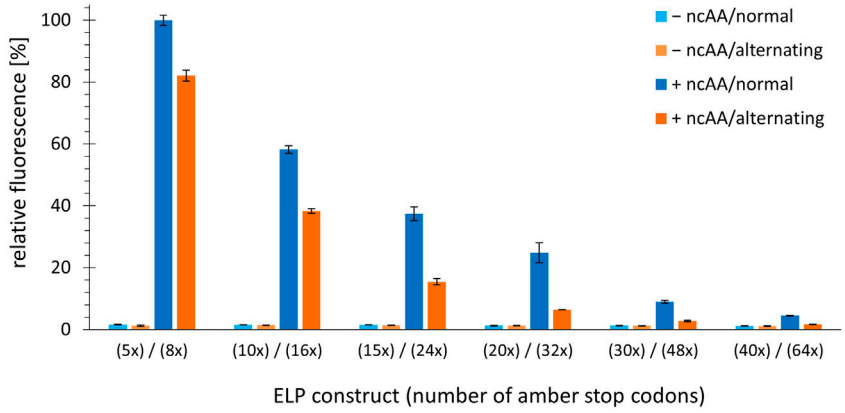


FIGURE 4
Comparison of OTS efficiency upon alteration of the number and spacing pattern of in-frame stop codons. Ribosomal incorporation of *m*-*o*NB-Dopa (1 mM) into ELP-sfGFP constructs containing a varying number of in-frame amber stop codons. Constructs are shown as pairs of identical ELP fragment length (e.g. ELP(5x amber) and ELP(8x amber)). Normal = one amber stop codon per three VPGXG repeats, equidistant placement. Alternating = alternating amber stop codons, see Supplementary Material for sequence details. Intact cell fluorescence of *E. coli* B-95.ΔA, endpoint measurements after 18 h of incubation. Data (incl. Standard deviation) represent the mean of three biological replicates.

leads to a corresponding decrease in the fluorescence signal. Even when the ELP repeat length and the number of stop codons reach high values (and the overall reporter signals are low), the same picture emerges. This can be seen, for example, when comparing the results of ELP(30x amber) (normal, equidistant placement) with ELP(32x amber) (alternating stop codon placement). Although the ELP part of the ELP(30x amber) construct is much larger (495 vs 335 AA), both constructs appear to be produced at comparable levels, suggesting that stop-codon readthrough is the most important determinant. As mentioned earlier, longer ELP length may also contribute to a reduction in the reporter fluorescence signal. However, this effect seems to be less pronounced in our pairwise comparison, e.g. for ELP(15x amber, normal) and ELP(16x amber, alternating). These differ only by one ncAA incorporation site but significantly in ELP fragment length (255 vs 175 AA). A comparable full-length reporter signal was observed, differences cannot be resolved. The high proportion of hydrophobic residues in the ELP part could increase the tendency of the fusion protein to aggregate once the ELP portion reaches a substantial length. At the ELP(15x amber) and ELP(24x amber) construct pair stage, the ELP and sfGFP parts are approximately equal in length. To sum up, constructs with a similar total number of stop codons result in a similar fluorescence signal in the cell, suggesting a comparable yield of full-length protein. To better understand this argument, we visualized the data from Figure 4 differently by omitting the columns without ncAA and sorting the constructs by the number of in-frame stop codons. In the resulting Supplementary Figure S3, we see that the reporter signal intensity follows the number of in-frame stop codons. Returning to the initial hypothesis: Increasing the distance between or changing the pattern of in-frame stop codons by the measures used herein does not facilitate ribosomal readthrough.

3.3 Analysis of recombinant ELP-sfGFP fusion protein

For a more detailed analysis of the protein species produced, the ELP(5x ONB-Dopa)-sfGFP construct was prepared with five ncAA incorporation sites at the scale of 50 ml shake flask cultures. The purified target protein samples were subjected to ESI-MS analysis. Figure 5 shows the successful incorporation of five *m*-*o*NB-Dopa (**1**) moieties into the full-length protein. Unfortunately, but not unexpectedly, a significant portion of the protein produced (approximately 50% of the total as estimated from peak intensities) contains one to five reduced nitro groups. For clarity, the measured mass shifts of the corresponding protein species in Figure 5 are shown and interpreted in Table 2. It can be seen that the observed intact protein masses agree very well with those predicted for possible nitroreduction species. The differences between the expected and observed masses become larger for the less abundant protein species as the measurement confidence is correlated. Attempts were made to analyze constructs containing more than five *m*-*o*NB-Dopa (**1**) moieties, but because of the increasing size of the ELP part, only insoluble fusion proteins were obtained. These were not compatible with the established ESI-MS analysis protocol. Methods exist to analyze such constructs, for example, by proteolytic digestion, but further efforts were

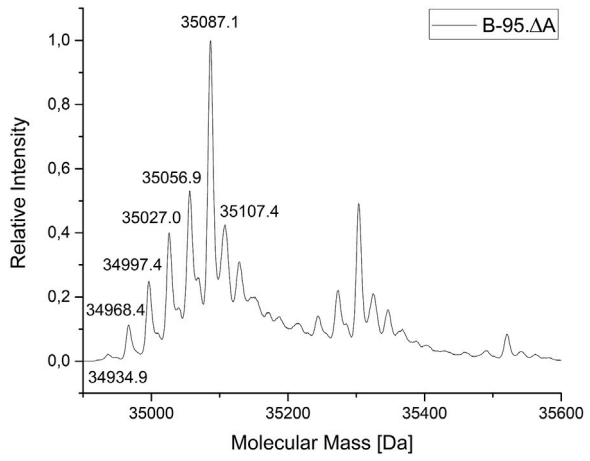


FIGURE 5
Mass profile (ESI-MS) of ELP(5x *m*-oNB-Dopa)-sfGFP produced in *E. coli* B-95.ΔA. The peak pattern shows a characteristic mass shift in line with nitro group reductions (with mass differences fitting to theoretical values for the reduction of the ncAA nitro group). Deconvoluted intact protein mass values and shifts are given in Table 1. The main peak (35087.1 Da) indicates the target species with five ncAAs and devoid of nitro group reduction. The corresponding chemistry is explained in the main text.

discontinued once we found better production methods for proteins with nitro group containing ncAAs, as described below (Israeli et al., 2021; Vanderschuren et al., 2022).

3.4 Engineering a bacterial strain with reduced nitroreductase activity

As initially outlined, bacterial host cells generally reduce aromatic nitro groups in peptides and proteins (Nguyen et al., 2014; Ren et al., 2015; Baumann et al., 2019; Böcker et al., 2019). For the ncAAs employed in this work, the nitro group of moiety **1** is reduced to the corresponding amine *in vivo*. This eliminates the photocleavable ability of the protecting group and ultimately the controlled generation of Dopa-containing recombinant proteins. After reviewing the genome sequence and prevalent literature, at least 11 *E. coli* wild-type genes can be linked to nitroreductase activity (Mercier et al., 2013; Copp et al., 2017). In particular, the NAD(P)H-dependent nitroreductases NfsA and NfsB have the highest ability to reduce nitro groups, e.g. with nitrofurazone (Zenno et al., 1996a; Valle et al., 2012; Copp et al., 2017). Unfortunately, deletion of *nfsA* and *nfsB* has been shown to be insufficient to prevent nitro group reduction for ncAAs in *E. coli* (Ren et al., 2015), even though biotransformation of certain other small molecules can be prevented by only these two deletions (Valle et al., 2012). Following this train of thought, we set out to inactivate six *E. coli* enzymes known to exert nitroreductase activity (Table 2). Gene knockouts were performed in the *E. coli* B-95.ΔA strain as a robust and efficient production platform for the target proteins, as seen above (Mukai et al., 2015).

For such a high number of targets, gene deletions were performed with a revised Lambda Red recombineering system (Jensen et al., 2015) using a single plasmid version of the established method of Datsenko and Wanner (Datsenko and Wanner, 2000). In contrast to the latter, the antibiotic resistance cassette with homologous DNA flanks was amplified from the genomic DNA of *E. coli* clones from the Keio Single Gene Knockout Collection rather than from a template plasmid (Baba et al., 2006). Since these single knockout clones contain the antibiotic resistance cassette with flanking FRT sites, the procedures of PCR amplification, knockout in the target strain and cassette excision are compatible. The workflow is shown in Figure 6 and has been described previously in different variants (Meuskens et al., 2017; Swings et al., 2018; Aedo et al., 2019; Rodríguez-Rojas et al., 2020). This hybrid approach has the advantage that the primer binding sites and the length of the homology regions (surrounding the targeted gene) can be chosen freely. It eliminates primer length constraints and helps to avoid bad primer metrics that otherwise compromise primer synthesis and PCR efficiency. Furthermore and notably, this benefits the recombineering efficiency, which is correlated to the length of the homology region (Murphy, 1998; Liu et al., 2003; Sharan et al., 2009). For one or a few deletions, sufficient knockout efficiencies

TABLE 1 ESI-MS analytics of ELP(5x *m*-oNB-Dopa)-sfGFP produced in *E. coli* B-95.ΔA.

Species[a]	Calculated mass [Da]	Found mass [Da]	Δ mass [Da]
mat., unreduced	35087.1	35087.1	0
mat., 1x reduced	35057.1	35056.9	0.2
mat., 2x reduced	35027.1	35027.0	0.1
mat., 3x reduced	34997.1	34997.4	0.3
mat., 4x reduced	34967.1	34968.4	1.3
mat., 5x reduced	34937.1	34934.9	2.2
Non-maturated, unreduced	35107.1	35107.4	0.3

[a]All protein species without the starting Met, mat. = sfGFP fluorophore is maturated.

TABLE 2 *E. coli* genes with known nitroreductase activity which were targeted for deletion.

Gene	Description	Nitroreductase activity[a]	References
nfsA	Major oxygen-insensitive nitroreductase	High	Zenno et al. (1996a), Copp et al. (2017)
nfsB	Minor oxygen-insensitive nitroreductase	Middle	Zenno et al. (1996b), Copp et al. (2017)
azoR	FMN dependent NADH:quinone oxidoreductase (NADH-azoreductase) which can reduce azo dyes. Nitroreductase activity was observed with ortholog from *Rhodobacter sphaeroides*	Middle/Low	Mercier et al. (2013), Copp et al. (2017)
ydja	One of the smallest nitroreductases from *E. coli*	Low	Choi et al. (2007), Copp et al. (2017)
nemA	Flavin-dependent oxidoreductases related to the old yellow enzyme family	Low	Williams et al. (2004), Copp et al. (2017)
rutE	Conserved nitroreductase domain type. Required for growth on uridine. Reduces malonic semialdehyde to 3-hydroxypropionic acid	Low	Kim et al. (2010), Copp et al. (2017), Loh et al. (2006)

[a]Estimation based on the reported activities.

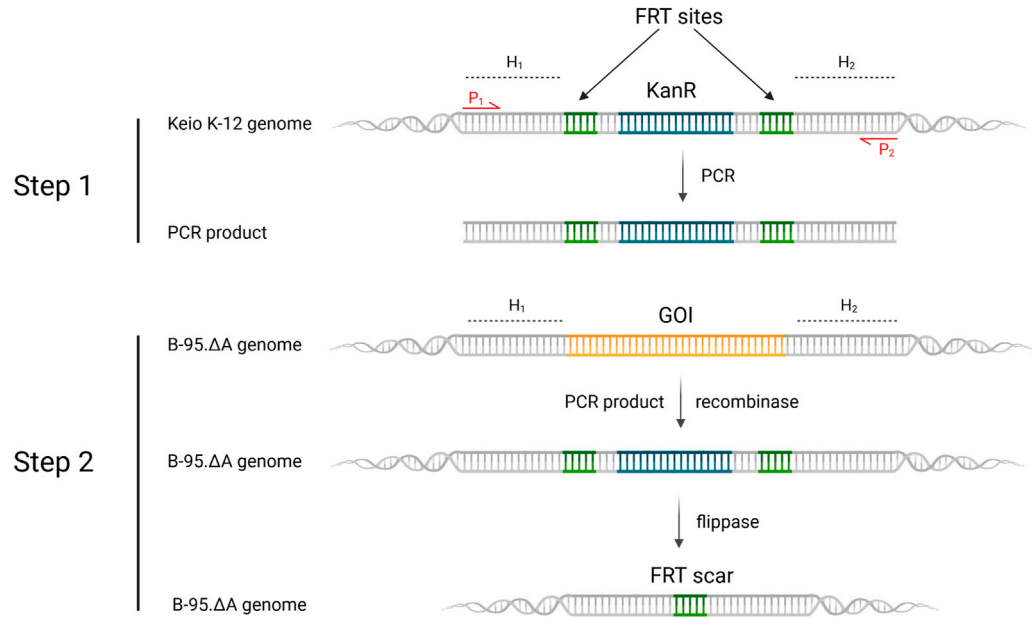


FIGURE 6
Revised lambda red recombineering workflow for gene deletions leading to the generation of the sextuple *E. coli* knockout strain B-95.ΔA NK53. $H_{1,2}$ = homology region 1 or 2; $P_{1,2}$ = Primer 1 or 2; KanR = kanamycin resistance cassette; GOI = gene of interest, which is targeted for deletion. Step 1 displays the PCR-amplification of the FRT-KanR-FRT-cassette from the Keio knockout strain. Step 2 depicts the recombinase induced replacement of the gene of interest with the FRT-KanR-FRT-cassette and its subsequent flippase catalyzed removal. Created with BioRender.com.

can be achieved with the standard homology region length of 35–50 bp. However, for each deleted gene, an FRT scar remains in the genome after cassette excision. This increases the likelihood that the next antibiotic cassette will be inserted at this scar site instead of the target gene, resulting in false positive knockout clones. Longer homology regions alleviate this effect. Another advantage is that the Keio knockouts were carefully designed to reduce negative effects that can occur when a deleted gene is part of an operon for example (Yamamoto et al., 2009). Homology lengths of 100–200 bp were chosen for our nitroreductase knockouts, which are easily and cost-effectively obtained using desalted primers. After each round of deletion, clones were probed *via* colony PCR and gel electrophoresis to verify target gene replacement and cassette excision, respectively. Sanger sequencing of the genomic PCR products was also performed.

Consecutive knockouts of the six genes from Table 2 yielded strain B-95.ΔA (ΔnfsA:FRT, ΔnfsB:FRT, ΔazoR:FRT, Δydja:FRT, ΔnemA:FRT, ΔrutE:FRT), herein referred

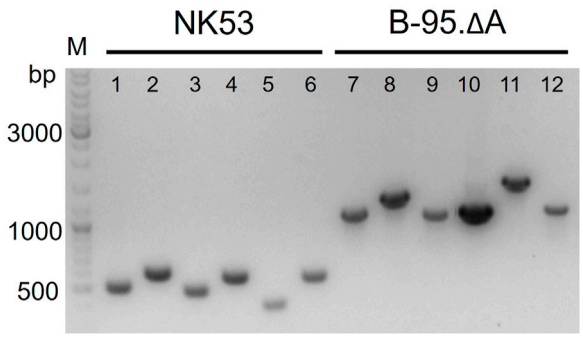


FIGURE 7
Verification of gene deletions by agarose gel electrophoresis of genomic PCR products. Smaller PCR product sizes indicate deletions within the amplified DNA region. Comparison of knockout strain B-95.ΔA NK53 to its progenitor *E. coli* B-95.ΔA used as source of genomic DNA. Lanes: 1) ΔnfsA:FRT; 2) ΔnfsB:FRT; 3) ΔazoR:FRT; 4) Δydja:FRT; 5) ΔnemA:FRT; 6) ΔrutE:FRT; lanes 7–12 are the corresponding wild-type equivalents.

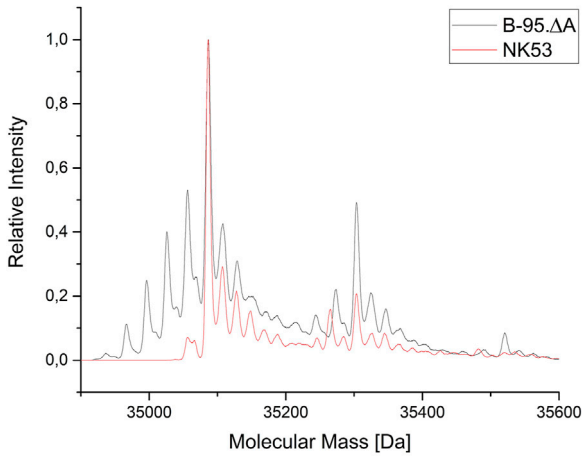


FIGURE 8
Mass spectrometric profile (ESI-MS deconvolution) of proteins produced in the *E. coli* B-95.ΔA NK53 strain with attenuated nitroreductase activity compared to its progenitor strain. ELP(5x *m*-oNB-Dopa)-sfGFP produced by progenitor strain B-95.ΔA (grey) indicates the reduction of several ncAA nitro groups. See Section 3.3 for detailed peak annotation of reduced protein species. Expected intact protein mass for five intact *m*-oNB-Dopa moieties = 35087.1 Da. Observed mass: 35086.9 Da. Expected mass upon nitro group reduction at one site: 35057.1 Da. Observed mass: 35057.5 Da.

to as B-95.ΔA NK53. This final strain was rescreened for all intended deletions, with the original strain serving as negative control (Figure 7, further exemplified for the *nfsB* deletion in Supplementary Figure S6). Figure 7 clearly shows that the genomic DNA amplification PCR products for all targeted genes are shorter than the equivalents of the original strain, indicating successful deletion of the genes. After six gene deletions, the potential impact on the fitness of the bacterial strain was evaluated by parallel growth assays in 24-well plates, monitoring OD_{600} as a proxy for cell density. To sample a variety of growth environments, strains B-95.ΔA, B-95.ΔA NK53 and BL21(DE3) were each grown in complex rich medium (LB, DYT), complex rich buffered medium (TB, ZYP5052) and chemically defined new minimal medium (NMM) without amino acid supplementation (Budisa et al., 1995; Studier, 2005). Interestingly, B-95.ΔA exhibited enhanced growth and also reached higher final densities than the other two strains. Strain B-95.ΔA NK53 was always on par with BL21(DE3), suggesting that the gene knockouts caused slight fitness deficits compared to the parent strain. But as mentioned earlier, the overall fitness still remains high on the level of BL21(DE3), which is the widely used gold standard strain for high-level protein production (Supplementary Figure S7, S8).

The strategy used herein to generate bacterial host strains with multiple gene knockouts turned out to be very efficient in practice. The generation of longer homology arms in the linear DNA fragment increased the efficiency for multiple knockouts in the same strain. Experience in our laboratory had shown that off-target integrations occur frequently when three or more FRT scars are present in the genome. For each targeted gene knockout by deletion, the selection and verification of only five clones was always sufficient to find a clone with the intended modification. It should be mentioned that there are a variety of genomic engineering strategies, including several scarless alternatives (Fels et al., 2020).

3.5 Elucidating the reduced nitroreduction by strain B-95.ΔA NK53

To evaluate the benefits of the performed gene deletions, the same reporter protein construct as in Figure 5 (ELP(5x *m*-oNB-Dopa)-sfGFP) was produced in the constructed sextuple knockout strain B-95.ΔA NK53, purified, and analyzed by ESI-MS. Protein yields ranged between 60–80 mg/L of culture for both strains. Superposition of Figure 5 and the MS spectrum for B-95.ΔA NK53 purified reporter resulted in Figure 8. Here, it is clear that almost no target protein suffering from nitro group reduction is present when produced by the knockout strain. Only a small fraction of target protein containing a reduced nitro group is detectable, rendering the genome engineering efforts of B-95.ΔA a success. Judging from the spectral peaks (which only provide a rough quantification), about 3% of the total protein contains one reduced nitro group. For this protein which contains five *m*-oNB-Dopa incorporation sites, each of which is susceptible to the undesired PTM, this corresponds to an overall reduction of 94% compared to the progenitor strain.

3.5.1 Applications of nitrated proteins and peptides

Next, the practical applicability of the B-95.ΔA NK53 strain for the incorporation of other nitro group-containing ncAAs was evaluated. Two types of experiments were conducted. First, we investigated whether differences in nitroreduction of another ncAA, ortho-nitrobenzyl-tyrosine (ONBY (**3**)), could be detected depending on the production strain. The orthogonal translation system consisting of tRNA and evolved aaRS for the incorporation of ONBY (**3**) has been described previously (Baumann et al., 2019). In the first experiment, ONBY (**3**) was incorporated at position Y63 (corresponding to the chromophore position Y66 in GFP) of amilCP, a chromoprotein from the coral *Acropora millepora* (Alieva et al., 2008). Photo-induced cleavage of the nitrobenzyl group should restore the caged Tyr and reinstate the maturation ability of the chromophore. It should therefore allow spatiotemporal control of chromophore maturation. Previously, a similar experiment was performed with a GFP derivative (Groff et al., 2010). Since amilCP has a similar tertiary protein structure, it was assumed that light-induced chromophore maturation would also be possible. The target protein amilCP(Y63ONBY) was produced both in B-95.ΔA and in the engineered B-95.ΔA NK53 strain. Purified proteins were normalized to the same concentration and irradiated with UV-light to deprotect ONBY (**3**) (Figure 9). Absorbance at 588 nm was monitored as a function of time as an indicator of amilCP chromophore maturation. As hypothesized, both protein batches allowed the photoinduced deprotection of ONBY (**3**), leading to autocatalytic chromophore formation of amilCP. To our surprise, no profound differences in chromoprotein absorption were detectable between proteins produced in B-95.ΔA or B-95.ΔA NK53, respectively. If there would have been a difference in the amount of reduced nitro groups of ONBY (**3**) depending on the expression strain, the observed chromophore absorption should have been significantly lower for amilCP(Y63ONBY) produced in the B-95.ΔA strain. We conclude that the reduction of ONBY (**3**) must occur after its ribosomal incorporation into the polypeptide chain, implying that the aaRS can discriminate between ONBY (**3**) and the corresponding reduced substrate (Baumann et al., 2019). It is plausible that the β-barrel structure of amilCP surrounding the chromophore motif can shield ONBY (**3**) from the enzymes present in the cytoplasm of the host that catalyze nitroreductions. This constellation may explain why reduction of ONBY (**3**) also failed to occur in the earlier GFP work (Groff et al., 2010). Confirming this hypothesis, ESI-MS measurements of proteins produced in both strains revealed that no detectable nitroreduction had occurred (Supplementary Figure S10, S11). Consequently, protection by the protein scaffold allows efficient production of ONBY-containing chromoproteins in both strains and nitroreductase knockouts do not confer any benefits.

The next recombinant target was the post-translationally modified antimicrobial peptide nisin. This lanthionine-containing peptide antibiotic (lantibiotic) is naturally produced by *Lactococcus lactis* (*L. lactis*). Indicator strain assays can be performed as a readout of the antimicrobial activity of this ribosomally synthesized and post-translationally modified peptide (RiPP; produced by natural hosts or recombinant systems) (Baumann et al., 2017; Nickling et al., 2018). Halos indicate growth inhibition zones on agar plates where the nisin-sensitive *L. lactis* indicator strain cannot grow. Here, we attempted to produce nitro group containing nisin variants in which ONBY (**3**) was incorporated at permissive sites. Due to the complex pathway of ribosomal precursor production, posttranslational modification by NisBC and NisP-catalyzed prepeptide cleavage, the contributions of production levels and specific antimicrobial potency to the observed activities are not easily dissected. Nevertheless, we set out to test whether photoactivatable nisin variants could be produced. By combining the recombinant nisin production system and the orthogonal translation system, positions I1, I4 and M17 of nisin were individually replaced with ONBY (**3**). A previous study by the Kuipers lab showed that these positions can be replaced by Trp and that the corresponding peptides continue to be cleaved by NisP (Zhou et al., 2016). Due to the presence of the photocleavable protection group, we anticipated a change in antimicrobial activity upon UV irradiation. Depending on the amount of reduced ONBY nitro groups, the ability to photocleave should be affected. Consequently, the activity of nisin variants with a high degree of nitroreduction should be less affected by irradiation. As shown in Figure 10, only the recombinant nisin produced in the nitroreductase-deficient strain showed a dramatic increase in antimicrobial activity upon UV-irradiation, which is associated with ONBY (**3**) decaging. Production in the progenitor strain B-95.ΔA revealed a low level of photoactivation of nisin (M17ONBY) and all other constructs yielded inactive preparations regardless of irradiation time. Therefore, only recombinant production of ONBY (**3**) modified nisin from the nitroreductase knockout strain will be discussed further. By far the strongest increase in activity upon irradiation is observed when ONBY (**3**) is introduced at nisin position 17, which is a Met residue in the native peptide and part of ring C (see (Rink et al., 2007) for an illustration of nisin structure and numbering). Mutational studies have shown that modulation of this ring leads to major changes in antimicrobial properties (van Kraaij et al., 2000). The physico-chemical properties of ring C amino acids have a major impact on the antimicrobial activity, which requires precise orientation to and interaction with the target cell membrane (Breukink et al., 1998). The corresponding side chain modification on nisin residue I4 resulted in moderate activity after radiation. Even before irradiation, a small halo was detected, indicating activity. Position 4 is part of nisin ring A

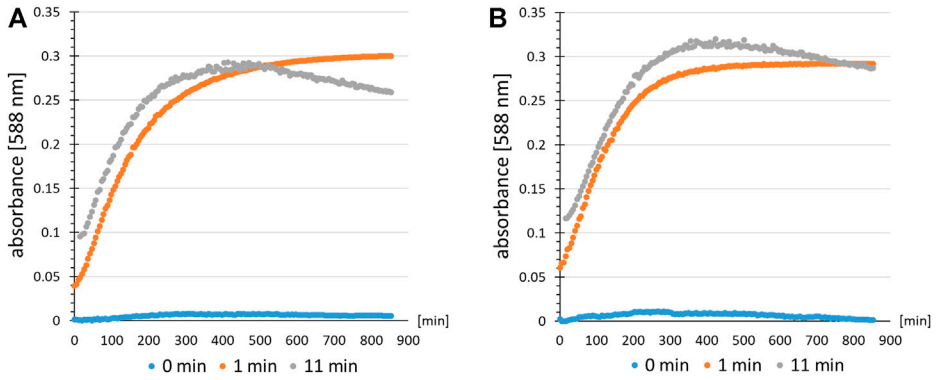


FIGURE 9
Time resolved chromophore maturation of amilCP(Y63ONBY) detected by an absorption assay after light-induced ncAA decaging. Colored dots indicate the time for which the protein solution of amilCP(Y63ONBY) was irradiated with UV-light prior to data recording (absorbance at λ = 588 nm). Controls (0 min, blue) are without irradiation. Comparison of proteins produced using *E. coli* strain **(A)** B-95.ΔA NK53 or **(B)** B-95.ΔA.

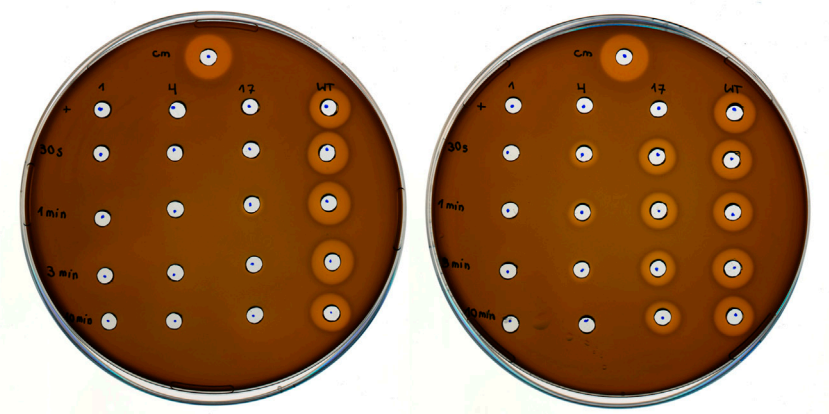


FIGURE 10
Antimicrobial activity of photocaged and decaged (UV-irradiated) nisin derivatives. *L. lactis* indicator strain assay using *E. coli* cell lysates containing nisin modified with ONBY (3) at position I1, I4 and M17, respectively, with recombinant wild-type nisin (WT) as reference. *E. coli* lysates were prepared with either B-95.ΔA (progenitor, left) or B-95.ΔA NK53 (nitroreductase knockout, right) as host strain. Cm = Chloramphenicol (400 μg/ml final) positive control, first row (+) samples were likewise prepared with ONBY (3) supplementation, but without UV-light irradiation of the *E. coli* cell lysate. Numbers from top to bottom indicate the duration of UV-light irradiation (30 s, 1, 3, 10 min). Numbers above the column indicate the position within the nisin peptide where ONBY (3) is introduced by amber suppression and genetic encoding of the ncAA.

and tolerates replacement by Phe without complete loss of activity (Rink et al., 2007). Nevertheless, the exchange of I4 with the ncAA moiety appears to be structurally too different to generate a highly active peptide or sufficient quantities thereof, respectively. When ONBY (**3**) was incorporated at position I1, no antimicrobial activity was observed regardless of irradiation, although a previous study showed that I1Y and I1W mutants of nisin can retain some activity (Lagedroste et al., 2019). This could be due to inefficient production, differences in assay sensitivity, or the fact that the nisin N-terminus is inserted into the target membrane surface, which requires a specific range of side chain hydrophobicity and size. We would like to note that the nisin experiment described above was intended as a snapshot for the applicability of the nitroreductase knockout strain. Because the purification and molecular analysis of ncAA-modified nisin peptides is challenging, the results obtained must be treated with caution. However, because the observed effects were large and consistent with our hypothesis and the data accumulated in this study, we felt confident to include these results.

This experiment demonstrates that, particularly for peptides lacking a tertiary structure, the bioactive function can be

efficiently masked (and subsequently activated) when expressed in strains that can protect the *o*NB group from reduction. The solvent accessibility of the ncAA side chain increases both the chance for the reduction of its nitro group by endogenous nitroreductases and the change in physico-chemical properties upon photocleavage. Our examples of ncAA-modified ELPs and a light-activatable lantibiotic show that in particular peptide-based biomolecules can benefit greatly from having a nitroreductase deficient *E. coli* strain available for recombinant production.

4 Conclusion

By merging three bioengineering streams (orthogonal translation system, protein scaffold and *E. coli* strain), we have created a platform for the efficient production of polypeptides containing multiple installations of nitro group containing non-canonical amino acids. Fluorescence assays show that ELPs containing *m*-oNB-Dopa (**1**) at up to 60 sites can be produced with as little as 1 mM ncAA supplied. The generated host strain with reduced nitroreductase activity has great potential for the general production of peptides and proteins containing non-reduced nitro groups, even beyond *m*-oNB-Dopa (**1**) and ONBY (**3**). Likewise, it can be used for cell-free protein synthesis, as reported for its progenitor strain (Seki et al., 2018). The strategy described herein can also serve as a template for dramatically reducing the nitroreductase activity of other *E. coli* laboratory strains. Sextuple knockouts can be generated within 1 month using standard lab equipment and consumables. While the genes chosen for knockout significantly reduce the endogenous reduction of nitro groups of solvent-exposed side chains of peptides and proteins, other nonessential genes with this activity remain. These can now be targeted to further reduce unwanted side chain conversions and to further increase the *in vivo* half-life of intact nitro groups. All types of applications where nitro groups in biomolecules are important for protection or mechanistic function can benefit from our engineered strain and approach. By using a robust RF1 knockout strain in combination with an efficient orthogonal translation system, we achieved a high level of amber suppression in ELP genes, resulting in satisfactory amounts of target protein. We envision that the high protein yield combined with the photocleavability of *m*-oNB-Dopa (**1**) and the thermo-responsiveness of ELPs will lead to promising spatio-temporal and temperature controlled smart materials.

Data availability statement

The original contributions presented in the study are included in the article and Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

NK carried out the experiments. AD performed one fluorescence assay. Nisin production and antimicrobial activity assays were conducted by JN. NK and TB designed, and NK and AD generated genetic constructs. NK, TB and NB analyzed the data. NK wrote the original draft and did the graphic visualization of the manuscript. TB, NB and NK performed review and editing. NK identified, designed and carried out the nitroreductase gene deletion strategy. TB and NB did the supervision and project management. All authors have given approval to the final version of the manuscript.

Funding

This research was supported by internal funds of the Biocatalysis Group at the TU Berlin and by the Canada Research Chairs Program (Grant No. 950-231971). We acknowledge support by the German Research Foundation and the Open Access Publication Fund of TU Berlin.

Acknowledgments

We thank Annette Poch (Technische Universität Berlin, Süssmuth group) for providing the Keio collection strains. The authors are grateful for the valuable comments and suggestions for improvement by the two reviewers.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.992748/full#supplementary-material

References

Aedo, S. J., Ma, H. R., and Brynildsen, M. P. (2019). "Checks and balances with use of the Keio collection for phenotype testing," in *Microbial metabolic engineering*. Editor A. P. Santos C (New York, NY: Humana Press), 125–138. doi:10.1007/978-1-4939-9142-6_9

Aerni, H. R., Shifman, M. A., Rogulina, S., O'Donoghue, P., and Rinehart, J. (2015). Revealing the amino acid composition of proteins within an expanded genetic code. *Nucleic Acids Res.* 43, e8. doi:10.1093/nar/gku1087

Agostini, F., Völler, J. S., Koksch, B., Acevedo-Rocha, C. G., Kubyshkin, V., and Budisa, N. (2017). Biocatalysis with unnatural amino acids: Enzymology meets xenobiology. *Angew. Chem. Int. Ed. Engl.* 56, 9680–9703. doi:10.1002/anie.201610129

Ahn, B. K., Das, S., Linstadt, R., Kaufman, Y., Martinez-Rodriguez, N. R., Mirshafian, R., et al. (2015). High-performance mussel-inspired adhesives of reduced complexity. *Nat. Commun.* 6, 8663–8667. doi:10.1038/ncomms9663

Alieva, N. O., Konzen, K. A., Field, S. F., Meleshkevitch, E. A., Hunt, M. E., Beltran-Ramirez, V., et al. (2008). Diversity and evolution of coral fluorescent proteins. *PLoS One* 3, e2680. doi:10.1371/journal.pone.0002680

Amiram, M., Haimovich, A. D., Fan, C., Wang, Y. S., Aerni, H. R., Ntai, I., et al. (2015). Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* 33, 1272–1279. doi:10.1038/nbt.3372

Antonczak, A. K., Simova, Z., Yonemoto, I. T., Bochtler, M., Piasecka, A., Czapińska, H., et al. (2011). Importance of single molecular determinants in the fidelity of expanded genetic codes. *Proc. Natl. Acad. Sci. U. S. A.* 108, 1320–1325. doi:10.1073/pnas.1012276108

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* 2, 2006.0008. doi:10.1038/msb4100050

Baumann, T., Nickling, J. H., Bartholomae, M., Buivydas, A., Kuipers, O. P., and Budisa, N. (2017). Prospects of *in vivo* incorporation of non-canonical amino acids for the chemical diversification of antimicrobial peptides. *Front. Microbiol.* 8, 124–129. doi:10.3389/fmicb.2017.00124

Baumann, T., Hauf, M., Richter, F., Albers, S., Möglich, A., Ignatova, Z., et al. (2019). Computational aminoacyl-tRNA synthetase library design for photocaged tyrosine. *Int. J. Mol. Sci.* 20, E2343. doi:10.3390/ijms20092343

Bellingham, C. M., Lillie, M. A., Gosline, J. M., Wright, G. M., Starcher, B. C., Bailey, A. J., et al. (2003). Recombinant human elastin polypeptides self-assemble into biomaterials with elastin-like properties. *Biopolymers* 70, 445–455. doi:10.1002/bip.10512

Böcker, J. K., Dörner, W., and Mootz, H. D. (2019). Light-control of the ultra-fast Gp41-1 split intein with preserved stability of a genetically encoded photo-caged amino acid in bacterial cells. *Chem. Commun.* 55, 1287–1290. doi:10.1039/c8cc09204d

Breukink, E., van Kraaij, C., van Dalen, A., Demel, R. A., Siezen, R. J., de Kruijff, B., et al. (1998). The orientation of nisin in membranes. *Biochemistry* 37, 8153–8162. doi:10.1021/bi9727971

Budisa, N., and Schneider, T. (2019). Expanding the DOPA universe with genetically encoded, mussel-inspired bioadhesives for material Sciences and medicine. *ChemBioChem* 20, 2163–2190. doi:10.1002/cbic.201900030

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (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, 788–796. doi:10.1111/j.1432-1033.1995.tb20622.x

Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013). A versatile platform for single- and multiple-unnatural amino acid mutagenesis in escherichia coli. *Biochemistry* 52, 1828–1837. doi:10.1021/bi4000244

Chemla, Y., Ozer, E., Algov, I., and Alfonta, L. (2018). Context effects of genetic code expansion by stop codon suppression. *Curr. Opin. Chem. Biol.* 46, 146–155. doi:10.1016/j.cbpa.2018.07.012

Choi, J. W., Lee, J., Kosuke, N., Jung, C. H., and Kim, J. S. (2007). Crystallization and preliminary X-ray diffraction analysis of ydjA, a minimal nitroreductase from *Escherichia coli* K12. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* 63, 1064–1066. doi:10.1107/S1744309107057636

Copp, J. N., Mowday, A. M., Williams, E. M., Guise, C. P., Ashoorzadeh, A., Sharrock, A. V., et al. (2017). Engineering a multifunctional nitroreductase for improved activation of prodrugs and PET probes for cancer gene therapy. *Cell Chem. Biol.* 24, 391–403. doi:10.1016/j.chembiol.2017.02.005

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6640–6645. doi:10.1073/pnas.120163297

de la Torre, D., and Chin, J. W. (2021). Reprogramming the genetic code. *Nat. Rev. Genet.* 22, 169–184. doi:10.1038/s41576-020-00307-7

Deiters, A., Groff, D., Ryu, Y., Xie, J., and Schultz, P. G. (2006). A genetically encoded photocaged tyrosine. *Angew. Chem. Int. Ed. Engl.* 45, 2728–2731. doi:10.1002/anie.200600264

Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015). Designing logical codon reassignment-Expanding the chemistry in biology. *Chem. Sci.* 6, 50–69. doi:10.1039/c4sc01534g

Fels, U., Gevaert, K., and Van Damme, P. (2020). Bacterial genetic engineering by means of recombineering for reverse genetics. *Front. Microbiol.* 11, 548410–548419. doi:10.3389/fmicb.2020.548410

Groff, D., Wang, F., Jockusch, S., Turro, N. J., and Schultz, P. G. (2010). A new strategy to photoactivate green fluorescent protein. *Angew. Chem. Int. Ed. Engl.* 49, 7677–7679. doi:10.1002/anie.201003797

Gueta, O., Sheinenzon, O., Azulay, R., Shalit, H., Strugach, D. S., Hadar, D., et al. (2022). Tuning the properties of protein-based polymers using high-performance orthogonal translation systems for the incorporation of aromatic non-canonical amino acids. *Front. Bioeng. Biotechnol.* 10, 913057–913112. doi:10.3389/fbioe.2022.913057

Hadar, D., Gelkop, S., Vaserman, L., and Amiram, M. (2021). Efficient incorporation of clickable unnatural amino acids enables rapid and biocompatible labeling of proteins *in vitro* and in bacteria. *Chembiochem.* 22, 1379–1384. doi:10.1002/cbic.202000663

Hale, M. K. M. C., Setton, L. a. D. P., and Chilkoti, A. (2005). Cartilaginous tissue repair. *Tissue Eng.* 11, 1768–1779. doi:10.1089/ten.2005.11.1768

Haller, C. M., Buerzle, W., Kivelio, A., Perrini, M., Brubaker, C. E., Gubeli, R. J., et al. (2012). Mussel-mimetic tissue adhesive for fetal membrane repair: An *ex vivo* evaluation. *Acta Biomater.* 8, 4365–4370. doi:10.1016/j.actbio.2012.07.047

Hauf, M., Richter, F., Schneider, T., Faidt, T., Martins, B. M., Baumann, T., et al. (2017). Photoactivatable mussel-based underwater adhesive proteins by an expanded genetic code. *ChemBioChem* 18, 1819–1823. doi:10.1002/cbic.201700327

Hoppmann, C., Wong, A., Yang, B., Li, S., Hunter, T., Shokat, K. M., et al. (2017). Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nat. Chem. Biol.* 13, 842–844. doi:10.1038/nchembio.2406

Israeli, B., Strugach, D. S., Gelkop, S., Weber, S., Gozlan, D. S., and Amiram, M. (2021). Genetically encoding light-responsive protein-polymers using translation machinery for the multi-site incorporation of photo-switchable unnatural amino acids. *Adv. Funct. Mater.* 31, 2011276. doi:10.1002/adfm.202011276

Italia, J. S., Peeler, J. C., Hillenbrand, C. M., Latour, C., Weerapana, E., and Chatterjee, A. (2020). Genetically encoded protein sulfation in mammalian cells. *Nat. Chem. Biol.* 16, 379–382. doi:10.1038/s41589-020-0493-1

Jang, H. S., Gu, X., Cooley, R. B., Porter, J. J., Henson, R. L., Willi, T., et al. (2020). Efficient site-specific prokaryotic and eukaryotic incorporation of halotyrosine amino acids into proteins. *ACS Chem. Biol.* 15, 562–574. doi:10.1021/acschembio.9b01026

Jensen, S. I., Lennen, R. M., Herrgård, M. J., and Nielsen, A. T. (2015). Seven gene deletions in seven days: Fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress. *Sci. Rep.* 5, 17874–17910. doi:10.1038/srep17874

Kaushik, N., Kaushik, N., Pardeshi, S., Sharma, J., Lee, S., and Choi, E. (2015). Biomedical and clinical importance of mussel-inspired polymers and materials. *Mar. Drugs* 13, 6792–6817. doi:10.3390/md13116792

Khusainov, R., and Kuipers, O. P. (2013). The presence of modifiable residues in the core peptide part of precursor nisin is not crucial for precursor nisin interactions with NisB- and NisC. *PLoS One* 8, e74890. doi:10.1371/journal.pone.0074890

Khusainov, R., Heils, R., Lubelski, J., Moll, G. N., and Kuipers, O. P. (2011). Determining sites of interaction between prenisin and its modification enzymes NisB and NisC. *Mol. Microbiol.* 82, 706–718. doi:10.1111/j.1365-2958.2011.07846.x

Kim, K. S., Pelton, J. G., Inwood, W. B., Andersen, U., Kustu, S., and Wemmer, D. E. (2010). The Rut pathway for pyrimidine degradation: Novel chemistry and toxicity problems. *J. Bacteriol.* 192, 4089–4102. doi:10.1128/JB.00201-10

Kim, E., Jeon, J., Zhu, Y., Hoppe, E. D., Jun, Y.-S., Genin, G. M., et al. (2021). A biosynthetic hybrid spidroin-amyloid-mussel foot protein for underwater adhesion on diverse surfaces. *ACS Appl. Mater. Interfaces* 13, 48457–48468. doi:10.1021/acsami.1c14182

Koch, N. G., Goettig, P., Rappsilber, J., and Budisa, N. (2021). Engineering pyrrolysyl-tRNA synthetase for the incorporation of non-canonical amino acids with smaller side chains. *Int. J. Mol. Sci.* 22, 11194. doi:10.3390/ijms222011194

Lagedroste, M., Reiners, J., Smits, S. H. J., and Schmitt, L. (2019). Systematic characterization of position one variants within the lantibiotic nisin. *Sci. Rep.* 9, 935–1011. doi:10.1038/s41598-018-37532-4

Liu, C. C., and Schultz, P. G. (2010). Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 79, 413–444. doi:10.1146/annurev.biochem.052308.105824

Liu, P., Jenkins, N. A., and Copeland, N. G. (2003). A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* 13, 476–484. doi:10.1101/gr.749203

Loh, K. D., Gyaneshwar, P., Papadimitriou, E. M., Fong, R., Kim, K. S., Parales, R., et al. (2006). A previously undescribed pathway for pyrimidine catabolism. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5114–5119. doi:10.1073/pnas.0600521103

Luo, X., Fu, G., Wang, R. E., Zhu, X., Zambaldo, C., Liu, R., et al. (2017). Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* 13, 845–849. doi:10.1038/nchembio.2405

Maier, G. P., Rapp, M. V., Waite, J. H., Israelachvili, J. N., and Butler, A. (2015). BIOLOGICAL ADHESIVES. Adaptive synergy between catechol and lysine promotes wet adhesion by surface salt displacement. *Science* 349, 628–632. doi:10.1126/science.aab0556

Martin, R. W., Des Soye, B. J., Kwon, Y.-C., Kay, J., Davis, R. G., Thomas, P. M., et al. (2018). Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nat. Commun.* 9, 1203. doi:10.1038/s41467-018-03469-5

Mercier, C., Chalansonnet, V., Orenga, S., and Gilbert, C. (2013). Characteristics of major *Escherichia coli* reductases involved in aerobic nitro and azo reduction. *J. Appl. Microbiol.* 115, 1012–1022. doi:10.1111/jam.12294

Meuskens, I., Michalik, M., Chauhan, N., Linke, D., and Leo, J. C. (2017). A new strain collection for improved expression of outer membrane proteins. *Front. Cell. Infect. Microbiol.* 7, 464–513. doi:10.3389/fcimb.2017.00464

Meyer, D. E., and Chilkoti, A. (2002). Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system. *Biomacromolecules* 3, 357–367. doi:10.1021/bm015630n

Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., et al. (2015). Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci. Rep.* 5, 9699–9. doi:10.1038/srep09699

Mukai, T., Lajoie, M. J., Englert, M., and Söll, D. (2017). Rewriting the genetic code. *Annu. Rev. Microbiol.* 71, 557–577. doi:10.1146/annurev-micro-090816-093247

Murphy, K. C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *J. Bacteriol.* 180, 2063–2071. doi:10.1128/JB.180.8.2063-2071.1998

Neumann, H., Hazen, J. L., Weinstein, J., Mehl, R. A., and Chin, J. W. (2008). Genetically encoding protein oxidative damage. *J. Am. Chem. Soc.* 130, 4028–4033. doi:10.1021/ja710100d

Nguyen, D. P., Mahesh, M., Elsässer, S. J., Hancock, S. M., Uttamapinant, C., and Chin, J. W. (2014). Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* 136, 2240–2243. doi:10.1021/ja412191m

Nickling, J. H., Baumann, T., Schmitt, F.-J., Bartholomae, M., Kuipers, O. P., Friedrich, T., et al. (2018). Antimicrobial peptides produced by selective pressure incorporation of non-canonical amino acids. *J. Vis. Exp.* 135, 57551. doi:10.3791/57551

O'Donoghue, P., Prat, L., Heinemann, I. U., Ling, J., Odoi, K., Liu, W. R., et al. (2012). Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. *FEBS Lett.* 586, 3931–3937. doi:10.1016/j.febslet.2012.09.033

Pagar, A. D., Patil, M. D., Flood, D. T., Yoo, T. H., Dawson, P. E., and Yun, H. (2021). Recent advances in Biocatalysis with chemical modification and expanded amino acid alphabet. *Chem. Rev.* 121, 6173–6245. doi:10.1021/acs.chemrev.0c01201

Porter, J. J., Jang, H. S., Van Fossen, E. M., Nguyen, D. P., Willi, T. S., Cooley, R. B., et al. (2019). Genetically encoded protein tyrosine nitration in mammalian cells. *ACS Chem. Biol.* 14, 1328–1336. doi:10.1021/acschembio.9b00371

Pott, M., Schmidt, M. J., and Summerer, D. (2014). Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids. *ACS Chem. Biol.* 9, 2815–2822. doi:10.1021/cb5006273

Registry of Standard Biological Parts (2022). Registry of standard biological parts. Available at: https://parts.igem.org/Part:pSB1C3 (Accessed July 8, 2022).

Ren, W., Ji, A., Wang, M. X., and Ai, H. (2015). Expanding the genetic code for a dinitrophenyl hapten. *ChemBioChem* 16, 2007–2010. doi:10.1002/cbic.201500204

Rink, R., Wierenga, J., Kuipers, A., Kluskens, L. D., Driessen, A. J. M., Kuipers, O. P., et al. (2007). Dissection and modulation of the four distinct activities of nisin by mutagenesis of rings A and B and by C-terminal truncation. *Appl. Environ. Microbiol.* 73, 5809–5816. doi:10.1128/AEM.01104-07

Roberts, S., Dzuricky, M., and Chilkoti, A. (2015). Elastin-like polypeptides as models of intrinsically disordered proteins. *FEBS Lett.* 589, 2477–2486. doi:10.1016/j.febslet.2015.08.029

Rodríguez-Rojas, A., Kim, J. J., Johnston, P. R., Makarova, O., Eravci, M., Weise, C., et al. (2020). Non-lethal exposure to H2O2 boosts bacterial survival and evolvability against oxidative stress. *PLoS Genet.* 16, e1008649. doi:10.1371/journal.pgen.1008649

Schwark, D. G., Schmitt, M. A., and Fisk, J. D. (2018). Dissecting the contribution of release factor interactions to amber stop codon reassignment efficiencies of the methanocaldococcus jannaschii orthogonal pair. *Genes (Basel)* 9, E546. doi:10.3390/genes9110546

Seki, E., Yanagisawa, T., and Yokoyama, S. (2018). Cell-free protein synthesis for multiple site-specific incorporation of noncanonical amino acids using cell extracts from RF-1 deletion *E. coli* strains. *Methods Mol. Biol.* 1728, 49–65. doi:10.1007/978-1-4939-7574-7_3

Seo, S., Das, S., Zalicki, P. J., Mirshafian, R., Eisenbach, C. D., Israelachvili, J. N., et al. (2015). Microphase behavior and enhanced wet-cohesion of synthetic copolyampholytes inspired by a mussel foot protein. *J. Am. Chem. Soc.* 137, 9214–9217. doi:10.1021/jacs.5b03827

Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* 4, 206–223. doi:10.1038/nprot.2008.227

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234. doi:10.1016/j.pep.2005.01.016

Swings, T., Marciano, D. C., Atri, B., Bosserman, R. E., Wang, C., Leysen, M., et al. (2018). CRISPR-FRT targets shared sites in a knock-out collection for off-the-shelf genome editing. *Nat. Commun.* 9, 2231. doi:10.1038/s41467-018-04651-5

Urry, D. W. (1997). Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers. *J. Phys. Chem. B* 101, 11007–11028. doi:10.1021/jp972167t

Valle, A., Le Borgne, S., Bolívar, J., Cabrera, G., and Cantero, D. (2012). Study of the role played by NfsA, NfsB nitroreductase and NemA flavin reductase from *Escherichia coli* in the conversion of ethyl 2-(2'-nitrophenoxy)acetate to 4-hydroxy-(2H)-1, 4-benzoxazin-3(4H)-one (D-DIBOA), a benzohydroxamic acid with interesting biological properties. *Appl. Microbiol. Biotechnol.* 94, 163–171. doi:10.1007/s00253-011-3787-0

van Kraaij, C., Breukink, E., Rollema, H. S., Bongers, R. S., Kosters, H. A., de Kruijff, B., et al. (2000). Engineering a disulfide bond and free thiols in the lantibiotic nisin Z. *Eur. J. Biochem.* 267, 901–909. doi:10.1046/j.1432-1327.2000.01075.x

Vanderschuren, K., Arranz-Gibert, P., Khang, M., Hadar, D., Gaudin, A., Yang, F., et al. (2022). Tuning protein half-life in mouse using sequence-defined biopolymers functionalized with lipids. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2103099119. doi:10.1073/pnas.2103099119

Vrhovski, B., and Weiss, A. S. (1998). Biochemistry of tropoelastin. *Eur. J. Biochem.* 258, 1–18. doi:10.1046/j.1432-1327.1998.2580001.x

Waite, J. H. (2017). Mussel adhesion - essential footwork. *J. Exp. Biol.* 220, 517–530. doi:10.1242/jeb.134056

Wan, W., Tharp, J. M., and Liu, W. R. (2014). Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* 1844, 1059–1070. doi:10.1016/j.bbapap.2014.03.002

Williams, R. E., Rathbone, D. A., Scrutton, N. S., and Bruce, N. C. (2004). Biotransformation of explosives by the old yellow enzyme family of flavoproteins. *Appl. Environ. Microbiol.* 70, 3566–3574. doi:10.1128/AEM.70.6.3566-3574.2004

Yamamoto, N., Nakahigashi, K., Nakamichi, T., Yoshino, M., Takai, Y., Touda, Y., et al. (2009). Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* 5, 335. doi:10.1038/msb.2009.92

Young, D. D., and Schultz, P. G. (2018). Playing with the molecules of life. *ACS Chem. Biol.* 13, 854–870. doi:10.1021/acschembio.7b00974

Zenno, S., Koike, H., Kumar, A. N., Jayaraman, R., Tanokura, M., and Saigo, K. (1996a). Biochemical characterization of NfsA, the *Escherichia coli* major nitroreductase exhibiting a high amino acid sequence homology to Frp, a Vibrio harveyi flavin oxidoreductase. *J. Bacteriol.* 178, 4508–4514. doi:10.1128/jb.178.15.4508-4514.1996

Zenno, S., Koike, H., Tanokura, M., and Saigo, K. (1996b). Gene cloning, purification, and characterization of NfsB, a minor oxygen-insensitive nitroreductase from *Escherichia coli*, similar in biochemical properties to FRase I, the major flavin reductase in Vibrio fischeri. *J. Biochem.* 120, 736–744. doi:10.1093/oxfordjournals.jbchem.a021473

Zhou, L., Shao, J., Li, Q., van Heel, A. J., de Vries, M. P., Broos, J., et al. (2016). Incorporation of tryptophan analogues into the lantibiotic nisin. *Amino Acids* 48, 1309–1318. doi:10.1007/s00726-016-2186-3

Zhu, P., Gafken, P. R., Mehl, R. A., and Cooley, R. B. (2019). A highly versatile expression system for the production of multiply phosphorylated proteins. *ACS Chem. Biol.* 14, 1564–1572. doi:10.1021/acschembio.9b00307

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 24 January 2023
DOI 10.3389/fmolb.2023.1096261

OPEN ACCESS

EDITED BY
Ned Budisa,
University of Manitoba, Canada

REVIEWED BY
Weimin Xuan,
Tianjin University, China
Xian Fu,
Beijing Genomics Institute (BGI), China
Chenguang Fan,
University of Arkansas, United States

*CORRESPONDENCE
Natalie Krahn,
natalie.krahn@yale.edu

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

RECEIVED 11 November 2022
ACCEPTED 11 January 2023
PUBLISHED 24 January 2023

CITATION
Morosky P, Comyns C, Nunes LGA, Chung CZ, Hoffmann PR, Söll D, Vargas-Rodriguez O and Krahn N (2023), Dual incorporation of non-canonical amino acids enables production of post-translationally modified selenoproteins. *Front. Mol. Biosci.* 10:1096261. doi: 10.3389/fmolb.2023.1096261

Dual incorporation of non-canonical amino acids enables production of post-translationally modified selenoproteins

Pearl Morosky[1], Cody Comyns[1], Lance G. A. Nunes[2], Christina Z. Chung[1], Peter R. Hoffmann[2], Dieter Söll[1,3], Oscar Vargas-Rodriguez[1] and Natalie Krahn[1]*

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, United States, [2]Department of Cell and Molecular Biology, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI, United States, [3]Department of Chemistry, Yale University, New Haven, CT, United States

Post-translational modifications (PTMs) can occur on almost all amino acids in eukaryotes as a key mechanism for regulating protein function. The ability to study the role of these modifications in various biological processes requires techniques to modify proteins site-specifically. One strategy for this is genetic code expansion (GCE) in bacteria. The low frequency of post-translational modifications in bacteria makes it a preferred host to study whether the presence of a post-translational modification influences a protein's function. Genetic code expansion employs orthogonal translation systems engineered to incorporate a modified amino acid at a designated protein position. Selenoproteins, proteins containing selenocysteine, are also known to be post-translationally modified. Selenoproteins have essential roles in oxidative stress, immune response, cell maintenance, and skeletal muscle regeneration. Their complicated biosynthesis mechanism has been a hurdle in our understanding of selenoprotein functions. As technologies for selenocysteine insertion have recently improved, we wanted to create a genetic system that would allow the study of post-translational modifications in selenoproteins. By combining genetic code expansion techniques and selenocysteine insertion technologies, we were able to recode stop codons for insertion of N_{ε}-acetyl-L-lysine and selenocysteine, respectively, into multiple proteins. The specificity of these amino acids for their assigned position and the simplicity of reverting the modified amino acid *via* mutagenesis of the codon sequence demonstrates the capacity of this method to study selenoproteins and the role of their post-translational modifications. Moreover, the evidence that Sec insertion technology can be combined with genetic code expansion tools further expands the chemical biology applications.

KEYWORDS

selenoproteins, acetyl-lysine, post-translational modifications, genetic code expansion, selenocysteine, synthetic biology

1 Introduction

Selenium is an essential micronutrient required for the biosynthesis of selenocysteine (Sec), which is exclusively found in selenoproteins. Selenoproteins are involved in many key cellular functions, including calcium and redox homeostasis, cell maintenance, and immune and inflammatory responses (Ye et al., 2022). Twenty-five selenoproteins have been identified in

humans, but their functions remain poorly characterized. This is partly due to the complicated and highly regulated Sec biosynthesis and its insertion into proteins. In eukaryotes, this process depends on *cis* and *trans*-acting factors in the mRNA and transportation of the tRNA to the ribosome, respectively (Pinkerton and Copeland, 2016). Similar to bacterial systems, the Sec-charged tRNA (Sec-tRNA^{Sec}) recognizes a dedicated UGA non-sense codon with the assistance of the Sec-specific elongation factor (eEFSec). Redefinition of the UGA codon as Sec is mediated by a hairpin element known as the Sec insertion sequence (SECIS) found up to 1,600 nucleotides away from the UGA codon in the 3′-untranslated region (UTR) of eukaryotic selenoprotein mRNA. eEFSec brings Sec-tRNA^{Sec} to the ribosome in response to the SECIS element in a process that involves an additional factor, the SECIS binding protein 2 (SBP2) (Copeland et al., 2000). SBP2 is predicted to bind to the SECIS element, bending the 3′-UTR to interact with eEFSec, though the detailed mechanism is still not fully understood (Kossinova et al., 2014; Hilal et al., 2022).

The complicated translation path of eukaryotic selenoproteins poses a challenge to overexpress and purify these proteins for functional characterization studies. Some strategies [e.g., pSelExpress1, orthogonal aminoacyl-tRNA synthetase (aaRS):tRNA pairs] have been developed to address this challenge in eukaryotes [reviewed in (Novoselov et al., 2007; Chung and Krahn, 2022)]. However, the bacterial Sec insertion system is simpler, only requiring a specialized elongation factor (SelB) to recognize Sec-tRNA^{Sec} and the SECIS element for insertion of Sec. As a result, more methodologies have been developed for recombinant Sec insertion in bacteria [reviewed in (Chung and Krahn, 2022)]. We have specifically focused on engineering SECIS-independent translation by removing the requirement for SelB and instead using the canonical elongation factor (EF-Tu) (Aldag et al., 2013). This method depends on allo-tRNAs, a tRNA species with an unusual cloverleaf structure (Mukai et al., 2017). We have since verified that some of the allo-tRNAs are recognized by EF-Tu (Mukai et al., 2018) with tertiary structures that facilitates accommodation in the *E. coli* ribosome (Prabhakar et al., 2022). Through engineering strategies, they have been altered for efficient Sec incorporation (Mukai et al., 2018).

As with standard protein expression, eukaryotic selenoproteins can also be expressed in bacteria but with minimal post-translational modifications (PTMs). To install PTMs into *E. coli* expressed proteins, genetic code expansion (GCE) utilizes orthogonal aaRS:tRNA synthetase pairs and non-canonical amino acids (ncAAs) (Chen et al., 2018; Porter and Mehl, 2018; Yang et al., 2018). The aaRS is engineered to accept diverse ncAAs typically for insertion at a non-sense or stop codon. Some commonly used orthogonal translation systems include the *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (TyrRS) and archaeal pyrrolysyl-tRNA synthetase (PylRS) systems. The tolerance of PylRS to anticodon changes in its cognate tRNA^{Pyl} makes it an attractive tool to recode any codon, while the codon choice for TyrRS can be more limiting due to its anticodon recognition (Crnković et al., 2016). These systems have been used to insert common PTMs such as acetylation (Lacoursiere et al., 2020), phosphorylation (Balasuriya et al., 2020), or both (Venkat et al., 2018).

The presence of three non-sense codons in the genetic code offers an opportunity to insert multiple ncAAs into a single protein by combining existing orthogonal systems. To this end, several studies have demonstrated the synthesis of proteins containing two and even three ncAAs (Neumann et al., 2010; Chatterjee et al., 2013; Wang et al., 2014; Dunkelmann et al., 2020; Tharp et al., 2021). Since tRNA^{Sec} is acylated with Ser by endogenous seryl-tRNA synthetase (SerRS) before conversion to Sec *via* dedicated enzymes, it may be compatible with many of these orthogonal translation systems. This was tested with the natural Sec machinery in *E. coli*. Using a natural eukaryotic selenoprotein with a penultimate Sec amino acid, mRNA engineering facilitated a SECIS element to be inserted into the 3′-UTR of the mRNA while still promoting recognition by endogenous SelB for Sec insertion. Combining this with AcKRS (an aaRS evolved from *Methanosarcinae* PylRS) for incorporation of N_ε-acetyl-L-lysine (AcK) showed the compatibility of these two systems and their capability of introducing AcK into a natural selenoprotein (Wright et al., 2018). However, the requirement of a SECIS element immediately downstream of the UGA codon limits this strategy to natural bacterial selenoproteins or eukaryotic selenoproteins with a Sec residue at the C-terminal end (penultimate or ultimate state).

Here we present a new strategy for the simultaneous insertion of Sec and AcK into super-folder green fluorescent protein (sfGFP) and human glutathione peroxidase 1 (GPx1). We achieved this by combining a previously engineered Sec-incorporation system to site-specifically insert Sec anywhere in a polypeptide chain (Mukai et al., 2018) with a *M. alvus* PylRS engineered for insertion of AcK (Figure 1) (Seki et al., 2020). With the correct choice of anticodon for each of these translation systems, we facilitated increased codon orthogonality and suppression to yield post-translationally modified selenoproteins. This further expands the genetic code expansion toolbox, enabling the capability to design novel proteins and study eukaryotic selenoproteins in *E. coli*.

2 Materials and methods

2.1 General

Enzymes for molecular cloning were purchased from New England Biolabs and Takara Bio. Gene fragments were purchased from Integrated DNA Technologies and Twist Bioscience. W.M. Keck Foundation at Yale University provided oligonucleotide synthesis (Supplementary Table S1) and DNA plasmid (Supplementary Table S2) sequencing. Antibiotics and media additives were used at the following final concentrations: ampicillin (amp), 100 μg/mL; spectinomycin (spec), 50 μg/mL; glucose, 1% (w/v); arabinose, 0.1% (w/v); sodium selenite, 10 μM; N_ε-acetyl-L-lysine (AcK), 10 mM; nicotinamide, 20 mM.

2.2 Plasmid construction

2.2.1 pB_sfGFP_2TGA (pB_PM02) and pB_sfGFP_151TAG (pB_PM03)

The stop codon of the sfGFP gene in the previously reported pB_sfGFP plasmid (Chung et al., 2022) was mutated to TAA using the primer pair PM01/02 (pB_PM01). PM03/04 and PM05/06 primers were used separately to introduce a TGA at position 2 (2TGA) or a TAG at position 151 (151TAG) of sfGFP in the pB_PM01 plasmid. This provided the plasmids (pB_PM02 and pB_PM03) to test each suppression system individually.

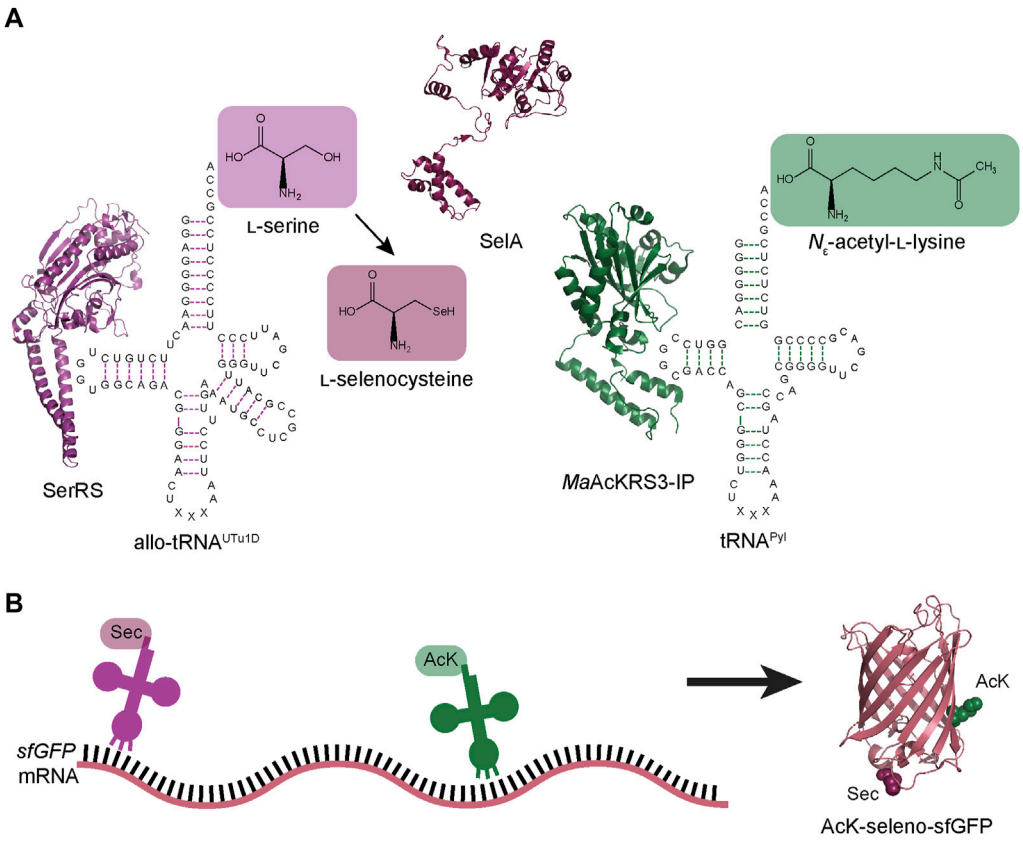


FIGURE 1
Schematic of dual ncAA incorporation to produce post-translationally modified seleno-sfGFP. **(A)** Endogenous *E. coli* seryl-tRNA synthetase (SerRS represented by PDB:1SER) aminoacylates allo-tRNA^{UTu1D} with L-serine, which is then converted to L-Sec by *A. salmonicida* selenocysteine synthase (SelA represented by PDB:3W1K). *M. alvus* PylRS containing mutations for recognition of N_ε-acetyl-L-lysine (AcK) (*Ma*AcKRS3-IP represented by PDB:4Q6G) aminoacylates *M. alvus* tRNA^{Pyl}. **(B)** Aminoacylated allo-tRNA^{UTu1D} and tRNA^{Pyl} insert Sec and AcK, respectively, at specific positions in the mRNA. These tRNAs are encoded to suppress stop codons (UAG or UGA), leading to the biosynthesis of AcK-seleno-sfGFP (represented by PDB:1EMB).

2.2.2 pB_sfGFP_2TGA-*Ma*AcLysRS3-tRNA$^{Pyl}_{UCA}$ (pB_PM06) and pB_sfGFP_151TAG-*Ma*AcLysRS3-tRNA$^{Pyl}_{UCA}$ (pB_PM07)

To insert the PylRS and tRNA^{Pyl} genes for AcK insertion into the previously designed plasmids, the primer pair PM07/08 was used to amplify the chemically synthesized *Ma*AcLysRS3-IP gene (Seki et al., 2020) and insert a proK promoter. Primers PM09/10 opened the pB_PM01, pB_PM02 or pB_PM03 plasmids downstream of the origin of replication with overhangs complementary to the amplified *Ma*AcLysRS3-IP fragment to insert in the reverse direction. The opened pB_PM01, pB_PM02 or pB_PM03 and *Ma*AcLysRS3-IP fragment were assembled with NEBuilder HiFi to generate pB_PM04, pB_PM05, and pB_PM06, respectively. These plasmids were then opened upstream of *Ma*AcLysRS3-IP with PM11/12 to insert the *M. alvus* tRNA$^{Pyl}_{UCA}$ fragment containing an Lpp promoter and rrcn terminator. The tRNA$^{Pyl}_{UCA}$ fragment was amplified with PM13/14 from a pBAD-sfGFP-tRNA plasmid (Tharp et al., 2020) containing the sequence for tRNA$^{Pyl}_{UCA}$. The opened pB_PM04, pB_PM05, and pB_PM06 plasmids and tRNA$^{Pyl}_{UCA}$ fragment were assembled with NEBuilder HiFi to generate pB_PM07, pB_PM08, and pB_PM09, respectively.

2.2.3 pB_sfGFP_2TGA_151TAG-*Ma*AcLysRS3-tRNA$^{Pyl}_{UCA}$ (pB_PM10)

Starting with the pB_PM08 plasmid, PM05/06 primers were used to add the second stop codon within the coding region of sfGFP for dual suppression studies.

2.2.4 pB_04_2TGA-*Ma*AcLysRS3- tRNA$^{Pyl}_{UCA}$ (pB_PM13)

The pB_04 plasmid containing sfGFP disrupted by the M86 DnaB mini-intein at position 204 was previously used to detect Sec insertion (Chung et al., 2022). The primer pair PM01/02 was used first to mutate the stop codon to TAA (pB_PM11). Due to the presence of a TAG already in the intein-sfGFP fusion, the second in-frame stop codon, TGA, was inserted at position 2 using the PM03/04 primers (pB_PM12). Insertion of the AcK-translation system (*Ma*AcLysRS3-mA17) was achieved as described in Section 2.2.2 (pB_PM13).

2.2.5 pB_sfGFP-*Ma*AcLysRS3-tRNA$^{Pyl}_{CUA}$ plasmids (pB_PM14, pB_PM15, pB_PM16, and pB_PM17)

The anticodon sequence of the tRNA$^{Pyl}_{UCA}$ gene in plasmids pB_PM07, pB_PM08, pB_PM09, and pB_PM10 plasmids was mutated to CUA to suppress the UAG codon using the primer pair PM15/16. This

generated plasmids pB_PM14, pB_PM15, pB_PM16, and pB_PM17, respectively.

2.2.6 pB_04TGA_2TAG-*Ma*AcLysRS3-tRNA$^{Pyl}_{CUA}$ (pB_PM18)

The anticodon sequence of tRNA$^{Pyl}_{UCA}$ from pB_PM13 was also converted to CUA following the strategy described in 2.2.5. However, since the intein is specific for Sec insertion, the codons at position 2 of sfGFP and position 204 (position 1 of the intein) had to also be switched. Primer pairs PM17/18 and PM19/20 were used to convert 2TGA to 2TAG and pB_04 to pB_04TGA, respectively.

2.2.7 pSecUGA

To recode UGA with Sec, the anticodon of allo-tRNA^{UTu1D} in the pSecUAG plasmid (Mukai et al., 2018; Chung et al., 2021) was changed to UCA. This was done using primer pair PM21/22 on the pSecUAG plasmid to generate pSecUGA.

2.2.8 pB_GPx1_49TAG-*Ma*AcLysRS3-tRNA$^{Pyl}_{CUA}$ (pB_PM19)

For GPx1 expression, the human GPx1 gene was inserted into the pB-*Ma*AcLysRS3-tRNA$^{Pyl}_{CUA}$ vector containing genes for AcK insertion at UAG. Primer pair PM23/24 amplified GPx1_49TAG from pET-GPx1 (Mukai et al., 2018), while primer pair PM25/26 was used to open pB_PM17. Digestion with DpnI removed any parental plasmid from the PCR product before NEBuilder HiFi assembled the two PCR purified samples (Gpx1_49TAG and pB-*Ma*AcLysRS3-tRNA$^{Pyl}_{CUA}$). This produced plasmid pB_PM19.

2.2.9 pB_GPx1_49TGA_114TAG-*Ma*AcLysRS3-tRNA$^{Pyl}_{CUA}$ (pB_PM21) and pB_GPx1_49TGA_148TAG-*Ma*AcLysRS3-tRNA$^{Pyl}_{CUA}$ plasmids (pB_PM22)

Mutagenesis of position 49 to encode a UGA was accomplished with primer pair PM27/28 on pB_PM19. This resulting plasmid (pB_PM20) was then subject to further mutagenesis to install UAG codons at position 114 or 148 (using primer pairs PM29/30 or PM31/32, respectively) for insertion of AcK. Final plasmids (pB_PM21 and pB_PM22) were then ready for dual insertion of Sec and AcK.

2.3 sfGFP fluorescence assay

Plasmids were transformed into electrocompetent *E. coli* cells (B-95.ΔA.Δ*fabR*Δ*selABC*) (Mukai et al., 2015) and plated on Luria Broth (LB) agar containing appropriate antibiotics and incubated overnight at 37°C. Single colonies were grown in 150 µL media at 37°C for 6 h in a 96-well black plate with clear bottoms. Unless otherwise noted, the media containing appropriate antibiotics included glucose, arabinose, sodium selenite, AcK, and nicotinamide was used. After 6 h, 75 µL from each well was transferred to a clean well in the same plate. The original wells were replenished with 75 µL of fresh media, while protein expression was induced in the new wells by adding 75 µL of media containing 2 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). Fluorescent measurements and analysis were performed as described previously (Chung et al., 2022) using a minimum of four biological replicates.

2.4 Post-translationally modified selenoprotein production

For protein expression, pB_PM17, pB_PM21, or pB_PM22 plasmids were co-transformed with pSecUGA into electrocompetent *E. coli* cells (B-95.ΔA.Δ*fabR*Δ*selABC*) (Mukai et al., 2015) before plated on LB agar containing appropriate antibiotics and incubated overnight at 37°C. Single colonies were grown overnight at 37°C in 25 mL LB containing the appropriate antibiotics. Precultures were transferred to 1 L of LB containing antibiotics, arabinose, and sodium selenite and grown at 37°C. At an OD_{600} = 0.6, protein expression was induced with 0.5 mM IPTG, following the addition of nicotinamide, AcK, and extra sodium selenite. Proteins were expressed at 20°C for 16 h before pelleting the cells for purification.

All purification steps were performed under anaerobic conditions (90% N_2, 5% H_2, 5% CO_2) in an anaerobic tent (Coy Laboratories). Each cell pellet was resuspended in 18 mL lysis buffer (50 mM sodium phosphate [pH 8.0], 300 mM NaCl, 30 mM imidazole, 10% (v/v) glycerol, 120 µg/mL lysozyme, 30 µg/mL Dnase, 0.5 mM phenylmethylsulfonyl fluoride [PMSF]) and 2 mL BugBuster® 10 X Extraction Reagent (EMD Millipore). The lysed cells were incubated at room temperature for 20 min before the lysate was ultracentrifuged at 45,000 rpm (150,000 x g) for 45 min at 4°C. The supernatant was 0.45 µm filtered before being loaded onto a 1 mL Ni-NTA column pre-equilibrated with wash buffer (50 mM sodium phosphate [pH 8.0], 300 mM NaCl, 30 mM imidazole, 10% (v/v) glycerol). The beads were washed with 50 mL of wash buffer and then eluted in 1 mL fractions with elution buffer (50 mM sodium phosphate [pH 8.0], 300 mM NaCl, 250 mM imidazole, 10% (v/v) glycerol). Protein fractions were combined, concentrated, and buffer exchanged into storage buffer (50 mM sodium phosphate [pH 8.0], 300 mM NaCl, 10% (v/v) glycerol) before being stored at −80°C.

2.5 Mass spectrometry analysis

LC-MS analysis of dithiothreitol reduced samples were performed on a ThermoFisher Scientific Orbitrap Exploris 240 mass spectrometer, equipped with a heated electrospray ionization source (H-ESI) in positive ion mode with a ThermoFisher Ultimate 3000 RSLCnano HPLC System. On H-ESI source, sheath gas was set to 2 arbitrary units (arb), and auxiliary gas was set to 6 arb. The ion transfer tube was set at 275°C and the vaporizer temp at 200°C. The sample was analyzed on a MAbPac RP, 4 µM, 3.0 mm × 50 mm analytical column (ThermoFisher Scientific), held at 60°C. The protein was eluted at a rate of 500 µL/min for a 10-min gradient, where 0–7 min: 10%–70% acetonitrile + 0.1% formic acid; 7–8.2 min: 95% acetonitrile + 0.1% formic acid, 8.2–10 min: 20% acetonitrile + 0.1% formic acid. MS spectra were acquired using full scans at 15,000 resolution in the orbitrap within a range of 700–2,200 m/z. The maximum injection time was set at auto with a standard AGC target. Ten micro scans were employed, and the RF lens was set to 100%. 15 V of insource CID was applied. Thermo BioPharma Finder 5.1 was used for intact mass deconvolution and peak identification.

10.3389/fmolb.2023.1096261

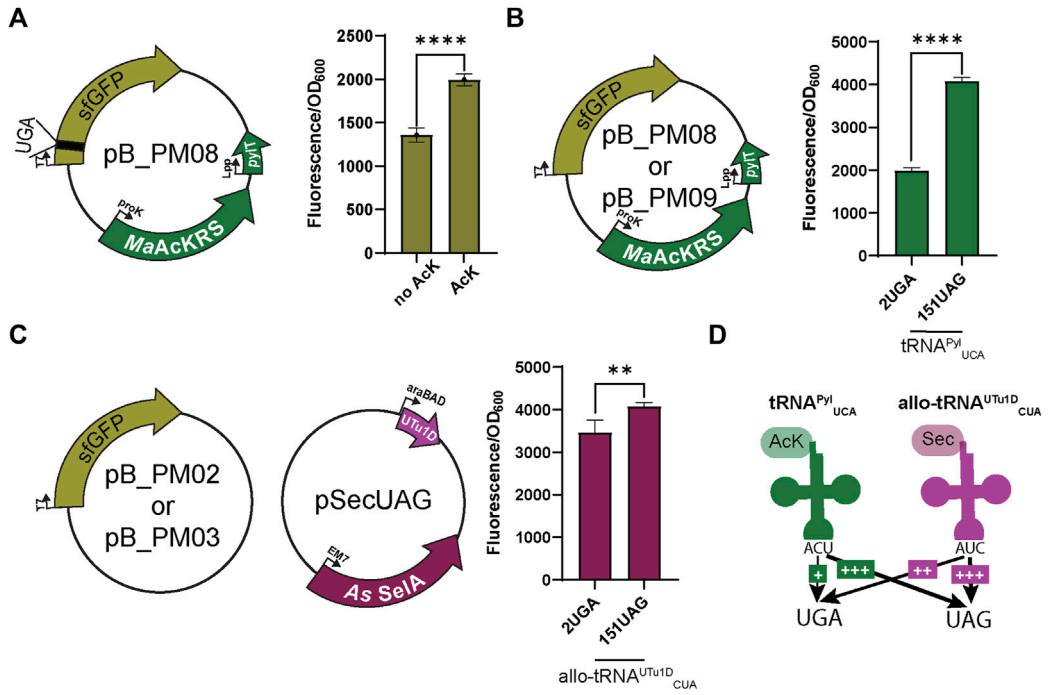


FIGURE 2
Fluorescence assay observes bias for UAG suppression. **(A)** To first test the specificity of the *Ma*AcKRS system for the presence of N_ε-acetyl-L-lysine (AcK), pB_PM06 was used. This plasmid contains sfGFP under the control of a T7 promoter encoding a UGA at position 2 (2UGA), *Ma*AcKRS3-IP under a proK promoter, and $tRNA^{Pyl}_{UCA}$ (*pylT*) under an Lpp promoter. The presence of AcK promoted sfGFP fluorescence ($p < 0.0001$, $n = 4$). **(B)** pB_PM06 and pB_PM07 plasmids encoding 2UGA or 151UAG, respectively, were tested for the specificity of $tRNA^{Pyl}_{UCA}$ for its cognate codon (UGA). Significant increase in readthrough fluorescence ($p < 0.0001$, $n = 4$) was observed for 151UAG, the non-cognate codon. **(C)** Combining plasmids pB_PM02 or pB_PM03 encoding 2UGA or 151UAG, respectively, and pSecUAG allowed a test for UAG suppression by allo-$tRNA^{UTu1D}_{CUA}$. This plasmid contains the machinery for insertion of Sec, namely allo-$tRNA^{UTu1D}_{CUA}$ (UTu1D) under the araBAD promoter and *A. salmonicida* (*As*) SelA under EM7. Similar readthrough fluorescence was observed for both codons with a preference for its cognate codon, UAG ($p < 0.01$, $n = 4$). **(D)** Combining this information, $tRNA^{Pyl}_{UCA}$ (green) poorly suppresses its cognate codon UGA but is much better at suppressing UAG, while allo-$tRNA^{UTu1D}_{CUA}$ (magenta) also suppresses both codons with a bias towards UAG. Increase in suppression efficiency is denoted by an increase in + and bolder arrow.

2.6 Western blot to identify acetylation

Approximately 5 µg of GPx1 protein samples were loaded onto a 4%–20% Mini-PROTEAN TGX stain-free gel (Bio-Rad) in reducing conditions. Samples were transferred to nitrocellulose and acetylation was detected with a primary mouse anti-acetylated lysine mAb (Novus Biologicals) and secondary anti-mouse IgG HRP-linked antibody (Cell Signaling Technology).

3 Results

3.1 Determination of suppression codon choice

Using our established pSecUAG plasmid system for insertion of Sec at UAG codons, we first engineered the anticodon of *M. alvus* $tRNA^{Pyl}$ to recode UGA codons ($tRNA^{Pyl}_{UCA}$). Before testing the orthogonality of each tRNA for its respective codon, we confirmed the specificity of *Ma*AcLysRS3-IP for AcK (Seki et al., 2020). *E. coli* cells expressing only pB_PM08 showed significantly higher fluorescence ($p < 0.0001$) in the presence of AcK compared to cells without the amino acid (Figure 2A). This suggests that AcK promotes readthrough of the UGA codon in sfGFP with the *Ma*AcLysRS3: $tRNA^{Pyl}_{UCA}$ pair.

When using multiple translation systems to suppress more than one stop codon, confirming the codon specificity and orthogonality of engineered suppressor tRNAs is imperative. Therefore, we tested the ability of each system to readthrough UAG or UGA independently. To first investigate the ability of $tRNA^{Pyl}_{UCA}$ to specifically suppress UGA, we expressed the plasmid pB_PM08 or pB_PM09 in the presence of AcK. Notably, a significant increase ($p < 0.0001$) in GFP fluorescence for readthrough of 151UAG relative to 2UGA was observed (Figure 2B), suggesting that $tRNA^{Pyl}_{UCA}$ has a propensity to decode UAG codons better than its cognate UGA codon. We also tested this for allo-$tRNA^{UTu1D}_{CUA}$ by expressing plasmids pB_PM02 or pB_PM03 together with pSecUAG in the presence of sodium selenite. Interestingly, allo-$tRNA^{UTu1D}_{CUA}$ translated both 2UGA and 151UAG, with a significant increase in fluorescence ($p < 0.01$) observed for the intended UAG codon (Figure 2C). Together, these results show that $tRNA^{Pyl}_{UCA}$ and allo-$tRNA^{UTu1D}_{CUA}$ can suppress both UGA and UAG codons in isolation, displaying a higher efficiency to recode UAG (Figure 2D). The propensity to recode UAG was also observed when testing the suppression efficiency of each codon individually (pB_PM08 or pB_PM09) and together (pB_PM10) compared to wild-type (pB_PM07) in the presence of AcK and

10.3389/fmolb.2023.1096261

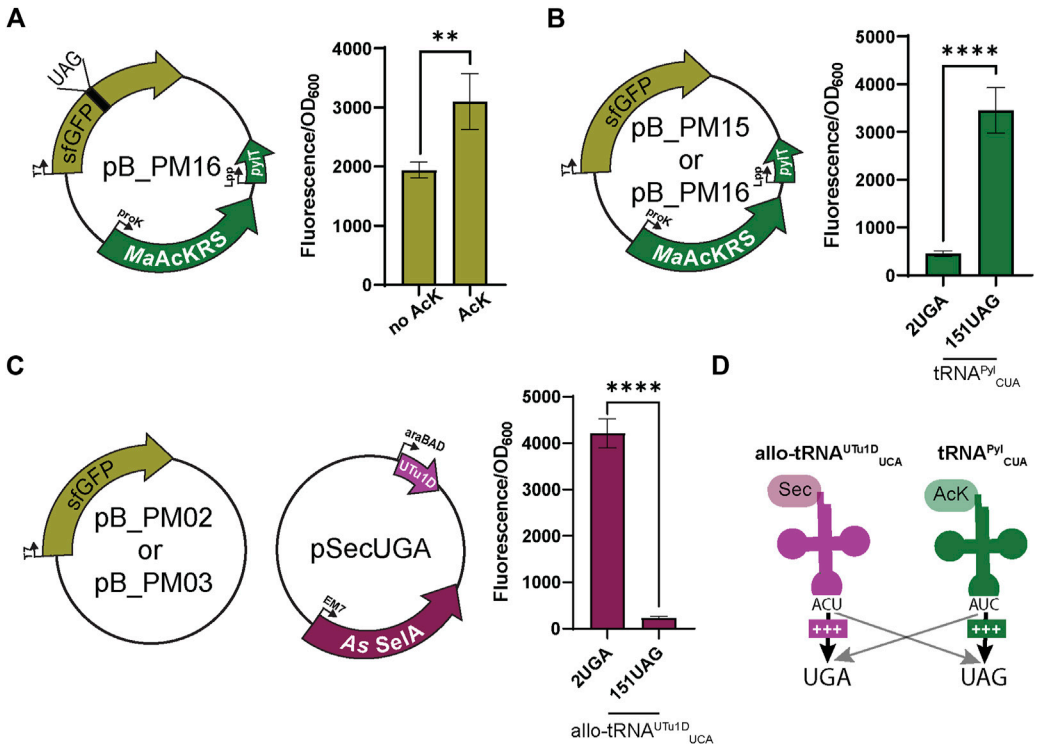


FIGURE 3
Anticodon swap creates codon orthogonality. **(A)** N_ε-acetyl-L-lysine (AcK) was still found to be required for significant fluorescence ($p < 0.0001$, $n = 4$) of sfGFP with plasmid pB_PM13. This plasmid contains sfGFP under the control of a T7 promoter encoding a UAG at position 151 (151UAG), *Ma*AcKRS3-IP under a proK promoter, and $tRNA^{Pyl}_{CUA}$ (*pylT*) under an Lpp promoter. **(B)** pB_PM12 and pB_PM13 encoding 2UGA or 151UAG, respectively, were tested for the specificity of $tRNA^{Pyl}_{CUA}$ for its cognate codon (UAG). Significant increase in readthrough fluorescence ($p < 0.0001$, $n = 4$) was observed for 151UAG, the cognate codon, compared to 2UGA. **(C)** Combining plasmids pB_PM02 or pB_PM03 encoding 2UGA or 151UAG, respectively, and pSecUGA allowed a test for UGA suppression by allo-$tRNA^{UTu1D}_{UCA}$. This plasmid contains the machinery for insertion of Sec, namely allo-$tRNA^{UTu1D}_{UCA}$ (UTu1D) under the araBAD promoter and *Aeromonas salmonicida* (*As*) SelA under EM7. Significant increase in readthrough fluorescence ($p < 0.0001$, $n = 4$) was observed for 2UGA, the cognate codon, compared to 151UAG. **(D)** Combining this information, allo-$tRNA^{UTu1D}_{UCA}$ (magenta) is efficient at UGA suppression, while $tRNA^{Pyl}_{CUA}$ (green) is efficient at UAG suppression, creating an orthogonal system for dual suppression. Increase in suppression efficiency is denoted by an increase in + and bolder arrow. Grey arrows denote that no suppression was observed.

sodium selenite (Supplementary Figure S1A). Thus, these two tRNA variants are not entirely orthogonal.

To test whether codon specificity can be attained, we swapped the anticodons of both tRNAs, resulting in $tRNA^{Pyl}_{CUA}$ and allo-$tRNA^{UTu1D}_{UCA}$. This led to the pB series (pB_PM14, pB_PM15, pB_PM16, and pB_PM17) and pSecUGA to decode UAG and UGA, respectively. We again confirmed that the *Ma*AcLysRS3:$tRNA^{Pyl}_{CUA}$ pair relies on AcK for efficient readthrough ($p < 0.01$) (Figure 3A). Importantly, we found that $tRNA^{Pyl}_{CUA}$ almost exclusively translates the UAG codon, with minimal decoding of UGA. Expressing only pB_PM15 or pB_PM16 we observed fluorescence when 151UAG is present but not in the presence of 2UGA ($p < 0.0001$) (Figure 3B). Similarly, allo-$tRNA^{UTu1D}_{UCA}$ displayed a higher specificity for UGA than UAG. Expressing pSecUGA with pB_PM02 or pB_PM03, we observed the opposite, fluorescence with 2UGA and not 151UAG ($p < 0.0001$) (Figure 3C). Furthermore, the suppression efficiencies of each codon individually (pB_PM15 or pB_PM16) and together (pB_PM17) were comparable and roughly 60%–70% of wild-type sfGFP (pB_PM14) in the presence of AcK and sodium selenite (Supplementary Figure S1B). These data suggest that allo-$tRNA^{UTu1D}_{UCA}$ and $tRNA^{Pyl}_{CUA}$ can simultaneously insert AcK at UAG and Sec at UGA into proteins efficiently (Figure 3D).

3.2 Dual insertion of AcK and Sec in a fluorescent reporter

Next, we combined the AcK and Sec incorporation systems to produce a synthetic post-translationally modified selenoprotein (AcK-seleno-sfGFP). We first tested the activities of the *Ma*AcLysRS3:$tRNA^{Pyl}_{CUA}$ and the Sec-translation systems (Figure 4A) in the presence and absence of AcK and arabinose (for allo-$tRNA^{UTu1D}_{UCA}$ expression). Significant sfGFP fluorescence ($p < 0.001$) was only observed in the presence of both AcK and arabinose but not if one or both is missing (Figure 4B), confirming the feasibility and compatibility of these two systems. Due to the possibility of Ser misincorporation by the pSecUGA system, we confirmed Sec incorporation using pB_PM18 (Figure 5A) which was engineered based on our previously developed Sec-specific sfGFP reporter (pB_04) (Chung et al., 2022). The sfGFP reporter is only functional (fluorescent) when Sec is inserted at amino acid 204 (position 1 of the M86 mini intein) to facilitate cleavage and splicing of the two sfGFP fragments (Appleby-Tagoe et al., 2011). Moreover, mutagenesis of codon 2 to UAG and the stop codon of the sfGFP to UAA allows for simultaneously monitoring the insertion of two ncAAs. By monitoring the fluorescence of the intein-sfGFP reporter, we only

10.3389/fmolb.2023.1096261

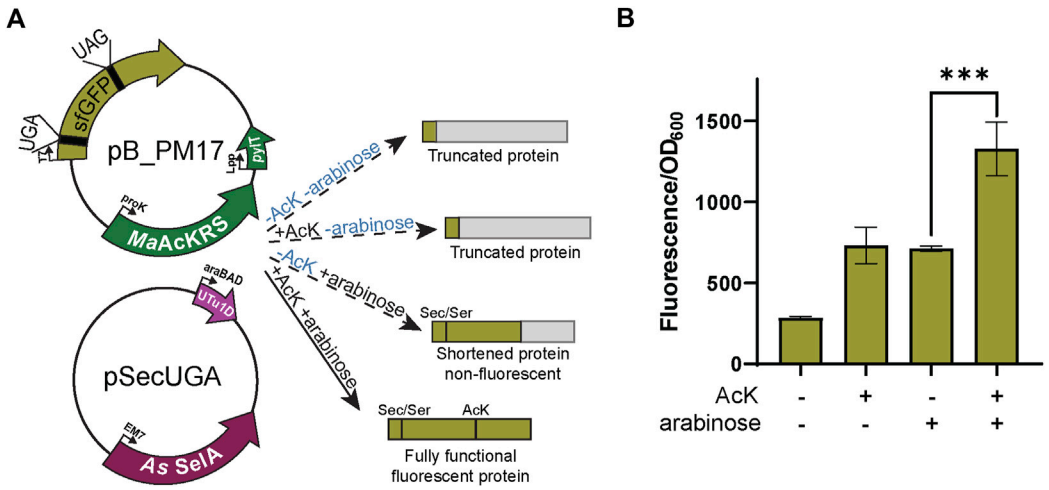


FIGURE 4
Fluorescence from dual suppression in sfGFP. **(A)** The assay was performed with a two-plasmid system: pB_PM14 and pSecUGA. The first plasmid contains the genes for *Ma*AcKRS3-IP and $tRNA^{Pyl}_{CUA}$ (*pylT*) for insertion of N_ε-acetyl-L-lysine (AcK) at a UAG and pSecUGA contains the machinery for insertion of Sec at a UGA. In the presence of AcK, arabinose, and sodium selenite, the entire sfGFP gene should be expressed to produce fluorescence. In the absence of AcK or arabinose (or both), translation stops resulting in truncated non-fluorescent protein. Without arabinose, only the first amino acid would be translated. Without AcK, translation would continue until the UAG at position 151 to produce a truncated protein. **(B)** sfGFP assay shows successful readthrough of both codons in the presence of AcK and Sec with a significant increase in fluorescence ($p < 0.001$) compared to when only one or none of the components were added. Data are shown as the average of at least four biological replicates with the corresponding standard deviation shown with error bars.

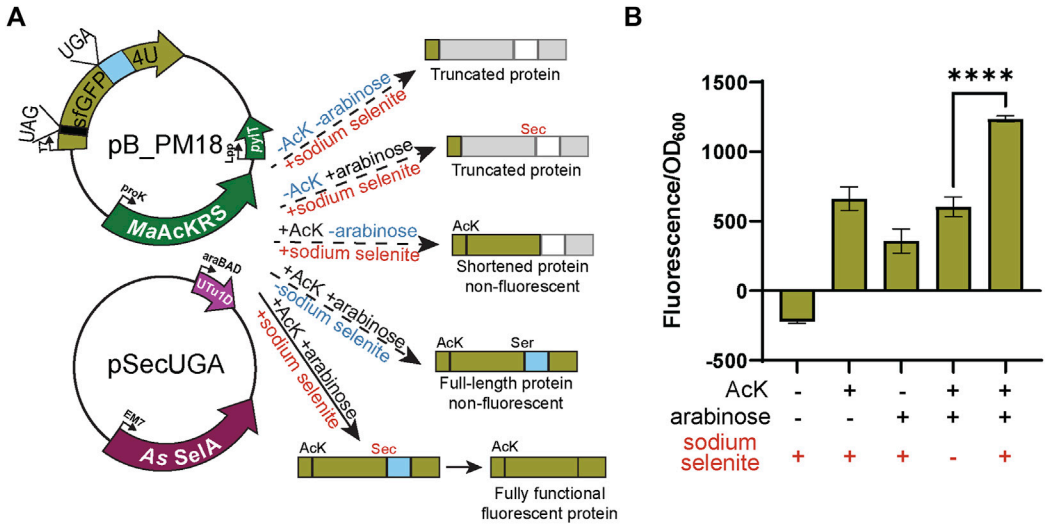


FIGURE 5
Confirmation of Sec insertion using intein reporter. **(A)** The assay was performed with a two-plasmid system: pB_PM15 and pSecUGA. The first plasmid contains the genes for *Ma*AcKRS3-IP and $tRNA^{Pyl}_{CUA}$ (*pylT*) for insertion of N_ε-acetyl-L-lysine (AcK) at a UAG and a modified pB_04 sfGFP intein reporter gene. This gene has been modified from the original (Chung et al., 2022) to encode a UAG at position 2 and a UGA at position 204, the first position of the M86 mini-intein. The second plasmid, pSecUGA, contains machinery for Sec insertion at UGA. In the presence of AcK, arabinose, and sodium selenite, the entire sfGFP gene should be expressed. When Sec is inserted at position 204, the intein will splice out to produce a fluorescent functional reporter. In the absence of sodium selenite alone, the entire sfGFP gene should be expressed except serine (Ser) should be inserted at position 204, which does not induce splicing. This results in a non-functional, non-fluorescent protein. In the absence of AcK or arabinose (or both), translation stops resulting in truncated non-fluorescent protein. Only the first amino acid would be translated without AcK, while without arabinose, translation would stop at 204UGA. **(B)** The sfGFP_intein assay shows successful readthrough of position 2 and splicing of the intein in the presence of AcK, arabinose, and sodium selenite, with a significant increase in fluorescence ($p < 0.0001$) compared to when any one of the components is missing or in the absence of them all. Data are shown as the average of at least four biological replicates with the corresponding standard deviation shown with error bars.

observed significant activity when both suppression systems and sodium selenite (Figure 5B) were present. These results corroborate that the fluorescence observed with the pB_PM17 plasmid (Figure 4B) is due to the presence of Sec. Moreover, the lack of fluorescence in the absence of AcK verifies its requirement and insertion at UAG (Figures 3A, 4B).

10.3389/fmolb.2023.1096261

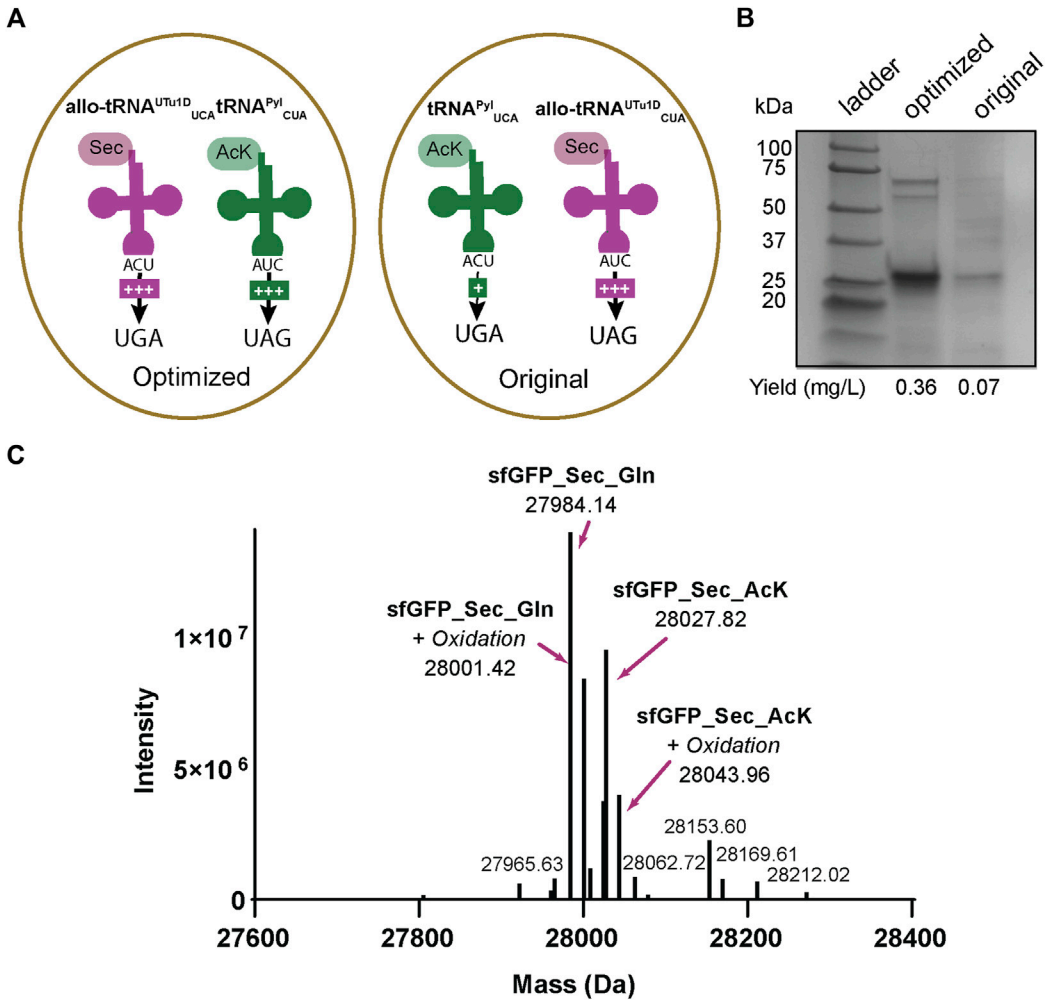


FIGURE 6
Production of AcK-seleno-sfGFP. **(A)** Schematic of tRNAs used and the respective codons they are expected to recode. Optimized conditions involved Sec insertion at position 2UGA with allo-tRNA$^{UTu1D}_{UCA}$ and N_ε-acetyl-L-lysine (AcK) at 151UAG with tRNA$^{Pyl}_{CUA}$. Original conditions involved AcK insertion at position 2UGA with tRNA$^{Pyl}_{UCA}$ and Sec insertion at 151UAG with allo-tRNA$^{UTu1D}_{CUA}$. An increase in the amount of + signs correspond to the increased suppression observed. **(B)** SDS-PAGE shows a strong band around the 25 kDa marker, corresponding to the AcK-seleno-sfGFP size of 28 kDa. An increase in protein yield was found for the anticodon-optimized conditions compared to the original. **(C)** Intact MS data has multiple peaks which can be assigned to AcK-seleno-sfGFP (37%) and Gln-seleno-sfGFP (63%) with or without an oxidation state.

3.3 Expression of post-translationally modified seleno-sfGFP in *E. coli*

The sfGFP readthrough and intein assays demonstrated that AcK and Sec could be inserted into a single protein using two different stop codons. Therefore, we applied our integrated plasmid setup and growth conditions to express and purify acetylated seleno-sfGFP from *E. coli*. The location of the His_6-tag on the C-terminus allowed the separation of full-length protein from truncated versions due to early termination (Figure 5A). This strategy produced a robust band at roughly 25 kDa correlating to our expected protein size (28 kDa) (Figure 6). We also compared our optimized system, which involves recoding UGA with Sec and UAG with AcK, to our original system (recoding UGA with AcK and UAG with Sec) (Figure 6A). In both systems, we obtained pure protein; however, our yields were increased roughly 5-fold with our optimized codon set compared to the original (Figure 6B). Mass spectrometry analysis of the optimized codon set confirmed the presence of acetylated seleno-sfGFP. Protein mass was also detected that corresponded to Gln instead of AcK at a proportion of 67%, suggesting that near-cognate suppression by tRNA^{Gln} can impact the final protein purity (Figure 6C). This was further verified with MS/MS data to show that indeed AcK or Gln was inserted at position 151 while Sec was always observed at position 2 (Supplementary Figure S2).

3.4 Recombinant purification of acetylated human selenoprotein GPx1

To prove the versatility and utility of our system, we produced an authentic selenoprotein, human GPx1. GPx1 is predicted to be acetylated at different Lys residues based on murine GPx1 MS studies (Rardin et al., 2013). We individually targeted K114 and K148 together with the Sec at position 49. The protein was expressed under the same conditions with the same plasmid

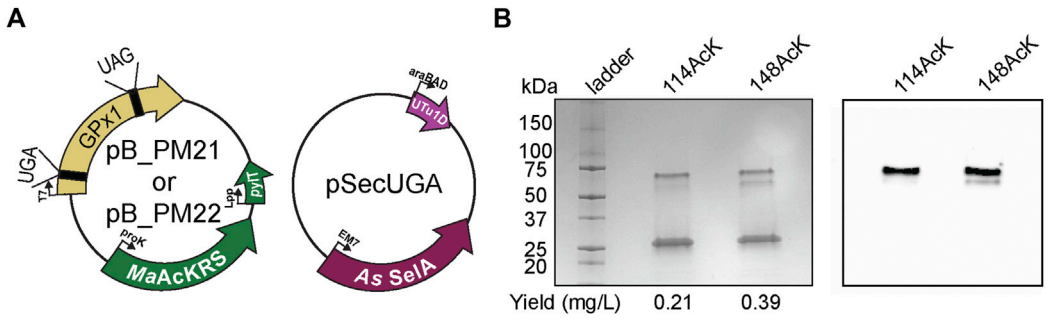


FIGURE 7
Production of post-translationally modified human GPx1. **(A)** Plasmids used for insertion of selenocysteine and N_ε-acetyl-L-lysine (AcK) into GPx1 at UGA and UAG, respectively. Two variants were made with AcK at either position 114 or 148. **(B)** 4%–20% PAGE shows three bands on the gel for each variant, with the strongest one at the 25-kDa marker corresponding to monomeric GPx1 (24 kDa). An increase in protein yield was observed for 148AcK. Western blot with anti-AcK antibody targeted the two higher molecular weight bands.

backbone as described for sfGFP (Figure 7A). Three bands were visible on the polyacrylamide gel, with the lower band corresponding to the expected size of monomeric GPx1 (25 kDa). In terms of protein yield, we found that GPx1_148AcK had roughly twice as much protein as GPx1_114AcK. This may be a result of codon context or the closer proximity of position 114 to 49Sec (Figure 7B). To confirm incorporation of AcK, a western blot with an anti-AcK antibody was used (Figure 7B). The western blot revealed the presence of acetylation in the higher molecular weight band which is approximately 30%–35% of the total protein (Supplementary Figure S3). We surmise that the lower molecular weight band corresponds to a GPx1 variant containing Gln (65%–70%) instead of AcK as suggested by the sfGFP MS data. This provides evidence that imply the potential role of AcK in GPx1 oligomerization, and emphasizes the necessity for a system which can study post-translationally modified selenoproteins.

4 Conclusion

Here we established a method for the simultaneous incorporation of selenocysteine and the ncAA N_ε-acetyl-L-lysine into a protein at two targeted positions in *E. coli*. This method integrates pSec_Evol and MaPylRS:tRNA^{Pyl} non-sense translation systems for the first time (Figure 1). This was accomplished by characterizing the decoding efficiency, specificity, and compatibility of these systems. We found that allo-tRNA^{UTu1D} and *Ma*AcLysRS3:tRNA^{Pyl} are compatible when UGA and UAG are assigned to Sec and AcK, respectively (Figures 3B, C). In contrast, allo-tRNA$^{UTu1D}_{CUA}$ and tRNA$^{Pyl}_{UCA}$ lacked decoding fidelity as they translated both UGA and UAG codons, albeit with different efficiencies (Figures 2B, C). Thus, our results indicate that the codon specificity of these two translation systems should be an essential consideration when developing GCE applications, particularly when combining them with other systems. Another important consideration is the position of the non-sense codons within each protein as mRNA context can influence the suppressor tRNAs as well as the proximity of the two ncAAs. The off-target activity of these tRNAs with non-specific anticodons may hinder the accurate incorporation of the desired amino acid and drastically decrease protein yields by impacting the fitness of the host organism. Using this knowledge, we demonstrated that pSecUGA and *Ma*AcLysRS3:tRNA$^{Pyl}_{CUA}$ systems could be combined in *E. coli* to create a platform for synthesizing proteins containing these two important amino acids. This provides the groundwork to produce acetylated natural or synthetic selenoproteins. The utility of our method can be expanded by combining the pSecUGA system with other existing GCE platforms to facilitate the synthesis of selenoproteins containing other important PTMs or ncAAs with reactive side chains to expand protein functionalization (Young and Schultz, 2018; Shandell et al., 2021). Ultimately, this technology can enable the investigation of PTMs in naturally occurring selenoproteins and the development of artificial proteins endowed with unique chemical properties.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

PM and NK performed the cloning and fluorescence assays. CC, LN, and NK performed protein expression and purification. PM, NK, and OV-R wrote the manuscript. OV-R, CZC, and NK conceptualized the experiments and edited the manuscript. DS and PH edited the manuscript and supported the research.

Funding

This work was supported by grants from the National Institute of General Medical Sciences (R35GM122560-05S1 to DS) and the National Institute of Allergy and Infectious Diseases (5R01AI147496 to PH). Genetic experiments were supported by the Department of Energy Office of Basic Energy Sciences (DE-FG0298ER2031 to DS). CZC holds a Postdoctoral Fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Acknowledgments

We are grateful to Kyle Hoffman (Bioinformatics Solutions Inc.) for intellectual and technical assistance with the mass spectrometry analyses. We also thank the Keck Oligo Synthesis Resource at Yale for assisting with nucleotide synthesis services and Kexin Meng for experimental advice and support.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1096261/full#supplementary-material

References

Aldag, C., Bröcker, M. J., Hohn, M. J., Prat, L., Hammond, G., Plummer, A., et al. (2013). Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. *Angew. Chem. Int. Ed. Engl.* 52 (5), 1441–1445. doi:10.1002/anie.201207567

Appleby-Tagoe, J. H., Thiel, I. V., Wang, Y., Wang, Y., Mootz, H. D., and Liu, X. Q. (2011). Highly efficient and more general cis- and trans-splicing inteins through sequential directed evolution. *J. Biol. Chem.* 286 (39), 34440–34447. doi:10.1074/jbc.M111.277350

Balasuriya, N., Davey, N. E., Johnson, J. L., Liu, H., Biggar, K. K., Cantley, L. C., et al. (2020). Phosphorylation-dependent substrate selectivity of protein kinase B (AKT1). *J. Biol. Chem.* 295 (24), 8120–8134. doi:10.1074/jbc.RA119.012425

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 (10), 1828–1837. doi:10.1021/bi4000244

Chen, H., Venkat, S., McGuire, P., Gan, Q., and Fan, C. (2018). Recent development of genetic code expansion for posttranslational modification studies. *Molecules* 23 (7), 1662. doi:10.3390/molecules23071662

Chung, C. Z., and Krahn, N. (2022). The selenocysteine toolbox: A guide to studying the 21st amino acid. *Arch. Biochem. Biophys.* 730, 109421. doi:10.1016/j.abb.2022.109421

Chung, C. Z., Miller, C., Söll, D., and Krahn, N. (2021). Introducing selenocysteine into recombinant proteins in *Escherichia coli*. *Curr. Protoc.* 1 (2), e54. doi:10.1002/cpz1.54

Chung, C. Z., Krahn, N., Crnković, A., and Söll, D. (2022). Intein-based design expands diversity of selenocysteine reporters. *J. Mol. Biol.* 434 (8), 167199. doi:10.1016/j.jmb.2021.167199

Copeland, P. R., Fletcher, J. E., Carlson, B. A., Hatfield, D. L., and Driscoll, D. M. (2000). A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J.* 19 (2), 306–314. doi:10.1093/emboj/19.2.306

Crnković, A., Suzuki, T., Söll, D., and Reynolds, N. M. (2016). Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion. *Croat. Chem. Acta* 89 (2), 163–174. doi:10.5562/cca2825

Dunkelmann, D. L., Willis, J. C. W., Beattie, A. T., and Chin, J. W. (2020). Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* 12 (6), 535–544. doi:10.1038/s41557-020-0472-x

Hilal, T., Killam, B. Y., Grozdanović, M., Dobosz-Bartoszek, M., Loerke, J., Burger, J., et al. (2022). Structure of the mammalian ribosome as it decodes the selenocysteine UGA codon. *Science* 376 (6599), 1338–1343. doi:10.1126/science.abg3875

Kossinova, O., Malygin, A., Krol, A., and Karpova, G. (2014). The SBP2 protein central to selenoprotein synthesis contacts the human ribosome at expansion segment 7L of the 28S rRNA. *RNA* 20 (7), 1046–1056. doi:10.1261/rna.044917.114

Lacoursiere, R. E., O'Donoghue, P., and Shaw, G. S. (2020). Programmed ubiquitin acetylation using genetic code expansion reveals altered ubiquitination patterns. *FEBS Lett.* 594 (7), 1226–1234. doi:10.1002/1873-3468.13702

Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., et al. (2015). Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci. Rep.* 5, 9699. doi:10.1038/srep09699

Mukai, T., Vargas-Rodriguez, O., Englert, M., Tripp, H. J., Ivanova, N. N., Rubin, E. M., et al. (2017). Transfer RNAs with novel cloverleaf structures. *Nucleic Acids Res.* 45 (5), 2776–2785. doi:10.1093/nar/gkw898

Mukai, T., Sevostyanova, A., Suzuki, T., Fu, X., and Söll, D. (2018). A facile method for producing selenocysteine-containing proteins. *Angew. Chem. Int. Ed. Engl.* 57 (24), 7215–7219. doi:10.1002/anie.201713215

Neumann, H., Slusarczyk, A. L., and Chin, J. W. (2010). De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *J. Am. Chem. Soc.* 132 (7), 2142–2144. doi:10.1021/ja9068722

Novoselov, S. V., Lobanov, A. V., Hua, D., Kasaikina, M. V., Hatfield, D. L., and Gladyshev, V. N. (2007). A highly efficient form of the selenocysteine insertion sequence element in protozoan parasites and its use in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 104 (19), 7857–7862. doi:10.1073/pnas.0610683104

Pinkerton, M. H., and Copeland, P. R. (2016). "Eukaryotic mechanisms of selenocysteine incorporation and its reconstitution *in vitro*," in *Selenium: its molecular biology and role in human health*. Fourth ed (Springer International Publishing), 13–24.

Porter, J. J., and Mehl, R. A. (2018). Genetic code expansion: A powerful tool for understanding the physiological consequences of oxidative stress protein modifications. *Oxid. Med. Cell Longev.* 2018, 7607463. doi:10.1155/2018/7607463

Prabhakar, A., Krahn, N., Zhang, J., Vargas-Rodriguez, O., Krupkin, M., Fu, Z., et al. (2022). Uncovering translation roadblocks during the development of a synthetic tRNA. *Nucleic Acids Res.* 50 (18), 10201–10211. doi:10.1093/nar/gkac576

Rardin, M. J., Newman, J. C., Held, J. M., Cusack, M. P., Sorensen, D. J., Li, B., et al. (2013). Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* 110 (16), 6601–6606. doi:10.1073/pnas.1302961110

Seki, E., Yanagisawa, T., Kuratani, M., Sakamoto, K., and Yokoyama, S. (2020). Fully productive cell-free genetic code expansion by structure-based engineering of *Methanomethylophilus alvus* pyrrolysyl-tRNA synthetase. *ACS Synth. Biol.* 9 (4), 718–732. doi:10.1021/acssynbio.9b00288

Shandell, M. A., Tan, Z., and Cornish, V. W. (2021). Genetic code expansion: A brief history and perspective. *Biochemistry* 60 (46), 3455–3469. doi:10.1021/acs.biochem.1c00286

Tharp, J. M., Ad, O., Amikura, K., Ward, F. R., Garcia, E. M., Cate, J. H. D., et al. (2020). Initiation of protein synthesis with non-canonical amino acids *in vivo*. *Angew. Chem. Int. Ed. Engl.* 59 (8), 3122–3126. doi:10.1002/anie.201914671

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

Venkat, S., Sturges, J., Stahman, A., Gregory, C., Gan, Q., and Fan, C. (2018). Genetically incorporating two distinct post-translational modifications into one protein simultaneously. *ACS Synth. Biol.* 7 (2), 689–695. doi:10.1021/acssynbio.7b00408

Wang, K., Sachdeva, A., Cox, D. J., Wilf, N. M., Lang, K., Wallace, S., et al. (2014). Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* 6 (5), 393–403. doi:10.1038/nchem.1919

Wright, D. E., Altaany, Z., Bi, Y., Alperstein, Z., and O'Donoghue, P. (2018). Acetylation regulates thioredoxin reductase oligomerization and activity. *Antioxid. Redox Signal* 29 (4), 377–388. doi:10.1089/ars.2017.7082

Yang, A., Cho, K., and Park, H. S. (2018). Chemical biology approaches for studying posttranslational modifications. *RNA Biol.* 15 (4-5), 427–440. doi:10.1080/15476286.2017.1360468

Ye, R., Huang, J., Wang, Z., Chen, Y., and Dong, Y. (2022). The role and mechanism of essential selenoproteins for homeostasis. *Antioxidants (Basel)* 11 (5), 973. doi:10.3390/antiox11050973

Young, D. D., and Schultz, P. G. (2018). Playing with the molecules of life. *ACS Chem. Biol.* 13 (4), 854–870. doi:10.1021/acschembio.7b00974

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 07 July 2023
DOI 10.3389/fmolb.2023.1214235

OPEN ACCESS

EDITED BY
Laurent Roberto Chiarelli,
University of Pavia, Italy

REVIEWED BY
Jeffery M. Tharp,
Indiana University Bloomington, United States
Richard Cooley,
Oregon State University, United States
Julia Shifman,
Hebrew University of Jerusalem, Israel

*CORRESPONDENCE
Gustavo Fuertes,
gustavo.fuertes@ibt.cas.cz,
gustavo.fuertesvives@gmail.com

RECEIVED 29 April 2023
ACCEPTED 27 June 2023
PUBLISHED 07 July 2023

CITATION
Pham PN, Zahradník J, Kolářová L, Schneider B and Fuertes G (2023), Regulation of IL-24/IL-20R2 complex formation using photocaged tyrosines and UV light.
Front. Mol. Biosci. 10:1214235.
doi: 10.3389/fmolb.2023.1214235

Regulation of IL-24/IL-20R2 complex formation using photocaged tyrosines and UV light

Phuong Ngoc Pham[1,2], Jiří Zahradník[3,4], Lucie Kolářová[1], Bohdan Schneider[1] and Gustavo Fuertes[1]*

[1]Laboratory of Biomolecular Recognition, Institute of Biotechnology of the Czech Academy of Sciences, Vestec, Czechia, [2]Faculty of Science, Charles University, Prague, Czechia, [3]First Faculty of Medicine, BIOCEV Center, Charles University, Prague, Czechia, [4]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

Human interleukin 24 (IL-24) is a multifunctional cytokine that represents an important target for autoimmune diseases and cancer. Since the biological functions of IL-24 depend on interactions with membrane receptors, on-demand regulation of the affinity between IL-24 and its cognate partners offers exciting possibilities in basic research and may have applications in therapy. As a proof-of-concept, we developed a strategy based on recombinant soluble protein variants and genetic code expansion technology to photocontrol the binding between IL-24 and one of its receptors, IL-20R2. Screening of non-canonical *ortho*-nitrobenzyl-tyrosine (NBY) residues introduced at several positions in both partners was done by a combination of biophysical and cell signaling assays. We identified one position for installing NBY, tyrosine70 of IL-20R2, which results in clear impairment of heterocomplex assembly in the dark. Irradiation with 365-nm light leads to decaging and reconstitutes the native tyrosine of the receptor that can then associate with IL-24. Photocaged IL-20R2 may be useful for the spatiotemporal control of the JAK/STAT phosphorylation cascade.

KEYWORDS

protein-protein interactions (PPI), interleukin-24, cytokines, optobinders, genetically encoded non-canonical amino acids (ncAA), photocaged proteins, ortho-nitrobenzyltyrosine (NBY), photoxenoprotein engineering

1 Introduction

IL-24 is a multifunctional cytokine playing key roles in immune response, host defense, tissue homeostasis, and cell proliferation (Ma et al., 2011; Rutz et al., 2014; Liu et al., 2023). It is a member of a broader family of IL-10 related cytokines including IL-10, IL-19, IL-20, IL-22, IL-24, IL-26, IL-28 and IL-29 (Akdis et al., 2011). An increase in the expression levels of IL-24 is connected to autoimmune diseases, such as psoriasis (Kragstrup et al., 2008), inflammatory bowel disease (Andoh et al., 2009), and rheumatoid arthritis (de Melo et al., 2012). Moreover, a large number of studies suggest anticancer properties for IL-24, such as stimulation of apoptosis and autophagy, or inhibition of angiogenesis, invasion and metastasis (Menezes et al., 2018). Given the anti-oncogenesis effects of IL-24, it became a pharmacological target and even reached phase I clinical trials (Cunningham et al., 2005). Accumulated evidence strengthened the concept of the "bystander effect" according to which secreted IL-24, either in normal or cancer cells, induces tumor apoptosis in presence of IL-20/IL-22 receptors (Su et al., 2005).

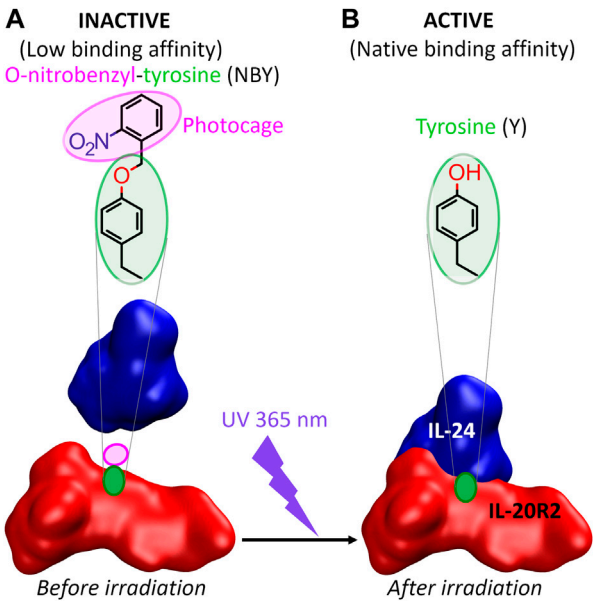


FIGURE 1
The concept of this study. **(A)** The association between IL-24 and IL-20R2 is diminished upon genetic encoding of a photocaged tyrosine (ortho-nitrobenzyl-tyrosine or NBY) into one of the two binding partners (IL-20R2 in the figure). The non-canonical amino acid must be introduced at a position, *a priori* unknown, which either blocks essential protein-protein interactions (e.g., hydrogen bonds) or causes steric clashes preventing a correct configuration of the interaction interface. **(B)** Irradiation with UV light (365 nm wavelength) releases the cage and reconstitutes a canonical tyrosine residue (and thus restores the native protein) that can now recognize the binding partner with the same affinity as the wild-type complex.

Membrane-bound receptors for IL-24 comprise IL-22R1, IL-20R1, and IL-20R2. IL-24/receptor interactions are mediated by the extracellular domains of these receptors. IL-24 signals through both the IL-20 receptor (IL-20R1/IL-20R2) and IL-20R1/IL-20R2 heterodimers (Dumoutier et al., 2001; Parrish-Novak et al., 2002; Wang et al., 2002). Upon binding and heterotrimer formation, IL-24 triggers a signaling cascade in the target cells through Janus Kinase/Signal Transducer and Activator of Transcription (JAK/STAT) pathway, which subsequently activates downstream transcription factors, like STAT3 and STAT1, through phosphorylation (Wang and Liang, 2005).

Since many physiological and pathophysiological roles of IL-24 critically depend on its interaction with cognate receptors (Wang et al., 2002; Wang et al., 2004), we hypothesize that controlling the binding between IL-24 and the shared receptor IL-20R2 could be useful for both basic research and therapeutics. The use of light to switch ON and OFF protein-protein interactions offers superior capabilities in terms of temporal and spatial resolution (Hoorens and Szymanski, 2018). We speculate that assays based on photocontrolled interleukins and/or their cognate receptors may be applied to the detection of specific interleukins, the screening of anti-interleukin/anti-receptor antibodies, or the screening of JAK/STAT inhibitors.

From the molecular engineering point of view, there are two major approaches to control protein functions by light (Kneuttinger, 2022). The first is hybrid protein optogenetics, where a chimera between an intrinsically light-responsive protein and a target protein is made. The second is photoxenoprotein engineering, where the target protein is modified at defined residues with light-responsive non-canonical amino acids (ncAA). NcAA can be incorporated into proteins by genetic code expansion technology through an orthogonal aminoacyl-tRNA synthetase (aaRS)/$tRNA_{CUA}$ pair (Manandhar et al., 2021). Several photocontrolled ncAA have been genetically encoded in *Escherichia coli* and mammalian cells, including analogs of tyrosine, lysine, cysteine, serine and histidine (Wu et al., 2004; Deiters et al., 2006; Lemke et al., 2007; Gautier et al., 2010; Arbely et al., 2012; Nguyen et al., 2014; Hauf et al., 2017; Luo et al., 2017; Baumann et al., 2019; Cheung et al., 2023). We choose to photocage tyrosines in the form of *ortho*-nitrobenzyl-tyrosine (NBY) for two reasons. First, NBY has been proposed as a universal proximal cage for the temporal blockage of protein activity (Wang et al., 2019; Wang et al., 2021). Second, NBY has been successfully employed to photocontrol antigen-antibody interactions (Bridge et al., 2019; Jedlitzke et al., 2019; Joest et al., 2021; Jedlitzke and Mootz, 2022; O'Shea et al., 2022; Yilmaz et al., 2022; Bridge et al., 2023).

Thus, as a first step towards the photocontrol *in vivo* of cell signaling pathways that rely on IL-24-mediated complex formation, we designed and successfully tested the following *in vitro* approach with purified proteins recombinantly expressed in *E. coli* (Zahradník et al., 2019) (Figure 1). Incorporation of photocaged NBY residues at positions critical for molecular recognition into either IL-24 or IL-20R2 would keep the protein partners incompetent for binding and unable to convey the signal. Following the removal of the cage by UV irradiation, the native protein structure and activity (binding and signaling) would be restored resulting in complex formation. In the future, the ability to switch OFF the binding between IL-24 and IL-20R2 in the dark and switch it ON by UV light may find applications in immunology and possibly also in therapeutics.

2 Materials and methods

2.1 Miscellaneous chemicals

pEVOL-NBY was a gift from Dr. Petr Schultz (Deiters et al., 2006). NBY was purchased from Accela. For a more complete list of reagents, please refer to Supplementary Table S0.

2.2 Gene cloning

An engineered IL-24 variant named IL-24B4 had been sequence-optimized to achieve soluble expression in *E. coli* following structural bioinformatics design (Zahradník et al., 2019). This sequence contains 29 mutations relative to wild-type IL-24 (Supplementary Table S1, residue numbering according to UniprotKB Q13007). A pHisSumo vector containing such interleukin-24 variant with an N-terminal HisSumo tag (Zahradník et al., 2019) was used as the starting material for cloning. We used the primers 1 and 2 (Supplementary Table S2) to sandwich the IL-24 gene between an N-terminal HisSumo-tag and a

C-terminal Strep-tag. The final construct was labeled as pHisSumo_IL-24B4_CStrep (Supplementary Figure S1A).

An IL-20R2 variant sequence, referred to as IL-20R2D, which was also optimized to be expressed and soluble in *E. coli*, contained 23 mutations relative to the wild-type (Supplementary Table S1 residue numbering according to UniprotKB Q6UXL0). The design details can be found in the main text and Supplementary Note S1. The ordered DNA string consists of a HisSumo tag and a Strep tag flanking IL-20R2 at the N-terminus and C-terminus, respectively. The final construct was labeled as pHisSumo_IL-20R2D_CStrep (Supplementary Figure S1B).

TAG triplets to incorporate non-canonical amino acids were introduced in the plasmid sequences by standard site-directed mutagenesis using primers 5 to 11 (Supplementary Table S2).

2.3 Protein expression

To express proteins with canonical amino acids, the pHisSumo_IL-24B4_CStrep and pHisSumo_IL-20R2D_CStrep plasmids were transformed into *E. coli* BL21 AI strain. The bacterial cells were grown in TB medium containing 1.2% tryptone, 2.4% yeast extract, 0.5% glycerol, 33.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.55 mM NaCl, 18.7 mM NH_4Cl, 2 mM $MgSO_4$/$MgCl_2$, and metal mix. The antibiotic kanamycin was supplied to the medium and cells were grown at 30°C with shaking at 200 rpm. When bacterial OD600 reached 0.6 to 0.8, cells were induced with IPTG and arabinose at 0.1 mM and 0.02% v/v final concentration, respectively, and grown at 16°C for another 20 h.

For proteins containing non-canonical amino acids, plasmids encoding the proteins-of-interest were co-transformed in BL21 AI with pEVOL-NBY, a plasmid encoding a tyrosyl tRNA synthetase derived from *Methanocaldococcus jannaschii* specific for *ortho*-nitrobenzyl-tyrosine (*Mj*NBYRS) and its cognate suppressor tRNA ($tRNA_{CUA}$) (Deiters et al., 2006).

Bacteria were grown at 30°C and shaking at 200 rpm in TB medium described above including antibiotics kanamycin and chloramphenicol. During culturing, we prepared a 50 mM stock of NBY by resuspending the amino acid in 50% DMSO heated at 70°C and adding NaOH dropwise until a homogeneous solution was obtained. When bacterial OD600 reached between 0.4 and 0.5, NBY was added to the culture at a final concentration of 1 mM. After 25 more minutes, the temperature was decreased to 16°C, and cells were induced by both IPTG and arabinose at 0.1 mM and 0.02% final concentration, respectively. Overnight-grown bacteria (20 h at 16°C) were separated from the medium culture by centrifugation at 5,000 *g* for 10 min at 4°C. All cell pellets were stored at −80°C until needed.

2.4 Protein purification

The cell pellet was suspended in cold washing buffer 50 mM Tris, pH = 8, 100 mM NaCl and sonicated 2 s on, 2 s off during 5 min with 50 W power on ice. The fraction of soluble proteins was collected by centrifuging at 40 000 g for 20 min at 4°C and the supernatant was passed over streptactin XT agarose beads equilibrated at room temperature by 50 mM Tris, pH = 8, 150 mM NaCl, 1 mM EDTA. After brief incubation for the proteins containing Strep tag to bind to the beads, the unbound proteins were washed with the same buffer. Commercial elution buffer BXT (100 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA, 50 mM biotin) was used to elute the bound Strep-tagged proteins. All the eluted proteins were further purified by size-exclusion chromatography (SEC) using a Superdex 75 Increase 10/300 GL column (GE Healthcare, Chicago, IL, United States) equilibrated with buffer 50 mM Tris, pH = 8, 100 mM NaCl.

IL-24B4 was digested with Sumo protease to remove the N-terminal HisSumo tag by at a ratio 1:200 (protease:protein) at room temperature overnight (16 h). The digestion mixture was passed over nickel-NTA agarose beads. Uncleaved IL-24 and the cleaved HisSumo fragment were retained on the beads, while cleaved IL-24 without the N-terminal HisSumo tag was recovered from the flow-through. Cleaved IL-24 was subsequently purified by SEC as described before, pooled and concentrated.

2.5 Protein characterization

The quality and quantity of all protein samples was checked by electrophoresis, UV/Visible spectroscopy, and mass spectrometry (de Marco et al., 2021). Proteins were analyzed by 15% sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) and stained with Coomassie Blue. UV/Vis spectra were taken in a Nanodrop spectrophotometer (DeNovix) using 10 mm pathlegnth cuvettes. The protein extinction coefficient at 280 nm, estimated from the sequence using web server (https://web.expasy.org/protparam/) was used to calculate protein concentration based on Beer–Lambert law.

For mass spectrometry analyses, proteins were diluted with 100 µL of 5% acetic acid in water and loaded onto Opti-trap C4 cartridge (Optimize Technologies), washed 4 times with 250 µL of 5% acetic acid in water and eluted with 100 µL of 80% acetonitrile, 5% acetic acid. Proteins were analyzed by direct infusion using syringe pump at a flow rate 3 µL/min connected with an electrospray ion source of TimsTOF Pro mass spectrometer (Bruker Daltonics). The mass spectrometer was externally calibrated using 0.1% (w/w) sodium trifluoracetate. Proteins were measured in positive mode. The data were processed using SNAP algorithm–a part of DataAnalysis 5.3 software (Bruker Daltonics). Deconvoluted spectra were generated using the UniDec 5.0.2 software (University of Oxford) (Marty et al., 2015). IL-20R2D and IL-20R2D Y70NBY were also measured on a high resolving power and high accuracy 15T SolariX XR FT-ICR mass spectrometer (Bruker Daltonics), Expected protein masses (average and monoisotopic) were calculated with 2M data acquisition, by submitting the protein sequences to the website https://web.expasy.org/compute_pi/.

The circular dichroism spectra were measured using a 0.1 cm quartz cuvette on ChirascanTM-plus spectrometer (Applied Photophysics). Proteins were diluted in buffer (Tris 50 mM NaCl 100 mM pH = 8.0) to final concentration of ~0.15 mg/mL. The data were recorded in the range of wavelengths of 190–260 nm with a step of 1 nm and time-per-point of 1 s at room temperature. The resulting spectra were water-subtracted and normalized to the concentration of the corresponding sample and are given as molar circular dichroism (Δε) vs wavelength. The data were

processed by CDNN 2.1 software and BestSel (https://bestsel.elte.hu/index.php) (Micsonai et al., 2022).

Label-free protein unfolding assays were done by detecting the temperature-dependent change in tryptophan fluorescence at emission wavelengths of 330 and 350 nm (Alexander et al., 2014). Melting temperatures were determined from the maximum of the first derivative of the fluorescence ratios (F350/F330). For thermal unfolding experiments, the proteins were diluted to a final concentration of ~0.15 mg/mL. The samples were loaded into UV capillaries (NanoTemper Technologies) and experiments were carried out using the Prometheus NT.48.

2.6 UV-decaging

200 µL of the proteins at 200 nM concentration were transferred to 1 mm path-length cylindrical cells (Hellma) and irradiated at 365 nm with an LED (Thorlabs M365L2) attached to a collimator (Thorlabs SM2F32-A) at 100 mW of power for 5 min at room temperature. Irradiated proteins were used immediately.

2.7 Affinity measurements by MST

We determine the binding affinity (dissociation constant) by MicroScale Thermophoresis (MST) using a Monolith NT.115 instrument (NanoTemper Technologies) (Jerabek-Willemsen et al., 2014). 200 nM of IL-20R2D, IL-20R2D Y70NBY, IL-20R2D Y74NBY, or IL-20R2D Y70NBY/Y74NBY were labeled with an NTA-conjugated fluorescent dye as recommended by the kit His-Tag Labeling RED-tris-NTA second generation. Binding assays were performed in buffer 50 mM Tris, pH = 8, 100 mM NaCl, 0.02% pluronic F127. The instrument settings were 20% MST power and 40% LED power. Data from triplicate experiments were fitted to mass-action kinetics by using the NT Analysis software version 1.5.41:

$$F(c) = A + B\left(\frac{c + c_T + K_d - \sqrt{(c + c_T + K_d)^2 - 4\,c\,c_T}}{2\,c_T}\right) \quad (1)$$

Where $F(c)$ is the fluorescence signal as a function of concentration; c is the concentration of IL-24B4 (unlabeled) c_T is the concentration of IL-20R2D variants (fluorescently labeled), which was held constant at 50 nM; and K_d is the dissociation constant; A is the titration curve baseline; B is the titration response range. To facilitate the comparison between datasets with distinct values of A and B, the fraction of bound molecules (normalized between zero and one) was calculated as:

$$Fraction\,bound = \frac{F(c) - A}{B} \quad (2)$$

2.8 Affinity measurements by yeast display

Yeast surface display of IL-24B4 and subsequent binding experiments were performed by using the enhanced yeast display platform pJYDC1 plasmid (Addgene ID: 162458) and *Saccharomyces cerevisiae* EBY100 strain (Zahradník et al., 2021a). The bait, an IL-24B4 version including an N-terminal linker of 17 amino acids, was cloned between *Nde*I and *Bam*HI sites (see full construct sequence in Supplementary Table S1), and transformed into yeast by lithium acetate method. Protein expression was achieved by cultivation in galactose-rich expression media 1/9 at 20°C for 24 h (Zahradník et al., 2021a). Prey proteins (IL-20R2D, IL-20R2D Y70NBY and their UV irradiated versions) were labeled by using CF®640R succinimidyl ester dye (Biotium) in 1:3 M excess.

The binding affinities of labeled IL-20R2D versions to IL-24B4 expressed on yeast surface were determined by flow cytometry titration experiments. We used 5 different concentrations (500, 100, 20, 4, and 0.1 nM) of labeled prey protein in PBS buffer supplemented by 1 g/L of Bovine Serum Albumin Fraction V (PBSB buffer, Sigma-Aldrich, Cat#10735078001), and 10 nM bilirubin (Sigma-Aldrich, Cat# 14370) for activation of eUnaG2 fluorescence (excitation at 498 nm, emission at 527 nm). To increase the control over the measurement, we enriched the suspension by addition of 0.5% of IL-24S, an engineered clone featuring 300 times tighter binding, which was obtained by yeast display affinity maturation (Supplementary Table S1 and Supplementary Figure S2).

Binding suspensions were incubated overnight at 4°C, washed twice with ice-cold PBSB and analyzed by using CytoFLEX S Flow Cytometer (Beckman Coulter, United States, Cat#. N0-V4-B2-Y4). The gating and analysis strategies were described previously (Zahradník et al., 2021a; Zahradník et al., 2021b). Briefly, the unspecific binding signals were subtracted from the binding signals and a Eq. 1 was fitted to the data by nonlinear least-squares regression using Python v3.7 (https://www.python.org).

2.9 JAK/STAT signaling assays in human cells

HeLa cells were gown in a media containing Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 1% of PenStrep and 1% of sodium pyruvate. To ensure sufficient amounts of the receptors IL-20R2 and IL-22R1, two plasmids encoding the native sequences (Supplementary Table S1) were constructed (oligos 12 to 15 of Supplementary Table S2) and co-transfected in HeLa cells (0.66 µg of IL-20R2 and 1.33 µg of IL-22R1) using the jetPrime transfection reagent (Polyplus 114–07) following manufacturer's instructions. The next day, the cells were treated for 30 min with the obtained interleukin variants at the indicated concentrations. After the incubation period, the plates were put on ice, the media discarded, and cells were washed with cold PBS buffer. Then, 100 µL of RIPA lysis buffer (Tris 50 mM pH = 8.0, NaCl 150 mM, EDTA 5 mM, IGEPAL CA-630 1%, sodium deoxycolate 10%, SDS 10%) supplemented with protease and phosphatase inhibitors was added to each well. Cells were scraped and put into Eppendorf tubes for 20 min on ice, followed by centrifugation at 4°C, 18,000 g for 30 min. Protein concentration in the supernatant fraction was quantified by the BCA assay (QPRO BCA kit from Cyanagen) using BSA as the standard. Equal protein amounts were loaded on SDS-PAGE gels. Western blotting to nitrocellulose membranes was done in a Transblot Turbo Bio-rad system. For immunoblotting we use two primary antibodies:

anti-phospho-STAT3 (Tyr705, D3A7) monoclonal antibody raised in rabbit, and anti *a*-tubulin monoclonal antibody raised in mouse (Sigma). As secondary antibodies we used anti-mouse or anti-rabbit peroxidase-conjugated IgG. One milliliter of peroxidase substrate was added to the membranes and the resulting chemiluminiscence was imaged using an Azure400 system. All original Western blot images can be found in the Supplementary Zip file. Band quantification was done in ImageJ (https://imagej.nih.gov/ij/index.html) as previously described (Stael et al., 2022). The band intensities of pSTAT3 were divided by the corresponding band intensities of tubulin and subtracted from a control experiment (untreated cells). Normalized and background-subtracted band intensities were plotted as a function of the concentration of added variant and a dose-response curve was fitted to the data:

$$\frac{pSTAT3}{tubulin} = \frac{Y_{max}}{1 + (EC_{50}/c)^{n_H}} \quad (3)$$

Where Y_{max} is the maximum observed signal, EC_{50} is the half maximal effective concentration, c is the concentration of added interleukin (IL-24B4, IL-20R2D and their NBY counterparts), and n_H is the Hill coefficient (which was fixed to 1, i.e., non-cooperative binding, to prevent overfitting due to the reduced number of points).

3 Results

3.1 Protein design combining canonical and non-canonical amino acid mutagenesis

In vitro testing of photo-induced changes in binding affinity requires pure and active protein partners. We designed engineered variants that are more amenable to downstream manipulations and may be used as surrogates of wild-type proteins after a scrutiny of their performance.

3.1.1 Enhancing solubility and stability with canonical amino acids

In our previous work, we successfully optimized the sequence of IL-24 based on canonical amino acid mutagenesis to achieve soluble expression in *E. coli* hosts (Zahradník et al., 2019). Here we have applied a similar approach to the high affinity receptor of IL-24, the IL-20R2. In order to obtain an stabilized soluble extracellular portion of IL-20R2 protein, we employed PROSS algorithm (Goldenzweig et al., 2016) (Supplementary Note S1, Supplementary Table S1, Supplementary Figure S2, and Supplementary Figure S3A). We then tested the expression of our newly designed IL-20R2D in *E. coli* BL21 (DE3) under the control of T7 promoter and compared it to IL-20R2 wild-type under the same conditions. Unlike the wild-type gene, our construct is expressed as a soluble protein in large quantities ~20 $mg.L^{-1}$ (Supplementary Figure S3B). Subsequent CD and melting measurements showed evidence of a folded protein (Supplementary Figure S3C) with a melting temperature of ~45°C (Supplementary Figure S3D). The protein showed considerable stability and retention of binding to IL-24B4 (Zahradník et al., 2019).

3.1.2 Adding photocontrol with photocaged ncAA

As mentioned before, we chose to photocage tyrosine residues with the o-nitrobenzyl moiety. We then looked for tyrosine residues at the binding interface between IL-24B4 and IL-20R2D. Since no high-resolution structure exists of the complex between the variants used in the present study, we used the X-ray crystal structure of the homologous native ternary complex (IL-24/IL-22R1/IL-20R2, PDB ID: 6DF3) (Lubkowski et al., 2018) (Figure 2A). The structure suggests that the main contribution to complex formation are the contacts between IL-24 and IL-20R2. The IL-24/IL-20R2 interface is shown schematically in 2D (Supplementary Figure S4A). The interface area covers ~900 $Å^2$ and is stabilized by several forces including 13 hydrogen bonds (H-bonds), 2 potential salt bridges, and a number of polar, water-mediated and hydrophobic contacts. We identified three tyrosine candidates: one in IL-24 (Y204), and two in IL-20R2 (Y70 and Y74). The phenolic hydroxyl group of tyr70 in IL-20R2 makes a hydrogen bond (H-bond) with the ε-amino group of lys135 in IL-24 (Figure 2B). The sidechain OH of tyr74 in IL-20R2 is H-bonded to the backbone CO group of leu117 in IL-24 (Figure 2C). Finally, tyr204 in IL-24 makes Van der Waals interactions with lys210 of IL-20R2 (Figure 2D) and is also involved in H-bond interactions with IL-22R1 (Supplementary Figures S4B–E). Considering all these data, we prepared three single mutants: IL-24B4 Y204NBY, IL-20R2D Y70NBY and IL-20R2D Y74NBY. In addition, we prepared the double mutant of IL-20R2D Y70NBY/Y74NBY.

3.2 Production and characterization of photocaged interleukins and receptors

For the design of the final constructs, we considered one additional phenomenon. Codon decoding by orthogonal aminoacyl-tRNA synthetase/$tRNA_{CUA}$ pairs is not 100% efficient, i.e., protein translation can also terminate at the introduced codon resulting in a mixture of truncated and full-length proteins that can be difficult to separate. To avoid such a problem, we added a C-terminal Strep-tag. All our final constructs express target proteins containing a HisSumo tag at the N-terminus and a Strep-tag at the C-terminus. First, we checked whether addition of the Strep tag alters the yield of IL-24B4 and IL-20R2D. The purity was assessed by SDS-PAGE (Figure 3A). In the case of IL-24B4, the HisSumo tag was removed to avoid interference with downstream affinity measurements by microscale thermophoresis (MST). This is because for MST, the His-tagged target protein is non-covalently labeled with a NTA-conjugated fluorescent dye. Mixing with another protein containing a His-tag will cause the dye to re-equilibrate between the two binding partners thus confounding the experiment. Sumo-free IL-24B4 is shown in Figure 3A (right panel).

To obtain proteins with NBY incorporated at specific positions, an evolved orthogonal aminoacyl tRNA-synthetase (*Mj*NBYRS) and suppressor tRNA ($tRNA_{CUA}$) pair was added to the native bacterial translation system (Deiters et al., 2006). Moreover, the bacteria cells were fed with the dissolved non-canonical amino acid NBY in TB culture medium. The expressed full-length non-canonical receptors were purified by one-step Strep-tag procedure as done for the parental proteins. In addition, the negative controls without NBY

10.3389/fmolb.2023.1214235

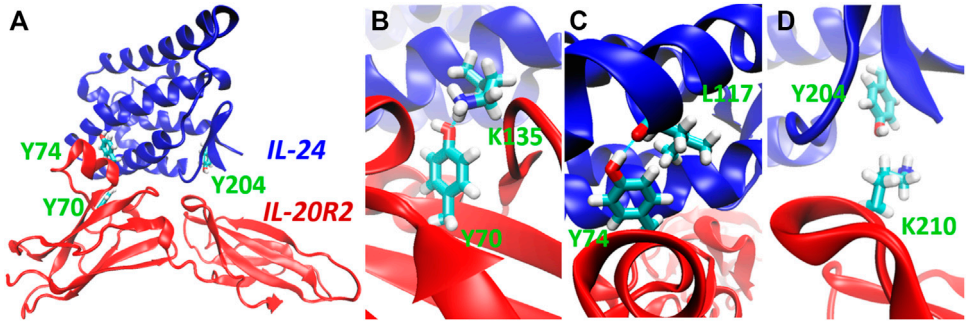


FIGURE 2
Selection of target residues for substitution with photocaged counterparts. **(A)** 3D model of the complex between IL-24 (blue) and IL-20R2 (red) based on X-ray crystallography (PDB 6DF3). The three chosen tyrosines (tyrosine70, tyrosine74, and tyrosine204) are highlighted in green. **(B)** Close-up view of the H-bond between tyrosine70 (IL-20R2) and lysine135 (IL-24). **(C)** Close-up view of the H-bond between tyrosine74 (IL-20R2) and leucine117 (IL-24). **(D)** Close-up view of Van der Waals contact between lysine210 (IL-20R2) and tyrosine204 (IL-24).

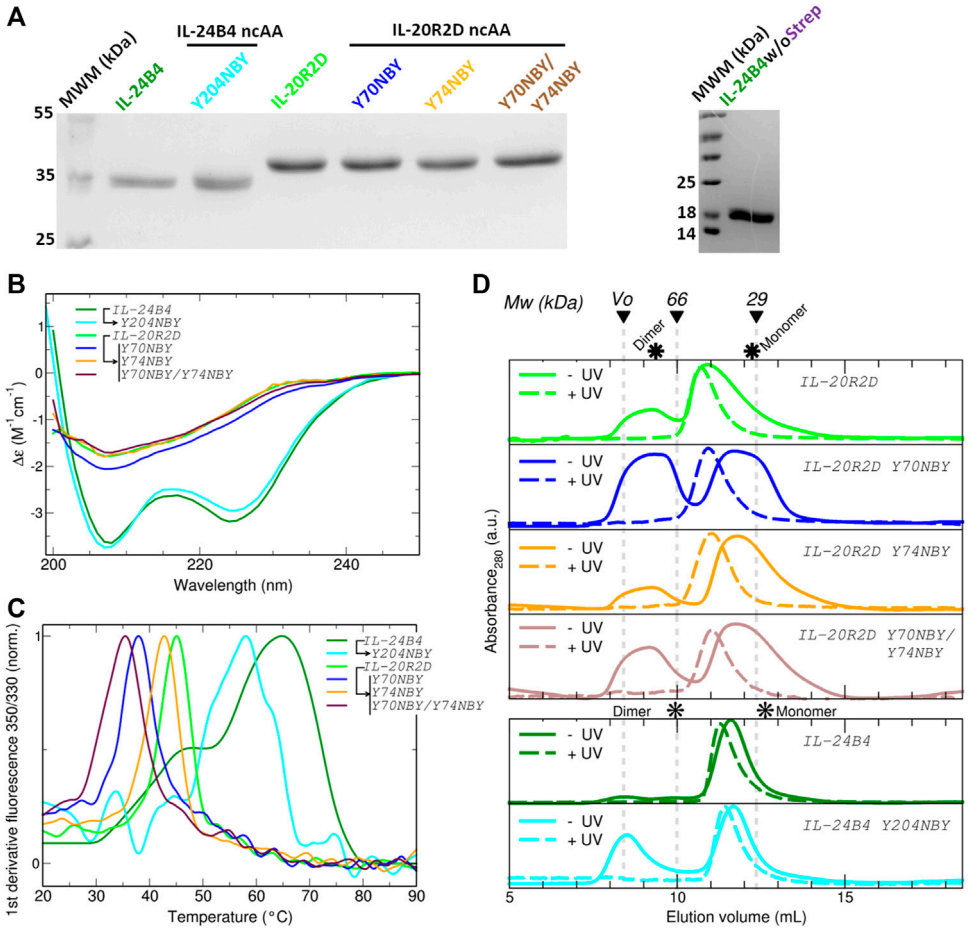


FIGURE 3
Expression, purification and characterization of parental and caged interleukins. **(A)** SDS-PAGE of all purified variants fused to SUMO (*left* panel) and SUMO-cleaved IL-24B4 (*right* panel). **(B)** Circular dichroism spectra expressed as per residue molar absorption (Δε). **(C)** Thermal denaturation curves expressed as the normalized first derivative of the fluorescence emission ratio between 350 nm and 330 nm. **(D)** Size-exclusion chromatograms of all studied variants before (solid lines) and after (dashed lines) UV irradiation. The molecular weight of gel filtration markers is shown on top. The asterisks indicate the expected elution volumes for both monomeric and dimeric assemblies. Data in panels **(B)**, **(C)** and **(D)** were taken in Tris 50 mM Nacl 100 mM pH = 8.

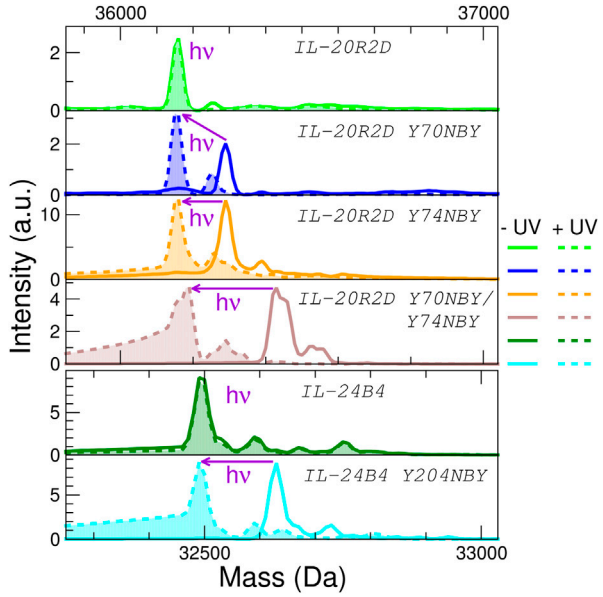


FIGURE 4
Analysis of the effect of protein UV irradiation by mass spectrometry. Deconvoluted mass spectra of the six interleukin variants used in this study before (solid lines) and after (dashed lines with filled areas) exposure to light (λ = 365 nm, 5 min at 100 mW) No significant mass changes are observed for the parental proteins (IL-20R2D and IL-24B4) upon UV irradiation. NBY-containing IL-20R2D and IL-24B4 are converted back to IL-20R2D and IL-24B4, respectively, by UV light. Average masses can be found in Supplementary Table S3.

supplement were performed for size comparison. Clear bands at the expected molecular weight were found in the case IL-24B4 Y204NBY (Supplementary Figure S5A) IL-20R2D Y70NBY, IL-20R2D Y74NBY (Supplementary Figure S5B) in the presence of ncAA. On the contrary, for the samples without added NBY there were essentially no proteins (Supplementary Figure S5). These results suggest that the ncAA NBY was incorporated into the interleukins and their receptors at selected residues only in *E. coli* cells containing both the non-canonical amino acid and the orthogonal translation machinery.

To check whether NBY was indeed present in the purified polypeptides, mass spectrometry was employed (Figure 4). The experimentally determined masses are very close to the theoretical masses (plus/minus 2 Da) (Supplementary Table S3). We did not detect mis-incorporation of phenylalanine residues or reduction of the nitro group, two common issues associated to the genetic encoding of NBY (Koch et al., 2022). These results confirm unambiguously that the non-canonical amino acid NBY was introduced into the target proteins at high levels.

We further characterized the obtained protein variants by circular dichroism (CD) and differential scanning fluorimetry (DSF). The CD spectra of the IL-24B4 variants showed two minima at ~208 nm and ~224 nm (Figure 3B). In opposition, the CD spectra of the IL-20R2D variants featured a single minimum around 207 nm (Figure 3B). Analysis of secondary structure content revealed that IL-24B4 variants were enriched in α-helices while IL-20R2D mutants contained predominantly antiparallel ß-sheets (Supplementary Table S4) in agreement with their crystal structures. Analysis of melting temperatures by DSF indicated significant differences in thermal stability among the mutants (Figure 3C). IL-24 variants were more stable than IL-20R2D variants, ~60°C vs. ~40°C, respectively (Supplementary Table S5). In both cases, NBY reduced the protein's melting temperature. The largest effect was observed in the double mutant (Y70NBY/Y74NBY) that had ~10°C lower melting temperature than the parental protein. Therefore, although the introduced NBY caused minimal structural changes, protein stability was reduced.

3.3 Testing the effect of UV irradiation

Supplementary Table S6 shows the employed UV wavelengths together with protein's UV/Visible absorption spectra. Prior to measuring protein affinities and cellular signaling, we checked the effect of UV light on protein oligomerization and molecular weight by size-exclusion chromatography (SEC) and mass spectrometry. According to SEC (see chromatograms in Figure 3D), the apparent molecular weights of all studied variants lay in between the expectation for a monomer and a dimer (Supplementary Table S6). Therefore, at least two interpretations are possible. Either the proteins are in monomer/dimer equilibrium or they feature an expanded ("open") monomeric conformation The NBY mutations tend to make the proteins more aggregation prone and substantial protein amounts were found in the void volume. Importantly, no major changes in the elution profiles were observed before/after UV and no large aggregates were found upon UV exposure. However, UV-irradiated samples eluted earlier than the non-irradiated counterparts (Supplementary Table S6). Although partial UV-induced dimerization cannot be completely ignored, we conclude that a molecular interaction with a 1:1 stoichiometry is a suitable analytical model.

Second, we ruled out any potential UV-induced photodamage to the proteins. The masses of IL-24B4 and IL-20R2D before and after illumination were virtually identical, suggesting that the irradiances used in this study are insufficient to cause significant residue modifications (Figure 4). Next we checked the efficiency of decaging. We shined UV light on the 4 NBY-containing variants and observed that the mass decreased ~135 Da (or ~270 Da in the case of the double mutant), which corresponds to that of the nitrobenzyl moiety, making it indistinguishable from the mass of canonical variants (Figure 4). For better visualization of the mass difference we also show detailed mass spectra of 35+ charge state (Supplementary Figure S7).

Thus, we prove that all photocaged interleukins/receptors can be efficiently decaged by UV irradiation, thereby regenerating the parental proteins natively containing a tyrosine residue (or residues in the case of the double mutant).

3.4 Measuring interleukin-receptor interactions and their effects

Having found conditions for efficient decaging, we next monitored the effect of the produced interleukin variants on IL-24/IL-20R2 complex formation in three environments of increasing

10.3389/fmolb.2023.1214235

TABLE 1 Binding affinities between IL-24B4 and IL-20R2D variants. Dissociation constant (K_d) and half-maximal effective concentration (EC_{50}) values are expressed as the mean ± s.d. (3 biological replicates for MST, 3 technical replicates for yeast display, and 2 technical replicates for Western blot).

	Method	*MST*	*Yeast display*	*Cell signaling (pSTAT3)*
Partner 1[a]	Partner 2	K_d (nM)	K_d[b] (nM)	EC_{50}[c] (nM)
IL-24B4	*IL-20R2D*[c]	630 ± 140	5.0 ± 0.6	0.28 ± 0.07
IL-24B4 Y204NBY	*IL-20R2D*[c]	N.P.	N.D.	0.9 ± 0.5
IL-24B4 Y204NBY + UV	*IL-20R2D*[c]	N.P.	N.D.	0.76 ± 0.24
IL-24B4	*IL-20R2D + UV*	3,700 ± 400	9.6 ± 1.6	N.P.
IL-24B4	*IL-20R2D Y70NBY*	>94,000	~2000	N.P.
IL-24B4	*IL-20R2D Y70NBY + UV*	3,300 ± 600	7.6 ± 0.4	N.P.
IL-24B4	*IL-20R2D Y74NBY*	430 ± 30	N.D.	N.P.
IL-24B4	*IL-20R2D Y70NBY/Y74NBY*	570 ± 60	N.D	N.P.

[a]For microscale thermophoresis (MST) experiments the partner 2 was held constant and partner 1 titrated, while the opposite is true for yeast display experiments.
[b]K_d, values obtained from yeast display should be considered *apparent* because they reflect not only affinity but also avidity due to the high protein densities on the cell surface.
[c]Production of phosphorylated STAT3 (pSTAT3) as a function of added partner 1. Note that in this case the partner 2 corresponds to IL-20R2/IL-22R1, expressed in HeLa cells.
N.D, means not determined; N.P, indicates a measurements that it is not possible due to low sample amounts or limitations of the cell assays (see main text more details).

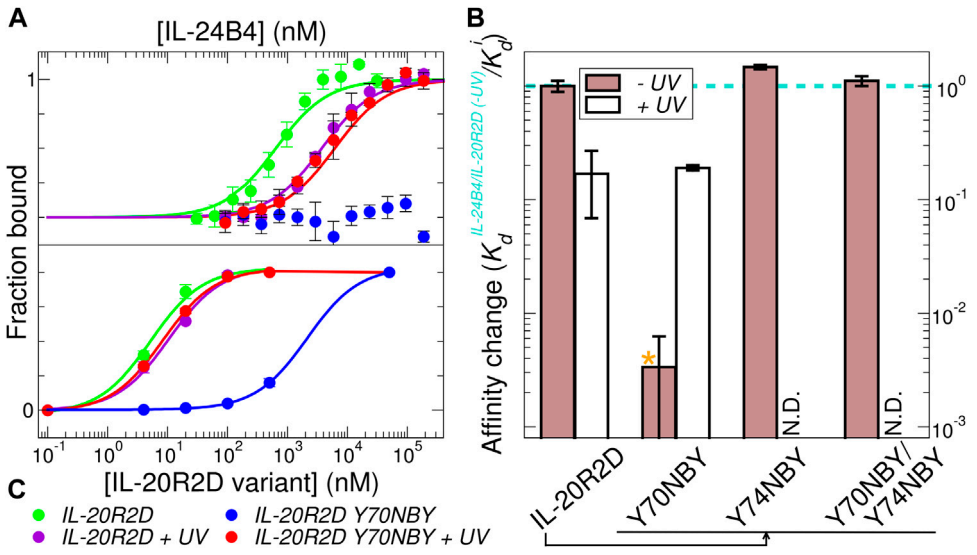


FIGURE 5
Estimation of binding affinities by microscale thermophoresis and yeast surface display. **(A)** Fraction of bound molecules as a function of IL-24B4 concentration according to MST experiments. **(B)** Fold-change in binding affinity relative to IL-24B4/IL-20R2D without UV irradiation of the four studied variants (IL-20R2D, IL-20R2D Y70NBY, IL-20R2D Y74NBY, and IL-20R2D Y70NBY/Y74NBY) by MST. The bars represent the average and standard deviation of 3 independent experiments. No significant binding was detected for the mixtures between IL-20R2D Y70NBY and IL-24B4 within the concentration range used. Therefore, the reported *apparent* K_d (orange asterisk) should be considered as lower boundary. **(C)** Fraction of bound molecules as a function of IL-20R2D concentration according to yeast display experiments. Solid lines in panels **(A)** and **(C)** are fits to Eq. 1 in order to retrieve the dissociation constants (see K_d values in Table 1).

complexity (Supplementary Figure S8). First, with both interacting partners free in solution by MST. Second, with one binding partner immobilized on the yeast cell surface and the other in solution by yeast display. Finally, with the receptors natively inserted in the membrane of human cells and the others exogenously added to the growth media. The first two methods directly report on IL-24/IL-20R2 binding affinities, while the latter is sensitive to the impact that such interactions have on signal transduction through the JAK/STAT pathway.

3.4.1 MST measurements

We determined binding affinities by MST, which measures temperature-induced changes in fluorescence intensity, using fluorescently labeled variants at a fixed concentration and mixing

with variable amounts of unlabeled binding partners. Unfortunately, the obtained amounts of IL-24B4 204NBY were too low. Hence, we concentrated on NBY-containing mutants of IL-20R2D for which we obtained sufficient yields for biophysical studies. The dissociation constant (K_d) of the complex between IL-24B4 and IL-20R2D is ~0.5 μM (see K_d values in Table 1 and raw data in Supplementary Figure S9) in agreement with the literature (Zahradník et al., 2019). Exposure to UV light slightly diminished the binding affinity by a factor of 6 (compare green and violet traces in Figure 5A). Of the three tested variants containing photocaged non-canonical amino acids, IL-20R2D Y74NBY and the double-mutant IL-20R2D Y70NBY/Y74NBY bound to IL-24B4 with similar affinities as their canonical counterparts (Figure 5B). Non-irradiated IL-20R2D Y70NBY showed virtually no binding to IL-24B4 within the concentration range used (Figure 5A), suggesting a decrease of binding affinity by a factor of at least 300 relative to the canonical complex (Figure 5B). Gratifyingly, photoconversion of IL-20R2D Y70NBY back to IL-20R2D by UV light shifted the binding curve towards the values observed for the canonical protein (Figure 5A). Thus, we found one residue position in IL-20R2D, Y70, where NBY significantly blocks the heterodimerization with IL-24B4. In such a case, complex formation can be observed after cage photolysis by UV light. Unexpectedly, the double-mutant displayed high binding affinity in the absence of UV irradiation suggesting non-additive effects of the two mutations where the inhibitory effect of Y70NBY is somehow compensated for by Y74NBY.

3.4.2 Yeast display measurements

To support the previous results in a more cellular context, we also estimated binding affinities using yeast display. The four purified and fluorescently labeled IL-20R2D variants (canonical and non-canonical, with and without UV stimulation) were incubated at different concentrations with yeast cells expressing on their surface IL-24B4 fused to a fluorescent protein. The fluorescent signal arising from each binding partner was then analyzed by flow cytometry (Supplementary Figure S10). In these plots, binding events appear along the diagonal. UV illumination did not cause a significant change (2-fold reduction) in the population of IL-20R2D/IL-24B4 complex (Figure 5C). In the case of IL-20R2D Y70NBY in the absence of UV light, complex formation with IL-24B4 was largely attenuated within the concentration range used (Figure 5C). Upon illumination, the number of heterodimers increased and reached similar levels as for the canonical complex (Figure 5C). The apparent K_d values, including both affinity and avidities, are listed in Table 1.

When comparing the relative trends, both MST and yeast display agree in suggesting low levels of complex formation between IL-20R2D Y70NBY and IL-24B4 in the absence of light. However, the binding affinities estimated by yeast display were higher (approximately 100-fold) than those determined by MST in diluted protein solutions, probably due to avidity (multiple interaction events) and crowding effects. In summary, the two methods point to two completely different binding scenarios. On one hand, low binding affinity of IL-20R2D Y70NBY (caged receptor) against IL-24B4 in the dark. On the other hand, high binding affinity (similar to the parental proteins) upon UV illumination and subsequent decaging.

3.4.3 Cell signaling assays

Eventually, we measured the ability of our engineered variants IL-20R2D and IL-24B4 to activate the JAK/STAT signaling cascade in a human cell line (HeLa). To this end, we quantified the levels of the transcription factor STAT3 phosphorylated at tyrosine705 (pSTAT3 for short), which is a well-known marker implicated in the signaling of IL-24 and other cytokines (Darnell et al., 1994). pSTAT3 levels after 30 min exposure with canonical or non-canonical interleukin variants were determined by Western blotting using an anti-pSTAT3 antibody and subsequent enhanced chemiluminiscence detection.

Initially, we checked whether our engineered versions (IL-24B4 and IL-20R2D) and the endogenous levels of native receptors (IL-20R2 and IL-22R1) present in HeLa cells can yield detectable amounts of pSTAT3. As a positive control we used interferon-α (IFNα), a cytokine that is known to induce rapid changes in JAK/STAT phosphorylation and initiate the signaling (Pellegrini et al., 1989). Little production of pSTAT3 was found in non-transfected cells regardless of the presence of IL-24B4 (Supplementary Figure S11A). On the contrary, significant amounts of pSTAT3 were found in transfected cells upon addition of IL-24B4 (Supplementary Figure S11A) suggesting that HeLa cells express limited amounts of endogenous receptors and require a boost of IL-24 receptors to enhance signal-to-noise. Similar amounts of pSTAT3 were found in the case of interferon-α stimulation (Supplementary Figure S11B). To test the capability of IL-20R2D to transduce biological signals, we created a chimeric receptor (IL-20R2DC) where the extracellular portion of full-length IL-20R2 was replaced by the designed IL-20R2D (Supplementary Table S1). The results (Supplementary Figure S11C) support the notion that IL-20R2DC can bind IL-24B4 and trigger signaling.

Surprisingly, we found that IL-20R2D signals through the same receptors and pathway although at high protein concentrations (Supplementary Figure S12A). Signal transduction through IL-20R2 was previously observed for IFNαR2 (Pattyn et al., 1999). Such a background level of pSTAT3 induction by IL-20R2D would complicate the interpretation of the results when both IL-24B4 and IL-20R2D are present. Taken together, these results confirm that the designed variants, IL-24B4 and IL-20R2D, recapitulate the role of native proteins, i.e., IL-24B4 binds its cognate membrane-bound receptors (IL-20R2, IL-22R1) and initiates the JAK/STAT pathway. However, a limitation of our cell signaling assays is that they do not unambiguously inform on the effect of IL-20R2D.

Therefore, we focused on the IL-24B4 Y204NBY mutant, which could be purified in sufficient amounts for the cell assays. We measured the intracellular levels of pSTAT3 reporter after incubation with different concentrations of IL-24B4 and IL-24B4 Y204NBY without and with UV irradiation (Supplementary Figure S12B). The more IL-24B4 was added to the cells the more pSTAT3 was obtained until saturating levels were reached. A non-cooperative Hill model was fitted to the dose-response curves (Supplementary Figure S12C) and the resulting half-maximal effective concentration (EC_{50}) values are reported in Table 1. The EC_{50} of caged IL-24B4

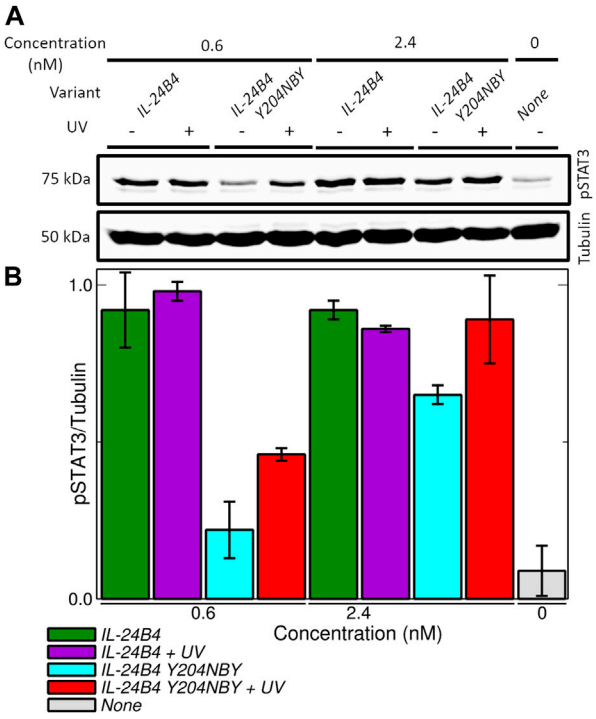


FIGURE 6
Signaling assays in HeLa cells transfected with plasmids encoding IL-20R2/IL-22R1. **(A)** Western blots showing the production of phosphorylated STAT3 (pSTAT3) upon induction of JAK/STAT signaling pathway by exogenously added amounts of IL-24B4 and its photocaged variant Y204NBY both in the presence and absence of UV irradiation (λ = 365 nm, 5 min at 100 mW). **(B)** Quantification of pSTAT3 (normalized by tubulin) as a function of the added variants and concentrations. The bars represent the average and standard deviation of 2 independent experiments.

Y204NBY was ~3 times larger than that of parental IL-24B4 but similar to the decaged counterpart considering experimental uncertainties. At concentrations higher than EC_{50}, all four IL-24B4 variants behaved similarly as expected (Figure 6). Only at concentrations near EC_{50}, the caged Y204NBY variant signaled less than the parental protein, and UV irradiation increased pSTAT3 levels but without reaching the amounts observed for the parental protein (Figure 6). We conclude that IL-24B4 Y204NBY is another promising candidate for future studies.

4 Discussion

Since the interaction between human interleukin-24 and its receptor IL-20R2 triggers the JAK/STAT signaling pathway, and acts on angiogenesis and cell proliferation, we are interested in controlling this interleukin-receptor system. In this work, we have genetically encoded in each binding partner *ortho*-nitrobenzyl-tyrosine, an unnatural photocaged analogue of tyrosine, to photoregulate IL-24/IL-20R2 interactions. Our main finding is that IL-20R2D Y70NBY and, to a lesser extent, IL-24B4 Y204 hamper complex formation and subsequent activation of the STAT3 cascade. Mild UV irradiation partially restores heteroassociation and cell signaling.

4.1 Photocontrolling protein-protein interactions involving interleukins with ncAA

Protein-protein interactions (PPI) lie at the heart of many biological processes ranging from metabolic pathways to signaling cascades. Moreover, abnormal PPI are associated with various diseases, including cancer (Lu et al., 2020). Therefore, the modulation of PPI is important not only for basic research but also for the development of new drugs. Research related to the control of PPI has focused primarily on antigen-antibody interactions (Lucchi et al., 2021).

Among several approaches to photocontrol PPI (Carrasco-López et al., 2020; Gil et al., 2020), genetically encoded photo-responsive non-canonical amino acids offer several advantages: i) a precise level of temporal and spatial control over protein binding; ii) the small size of the modification (typical caging groups are less than 500 Da in size); and iii) the possibility of generating the modified protein directly by the biosynthetic machinery within the cell (Courtney and Deiters, 2018). However, there are also challenges associated with their use. First, the suppression efficiencies are variable, context-dependent and may result in severe drops of protein yields, particularly in the case of multi-site incorporation (Amiram et al., 2015). Second, the introduction of non-native side-chains may affect protein folding, stability, binding and dynamics (Rogers et al., 2018; Kesgin-Schaefer et al., 2019; Chaudhari et al., 2023). Third, due to the usually large interfacial areas that are stabilized by multiple non-covalent interactions (hydrogen bonds, hydrophobic interactions, *etc.*), targeting PPI with single-residue perturbations is not straightforward, typically requiring extensive screening campaigns (Bridge et al., 2019; O'Shea et al., 2022). This contrasts with the photocontrol of interactions between proteins and small ligands or substrates, which typically depend on one or a reduced number of residues closely localized in space, e.g., an enzyme's catalytic site (Nguyen et al., 2014; Zhou et al., 2020).

The interleukin 10 family, which includes IL-24, plays important roles in cell differentiation, apoptosis, and tumor growth, and thus represents an attractive therapeutic target (Chada et al., 2005; Cunningham et al., 2005). Disrupted cytokine-receptor signaling and JAK/STAT phosphorylation cascade may lead to a variety of diseases, such as skin conditions, cancers, and disorders affecting the immune system (Aaronson and Horvath, 2002; Hu et al., 2021). Members of the IL-10 are also involved in host-pathogen interactions, wound healing, and have potential in the treatment of respirator, inflammatory and autoimmune diseases (Iyer and Cheng, 2012; Traupe, 2017; Liu et al., 2023; Reis et al., 2023). To the best of our knowledge, there is only one example of inhibition of the interaction between IL-24 and its receptors, which is based on antibody binding to IL-20R1 and IL-22R1 (Chada et al., 2004). The light-dependent assembly between IL-24 and its receptors may be an alternative route toward the development of a new therapeutic strategy to combat such diseases and conditions.

4.2 Heterologous expression of interleukins and receptors containing ncAA

To obtain IL-24, the receptor IL-20R2, and their photoactivatable variants, we combine canonical and ncAA mutagenesis.

Depending on protein sequence and need for post-translational modifications, obtaining sufficient yields of stable proteins can be a serious bottleneck. Obtaining "difficult" proteins requires multiple expression attempts (Peleg and Unger, 2012) or cumbersome and cost-ineffective expression systems (Hopkins et al., 2010). On top of these effects, the incorporation of ncAA may cause further decreases in protein yields and alter protein stability or function in ways that are difficult to predict (Chaudhari et al., 2023). Indeed, all our proteins of interest are naturally N-glycosylated and bear disulfide bonds, which make them hard to produce in traditional *E. coli* expression systems. Production of wild-type IL-24 in high quantities is only possible with the co-expression of its interacting partner (Lubkowski et al., 2018). Overall, protein production renders the IL-24 signaling complex a very challenging target for (photo) xenoprotein engineering. Therefore, we adopted an alternative strategy to obtain active proteins while avoiding long and severe optimization trials based on structure- and sequence-based protein design using the 20 canonical amino acids (Goldenzweig et al., 2016). Such a bioinformatics design optimizes the amino acid sequence of IL-24 and IL-20R2 to stabilize their expression in bacteria, rendering them more suitable for functional studies. Both engineered variants, named IL-24B4 and IL-20R2D, are biologically active and support signal transduction through the JAK/STAT pathway.

Next, the light-inducible variants are made by adding a photocaged amino acid, NBY, into IL-24B4 and IL-20R2D by genetic code expansion technology. We looked for candidate residues in both interacting partners and found three target tyrosines. The single tyrosine mutant of IL-24B4, Y204NBY, did not express at sufficient and cost-effective levels for subsequent biophysical investigations by MST. Thus, IL-24B4 Y204NBY could only be studied in cell assays, which require relatively low amounts of material. In the case of the receptor, all ncAA variants including two single mutants and one double mutant were purified at the required amounts for all three methods employed in this work (MST, yeast display and human cell assays). NBY did not substantially affect protein conformation but it had a negative impact on protein's thermal stability, particularly in the case of the double mutant.

4.3 Monitoring IL-24 – IL-20R2 interactions by biophysical methods and cell assays

IL-24B4 Y204NBY exhibits minimal albeit reproducible differences in cell signaling assays when comparing parental and caged variants.

We find one clear "hit" confirmed by both MST and yeast display: IL-20R2D Y70NBY shows low affinity for IL-24B4 and the interaction can be activated by UV light to reach native-like levels. This in line with previous studies reporting that a single NBY residue is sufficient to diminish PPI (Bridge et al., 2019; Jedlitzke et al., 2019; Joest et al., 2021; Jedlitzke and Mootz, 2022). Here we used an innovative approach of yeast surface display to gain better control over our experiment and see the residual concentrations. We mixed yeasts expressing IL-24B4 with a low amount of yeasts expressing its affinity matured counterpart, which has more than two orders of magnitude higher affinity (IL-24S). The analysis showed detectable but low, below saturation, binding signals of IL-24S with IL-20R2D Y70NBY (Supplementary Figure S10). Given the difference in binding affinity, the concentration of binding-capable ligands was below 1% of total concentration.

Unfortunately, due to limitations in our signaling assays, including both the choice of synthetase and the background signaling effect of IL-20R2D, we could not test the effect of IL-20R2D Y70NBY in human cells. It was previously reported that dimerization of IFNαR2 is sufficient for induction of interferon-regulated genes but not for full activity (Pattyn et al., 1999). Here, we show that purified IL-20R2D added to the cells can trigger signaling albeit at high concentrations, This problem may be partially circumvented by using general reporter cell lines and/or reporter assays based on gene expression (Mock et al., 2020; Cho et al., 2023) with further research needed. As an alternative to the genetic encoding of NBY in sequence-optimized truncated receptors recombinantly expressed in *E. coli*, one could biosynthetically incorporate photocaged tyrosines in full-length IL-20R2 receptors expressed in HeLa cells by using an orthogonal translation system derived from *Methanosarcina barkeri* (Arbely et al., 2012).

The more effective blocking of heterocomplex assembly with Y70NBY compared to Y204NBY can be rationalized in structural terms: Y70 of IL-20R2 is found at the core of the IL-24/IL-20R2 interface while Y204 of IL-24 is located in a flexible region at one edge of the interaction surface. We notice that the use of NBY to photocontrol IL-24-dependent PPI necessitates further improvements. Y70NBY and particularly Y204NBY do not fully prevent complex assembly, but rather shift the binding equilibrium towards higher concentrations (lower affinities). IL-24/IL-20R2 complex may still be formed and signal transduction started if large concentrations of photocaged binding partners are employed. In principle, substantial affinity differences between native and caged variants are preferred. The dynamical range is much narrower for IL-24B4 Y204NBY (3-fold difference in EC_{50}) than for IL-20R2D Y70NBY (at least 300-fold difference in K_d according to both MST and yeast display). Larger dynamical ranges could be achieved through the installment of photocages on each partner, for instance Y70NBY in IL-20R2D and Y204NBY in IL-24B4. Although residues at the binding interface seem the natural choice for replacement, other distant sites may also influence PPI through allosteric communication, which has been reported for some interleukins (Bowman and Geissler, 2012; De Paula et al., 2020).

Strikingly, the two single mutants (IL-20R2 Y70 NBY and IL-20R2D Y74NBY) and the double mutant (IL-20R2D Y70NBY/Y74NBY) show clear signs of non-additive effects on protein binding. Y70NBY inhibits association with IL-24B4 while Y74NBY and Y70NBY/Y74NBY do not significantly alter protein-protein interactions. Non-additive interactions between mutations (epistasis) are common in proteins and play a crucial

role in protein engineering (Reetz, 2013). There is no clear picture of the mechanisms that cause epistasis, although effects on protein stability, conformation and dynamics have all been invoked (Starr and Thornton, 2016). Our CD and thermal denaturing experiments suggest minimal differences among the mutants. All variants show similar secondary structure content. The melting temperature of the mutants is reduced with respect to the parental proteins and the Tm of the double mutant Y70NBY/Y74NBY is further reduced thus suggesting additive effects. Therefore, protein structure and thermal stability do not seem to be the major causes of the negative (antagonistic/deleterious) epistasis observed when combining Y70NBY and Y74NBY mutations. We hypothesize that the two mutations together alter IL-20R2D structural dynamics, a phenomenon that has been described for other proteins (Wagner et al., 1995; Acevedo-Rocha et al., 2021).

Interestingly, we found evidence that in some cases UV-irradiation alone reduced the binding affinity. Although we did not detect clear signs of photoinduced chemical modifications, e.g., photo-oxidation (the masses of IL-24B4 and IL-20R2D before/after UV are virtually identical), the affinity between IL-24B4 and IL-20R2 was reduced 6-fold (MST-based experiments) or 2-fold (yeast surface display experiments). This result implies that our UV irradiation protocol damages the proteins to a certain extent. We suggest that by carefully controlling the illumination conditions (wavelength, power density and time) such an undesired effect may be eliminated. Moreover, UV light can potentially damage nucleic acids depending on the dose used and it does not penetrate deeply into tissues (Kielbassa et al., 1997). Consequently, for *in vivo* applications requiring native-like binding affinities it would be beneficial to use other photocages. Suitable candidates would be the nitropiperonyl moiety, which features longer absorption wavelengths (Gautier et al., 2010; Luo et al., 2017) or coumarins, which can be activated by two-photon excitation in the near infrared range (Luo et al., 2014). Apart from photocaged ncAAs, which enable unidirectional OFF-to-ON switch of protein binding by light, one could use other ncAA. For instance, in applications requiring bidirectional photocontrol of protein binding (Jankovic et al., 2019; Myrhammar et al., 2020; Zhang et al., 2022), photoswitchable ncAA like azobenzene-phenylalanine constitute an excellent alternative (Bose et al., 2005; Luo et al., 2018; Israeli et al., 2021).

Overall, our results suggest the feasibility of regulating interactions involving protein partners from the interleukin family, IL-24 and IL-20R2, with UV light and photocaged tyrosines introduced at certain residue positions. In the future, other cytokine/receptor pairs and non-antibody-based protein scaffolds (Kolářová et al., 2021; Pham et al., 2021; Huličiak et al., 2023) may benefit from a similar approach and offer excellent interaction control.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.5281/zenodo.7877998.

Author contributions

Conceptualization, GF; methodology and investigation, PP and JZ; resources, LK and GF; writing—original draft preparation, PP and GF; writing—review and editing, PP, JZ, BS and GF; project administration and supervision, GF; funding acquisition, JZ and BS. All authors contributed to the article and approved the submitted version.

Funding

The work was supported by the project ELIBIO: Structural dynamics of biomolecular systems (CZ.02.1.01/0.0/0.0/15_003/0000447) from the European Regional Development Fund and the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic (CR). The Institute of Biotechnology of the Czech Academy of Sciences acknowledges the institutional grant RVO 86652036. We acknowledge CF Biophysic, CF SMS of CIISB, Instruct-CZ Centre, supported by MEYS CR (LM2023042) and European Regional Development Fund-Project "UP CIISB" (No. CZ.02.1.01/0.0/0.0/18_046/0015974). We acknowledge Charles University grant PRIMUS no. 23/MED/002. This paper was funded by the European Union - Next Generation EU (National Institute for Cancer Research, programme EXCELES, Project No. LX22NPO5102).

Acknowledgments

We thank Petr Pompach and Pavla Vaňková for the mass spectrometry measurements. We are also grateful to Tania Charnavets and Volha Dzmitruk for assistance with MST analysis.

We thank Aditi Chatterjee for calibrating the gel filtration column.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1214235/full#supplementary-material

References

Aaronson, D. S., and Horvath, C. M. (2002). A road map for those who don't know JAK-STAT. *Science* 296 (5573), 1653–1655. doi:10.1126/science.1071545

Acevedo-Rocha, C. G., Li, A., D'Amore, L., Hoebenreich, S., Sanchis, J., Lubrano, P., et al. (2021). Pervasive cooperative mutational effects on multiple catalytic enzyme traits emerge via long-range conformational dynamics. *Nat. Commun.* 12 (1), 1621. doi:10.1038/s41467-021-21833-w

Akdis, M., Burgler, S., Crameri, R., Eiwegger, T., Fujita, H., Gomez, E., et al. (2011). Interleukins, from 1 to 37, and interferon-γ: Receptors, functions, and roles in diseases. *J. Allergy Clin. Immunol.* 127 (3), 701–721.e1-70. doi:10.1016/j.jaci.2010.11.050

Alexander, C. G., Wanner, R., Johnson, C. M., Breitsprecher, D., Winter, G., Duhr, S., et al. (2014). Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings. *Biochimica Biophysica Acta (BBA) - Proteins Proteomics* 1844 (12), 2241–2250. doi:10.1016/j.bbapap.2014.09.016

Amiram, M., Haimovich, A. D., Fan, C., Wang, Y.-S., Aerni, H.-R., Ntai, I., et al. (2015). Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* 33 (12), 1272–1279. doi:10.1038/nbt.3372

Andoh, A., Shioya, M., Nishida, A., Bamba, S., Tsujikawa, T., Kim-Mitsuyama, S., et al. (2009). Expression of IL-24, an activator of the JAK1/STAT3/SOCS3 cascade, is enhanced in inflammatory bowel disease. *J. Immunol.* 183 (1), 687–695. doi:10.4049/jimmunol.0804169

Arbely, E., Torres-Kolbus, J., Deiters, A., and Chin, J. W. (2012). Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* 134 (29), 11912–11915. doi:10.1021/ja3046958

Baumann, T., Hauf, M., Richter, F., Albers, S., Möglich, A., Ignatova, Z., et al. (2019). Computational aminoacyl-tRNA synthetase library design for photocaged tyrosine. *Int. J. Mol. Sci.* 20 (9), 2343. doi:10.3390/ijms20092343

Bose, M., Groff, D., Xie, J., Brustad, E., and Schultz, P. G. (2005). The incorporation of a photoisomerizable amino acid into proteins in *E. coli*. *J. Am. Chem. Soc.* 128 (2), 388–389. doi:10.1021/ja055467u

Bowman, G. R., and Geissler, P. L. (2012). Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci.* 109 (29), 11681–11686. doi:10.1073/pnas.1209309109

Bridge, T., Shaikh, S. A., Thomas, P., Botta, J., McCormick, P. J., and Sachdeva, A. (2019). Site-specific encoding of photoactivity in antibodies enables light-mediated antibody–antigen binding on live cells. *Angew. Chem. Int. Ed.* 58 (50), 17986–17993. doi:10.1002/anie.201908655

Bridge, T., Wegmann, U., Crack, J. C., Orman, K., Shaikh, S. A., Farndon, W., et al. (2023). Site-specific encoding of photoactivity and photoreactivity into antibody fragments. *Nat. Chem. Biol.* 19, 740–749. doi:10.1038/s41589-022-01251-9

Carrasco-López, C., Zhao, E. M., Gil, A. A., Alam, N., Toettcher, J. E., and Avalos, J. L. (2020). Development of light-responsive protein binding in the monobody non-immunoglobulin scaffold. *Nat. Commun.* 11 (1), 4045. doi:10.1038/s41467-020-17837-7

Chada, S., Mhashilkar, A. M., Ramesh, R., Mumm, J. B., Sutton, R. B., Bocangel, D., et al. (2004). Bystander activity of Ad-mda7: Human MDA-7 protein kills melanoma cells via an IL-20 receptor-dependent but STAT3-independent mechanism. *Mol. Ther.* 10 (6), 1085–1095. doi:10.1016/j.ymthe.2004.08.020

Chada, S., Mhashilkar, A. M., Liu, Y., Nishikawa, T., Bocangel, D., Zheng, M., et al. (2005). mda-7 gene transfer sensitizes breast carcinoma cells to chemotherapy, biologic therapies and radiotherapy: correlation with expression of bcl-2 family members. *Cancer Gene Ther.* 13 (5), 490–502. doi:10.1038/sj.cgt.7700915

Chaudhari, A. S., Chatterjee, A., Domingos, C. A. O., Andrikopoulos, P. C., Liu, Y., Andersson, I., et al. (2023). Genetically encoded non-canonical amino acids reveal asynchronous dark reversion of chromophore, backbone and side-chains in EL222. *Protein Sci.* 32, e4590. doi:10.1002/pro.4590

Cheung, J. W., Kinney, W. D., Wesalo, J. S., Reed, M., Nicholson, E. M., Deiters, A., et al. (2023). Genetic encoding of a photocaged histidine for light-control of protein activity. *ChemBioChem* 24, e202200721. doi:10.1002/cbic.202200721

Cho, O., Lee, J.-W., Kim, H.-S., Jeong, Y.-J., and Heo, T.-H. (2023). Chelerythrine, a novel small molecule targeting IL-2, inhibits melanoma progression by blocking the interaction between IL-2 and its receptor. *Life Sci.* 320, 121559. doi:10.1016/j.lfs.2023.121559

Courtney, T., and Deiters, A. (2018). Recent advances in the optical control of protein function through genetic code expansion. *Curr. Opin. Chem. Biol.* 46, 99–107. doi:10.1016/j.cbpa.2018.07.011

Cunningham, C. C., Chada, S., Merritt, J. A., Tong, A., Senzer, N., Zhang, Y., et al. (2005). Clinical and local biological effects of an intratumoral injection of mda-7 (IL24; INGN 241) in patients with advanced carcinoma: a phase I study. *Mol. Ther.* 11 (1), 149–159. doi:10.1016/j.ymthe.2004.09.019

Darnell, J. E., Kerr, I. M., and Stark, G. R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 264 (5164), 1415–1421. doi:10.1126/science.8197455

de Marco, A., Berrow, N., Lebendiker, M., Garcia-Alai, M., Knauer, S. H., Lopez-Mendez, B., et al. (2021). Quality control of protein reagents for the improvement of research data reproducibility. *Nat. Commun.* 12 (1), 2795. doi:10.1038/s41467-021-23167-z

de Melo, C. B., Uto-Konomi, A., Miyauchi, K., Ozaki, N., Motomura, Y., Suzuki, Y., et al. (2012). Dysregulation of suppressor of cytokine signaling 3 in keratinocytes causes skin inflammation mediated by interleukin-20 receptor-related cytokines. *PLoS ONE* 7 (7), e40343. doi:10.1371/journal.pone.0040343

De Paula, V. S., Jude, K. M., Nerli, S., Glassman, C. R., Garcia, K. C., and Sgourakis, N. G. (2020). Interleukin-2 druggability is modulated by global conformational transitions controlled by a helical capping switch. *Proc. Natl. Acad. Sci.* 117 (13), 7183–7192. doi:10.1073/pnas.2000419117

Deiters, A., Groff, D., Ryu, Y., Xie, J., and Schultz, P. G. (2006). A genetically encoded photocaged tyrosine. *Angew. Chem. Int. Ed.* 45 (17), 2728–2731. doi:10.1002/anie.200600264

Dumoutier, L., Leemans, C., Lejeune, D., Kotenko, S. V., and Renauld, J.-C. (2001). Cutting edge: STAT activation by IL-19, IL-20 and mda-7 through IL-20 receptor complexes of two types. *J. Immunol.* 167 (7), 3545–3549. doi:10.4049/jimmunol.167.7.3545

Gautier, A., Nguyen, D. P., Lusic, H., An, W., Deiters, A., and Chin, J. W. (2010). Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* 132 (12), 4086–4088. doi:10.1021/ja910688s

Gil, A. A., Carrasco-López, C., Zhu, L., Zhao, E. M., Ravindran, P. T., Wilson, M. Z., et al. (2020). Optogenetic control of protein binding using light-switchable nanobodies. *Nat. Commun.* 11 (1), 4044. doi:10.1038/s41467-020-17836-8

Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., et al. (2016). Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell* 63 (2), 337–346. doi:10.1016/j.molcel.2016.06.012

Hauf, M., Richter, F., Schneider, T., Faidt, T., Martins, B. M., Baumann, T., et al. (2017). Photoactivatable mussel-based underwater adhesive proteins by an expanded genetic code. *ChemBioChem* 18 (18), 1819–1823. doi:10.1002/cbic.201700327

Hoorens, M. W. H., and Szymanski, W. (2018). Reversible, spatial and temporal control over protein activity using light. *Trends Biochem. Sci.* 43 (8), 567–575. doi:10.1016/j.tibs.2018.05.004

Hopkins, R., Esposito, D., and Gillette, W. (2010). Widening the bottleneck: Increasing success in protein expression and purification. *J. Struct. Biol.* 172 (1), 14–20. doi:10.1016/j.jsb.2010.07.005

Hu, X., li, J., Fu, M., Zhao, X., and Wang, W. (2021). The JAK/STAT signaling pathway: From bench to clinic. *Signal Transduct. Target. Ther.* 6 (1), 402. doi:10.1038/s41392-021-00791-1

Huličiak, M., Biedermanová, L., Berdár, D., Herynek, Š., Kolářová, L., Tomala, J., et al. (2023). Combined *in vitro* and cell-based selection display method producing specific binders against IL-9 receptor in high yields. *FEBS J.* 290, 2993–3005. doi:10.1111/febs.16726

Israeli, B., Strugach, D. S., Gelkop, S., Weber, S., Gozlan, D. S., and Amiram, M. (2021). Genetically encoding light-responsive protein-polymers using translation machinery for the multi-site incorporation of photo-switchable unnatural amino acids. *Adv. Funct. Mater.* 31 (44), 2011276. doi:10.1002/adfm.202011276

Iyer, S. S., and Cheng, G. (2012). Role of interleukin 10 transcriptional regulation in inflammation and autoimmune disease. *Crit. Reviews™ Immunol.* 32 (1), 23–63. doi:10.1615/CritRevImmunol.v32.i1.30

Jankovic, B., Gulzar, A., Zanobini, C., Bozovic, O., Wolf, S., Stock, G., et al. (2019). Photocontrolling protein–peptide interactions: From minimal perturbation to complete unbinding. *J. Am. Chem. Soc.* 141 (27), 10702–10710. doi:10.1021/jacs.9b03222

Jedlitzke, B., and Mootz, H. D. (2022). A light-activatable photocaged variant of the ultra-high affinity ALFA-tag nanobody. *ChemBioChem* 23 (12), e202200079. doi:10.1002/cbic.202200079

Jedlitzke, B., Yilmaz, Z., Dörner, W., and Mootz, H. D. (2019). Photobodies: Light-Activatable single-domain antibody fragments. *Angew. Chem. Int. Ed.* 59 (4), 1506–1510. doi:10.1002/anie.201912286

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., et al. (2014). MicroScale thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* 1077, 101–113. doi:10.1016/j.molstruc.2014.03.009

Joest, E. F., Winter, C., Wesalo, J. S., Deiters, A., and Tampé, R. (2021). Light-guided intrabodies for on-demand *in situ* target recognition in human cells. *Chem. Sci.* 12 (16), 5787–5795. doi:10.1039/d1sc01331a

Kesgin-Schaefer, S., Heidemann, J., Puchert, A., Koelbel, K., Yorke, B. A., Huse, N., et al. (2019). Crystal structure of a domain-swapped photoactivatable sfGFP variant provides evidence for GFP folding pathway. *FEBS J.* 286 (12), 2329–2340. doi:10.1111/febs.14797

Kielbassa, C., Roza, L., and Epe, B. (1997). Wavelength dependence of oxidative DNA damage induced by UV and visible light. *Carcinogenesis* 18 (4), 811–816. doi:10.1093/carcin/18.4.811

Kneuttinger, A. C. (2022). A guide to designing photocontrol in proteins: Methods, strategies and applications. *Biol. Chem.* 403 (5-6), 573–613. doi:10.1515/hsz-2021-0417

Koch, N. G., Baumann, T., Nickling, J. H., Dziegielewski, A., and Budisa, N. (2022). Engineered bacterial host for genetic encoding of physiologically stable protein nitration. *Front. Mol. Biosci.* 9, 992748. doi:10.3389/fmolb.2022.992748

Kolářová, L., Zahradník, J., Huličiak, M., Mikulecký, P., Peleg, Y., Shemesh, M., et al. (2021). De novo developed protein binders mimicking Interferon lambda signaling. *FEBS J.* 289 (9), 2672–2684. doi:10.1111/febs.16300

Kragstrup, T. W., Otkjaer, K., Holm, C., Jørgensen, A., Hokland, M., Iversen, L., et al. (2008). The expression of IL-20 and IL-24 and their shared receptors are increased in rheumatoid arthritis and spondyloarthropathy. *Cytokine* 41 (1), 16–23. doi:10.1016/j.cyto.2007.10.004

Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M., and Schultz, P. G. (2007). Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* 3 (12), 769–772. doi:10.1038/nchembio.2007.44

Liu, S., Hur, Y. H., Cai, X., Cong, Q., Yang, Y., Xu, C., et al. (2023). A tissue injury sensing and repair pathway distinct from host pathogen defense. *Cell* 186 (10), 2127–2143.e22. doi:10.1016/j.cell.2023.03.031

Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., et al. (2020). Recent advances in the development of protein–protein interactions modulators: Mechanisms and clinical trials. *Signal Transduct. Target. Ther.* 5 (1), 213. doi:10.1038/s41392-020-00315-3

Lubkowski, J., Sonmez, C., Smirnov, S. V., Anishkin, A., Kotenko, S. V., and Wlodawer, A. (2018). Crystal structure of the labile complex of IL-24 with the extracellular domains of IL-22r1 and IL-20r2. *J. Immunol.* 201 (7), 2082–2093. doi:10.4049/jimmunol.1800726

Lucchi, R., Bentanachs, J., and Oller-Salvia, B. (2021). The masking game: Design of activatable antibodies and mimetics for selective therapeutics and cell control. *ACS Central Sci.* 7 (5), 724–738. doi:10.1021/acscentsci.0c01448

Luo, J., Uprety, R., Naro, Y., Chou, C., Nguyen, D. P., Chin, J. W., et al. (2014). Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation. *J. Am. Chem. Soc.* 136 (44), 15551–15558. doi:10.1021/ja5055862

Luo, J., Torres-Kolbus, J., Liu, J., and Deiters, A. (2017). Genetic encoding of photocaged tyrosines with improved light-activation properties for the optical control of protease function. *ChemBioChem* 18 (14), 1442–1447. doi:10.1002/cbic.201700147

Luo, J., Samanta, S., Convertino, M., Dokholyan, N. V., and Deiters, A. (2018). Reversible and tunable photoswitching of protein function through genetic encoding of azobenzene amino acids in mammalian cells. *ChemBioChem* 19 (20), 2178–2185. doi:10.1002/cbic.201800226

Ma, Y., Chen, H.-D., Wang, Y., Wang, Q., Li, Y., Zhao, Y., et al. (2011). Interleukin 24 as a novel potential cytokine immunotherapy for the treatment of *Mycobacterium tuberculosis* infection. *Microbes Infect.* 13 (12-13), 1099–1110. doi:10.1016/j.micinf.2011.06.012

Manandhar, M., Chun, E., and Romesberg, F. E. (2021). Genetic code expansion: Inception, development, commercialization. *J. Am. Chem. Soc.* 143 (13), 4859–4878. doi:10.1021/jacs.0c11938

Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K. A., Benesch, J. L. P., and Robinson, C. V. (2015). Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* 87 (8), 4370–4376. doi:10.1021/acs.analchem.5b00140

Menezes, M. E., Bhoopathi, P., Pradhan, A. K., Emdad, L., Das, S. K., Guo, C., et al. (2018). Role of MDA-7/IL-24 a multifunction protein in human diseases. *Adv. Cancer Res.* 138, 143–182. doi:10.1016/bs.acr.2018.02.005

Micsonai, A., Moussong, É., Wien, F., Boros, E., Vadászi, H., Murvai, N., et al. (2022). BeStSel: Webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Res.* 50 (W1), W90–W98. doi:10.1093/nar/gkac345

Mock, J., Pellegrino, C., and Neri, D. (2020). A universal reporter cell line for bioactivity evaluation of engineered cytokine products. *Sci. Rep.* 10 (1), 3234. doi:10.1038/s41598-020-60182-4

Myrhammar, A., Rosik, D., and Karlström, A. E. (2020). Photocontrolled reversible binding between the protein A-derived Z domain and immunoglobulin G. *Bioconjugate Chem.* 31 (3), 622–630. doi:10.1021/acs.bioconjchem.9b00786

Nguyen, D. P., Mahesh, M., Elsässer, S. J., Hancock, S. M., Uttamapinant, C., and Chin, J. W. (2014). Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* 136 (6), 2240–2243. doi:10.1021/ja412191m

O'Shea, J. M., Goutou, A., Brydon, J., Sethna, C. R., Wood, C. W., and Greiss, S. (2022). Generation of photocaged nanobodies for intracellular applications in an animal using genetic code expansion and computationally guided protein engineering. *ChemBioChem* 23 (16), e202200321. doi:10.1002/cbic.202200321

Parrish-Novak, J., Xu, W., Brender, T., Yao, L., Jones, C., West, J., et al. (2002). Interleukins 19, 20, and 24 signal through two distinct receptor complexes. Differences in receptor-ligand interactions mediate unique biological functions. *J. Biol. Chem.* 277 (49), 47517–47523. doi:10.1074/jbc.M205114200

Pattyn, E., Van Ostade, X., Schauvliege, L., Verhee, A., Kalai, M., Vandekerckhove, J., et al. (1999). Dimerization of the interferon type I receptor IFNaR2–2 is sufficient for induction of interferon effector genes but not for full antiviral activity. *J. Biol. Chem.* 274 (49), 34838–34845. doi:10.1074/jbc.274.49.34838

Peleg, Y., and Unger, T. (2012). "Resolving bottlenecks for recombinant protein expression in *E. coli*," in *Chemical genomics and proteomics*, 173–186.

Pellegrini, S., John, J., Shearer, M., Kerr, I. M., and Stark, G. R. (1989). Use of a selectable marker regulated by alpha interferon to obtain mutations in the signaling pathway. *Mol. Cell. Biol.* 9 (11), 4605–4612. doi:10.1128/mcb.9.11.4605-4612.1989

Pham, P. N., Huličiak, M., Biedermannová, L., Černý, J., Charnavets, T., Fuertes, G., et al. (2021). Protein binder (ProBi) as a new class of structurally robust non-antibody protein scaffold for directed evolution. *Viruses* 13 (2), 190. doi:10.3390/v13020190

Reetz, M. T. (2013). The importance of additive and non-additive mutational effects in protein engineering. *Angew. Chem. Int. Ed.* 52 (10), 2658–2666. doi:10.1002/anie.201207842

Reis, G., Moreira Silva, E. A. S., Medeiros Silva, D. C., Thabane, L., Campos, V. H. S., Ferreira, T. S., et al. (2023). Early treatment with pegylated interferon lambda for covid-19. *N. Engl. J. Med.* 388 (6), 518–528. doi:10.1056/NEJMoa2209760

Rogers, J. M., Passioura, T., and Suga, H. (2018). Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proc. Natl. Acad. Sci.* 115 (43), 10959–10964. doi:10.1073/pnas.1809901115

Rutz, S., Wang, X., and Ouyang, W. (2014). The IL-20 subfamily of cytokines — From host defence to tissue homeostasis. *Nat. Rev. Immunol.* 14 (12), 783–795. doi:10.1038/nri3766

Stael, S., Miller, L. P., Fernández-Fernández, Á. D., and Van Breusegem, F. (2022). "Detection of damage-activated metacaspase activity by western blot in plants," in *Plant proteases and plant cell death*, 127–137.

Starr, T. N., and Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Sci.* 25 (7), 1204–1218. doi:10.1002/pro.2897

Su, Z., Emdad, L., Sauane, M., Lebedeva, I. V., Sarkar, D., Gupta, P., et al. (2005). Unique aspects of mda-7/IL-24 antitumor bystander activity: Establishing a role for secretion of MDA-7/IL-24 protein by normal cells. *Oncogene* 24 (51), 7552–7566. doi:10.1038/sj.onc.1208911

Traupe, H. (2017). Psoriasis and the interleukin-10 family: Evidence for a protective genetic effect, but not an easy target as a drug. *Br. J. Dermatology* 176 (6), 1438–1439. doi:10.1111/bjd.15158

Wagner, C. R., Huang, Z., Singleton, S. F., and Benkovic, S. J. (1995). Molecular basis for nonadditive mutational effects in *Escherichia coli* dihydrofolate reductase. *Biochemistry* 34 (48), 15671–15680. doi:10.1021/bi00048a011

Wang, M., and Liang, P. (2005). Interleukin-24 and its receptors. *Immunology* 114 (2), 166–170. doi:10.1111/j.1365-2567.2005.02094.x

Wang, M., Tan, Z., Zhang, R., Kotenko, S. V., and Liang, P. (2002). Interleukin 24 (MDA-7/MOB-5) signals through two heterodimeric receptors, IL-22r1/IL-20r2 and IL-20r1/IL-20r2. *J. Biol. Chem.* 277 (9), 7341–7347. doi:10.1074/jbc.M106043200

Wang, M., Tan, Z., Thomas, E. K., and Liang, P. (2004). Conservation of the genomic structure and receptor-mediated signaling between human and rat IL-24. *Genes &Immun.* 5 (5), 363–370. doi:10.1038/sj.gene.6364101

Wang, J., Liu, Y., Liu, Y., Zheng, S., Wang, X., Zhao, J., et al. (2019). Time-resolved protein activation by proximal decaging in living systems. *Nature* 569 (7757), 509–513. doi:10.1038/s41586-019-1188-1

Wang, J., Liu, Y., Liu, Y., Wang, C., and Chen, P. R. (2021). CAGE-prox: A unified approach for time-resolved protein activation in living systems. *Curr. Protoc.* 1 (6), e180. doi:10.1002/cpz1.180

Wu, N., Deiters, A., Cropp, T. A., King, D., and Schultz, P. G. (2004). A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* 126 (44), 14306–14307. doi:10.1021/ja040175z

Yilmaz, Z., Jedlitzke, B., and Mootz, H. D. (2022). "Design and preparation of photobodies: Light-activated single-domain antibody fragments," in *Single-domain antibodies*, 409–424.

Zahradník, J., Kolářová, L., Peleg, Y., Kolenko, P., Svidenská, S., Charnavets, T., et al. (2019). Flexible regions govern promiscuous binding ofIL-24 to receptorsIL-20R1 andIL-22R1. *FEBS J.* 286 (19), 3858–3873. doi:10.1111/febs.14945

Zahradník, J., Dey, D., Marciano, S., Kolářová, L., Charendoff, C. I., Subtil, A., et al. (2021a). A protein-engineered, enhanced yeast display platform for rapid evolution of challenging targets. *ACS Synth. Biol.* 10 (12), 3445–3460. doi:10.1021/acssynbio.1c00395

Zahradník, J., Marciano, S., Shemesh, M., Zoler, E., Harari, D., Chiaravalli, J., et al. (2021b). SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD *in vitro* evolution. *Nat. Microbiol.* 6 (9), 1188–1198. doi:10.1038/s41564-021-00954-4

Zhang, X., Pan, Y., Kang, S., and Gu, L. (2022). Combinatorial approaches for efficient design of photoswitchable protein-protein interactions as *in vivo* actuators. *Front. Bioeng. Biotechnol.* 10, 844405. doi:10.3389/fbioe.2022.844405

Zhou, W., Hankinson, C. P., and Deiters, A. (2020). Optical control of cellular ATP levels with a photocaged adenylate Kinase. *ChemBioChem* 21 (13), 1832–1836. doi:10.1002/cbic.201900757

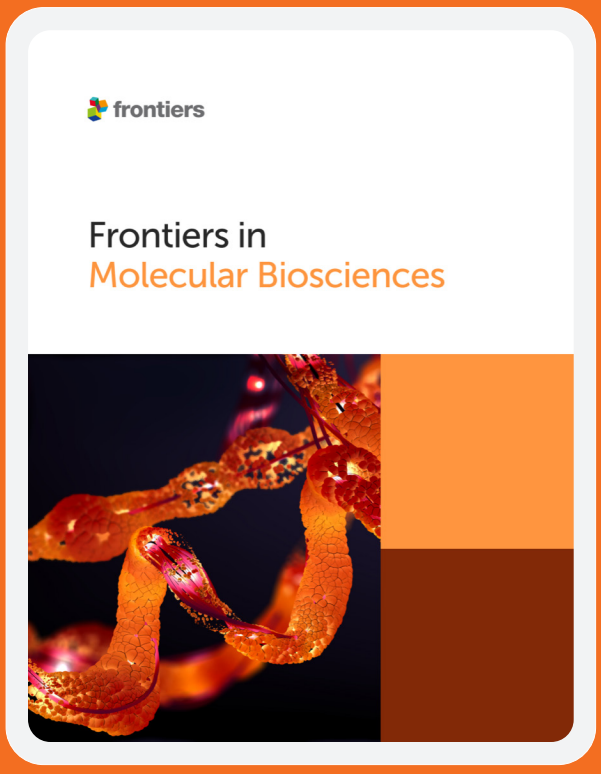
Frontiers in
Molecular Biosciences

Explores biological processes in living organisms on a molecular scale

Focuses on the molecular mechanisms underpinning and regulating biological processes in organisms across all branches of life.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact