VIEWS & NEWS

Enhanced quadruplet suppression in *E. coli* for expanding the genetic code

Chintan Soni & Abhishek Chatterjee*

Department of Chemistry, Boston College, Chestnut Hill, MA 02467, USA

*Correspondence: abhishek.chatterjee@bc.edu

A report from Costello and colleagues at the Scripps Research Institute uncovers codon usage as an important component for genetic code expansion using non-native quadruplet codons.

Using engineered tRNA/aminoacyl-tRNA synthetase (aaRS) pairs, genetic code expansion (GCE) technology enables the co-translation, site-specific incorporation of non-canonical amino acids (ncAAs) into proteins in living cells, creating exciting new ways to probe and engineer protein function.[1,2] This technology requires a unique codon to specify the site of ncAA incorporation. Traditionally, reassigned nonsense codons (mainly UAG) have been used for this purpose, but this approach can only support the incorporation of up to three distinct ncAAs at the same time due to the lack of additional 'blank' triplet codons.[3,4] The development of strategies to simultaneously introduce a larger number of distinct building blocks into peptides can create genetically encoded evolvable non-natural polymers with novel properties, impacting diverse disciplines from material science to medicine.

Remarkable progress has been made over the past decade to address this 'codon problem' for simultaneously incorporating multiple distinct ncAAs. The entire *Escherichia coli* genome was recently recoded to reduce the codon degeneracy of certain amino acids and reassign the 'liberated' sense codons to ncAAs.[5,6] However, extensive genome engineering can lead to unforeseen deleterious effects compromising host fitness.[7] An orthogonal unnatural base pair has also been introduced into the *E. coli* genetic alphabet, which can provide access to a significantly expanded codon space.[8] However, this strategy is technically demanding and the robustness of the resulting expression host is not yet fully established.

A simpler alternative strategy to overcome the codon problem involves the use of four-base or 'quadruplet' codons.[9-14] Unlike stop codons, quadruplet codons do not compete with release factors, and tapping into the four-base genetic code would greatly expand the number of coding

channels to accommodate novel ncAAs. Additionally, this strategy could be potentially more accessible, since it can be implemented in regular *E. coli* strains, without the need for complex genome engineering.

Although the feasibility of using quadruplet codons for ncAA incorporation has been well-established, broader adoption of this strategy has been precluded by poor efficiency. Major factors contributing to its low efficiency include competing recognition of their first three codons by endogenous tRNAs, and suboptimal decoding of the quadruplet codons by the native translation system (e.g., tRNA, ribosome). Not surprisingly, the quadruplet codons that do exhibit high suppression efficiency in *E. coli* are based on low-usage triplet codons, such as AGGN (where AGG is rare).[9]

Expanding the Bases

To improve the efficiency of four-base codon suppression, efforts have been made to engineer enhanced quadruplet-decoding tRNAs as well as the *E. coli* ribosome.[10,12,15] Nonetheless, their performance has remained subpar relative to UAG-mediated ncAA incorporation. A recent report by Alan Costello and colleagues at the Scripps Research Institute takes a systematic approach to enhance the scope of quadruplet decoding in *E. coli*, ultimately enabling multiplexed quadruplet codon suppression systems using wild-type ribosomes and unedited *E. coli* strains.[16]

It is well-established that mRNA sequence context can play a key role in determining ncAA incorporation efficiency.[17] Prior reports exploited these aspects to identify beneficial sequence contexts for quadruplet suppression.[13] Costello *et al*. systematically investigated context dependence of quadruplet suppression using a green fluorescent protein reporter, where the permissive Y151 site was replaced with quadruplet codons UAGA or AGGA, and the surrounding five codons were randomized to all synonymous codons.[16] These quadruplet codons were decoded in *E. coli* using established orthogonal tRNA/aaRS pairs.

Through a combination of fluorescence-assisted cell sorting and next-generation sequencing, the authors assessed how sequence context affected quadruplet decoding efficiency. Their analyses revealed that quadruplet codon suppression significantly improves when immediately followed by a high-usage triplet codon.[16] Furthermore, they found that off-target suppression by quadruplet-decoding tRNAs, which can cause frameshift during translation, negatively impacts reporter expression. This negative impact can be mitigated by systematically removing off-target appearances of the target quadruplet codon from the open reading frames in the expression plasmid through codon optimization.

The authors next focused on developing a set of mutually orthogonal tRNA/aaRS pairs to enable simultaneous incorporation of multiple different ncAAs into a protein in response to distinct quadruplet codons. The authors carefully cross-tested many of the tRNA/aaRS pairs that have been reported to date, and identified a set of 12 pairs that do not cross-react with each other. A subset of these pairs were further subjected to directed evolution to significantly enhance their

suppression efficiency to distinct quadruplet codons. Another challenge for multi-ncAA incorporation is substrate-level cross-reactivity, where the intended ncAA of one aaRS is inadvertently accepted by another, leading to a loss of fidelity.[18] To avoid this challenge, the substrate scope of each of the candidate tRNA/aaRS pairs were systematically screened. Finally, these mutually orthogonal quadruplet-decoding tRNA/aaRS pairs were combined with split-intein circular ligation (SICLOPPS)[19] technology to generate macrocyclic peptides containing up to three distinct ncAAs in cells.

In summary, this study offers an attractive alternative to multi-ncAA incorporation using quadruplet codons. It provides valuable insights into factors that enhance quadruplet decoding efficiency, such as context dependence and expression system optimizations. The careful evaluation of the cross-reactivities of available tRNA/aaRS pairs will also be a useful resource for developing high-fidelity multi-ncAA incorporation systems in the future.

However, the study also highlights the current limitation of multi-ncAA incorporation efforts. Despite the genetic encoding of hundreds of ncAAs, nearly all use either archaeal tyrosyl or pyrrolysyl pairs, and many synthetases cross-react with diverse ncAA substrates. While substrate promiscuity is typically beneficial, allowing rapid incorporation of numerous ncAAs without resource-intensive evolution campaigns, it compromises the fidelity of multi-ncAA incorporation experiments. Expanding the ncAA substrate scope of additional tRNA/aaRS pairs and creating aaRS variants with more fine-tuned substrate scopes are necessary to overcome these limitations.

Additionally, adding diverse non-α-amino acid monomers to the genetic code would further expand the scope of such technology, enabling ribosomal synthesis of novel polymers.[20,21] Harnessing the full potential of the multi-ncAA incorporation technology in this manner will have countless high-impact applications across diverse disciplines.

Acknowledgements

This work was supported by NIGMS (R35GM136437 to AC) and NSF Center for Genetically Encoded Materials (C-GEM; CHE 2002182)

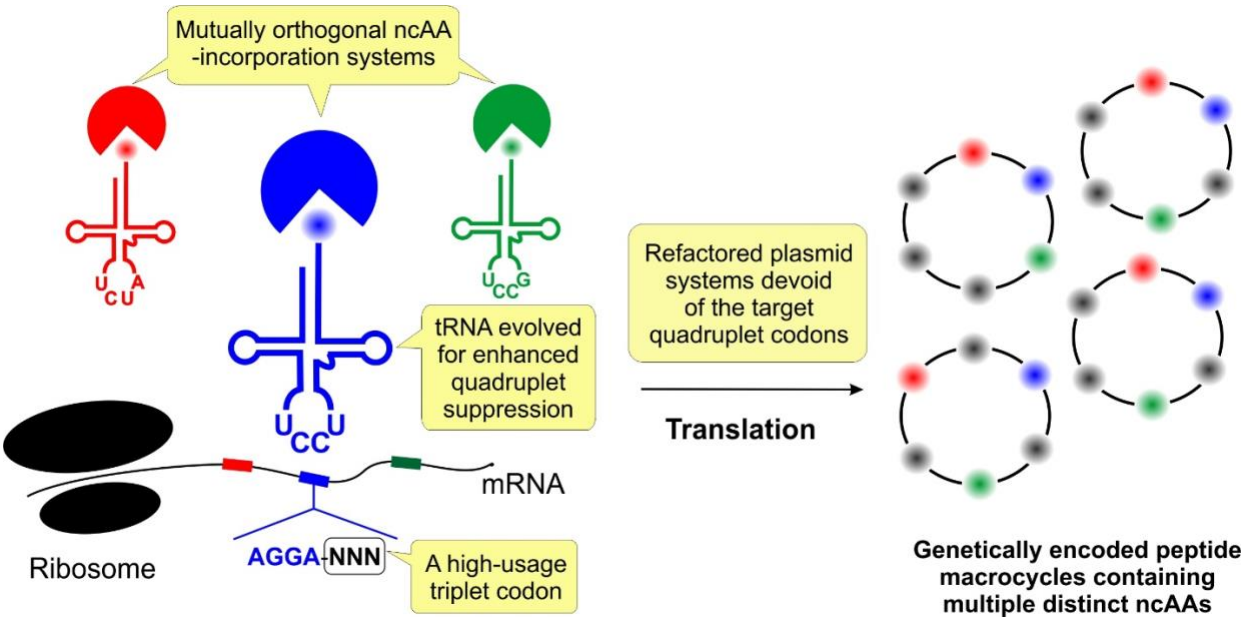

Fig. 1 Costello *et al*. improved the efficiency of ncAA incorporation at quadruplet codons by optimizing mRNA sequence contexts and the suppressor tRNA sequence. The development of multiple mutually orthogonal quadruplet-suppressing systems in this manner enabled ribosomal synthesis of peptide macrocycles containing multiple ncAAs.

References

1. Chin JW. Expanding and reprogramming the genetic code. Nature 2017;550(7674):53-60, doi:10.1038/nature24031
2. Young DD, Schultz PG. Playing with the Molecules of Life. ACS Chem Biol 2018;13(4):854-870, doi:10.1021/acschembio.7b00974
3. Italia JS, Addy PS, Erickson SB, et al. Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. Journal of the American Chemical Society 2019;141(15):6204-6212, doi:10.1021/jacs.8b12954
4. Osgood AO, Zheng Y, Roy SJS, et al. An Efficient Opal-Suppressor Tryptophanyl Pair Creates New Routes for Simultaneously Incorporating up to Three Distinct Noncanonical Amino Acids into Proteins in Mammalian Cells. Angewandte Chemie (International ed in English) 2023;62(19):e202219269, doi:10.1002/anie.202219269
5. Fredens J, Wang K, de la Torre D, et al. Total synthesis of Escherichia coli with a recoded genome. Nature 2019;569(7757):514-518, doi:10.1038/s41586-019-1192-5
6. Ostrov N, Landon M, Guell M, et al. Design, synthesis, and testing toward a 57-codon genome. Science (New York, NY) 2016;353(6301):819-22, doi:10.1126/science.aaf3639
7. Nyerges A, Chiappino-Pepe A, Budnik B, et al. Synthetic genomes unveil the effects of synonymous recoding. bioRxiv : the preprint server for biology 2024, doi:10.1101/2024.06.16.599206
8. Zhang Y, Ptacin JL, Fischer EC, et al. A semi-synthetic organism that stores and retrieves increased genetic information. Nature 2017;551(7682):644-647, doi:10.1038/nature24659

9. Magliery TJ, Anderson JC, Schultz PG. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. Journal of molecular biology 2001;307(3):755-69, doi:10.1006/jmbi.2001.4518
10. Anderson JC, Wu N, Santoro SW, et al. An expanded genetic code with a functional quadruplet codon. Proceedings of the National Academy of Sciences of the United States of America 2004;101(20):7566-71, doi:10.1073/pnas.0401517101
11. Dunkelmann DL, Oehm SB, Beattie AT, et al. A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design. Nature chemistry 2021;13(11):1110-1117, doi:10.1038/s41557-021-00764-5
12. DeBenedictis EA, Carver GD, Chung CZ, et al. Multiplex suppression of four quadruplet codons via tRNA directed evolution. Nature communications 2021;12(1):5706, doi:10.1038/s41467-021-25948-y
13. Chen Y, He X, Ma B, et al. Noncanonical amino acid mutagenesis in response to recoding signal-enhanced quadruplet codons. Nucleic acids research 2022;50(16):e94, doi:10.1093/nar/gkac474
14. Guo J, Niu W. Genetic Code Expansion Through Quadruplet Codon Decoding. Journal of molecular biology 2022;434(8):167346, doi:10.1016/j.jmb.2021.167346
15. Neumann H, Wang K, Davis L, et al. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature 2010;464(7287):441-4, doi:10.1038/nature08817
16. Costello A, Peterson AA, Lanster DL, et al. Efficient genetic code expansion without host genome modifications. Nature biotechnology 2024, doi:10.1038/s41587-024-02385-y
17. Chemla Y, Ozer E, Algov I, et al. Context effects of genetic code expansion by stop codon suppression. Current opinion in chemical biology 2018;46(146-155, doi:10.1016/j.cbpa.2018.07.012
18. Zheng Y, Mukherjee R, Chin MA, et al. Expanding the Scope of Single- and Double-Noncanonical Amino Acid Mutagenesis in Mammalian Cells Using Orthogonal Polyspecific Leucyl-tRNA Synthetases. Biochemistry 2018;57(4):441-445, doi:10.1021/acs.biochem.7b00952
19. Scott CP, Abel-Santos E, Wall M, et al. Production of cyclic peptides and proteins in vivo. Proceedings of the National Academy of Sciences of the United States of America 1999;96(24):13638-43, doi:10.1073/pnas.96.24.13638
20. Dunkelmann DL, Piedrafita C, Dickson A, et al. Adding α,α-disubstituted and β-linked monomers to the genetic code of an organism. Nature 2024;625(7995):603-610, doi:10.1038/s41586-023-06897-6
21. Soni C, Prywes N, Hall M, et al. A Translation-Independent Directed Evolution Strategy to Engineer Aminoacyl-tRNA Synthetases. ACS central science 2024;10(6):1211-1220, doi:10.1021/acscentsci.3c01557