Recoded gene circuits for multiplexed genetic code expansion

Quadruplet codons allow multiplexing of non-canonical amino acids within single polypeptides in living cells. We show that including high-usage triplet codons after quadruplet codons can improve their decoding efficiency in genetic circuits, which allowed us to develop a system for the programmable biosynthesis of exotic macrocyclic peptides in cells.

This is a summary of:
Costello, A. et al. Efficient genetic code expansion without host genome modifications. *Nat. Biotechnol*. https://doi.org/10.1038/s41587-024-02385-y (2024).

Published online: 11 September 2024

The mission

Expanding the range of proteinaceous amino acids beyond the approximately 20 canonical ones has the potential to fundamentally change the physicochemical parameters of genetically encoded proteins. Technologies developed since the 1990s to incorporate non-canonical amino acids (ncAAs) into the proteins of living cells have typically depended on the suppression of amber (TAG) stop codons[1]. Although amber codon suppression is a well-established tool, to incorporate multiple distinct ncAAs into a single polypeptide requires the development of additional codons for ncAA assignment[2,3]. Quadruplet codons offer such an opportunity for diversification by encoding 255 possible ncAAs (plus 1 stop codon). However, previous efforts to use quadruplet codons with orthogonal aminoacyl-tRNA synthetase–tRNA pairs have suffered from the low efficiency of ncAA incorporation and an unexplained amino acid sequence context dependency. Thus, we sought to develop multiple such synthetase–tRNA pairs capable of high-efficiency ncAA incorporation using quadruplet codons within a broadly applicable framework and easily multiplexable setting.

The solution

We first explored the mRNA sequence context effects of quadruplet decoding by examining the impact of triplet codon usage at the residues immediately surrounding a quadruplet codon (Fig. 1a). We sorted library members with either high or low superfolder GFP expression using fluorescence-activated cell sorting and then identified triplet codon usage biases that may influence amber or quadruplet codon decoding using next-generation sequencing. To our surprise, we found a preference for high-usage codons immediately downstream of quadruplet codons. This observation inspired us to pursue a synthetic genetic recoding approach in which we exclusively used a single, high-usage codon for each amino acid in the protein-coding elements of our genetic circuits — a single triplet codon for each of the 20 canonical amino acids and a quadruplet codon for each ncAA. We observed robust quadruplet decoding in these circuits and demonstrated that competition between quadruplet codons and overlapping triplet codons decreased quadruplet suppression.

The ability to multiplex ncAA incorporation depends on all of the synthetase–tRNA pairs being mutually orthogonal to one another and to the host machinery[4]. To broaden the available synthetase–tRNA pairs, we assayed more than 800 combinations of previously described synthetases and tRNAs to establish their activities and mutual orthogonality. We selected five mutually orthogonal pairs for quadruplet decoding optimization. For each of the selected tRNAs we created targeted mutagenesis libraries, which we subjected to iterative rounds of selective pressure to increase their quadruplet decoding and ncAA incorporation efficiencies. We then optimized the co-expression of five of these evolved tRNAs by building a tRNA operon, which further improved ncAA-dependent decoding. Last, we screened for synthetases capable of aminoacylating their cognate tRNAs with a broad substrate repertoire. Together, our genetic circuits could then collectively incorporate 47 unique ncAAs into target proteins.

To demonstrate the utility of our system in cells, we focused on the biosynthesis of programmable macrocyclic peptides. Despite their therapeutic potential across disease modalities, discovery of macrocycles with user-defined bioactivities with existing technologies remains challenging. We first adapted and optimized a split-intein system for circular ligation of peptides and proteins[5] for peptide macrocyclization to function alongside our ncAA-incorporating genetic circuits. We used this to demonstrate the incorporation of up to three distinct ncAAs into biosynthesized macrocycles, which altered their chemical properties (Fig. 1b).

The implications

The remarkable protein diversity of the natural world is accomplished by combining about 20 amino acids. Our ability to solve global problems with biologically inspired solutions may be accelerated by strategies to easily build novel, chemically defined biopolymers. The strategies reported in our study will complement established and emerging applications of directed evolution and protein design.

Translating multiple quadruplet codons opens the door to a large combinatorial space whereby protein structure and function can be probed and expanded in high throughput. Indeed, a single transcript encoding the five quadruplet codons we report could yield nearly 70,000 unique peptides when these ncAAs are used in combination with one another. We envision even more exotic monomers — including alternative backbone chemistries — increasing the broad applicability of our system and the development of increasingly diverse macrocycles, hard-to-evolve enzymes and new-to-nature bioactivities.

David L. Lanster, Alan Costello & Ahmed H. Badran
Scripps Research, La Jolla, CA, USA.

EXPERT OPINION

"This manuscript's key insight is that, by using codon compression in a template (removing rare codons), one can more efficiently enforce quadruplet codon readthrough. It also makes sense that putting higher usage codons on either side of the quadruplet codon would help reinforce the reading frameshift. The beauty of this concept is that it is very easy to incorporate into many types of workflows as it just requires thoughtful gene design." **Matthew Hartman, Virginia Commonwealth University, Richmond, VA, USA.**

FIGURE

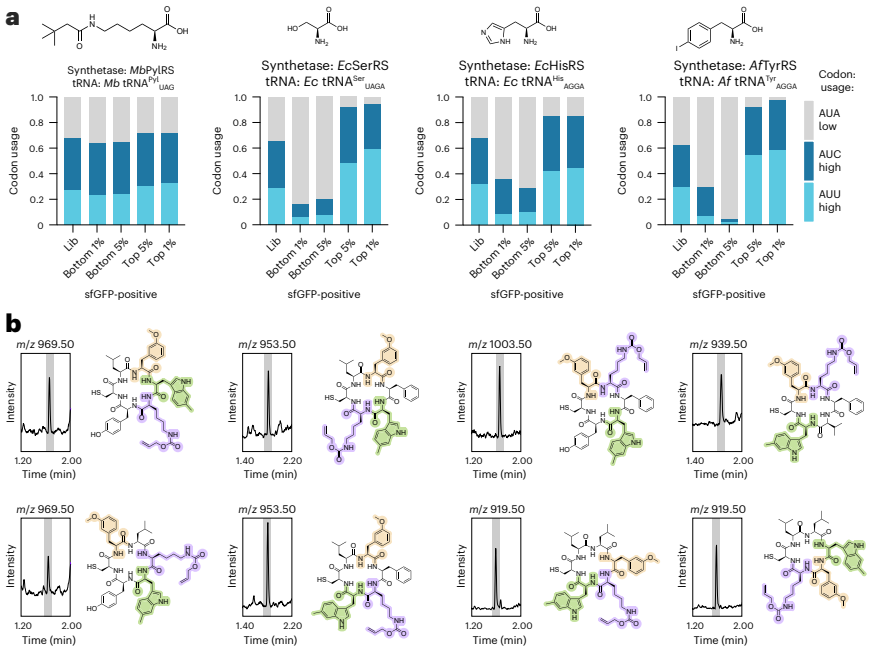

Fig. 1 | Context effects of quadruplet decoding allow the biosynthesis of non-canonical macrocycles. **a**, Using synthetase–tRNA pairs from *Methanosarcina barkeri* (*Mb*), *Escherichia coli* (*Ec*) and *Archaeoglobus fulgidus* (*Af*), we investigated how usage of codons immediately downstream of the quadruplet codons (UAGA and AGGA) improves suppression efficiency, by measuring the abundance of high- or low-usage codons using next-generation sequencing and superfolder GFP (sfGFP) expression. A UAG decoding control helped identify rules specific to quadruplet decoding by comparing the top and bottom 1% or 5% of sfGFP-expressing cells. **b**, Biosynthesis of macrocycles in expression strains that contain up to three non-canonical amino acids demonstrates the diverse chemical substituents that can be achieved using our system. Lib, library; *m/z*, mass-to-charge ratio. © 2024, Costello, A. et al.

BEHIND THE PAPER

Our lab's long-term goal is to develop proteins and enzymes with activities beyond those found in nature. We envision this being possible by combining genetic code expansion with directed evolution and modern protein design paradigms. To achieve this, genetic code expansion efficiency must be sufficiently high to enable the discovery of new-to-nature bioactivities.

We are excited by the impact of quadruplet codons since they allow multiplexing ncAAs and implementation into engineered strains with minimal genomic changes. While quadruplet decoding can be inefficient, quadruplet-decoding tRNA evolution can improve this process. This led us to question the contribution of the mRNA and codon usage to quadruplet decoding, which we investigated using the synonymous codon approach outlined in the manuscript. Establishing the impact of codon usage on quadruplet decoding was a key moment in this study that propelled the bioengineering and directed evolution efforts that followed. **A.H.B. & D.L.L.**

REFERENCES

1. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989).
   This seminal paper from the Schultz Lab describes the first use of amber suppression to site-specifically incorporate a non-canonical amino acid.
2. Dunkelmann, D. L., Oehm, S. B., Beattie, A. T. & Chin, J. W. A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design. *Nat. Chem.* **13**, 1110–1117 (2021).
   This paper describes quadruplet decoding with orthogonal ribosomes.
3. DeBenedictis, E. A., Carver, G. D., Chung, C. Z., Soll, D. & Badran, A. H. Multiplex suppression of four quadruplet codons via tRNA directed evolution. *Nat. Commun.* **12**, 5706 (2021).
   This paper describes our previous efforts to decode quadruplet codons with *E. coli* tRNAs.
4. Costello, A. & Badran, A. H. Synthetic biological circuits within an orthogonal central dogma. *Trends Biotechnol.* **39**, 59–71 (2021).
   This perspective article highlights the importance of orthogonality in genetic circuits aiming to expand the central dogma.
5. Townend, J. E. & Tavassoli, A. Traceless production of cyclic peptide libraries in *E. coli*. *ACS Chem. Biol.* **11**, 1624–1630 (2016).
   This paper developed the split-intein system for circular ligation of peptides and proteins we adapted for peptide macrocyclization.

FROM THE EDITOR

"The manuscript by Costello et al. uses a simple but elegant approach to efficiently expand the genetic code. This includes optimizing the codon usage and using codon compression. This approach will hopefully move the field forward substantially." **Editorial Team, *Nature Biotechnology*.**