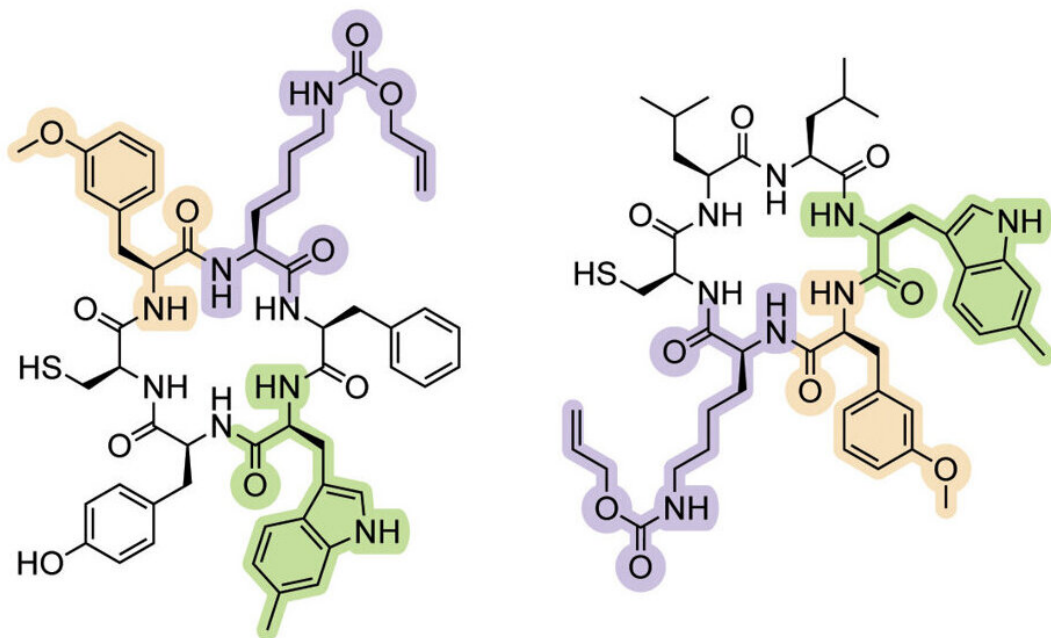# Scientists expand the genetic alphabet to create new proteins

September 11 2024

Examples of the >100 macrocycles generated in this study. Colored components represent new-to-nature amino acids that were incorporated into either peptide. Credit: Scripps Research

It's a dogma taught in every introductory biology class: Proteins are composed of combinations of 20 different amino acids, arranged into

diverse sequences like words. But researchers trying to engineer biologic molecules with new functions have long felt limited by those 20 basic building blocks and strived to develop ways of putting new building blocks—called non-canonical amino acids—into their proteins.

Now, scientists at Scripps Research have designed a new paradigm for easily adding non-canonical amino acids to proteins. Their approach, described in *Nature Biotechnology* on September 11, 2024, revolves around using four RNA nucleotides—rather than the typical three—to encode each new amino acid.

"Our goal is to develop proteins with tailored functions for applications in fields spanning bioengineering to drug discovery," says senior author Ahmed Badran, Ph.D., an assistant professor of chemistry at Scripps Research. "Being able to incorporate non-canonical amino acids into proteins with this new method gets us closer to that goal."

For a cell to produce any given protein, it must translate a strand of RNA into a string of amino acids. Every three nucleotides of RNA, called a codon, correspond to one amino acid. But many amino acids have more than one possible codon; for instance, RNA reading the sequences UAU and UAC both correspond to the amino acid tyrosine. It's the job of small molecules called transfer RNAs (tRNAs) to link each amino acid to its corresponding codons.

Recently, researchers aiming to add completely new amino acids to a protein have created strategies to reassign a codon. For instance, the UAU codon could be linked to a new amino acid by changing the tRNA for UAU; this would result in UAU being read by the cell as corresponding to a building block other than tyrosine. But at the same time, every instance of UAU in the cell's genome would need to become UAC, in order to prevent the new amino acid from being integrated into thousands of other proteins where it doesn't belong.

"Creating free codons by whole genome recoding can be a powerful strategy, but it can also be a challenging undertaking since it requires considerable resources to build new genomes," says Badran. "For the organism itself, it can be difficult to predict how such codon changes influence genome stability and host protein production."

Badran and his colleagues wanted to create an efficient plug-and-play strategy that would only incorporate the chosen non-canonical amino acid(s) into specific sites in a target protein, without disrupting the cell's normal biology or requiring the entire genome to be edited. That meant using tRNA that wasn't already assigned to an amino acid. Their solution: a four-nucleotide codon.

The team knew that in a few situations—such as bacteria quickly adapting to resist drugs—four-nucleotide codons had naturally evolved. So, in their new work, the researchers studied what caused cells to use a codon with four nucleotides rather than three. They discovered that the identities of the sequences nearby to the four-base codon were critical—frequently used codons enhanced how the cell could read a four-nucleotide codon to incorporate a non-canonical amino acid.

Badran's group then tested whether they could alter the sequence of a single gene so that it had a new four-nucleotide codon that would be correctly used by the cell.

The method worked: When the researchers surrounded a target site with three-letter, frequently used codons and maintained sufficient levels of the four-nucleotide tRNA, the cell incorporated any new amino acid that was attached to the corresponding four-letter tRNA. The research team repeated the experiment with 12 different four-nucleotide codons and then used the technique to design more than 100 new cyclic peptides—called macrocycles—with up to three non-canonical amino acids in each.

"These cyclic peptides are reminiscent of bioactive small molecules that one might find in nature," says Badran. "By capitalizing on the programmability of protein synthesis and the diversity of building blocks accessible by this approach, we can create new-to-nature small molecules that will have exciting applications in drug discovery."

He adds that, compared with previous approaches to non-canonical amino acid incorporation, this new method is easy to use since it involves altering only one gene rather than a cell's entire genome. Additionally, more non-canonical amino acids could be used in a single protein since there are more possible four-nucleotide codons than three-nucleotide ones.

"Our results suggest that one can now easily and effectively incorporate non-canonical amino acids at diverse sites in a wide array of proteins," says Badran. "We're excited about these possibilities for our ongoing work and to provide this capability to the broader community."

He notes that the technique could be used to re-engineer existing proteins—or create entirely new ones—that have utility in a range of sectors, including medicine, manufacturing and chemical sensing.

In addition to Badran, authors of the study, "Efficient Genetic Code Expansion Without Host Genome Modifications," are Alan Costello, Alexander Peterson, David Lanster and Zhiyi Li of Scripps; and Gavriela Carver of Princeton University.

More information: Alan Costello et al, Efficient genetic code expansion without host genome modifications, *Nature Biotechnology* (2024). DOI: 10.1038/s41587-024-02385-y