Despite a long history for flavin chemistry, unprecedented reactions continue to emerge. Notable recent discoveries include flavin prenylations that support enzyme-catalyzed (de)carboxylation[4]; photochemical activation of flavin for decarboxylation[5] and stereoselective cyclizations[6]; formation of superoxidized flavin N5-oxides for the 4-electron oxidation of polyketides[7] and flavin N5-centred C1 transfer reactions[8]. This catalytic and chemical versatility of flavin expands the utility of flavoenzymes as (photo)biocatalysts, broadening the reaction types available to biocatalysis toolkits. It also extends the capabilities of natural enzymes to engage in a wide range of cellular transformations, including light emission and light-dependent signaling[9,10]. Further surprises will no doubt come to light to illustrate yet more chemical versatility, especially as we continue to discover new flavoenzymes and engineer them toward novel function. That selected flavoproteins can use a flavin-N5-aminoperoxide adduct for monooxygenation is especially noteworthy. This finding breaks with dogma; flavoprotein monooxygenase reactions were inferred to progress exclusively through a covalent flavin-C4a-(hydro)peroxide intermediate prior to this discovery.

The work of Matthews et al.[2] extends a previous study of flavin N-oxide identification in the enzyme EncM[7] by showing that the N5 atom can also support flavin-N5-aminoperoxide formation. The authors investigated RutA as a model enzyme, which—alongside two other enzymes, DszA and HcbA1—had previously been suggested to catalyze substrate C–heteroatom cleavage, forming a flavin N5-oxide coproduct. The original catalytic mechanism had invoked the use of C4a-flavin-hydroperoxide, in-line with current dogma relating to flavin-dependent monooxygenases. The study now reported in this issue fundamentally challenges this assertion and suggests that monooxygenation can occur through an alternative flavin-N5-aminoperoxide intermediate[2]. Matthews et al.[2] used O_2-pressurized X-ray crystallography to show O_2 binding close to the N5 atom of the flavin mononucleotide (FMN) isoalloxazine ring at high pressure, whereas electron density for the modeled O_2 was less clear at lower pressures and absent under anaerobic conditions. Unusually, the proposed hydrophobic O_2-binding site was found on the *re*-side of the flavin, whereas the RutA substrate uracil was found to bind on the opposite *si*-face. There are therefore profound implications of this mode of binding: following formation of the flavin-N5-aminoperoxide intermediate, flavin N5-inversion is needed to position the reactive peroxide species close to the acceptor uracil bound on the *si*-face of FMN. Computations showed a small free-energy barrier for N5-inversion and revealed near attack conformers for O-atom insertion into uracil. Overall, this new reaction mechanism was consistent with detailed density functional theory (DFT) computations, mutagenesis and rapid mixing spectroscopy studies. Furthermore, the authors showed that the proposed O_2-binding pocket is conserved in closely related group C flavin-dependent monooxygenases, but not in other classes of flavoprotein monooxygenase, indicating that this mechanism is likely conserved in other enzymes, but not across the entire flavin-dependent monooxygenase family.

Direct observation of flavin-peroxide adducts is non-trivial, and the work of Matthews et al. infers existence of the presumably highly unstable N5-aminoperoxide species[2]. Unlike the canonical C4a-(hydro)peroxide-based pathway, the N5-peroxide mediated oxygen insertion does not occur with formal oxidation of the product, yielding the 4-electron oxidized flavin N5-oxide instead. The redox neutral C-X bond breakage catalyzed by RutA, DszA and HcbA1 therefore mimics a hydrolysis reaction and requires four electrons to return flavin to the reduced starting state. The inherent chemical repertoire associated with N5-oxidative chemistry and the extent to which nature makes use of this remains unclear. The future will tell whether it can mimic the wealth of C4a-based transformations. ❐

David Leys and Nigel S. Scrutton
Manchester Institute of Biotechnology and Department of Chemistry, School of Natural Sciences, The University of Manchester, Manchester, UK.
e-mail: david.leys@manchester.ac.uk; nigel.scrutton@manchester.ac.uk

Published online: 22 April 2020
https://doi.org/10.1038/s41589-020-0523-z

References
1. Romero, E., Gómez Castellanos, J. R., Gadda, G., Fraaije, M. W. & Mattevi, A. *Chem. Rev.* **118**, 1742–1769 (2018).
2. Matthews, A. et al. *Nat. Chem. Biol.* https://doi.org/10.1038/s41589-020-0476-2 (2020).
3. Leys, D. & Scrutton, N. S. *Curr. Opin. Struct. Biol.* **41**, 19–26 (2016).
4. Payne, K. A. P. et al. *Nature* **522**, 497–501 (2015).
5. Sorigué, D. et al. *Science* **357**, 903–907 (2017).
6. Black, M. J. et al. *Nat. Chem.* **12**, 71–75 (2020).
7. Teufel, R. et al. *Nature* **503**, 552–556 (2013).
8. Mishanina, T. V. et al. *Science* **351**, 507–510 (2016).
9. Conrad, K. S., Manahan, C. C. & Crane, B. R. *Nat. Chem. Biol.* **10**, 801–809 (2014).
10. Kutta, R. J., Archipowa, N. & Scrutton, N. S. *Phys. Chem. Chem. Phys.* **20**, 28767–28776 (2018).

Competing interests
The authors declare no competing interests.

Check for updates

GENETIC CODE EXPANSION

Ribosomes read new languages

Using new letters of DNA to encode information promises to expand the genetic code in a transformative way. A new semisynthetic organism has been created that uses 67 codons for protein biosynthesis, with three new codons based on unnatural nucleotides.

Michael C. Jewett and Michael J. Hammerling

The process of molecular translation enables sequence-defined polymerization of amino acids into proteins from a messenger RNA (mRNA) template. The incredible catalytic capability of the protein biosynthesis machinery motivates efforts to develop and harness engineered versions for biotechnological application. In nature, however, a limited set of amino acid monomers is utilized, constraining the diversity of possible proteins. Expanding the set of

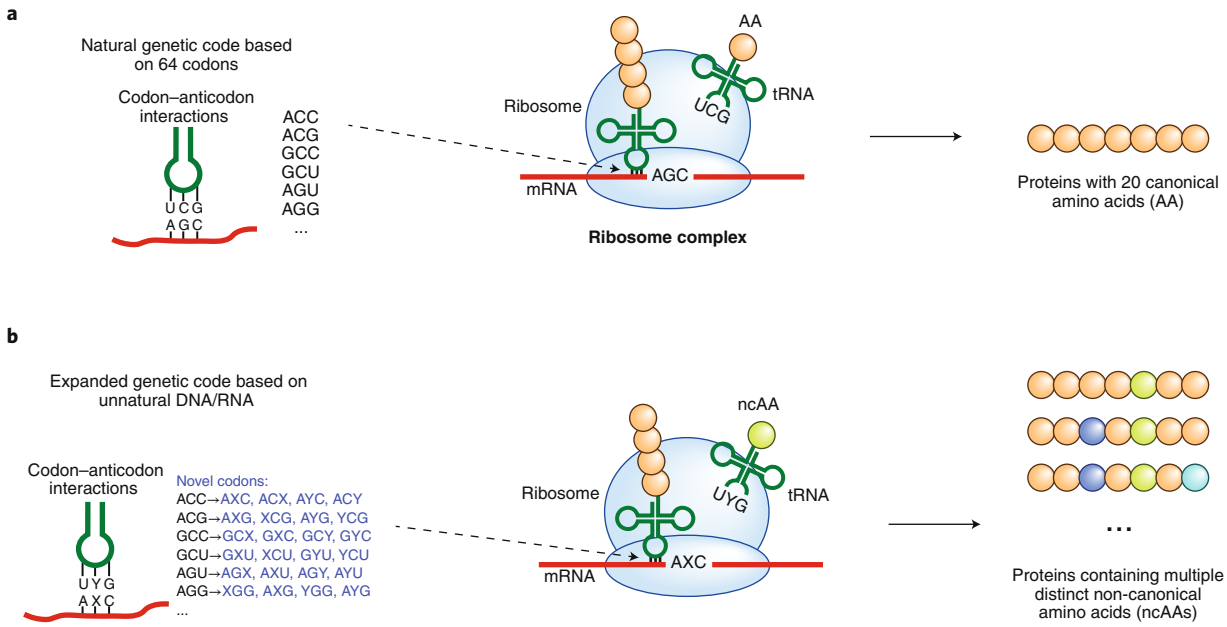

Fig. 1 | Unnatural base pairs expand the genetic code to facilitate synthesis of proteins with multiple, distinct non-canonical amino acids (ncAAs). a, Nature uses a near-universal genetic code to define how 64 triplet codons are decoded in a template-directed manner on the ribosome to synthesize proteins. All proteins are synthesized from a limited set of building blocks: the 20 canonical amino acids. **b**, Fischer et al.[3] have created a semisynthetic organism that uses 67 triplet codons, some including unnatural base pairs from an expanded genetic alphabet, to make proteins with multiple, distinct ncAAs. dNaM = X. dTPT3 = Y.

monomers could yield new classes of designer therapeutics, genetically encoded materials, and artificial enzymes, as well as organisms with new forms, functions, and biocontainment strategies[1,2]. In this issue, Fischer et al. describe the first living organism with 67 triplet-based codons that can simultaneously decode three unnatural base pairs and incorporate multiple non-canonical amino acids (ncAAs) into one protein[3] (Fig. 1).

The genetic code defines the relationship between DNA bases in a gene and the protein sequence it encodes. Notably, the code is degenerate (i.e., the 64 codons code for 20 canonical amino acids and 3 stop codons). This means that individual codons can be assigned a new meaning while the organism retains the ability to incorporate all canonical amino acids. This feature has inspired synthetic biologists to explore methods to engineer the code and invent rules for genetic reprogramming.

Several strategies have been used to reprogram the genome to facilitate incorporation of ncAAs into proteins. In one approach, all occurrences of the amber stop codon (TAG) have been genomically recoded to the TAA codon, which permits the deletion of release factor 1 and reassignment of the amber codon translation function to a ncAA[4]. To produce proteins with multiple, distinct ncAAs, recoding efforts are underway to generate strains with multiple open codons for efficient ncAA incorporation[5,6]. However, such strategies may require substantial re-engineering of natural transfer RNA (tRNA) charging systems, confounding these efforts. Quadruplet codons can alternatively be used[7], but the translation apparatus has been delicately tuned by evolution to perform optimally with a triplet codon.

A compelling strategy to surmount these challenges involves expanding the genetic alphabet. Leading up to this paper, an unnatural base pair was developed that relies on hydrophobic and packing interactions arising from the unnatural nucleotides dNaM and dTPT3. This base pair laid the foundation for a semisynthetic organism that could store increased information in DNA and faithfully retrieve it as mRNA[8]. Although an unnatural base pair has been used to encode and translate genetic information into proteins[9], the ability of the ribosome to efficiently and dependably decode multiple unnatural codons has yet to be assessed.

Here, the authors carry out a systematic analysis of the translation of unnatural codons in a semisynthetic organism. They first determined which unnatural codons can be efficiently decoded by the ribosome and found that dNaM and dTPT3 used at the second or third position could be efficiently decoded. Of note, there was also a requirement for at least one G–C pair in the codon, perhaps to compensate for the weaker interaction of the unnatural bases. Taken together, the analyses identified 9 unnatural codon–anticodon pairs that could be used for molecular translation in the semisynthetic organism, resulting in a possible genetic code with 73 available codons.

The authors identified three unnatural base pairs that are orthogonal and can be used for site-specific incorporation of multiple ncAAs into one protein. About 96% of protein produced using three unnatural codons had the correct amino acids incorporated at these positions, demonstrating the efficiency and orthogonality of their optimized system.

This elegant work by the authors represents an exciting frontier for using an expanded genetic alphabet in a living organism, but it also raises many questions. For example, how far can the technology be pushed to enable high-yielding expression of proteins with multiple, distinct ncAAs? Additionally, what is the maximal number of codons that can function in a living organism? Could more than 100 codons be decoded in a single organism? What about 200? Recently, a genetic system based on 'hachimoji' DNA and RNA was developed, comprising two new base pairs, named S and B as well as P and Z, on the basis of hydrogen bonding rules[10]. The 4 additional bases made available by this advance

theoretically creates 512 potential codons (8^3), moving such questions out of being purely hypothetical and one step closer to being possible.

Though many interesting questions remain, the potential of this technology for repurposing molecular translation is clear. The unnatural base pair technology described liberates the genetic code from evolutionary constraints imposed by the decoding needs of the host organism. This offers exciting opportunities to dramatically increase the resolution at which we can manipulate the protein biosynthesis machinery to create new design rules for constructive biology that look beyond what does exist to what can exist. We can even begin to imagine a world in which organisms are engineered to decipher entirely alternative genetic codes. ❐

Michael C. Jewett ✉ and Michael J. Hammerling

Department of Chemical and Biological Engineering and Center for Synthetic Biology, Northwestern University, Evanston, IL, USA.

✉*e-mail: m-jewett@northwestern.edu*

Published online: 6 April 2020
https://doi.org/10.1038/s41589-020-0522-0

References

1. Hammerling, M. J., Krüger, A. & Jewett, M. C. *Nucleic Acids Res.* **48**, 1068–1083 (2020).
2. Arranz-Gibert, P., Vanderschuren, K. & Isaacs, F. J. *Curr. Opin. Chem. Biol.* **46**, 203–211 (2018).
3. Fischer, E. C. et al. *Nat. Chem. Biol.* https://doi.org/10.1038/s41589-020-0507-z (2020).
4. Lajoie, M. J. et al. *Science* **342**, 357–360 (2013).
5. Ostrov, N. et al. *Science* **353**, 819–822 (2016).
6. Fredens, J. et al. *Nature* **569**, 514–518 (2019).
7. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. *Nature* **464**, 441–444 (2010).
8. Malyshev, D. A. et al. *Nature* **509**, 385–388 (2014).
9. Zhang, Y. et al. *Nature* **551**, 644–647 (2017).
10. Hoshika, S. et al. *Science* **363**, 884–887 (2019).

Competing interests

The authors declare no competing interests.