HHS Public Access

Author manuscript

*Biochemistry*. Author manuscript; available in PMC 2019 April 17.

Published in final edited form as:

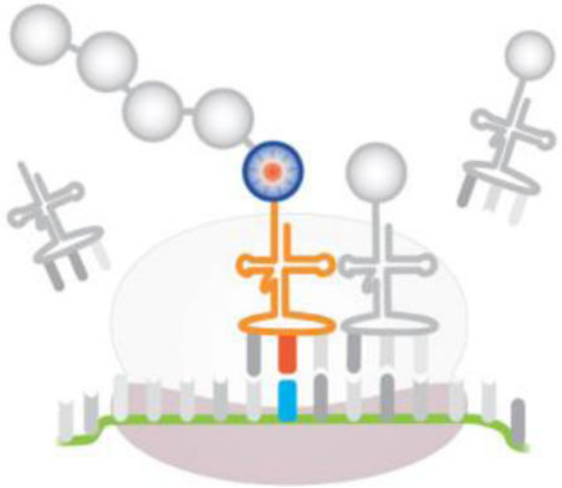
*Biochemistry*. 2018 April 17; 57(15): 2177–2178. doi:10.1021/acs.biochem.8b00013.

# Semi-Synthetic Organisms with Expanded Genetic Codes

**Yorke Zhang** and **Floyd E. Romesberg**

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 USA

## Graphical Abstract



Almost 20 years ago, Peter Schultz revolutionized the role of chemists in the biological sciences by developing the first cells with an expanded genetic code. These *Escherichia coli* cells expressed an orthogonal tRNA–synthetase pair: a Tyr tRNA, with an anticodon recoded to suppress an amber stop codon; and a cognate synthetase from *Methanocaldococcus jannaschii*, which was evolved to selectively charge the tRNA with Tyr analogs, thus enabling the site-specific incorporation of non-canonical amino acids (ncAAs) into proteins. The method has since been expanded to include suppression of up to two, different stop codons and the use of several other orthogonal tRNA–synthetase pairs (most notably the Pyl tRNA–synthetase pair from *Methanosarcina barkeri/mazei*) which expands the scope of ncAAs that may be incorporated into proteins, in both prokaryotic and eukaryotic host organisms. By virtue of unique chemical or physical properties, these ncAAs can bestow proteins with novel properties and enable applications ranging from the development of novel protein therapeutics to the creation of cells with new forms and functions.

As enabling as the method has been, stop codon suppression comes with a limitation: competition with endogenous release factors results in difficult-to-predict sequence- and

*Corresponding Author floyd@scripps.edu.

Notes

F.E.R. has a financial interest (shares) in Synthorx, Inc., a company that has commercial interests in the UBP.

position-dependent suppression efficiencies, as well as deleterious misincorporation of ncAAs into endogenous proteins. This has been overcome by eliminating all instances of the amber stop codon throughout the *E. coli* genome, allowing for the deletion of the competing release factor and unambiguous dedication of the stop codon to ncAA incorporation.[1] To liberate more codons, efforts are now focused on complete genome re-synthesis and the elimination of all instances of specific sense codons, allowing for their reassignment to ncAAs as well.[2] However, sense codons are not truly degenerate due to their contribution to the kinetics of transcription and translation. Moreover, total genome synthesis remains a challenge and is cost-intensive, especially for the larger genomes of eukaryotes. In contrast to natural codon reassignment, the expansion of the genetic alphabet could create a virtually unlimited number of genuinely new codons, free of any natural function, for an unrestricted expansion of the genetic code without requiring genome synthesis.

For nearly 20 years, we have worked to identify a pair of synthetic nucleotides that act as fifth and sixth letters of the genetic alphabet via the formation of an unnatural base pair (UBP). After the synthesis and evaluation of nearly 200 unnatural nucleotides, we identified a family of UBPs exemplified by d**NaM**–d**TPT3** (Fig. 1). The unnatural nucleotides of these UBPs selectively interact via hydrophobic and packing forces, as opposed to the complementary hydrogen bonding that underlies the formation of the natural base pairs. In 2014, we reported that when the corresponding unnatural triphosphates are imported into *Escherichia coli* via expression of the algal nucleoside triphosphate transporter *Pt*NTT2, they are used by endogenous polymerases to replicate a plasmid containing the UBP (See Ref. 14 of Zhang et al.[3]), thus giving rise to the first semi-synthetic organism (SSO) capable of increased information storage. We have also reported an error elimination mechanism based on the expression of Cas9 directed to cleave, and thus degrade, DNA sequences that have lost the UBP.[3] Most recently, we demonstrated that DNA containing the UBP is replicated by the main replicative polymerase, polymerase III, at times aided by polymerase II, and that the major route to UBP loss is recombinational repair of stalled forks mediated by RecA.[4] These mechanistic observations allowed us to create SSOs with error avoidance mechanisms that allow them to store the UBP in virtually any sequence context and even for the first time in the chromosome of the SSO.

Near the end of 2017, we reported the culmination of our work with the first generation SSO, demonstrating that the expanded genetic alphabet could be used as the basis of an expanded genetic code (Fig. 2). The SSO transcribes DNA containing the UBP into mRNAs and tRNAs with cognate unnatural codon–anticodon pairs that efficiently direct the incorporation of natural or ncAAs into proteins.[5] Using an *E. coli* Ser tRNA recoded with an unnatural anticodon, as well as recoded tRNA–synthetase pairs from *M. jannaschii* and *M. barkeri*/*mazei*, we site-specifically incorporated serine or either of the ncAAs, *p*-azido-phenylalanine or *N*-propargyl-lysine, into the protein superfolder GFP. The proteins were produced at near natural-like levels with up to 98% containing the amino acid encoded by the unnatural codon. In addition to opening a new and possibly less restricted route to the expansion of the genetic code, the results are conceptually intriguing, as they demonstrate that hydrophobic UBPs, despite being completely devoid of the complementary hydrogen

bonds that are so central to their natural counterparts, can participate in every step of information storage and retrieval with both high efficiency and high fidelity.

The production of proteins via the decoding of a six-letter alphabet marks a distinct transition from every other protein ever produced by an organism. While we only reported the use of two unnatural codons, AXC and GXC (X = NaM), the UBP should enable the creation of a practically unlimited number of new codons, which we are currently exploring, including with SSOs benefiting from the replication error elimination and/or avoidance systems to increase the fidelity of ncAA incorporation.[3,4] Our next challenge is to take advantage of this abundance of new codons and develop new, orthogonal tRNA–synthetase pairs, which will increase the chemical diversity of ncAAs that may be accessed and enable the facile labeling of proteins with multiple, distinct ncAAs. Finally, we are exploring the deployment of the expanded genetic alphabet in other organisms, including eukaryotes. Thus, the reported SSO is likely just the first form of semi-synthetic life able to access a broad range of forms and functions not available to natural organisms.

## Acknowledgments

Funding Sources

This work was supported by the National Institutes of Health (GM118178 to F.E.R. and GM123735 to Y.Z.).

## REFERENCES

(1). Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, and Isaacs FJ (2013) Genomically recoded organisms expand biological functions. Science 342, 357–360. [PubMed: 24136966]

(2). Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M, Shrock E, Pruitt BW, Conway N, Goodman DB, Gardner CL, Tyree G, Gonzales A, Wanner BL, Norville JE, Lajoie MJ, and Church GM (2016) Design, synthesis, and testing toward a 57-codon genome. Science 353, 819–822. [PubMed: 27540174]

(3). Zhang Y, Lamb BM, Feldman AW, Zhou AX, Lavergne T, Li L, and Romesberg FE (2017) A semisynthetic organism engineered for the stable expansion of the genetic alphabet. Proc. Natl. Acad. Sci. U.S.A 114, 1317–1322. [PubMed: 28115716]

(4). Ledbetter MP, Karadeema RJ, and Romesberg FE (2018) Reprograming the replisome of a semi-synthetic organism for the expansion of the genetic alphabet. J. Am. Chem. Soc 140:758–765. [PubMed: 29309130]

(5). Zhang Y, Ptacin JL, Fischer EC, Aerni HR, Caffaro CE, San Jose K, Feldman AW, Turner CR, and Romesberg FE (2017) A semi-synthetic organism that stores and retrieves increased genetic information. Nature 551, 644–647. [PubMed: 29189780]

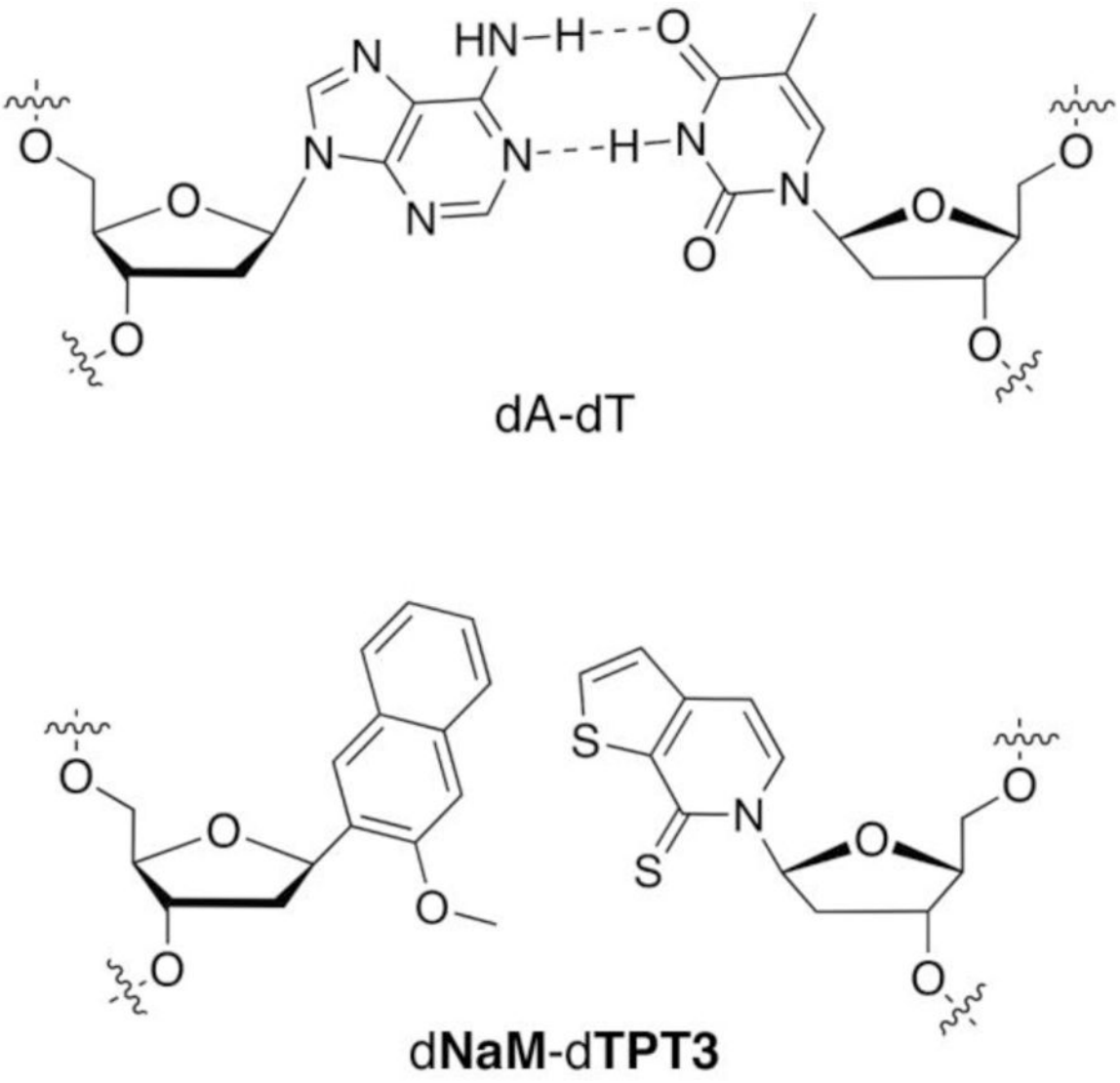



**Figure 1.**
Structure of a natural dA-dT base pair and the d**NaM**-d**TPT3** UBP.

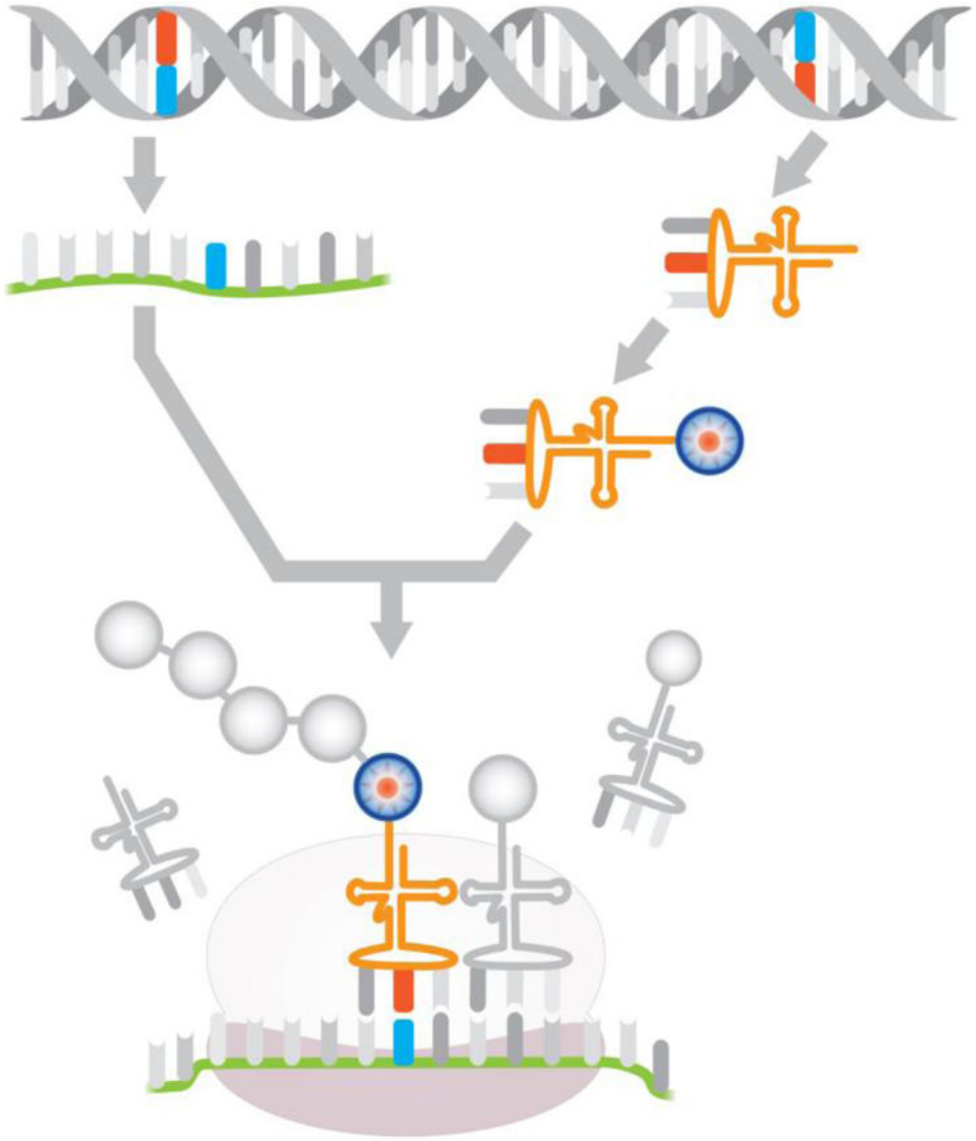


**Figure 2.**

The SSO replicates DNA containing the d**NaM**-d**TPT3** UBP (shown in blue and red, respectively), transcribes mRNA and tRNA with complementary codons and anticodons containing **NaM** or **TPT3**, uses an orthogonal synthetase to charge the tRNA with an ncAA, and uses them to translate proteins containing ncAAs.