ESSAY

AMERSHAM PRIZE WINNER

Expanding the Genetic Code

Lei Wang

The canonical genetic code includes 64 codons encoding 20 amino acids and three stop signals. It is preserved in three kingdoms of life. The origin of the genetic code, whether a "frozen accident" or an expansion from a primordial code with fewer amino acids, remains an enigma (*1*, *2*). Although proteins carry out most of the complex processes of life, there is clearly a need for additional building blocks: Some functions are dependent on posttranslational modification or cofactors, and many important peptides containing unusual amino acids are synthesized nonribosomally (*3*). Why only 20, and why were these 20 amino acids in particular chosen for the code? Nature encodes two additional amino acids, selenocysteine (Sec) and pyrrolysine (Pyl), in limited proteins but in a distinctive way: The Sec-tRNA^{Sec} is converted from a preloaded Ser-tRNA^{Sec}. Special mRNA elements and elongation factors are also required for Sec incorporation into proteins (*4*). Pyl is likely incorporated similarly (*5*). Why does nature not simply employ the standard mechanism of directly loading amino acids onto their cognate tRNA to incorporate Sec and Pyl?

In my graduate research, I explored whether the genetic code can be expanded to accommodate additional amino acids, using a strategy that mimics the way that the common amino acids are encoded. To do this, a novel tRNA-codon pair and an aminoacyl–tRNA synthetase (aaRS) need to be generated that uniquely incorporate an unnatural amino acid. The new components should be orthogonal to the endogenous ones to avoid crosstalk and should function efficiently with the translational apparatus (see the figure) (*6*).

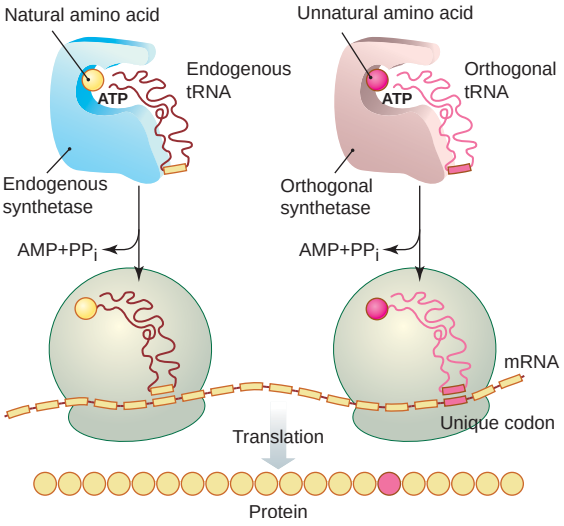
Design of a new codon-tRNA-aaRS set from scratch would be nearly impossible, considering their delicate interactions evolved to ensure translational accuracy. My approach was therefore borrowing and engineering. *Escherichia coli* was chosen as the host organism, and the amber nonsense codon (UAG) was hijacked to encode an unnatural amino acid. I first generated an orthogonal amber suppressor tRNA$^{Tyr}_{CUA}$/TyrRS pair in *E. coli* by importing a tRNA^{Tyr}/TyrRS pair from the archaebacterium *Methanococcus jannaschii* (*Mj*), after testing various tRNA/aaRS pairs from different organisms (*7*). To optimize this pair, I thendeveloped a general strategy consisting of negative and positive selections of a mutant suppressor tRNA library (*8*). Eleven nucleotides of *Mj*-tRNA$^{Tyr}_{CUA}$ were randomly mutated, and from the resulting library a mutant (mutRNA$^{Tyr}_{CUA}$) was identified that has almost no affinity for *E. coli* synthetases and is still charged efficiently by the orthogonal *Mj*-TyrRS with tyrosine. The next step was to alter the amino acid specificity of the *Mj*-TyrRS so that it aminoacylated the mutRNA$^{Tyr}_{CUA}$ with an unnatural amino acid only. A combinatorial approach was pursued, in which a pool of mutant synthetases was generated from the framework of the wild-type synthetase and then mutants were selected based on their specificity for an unnatural amino acid relative to the common 20.

Five active-site residues of *Mj*-TyrRS were randomly mutated to generate the synthetase library. After two rounds of selection, a synthetase was evolved that, when coexpressed with the mutRNA$^{Tyr}_{CUA}$, incorporates *O*-methyl-L-tyrosine into proteins in response to the amber codon with translational fidelity and yield rivaling those of natural amino acids (*9*). Thus, the genetic code of *E. coli* was expanded for the first time. I subsequently evolved a second mutant synthetase that is capable of selectively inserting L-3-(2-naphthyl)-alanine, an amino acid structurally distinct from tyrosine, suggesting that this methodology should be generalizable to various unnatural amino acids (*10*).

These results show that the genetic code can indeed be expanded further using nature's technique—loading amino acid to cognate tRNA via synthetase—although nature did not repeat this method for a 21st amino acid. This expansion may recapitulate how some of the common amino acids were added to the genetic code, suggesting that an incremental expansion was involved in the code's origin. The fact that no toxic side effects were observed in *E. coli* cells with UAG encoding an unnatural amino acid supports the codon reassignment hypothesis (*2*). To investigate the evolutionary consequences of adding novel amino acids to the genetic repertoire, a completely autonomous 21–amino acid bacterium was generated, which biosynthesizes *p*-amino-L-phenylalanine from basic carbon sources and incorporates it in response to the UAG codon (*11*). Directed evolution of such organisms under selective pressure is under way and may shed light on whether additional amino acids give an evolutionary advantage.

Genetically encoding new amino acids makes it possible to tailor changes in proteins in live cells, and therefore protein structure and function can be studied directly in vivo in addition to in vitro. Using the same method and system, we subsequently encoded more than 13 unnatural amino acids with novel functionalities in *E. coli* (*12*). For instance, the versatile keto group

Amersham Biosciences and *Science* are pleased to present the prize-winning essay by Lei Wang, a regional winner from North America, who is the 2003 Grand Prize winner of the Amersham Biosciences and *Science* Prize for young scientists.



New building blocks. A general method for genetically encoding unnatural amino acids into proteins.

CREDIT: PRESTON HUEY/*SCIENCE*

The author is in the Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA. E-mail: lewang@ucsd.edu

was genetically encoded in the form of *p*-acetyl-L-phenylalanine (*13*). It served as a unique chemical handle, through which proteins were selectively labeled with fluorophores for imaging, with biotin for detection, and with carbohydrates for generation of homogeneous glycoprotein mimetics (*14*). Other agents such as spin labels, metal chelators, cross-linking agents, polyethers, fatty acids, and toxins can be attached similarly. Two heavy atom–containing amino acids (*p*-bromo and *p*-iodo-L-phenylalanine) were site-specifically incorporated into proteins, providing a reliable method for preparing isomorphous heavy-atom derivatives of proteins for crystallography (*12*).

The availability of novel building blocks may lead to protein properties that never existed before. In an initial test, Tyr^{66} of the green fluorescent protein (GFP) was substituted with several tyrosine analogs, resulting in mutant GFPs with emissions ranging from blue to cyan to green, as well as other new spectral properties (*15*). In vivo unnatural amino acid mutagenesis by rational design or directed protein evolution should greatly expand the scope and power of protein engineering.

In summary, my thesis research demonstrated that the genetic code can be expanded to include new amino acids. The methodology is generalizable to different amino acids as well as cell types (*16*). It provides a new means for evolutionary study of the genetic code, and powerful tools for molecular and cellular biologists to dissect protein and cellular function both in vitro and in vivo. With additional building blocks genetically encoded, proteins and even organisms with enhanced or novel properties may be evolved.

References

1. F. H. Crick, *J. Mol. Biol.* **38**, 367 (1968).
2. R. D. Knight, S. J. Freeland, L. F. Landweber, *Nature Rev. Genet.* **2**, 49 (2001).
3. C. T. Walsh *et al.*, *Curr. Opin. Chem. Biol.* **5**, 525 (2001).
4. A. Bock *et al.*, *Mol. Microbiol.* **5**, 515 (1991).
5. G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **296**, 1459 (2002).
6. L. Wang, P. G. Schultz, *Chem. Commun.* **1** (2002).
7. L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, *J. Am. Chem. Soc.* **122**, 5010 (2000).
8. L. Wang, P. G. Schultz, *Chem. Biol.* **8**, 883 (2001).
9. L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **292**, 498 (2001).
10. L. Wang, A. Brock, P. G. Schultz, *J. Am. Chem. Soc.* **124**, 1836 (2002).
11. R. A. Mehl *et al.*, *J. Am. Chem. Soc.* **125**, 935 (2003).
12. L. Wang, thesis, University of California at Berkeley (2002).
13. L. Wang, Z. Zhang, A. Brock, P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 56 (2003).
14. H. Liu, L. Wang, A. Brock, C. H. Wong, P. G. Schultz, *J. Am. Chem. Soc.* **125**, 1702 (2003).
15. L. Wang, J. Xie, A. A. Deniz, P. G. Schultz, *J. Org. Chem.* **68**, 174 (2003).
16. K. Sakamoto *et al.*, *Nucleic Acids Res.* **30**, 4692 (2002).

2003 Grand Prize Winner

Lei Wang was born in Tonggu, China. He attended Peking University and received his bachelor's degree in organic chemistry in 1994 and his master's degree in physical chemistry in 1997. At Peking University, working in Dr. Zhongfan Liu's laboratory, he used scanning probe microscopy to investigate the properties of nanoparticles. He left China for the United States to pursue graduate studies at the University of California at Berkeley. Under the guidance of Dr. Peter G. Schultz, Dr. Wang developed a general method for genetically encoding unnatural amino acids into proteins in live cells. After receiving his Ph.D. in 2002, Dr. Wang joined Dr. Roger Y. Tsien's group at the University of California, San Diego, for postdoctoral training as a Merck Fellow of the Damon Runyon Cancer Research Foundation.

Regional Winners

North America: Jeff Levsky, for his essay "Simple Single Cells," based on his Ph.D. research in the laboratory of Dr. Robert Singer at the Albert Einstein College of Medicine, New York. Dr. Levsky was born in Washington, DC, in 1978. After receiving a bachelor's degree from Northwestern University, Illinois, in 1998 he joined the Medical Scientist Training Program at Albert Einstein University. In Dr. Singer's group, he studied the development of single-cell gene expression profiling technology. Dr. Levsky is currently completing his clinical training and is searching for a residency position to pursue a physician-scientist career in imaging and computing.

Europe: Rut Carbadillo-Lopez, for her essay "Shaping Bacteria: The Actin-Like Prokaryotic Cytoskeleton," based on research performed under the guidance of Prof. Jeff Errington at the Sir William Dunn School of Pathology, University of Oxford, UK. Dr. Carbadillo-Lopez left her home town of Barcelona, Spain, to attend the National Institute of Applied Sciences (INSA) in Lyon, France, and graduated in 1996 with an engineering degree in biochemistry. She spent 17 months working in industry, and then obtained a master's degree in general microbiology from the Pasteur Institute of Paris before joining the University of Oxford as a graduate student. After receiving her Ph.D. in 2002, she was awarded a Long-Term Fellowship from the Human Frontiers Science Program Organization for postdoctoral training.

Ravi Kamath, for his essay "Functional Genomics in *C. elegans* Using RNAi," based on his Ph.D. research performed in the laboratory of Dr. Julie Ahringer at the Wellcome Trust/Cancer Research UK Institute of Cancer and Developmental Biology. Dr. Kamath was born in Ohio. After completing an undergraduate degree at Harvard University, he entered Harvard Medical School in 1997, but took a break to pursue graduate studies at the University of Cambridge as a Howard Hughes Medical Institute Predoctoral Fellow. He completed his Ph.D. degree in 2002 and has since returned to complete his medical degree at Harvard, where he is currently in his final year.

All Other Countries: Qing Chen, for her essay "Induction of bgl Operon Expression in *E. coli*: Novel Insights into Sensor Stimulation and Signaling," based on research in the laboratory of Prof. Orna Amster-Choder at the Hebrew University Medical School, Israel. Dr. Chen was born in 1964 in Jinan, China. She received her M.D. degree in 1986, and a master's degree in biochemistry in 1991 from the Medical School of Shandong University. She worked as a lecturer at the Medical School of Beijing University for 2 years before going to Jerusalem to pursue Ph.D. studies. She was awarded her Ph.D. degree in 2002. In 1998, she left Israel for a research associate position in the laboratory of Prof. Robert Kadner at the University of Virginia. She is currently a research scientist in the laboratory of Dr. Malabi Venkatensan at the Walter Reed Army Institute of Research, Maryland.

David Lando, for his essay "The Huff and Puff of HIF Regulation," based on his Ph.D. research carried out in the laboratory of Dr. Murray Whitelaw in the Department of Molecular Biosciences at the University of Adelaide, Australia. Dr. Lando was born in 1969 in Mildura, a small Australian country town. He obtained his B.Sc. degree in 1990 from Flinders University of South Australia. After spending 6 years working for a local biotechnology company, Dr. Lando returned to university to work toward a doctoral degree. After completing his Ph.D., he joined the laboratory of Dr. Tony Kouzarides at the Wellcome Trust Cancer Research UK Institute in Cambridge, where he currently holds a C. J. Martin Fellowship from the the National Health and Medical Research Council of Australia.

For the full text of essays by the regional winners and for information about applying for next year's awards, see *Science* Online at http://www.sciencemag.org/feature/data/pharmacia/prize/apbprize.shl.

CREDIT: THE COLLEGIATE INVENTORS COMPETITION