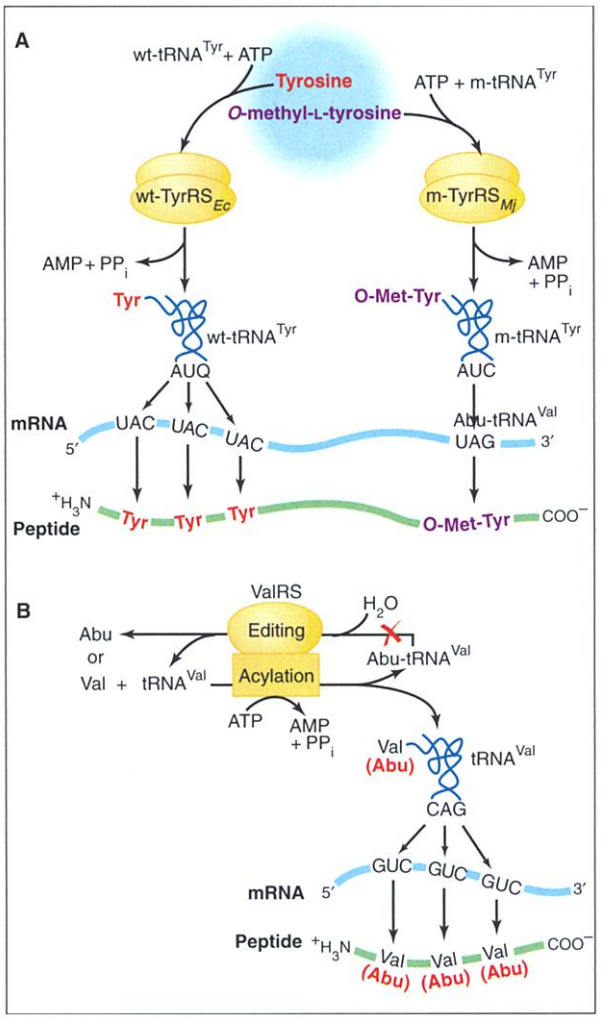
Invading the Genetic Code

August Böck

One of the most intriguing aspects in the evolution of the genetic code—the 61 nucleotide sense codons in the DNA that encode the amino acids of all organisms—is why the amino acid complement is limited to the magic number of 20 (*1*). This number is invariant—from archaea and bacteria to eukaryotes—irrespective of their evolutionary complexity and the chemical and physical environment in which they live. For example, in halophilic (salt-loving) archaea, the 20 amino acids that are the building blocks for constructing proteins are the same as those in nonhalophilic organisms. Equally, proteins of psychrophilic (cold-loving) bacteria—which grow at temperatures of 20°C to below freezing point—are composed of the same 20 amino acids as those of hyperthermophilic bacteria, which exist at temperatures of 100°C or more. Bacteria adapt to their extreme environments, not by the design and incorporation of unnatural amino acids into "new" proteins, but by shifting the proportions of the 20 amino acids in existing proteins, which carry out the same tasks regardless of where the bacteria live. Nonstandard amino acids such as L-ornithine and L-citrulline, intermediates of L-arginine biosynthesis found in every cell, have not gained access to the genetic code. In the cases where nonstandard amino acids are present in proteins, they are generated by modifying one of the 20 standard amino acids after their incorporation into protein rather than by alteration of the genetic code itself.

This scenario is set to change with the reports by Wang *et al.* on page 498 (*2*) and Döring *et al.* (*3*) on page 501 of this week's *Science*. These two groups have designed separate strategies—site-specific replacement and general replacement—for incorporating unnatural amino acids into the proteins of bacteria. Their work sheds light on why nonstandard amino acids are not maintained in the genetic code and how cells prevent their incorporation into proteins. Even more importantly, these studies pave the way for engineering and selecting organisms that encode unnatural amino acids in their DNA and for replacing the standard 20 amino acids in proteins with new residues.



Expanding the genetic code. (**A**) Site-specific insertion of the unnatural amino acid *O*-methyl-L-tyrosine catalyzed by mutant variants of tyrosyl–tRNA synthetase (m-TyrRS$_{Mj}$) and tRNA^{Tyr} (m-tRNA^{Tyr}) from the thermophilic bacterium *M. jannaschii* (*4*). The orthologous tyrosyl–tRNA synthetase and tRNA^{Tyr} from *E. coli* (wt-TyrRS$_{Ec}$ and wt-tRNA^{Tyr}, respectively) are responsible for the correct incorporation of tyrosine into the growing polypeptide chain. (**B**) In an *E. coli* mutant defective in the editing of valyl–tRNA synthetase (ValRS), 20% of valines in the polypeptide are replaced by the compound aminobutyrate (Abu). PP$_i$, inorganic pyrophosphate.

The author is at der Universitat, Institut fur Genetik und Mikrobiologie, Munich 19 80638, Germany. E-mail: august.boeck@lrz.uni-muenchen.de

What are the requirements for achieving incorporation of unnatural amino acids into proteins? Besides having to penetrate the cell membrane by passive diffusion or active transport, unnatural amino acids must bypass the fidelity mechanisms that guard the precision of protein synthesis, ensuring that only the correct amino acids enter the cell's translation machine, the ribosome. Two crucial steps in protein synthesis are essential for the correct translation of codons in the mRNA into amino acids in the growing polypeptide chain. The first step depends on enzymes called aminoacyl–tRNA synthetases that load adapter molecules—the transfer RNAs (tRNAs)—with their correct amino acids. The second step is the selection of the correct aminoacyl-tRNA (the tRNA with its attached amino acid) by the A-site in the ribosome, which reads the mRNA codons. The precision of tRNA aminoacylation (10^{-4} to 10^{-5} error frequency) is at least an order of magnitude greater than the selection of the correct tRNA by the ribosomal A-site. This precision is necessary because once an incorrect amino acid is attached to a tRNA, it is (with a few exceptions) incorporated into the protein. Thus, it is principally the aminoacylation reaction that prevents incorrect amino acids from accessing the ribosome (*4*, *5*).

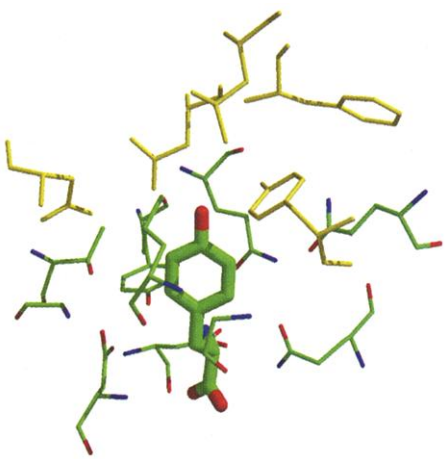
The site-specific replacement strategy of Wang *et al.* (*2*, *6*)—first postulated 20 years ago by Wong and colleagues (*7*)—involves importing the genes encoding a tRNA and its aminoacyl–tRNA synthetase from another organism. The selected tRNA and its synthetase must not exhibit "cross talk" with the homologous pair of the host bacterium. Wang and co-workers selected genes encoding tRNA^{Tyr} (a tRNA carrying the amino acid tyrosine) and its tyrosyl–tRNA synthetase from *Methanococcus jannaschii*, an extreme thermophilic archaeon, and inserted them into the host bacterium, *Escherichia coli*. The tRNA^{Tyr} from *M. jannaschii* is not a good substrate for *E. coli* tyrosyl–tRNA synthetase and—vice versa—the archaeal enzyme does not attach tyrosine to *E. coli* tRNA^{Tyr} (see the figure). The investigators converted the anticodon of the archaeal tRNA^{Tyr} into CUA, which only recognizes UAG, a nonsense codon in mRNA that does not code for any of the 20 amino acids. The modified tRNA translates (suppresses) the nonsense codon and inserts the amino acid that it carries into the

growing polypeptide chain. With the three-dimensional (3D) structure of the tyrosyl–tRNA synthetase/$tRNA^{Tyr}$ complex as a guide (*8*), they mutated the amino acid residues around the tyrosine-binding domain of the synthetase (see figure, this page) and then tested the variants for their ability to suppress the UAG nonsense codon in the presence of various tyrosine analogs. One of the mutant synthetases was able to attach the tyrosine analog *O*-methyl-L-tyrosine to $tRNA^{Tyr}$, which was then able to suppress a nonsense codon in the mRNA. With mass spectrometry, the authors showed that *O*-methyl-L-tyrosine was incorporated both specifically and efficiently into the growing polypeptide chain.

The Wang *et al.* strategy resembles the approach evolved by nature (suppression of an in-frame nonsense codon) to accomplish the incorporation of the nonstandard amino acid selenocysteine into proteins. A conspicuous difference, however, is the elaborate machinery—a 3D structural element of the mRNA together with a special elongation factor—required for selenocysteine insertion. The purpose of this complexity (rather than simple suppression of a nonsense codon) may lie in the need for a stricter discrimination between "captured" nonsense codons and genuine stop codons (*9*).

In their study, Döring and colleagues (*3*) take a different tack: They manipulate the codons encoding the amino acid valine. The specificity of tRNA aminoacylation by the valyl–tRNA synthetase and the isoleucyl–tRNA synthetase has attracted attention because of the chemical similarity of their amino acid substrates (*4*, *5*). Despite this similarity, the synthetases attach valine and isoleucine to their respective tRNA species in a highly specific manner that cannot be explained solely by the stereochemistry of their amino acid–binding pockets. Proofreading of the isoleucyl–tRNA synthetase, resulting in removal of any erroneously attached valines through hydrolysis, explains this specificity. Similarly, valyl–tRNA synthetase can accidentally attach L-cysteine, L-threonine, or L-aminobutyrate to the valyl-tRNA, but the incorrect amino acids are not incorporated into protein at valine positions because this enzyme is able to edit out (hydrolyze) cysteyl-$tRNA^{Val}$ or aminobutyryl-$tRNA^{Val}$. This proofreading (editing) activity is located in a different domain of the synthetase than the active site (*10*).

With an elegant genetic screening strategy, Doring *et al.* were able to select *E. coli* in which this proofreading activity had been disabled. As a result, more than 20% of the valine residues in proteins were replaced by aminobutyrate (see the figure). In their screen, they replaced the cysteine codon at position 146 of thymidylate synthase with each of the other 63 codons. This enzyme is required for synthesis of thymidylate (one of the building blocks of DNA), and cysteine 146 is indispensable for its activity. The authors observed low-level suppression of the editing defect when valine codons were introduced and when high cysteine concentrations were supplied in the medium. These findings indicated that cysteine can be erroneously attached to $tRNA^{Val}$ by valyl–tRNA synthetase. Mutagenesis and selection of bacteria with impaired editing allowed the recovery of valyl–tRNA synthetase mutants with an increased efficiency for attaching L-cysteine to $tRNA^{Val}$ (*11*, *12*). These mutant synthetases attached not only L-cysteine but also L-threonine and L-aminobutyrate (which are sterically similar to cysteine) with greater efficiency to the $tRNA^{Val}$. When *E. coli* harboring a valyl–tRNA synthetase with a defective editing activity were grown in the presence of L-aminobutyrate, they incorporated this amino acid into proteins as long as the mixtures of valine and aminobutyrate were carefully adjusted. Mass spectrometric analysis of fragmented proteins synthesized by defective editing mutants revealed that about 24% of valine residues were replaced by aminobutyrate molecules at each valine position.



A brand-new enzyme. Representation of an amino acid (the thick green and red structure) in the active site of the new tRNA synthetase generated by Wang *et al.* (*2*). The mutated amino acids (shown in yellow) allow the incorporation into newly synthesized peptides of the unconventional amino acid *O*-methyl-L-tyrosine, rather than the tyrosine shown.

Both site-specific insertion and general replacement are feasible strategies for the incorporation of unnatural amino acids into proteins. Site-specific insertion depends on the availability of an orthologous tRNA/aminoacyl–tRNA synthetase pair and on the tRNA anticodon being altered so that it can pair with a nonsense codon. The context of the particular nonsense codon should be such that continuation of protein synthesis outcompetes termination. The amino acid–binding pocket of the synthetase also needs to be altered so that it preferentially accepts an unnatural amino acid over the correct residue. Most importantly, a prerequisite for both the site-specific and general replacement strategies is the absence of editing activity. In nature, tyrosyl–tRNA synthetase is devoid of editing activity, whereas in other tRNA synthetases this activity has to be silenced through mutation.

The ability to engineer organisms with an expanded amino acid repertoire will be hugely beneficial to biotechnology. Sequence position–dependent incorporation of unnatural amino acids with chemically or physically reactive side chains now should be possible. Equally fascinating are the evolutionary implications of the general replacement strategy. What are the limiting steps, and is it possible to use selection to shift the degree of replacement quantitatively or even so that there is complete incorporation of the unnatural amino acids? What are the consequences for protein folding and for protein degradation pathways—will these pathways accommodate the "new" proteins? Will it be possible to create organisms that are dependent on an unnatural amino acid for survival and growth? There are already indications that selection can "press" an organism to become dependent on an unnatural amino acid. When *Bacillus subtilis* was cultured for many generations in the presence of the tryptophan analog fluorotryptophan, it became dependent on this compound, and the natural amino acid tryptophan could not substitute for the unnatural compound (*13*). From the Wang *et al.* and Döring *et al.* work emerges a new realm of biology, bordering the world of chemistry, which will allow experimenters to explore ideas about completely new proteins that were once inconceivable.

References

1. S. Osawa, *Evolution of the Genetic Code* (Oxford Univ. Press, Oxford, 1995).
2. L. Wang *et al.*, *Science* **292**, 458 (2001).
3. V. Döring *et al. Science* **292**, 501 (2001).
4. J. Parker, *Microbiol. Rev.* **53**, 273 (1989).
5. H. Jakubowski, E. Goldman. *Microbiol. Rev.* **56**, 412 (1992).
6. C. J. Noren *et al.*, *Science* **244**, 182 (1989).
7. Y. Kwok, J. T. Wong. *Can. J. Biochem.* **58**, 213 (1980).
8. P. Brick *et al.*, *J. Mol. Biol.* **208**, 83 (1988).
9. A. Böck *et al.*, *Mol. Microbiol.* **5**, 515 (1991).
10. L. Lin *et al.*, *Nature* **384**, 33 (1996).
11. L. Lin, P. Schimmel. *Biochemistry* **35**, 5596 (1996).
12. O. Nureki *et al.*, *Science* **280**, 578 (1998).
13. J. T. Wong. *Microbiol. Sci.* **5**, 174 (1988).

CREDIT: ADAPTED FROM (*8*)