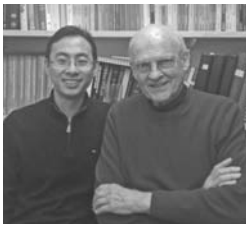
The Genetic Code: Yesterday, Today, and Tomorrow

Jiqiang Ling and Dieter Söll



Jiqiang Ling (left) is a Postdoctoral Associate in the Department of Molecular Biophysics and Biochemistry at Yale University. His research interests include understanding the mechanism of protein synthesis quality control, and its impact on human diseases.

Dieter Söll (right) is Sterling Professor of Molecular Biophysics and Biochemistry at Yale University. He is currently interested in expanding the amino acid repertoire of the genetic code and in recoding organisms in the realm of synthetic biology.

Keywords
Protein synthesis, aminoacyl-tRNA synthetase, transfer RNA, genetic code evolution, synthetic biology.

This issue is a tribute to Har Gobind Khorana who received the Nobel Prize in Physiology or Medicine in 1968 for the elucidation of the Genetic Code. Here we present our view of the changing and ever-challenging perception of the genetic code over the last 50 years.

In October 1962, I (DS) arrived from Germany at the Enzyme Institute of the University of Wisconsin in Madison. I was a young postdoc keen to learn a new trade. I realized that this was a very exciting time for molecular biology. Several fundamental discoveries were about to emerge. The structure of the genetic material (DNA) was known, and an early understanding of RNA polymerase and of the genetic code (the Rosetta stone between the nucleic acid and the amino acid languages) was developing. This vibrant field of molecular biology cast a lasting spell over me. In Har Gobind Khorana's laboratory, I joined the group that used chemistry to unravel the genetic code, and there I also became acquainted with transfer RNA (tRNA). These two topics define my research interest to this day. Below we describe our changing understanding of the genetic code over the last fifty years [1] and expectations for the future.

Yesterday: The Genetic Code and its Discovery

The central dogma of biology states that the genetic information in living organisms is passed from DNA through RNA to protein. The accuracy of information flow depends on several processes, including DNA replication, messenger RNA (mRNA) transcription, and translation, i.e., protein synthesis on the ribosome using mRNA as the template (*Figure* 1). Francis Crick proposed adapter molecules [2] that would connect the 20 canonical amino acids and the mRNA codon; these adaptors turned out to be the tRNAs. Catalyzed by aminoacyl-tRNA synthetases each tRNA accepts a

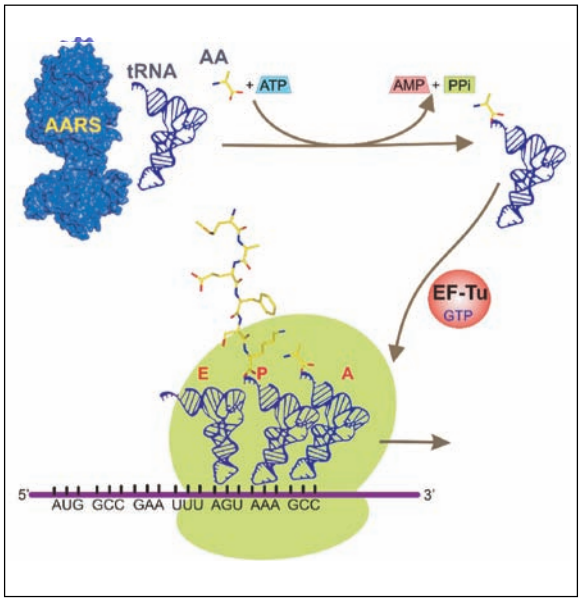


Figure 1. Expression of the genetic code from messenger RNA to protein. The genetic information is transferred from DNA to mRNA through transcription by RNA polymerases, and from mRNA to protein through a process named translation. Translation starts from pairing of amino acids and tRNAs by aminoacyl-tRNA synthetases (AARS). The resulting aminoacyl-tRNAs (aa-tRNA) are then delivered by an elongation factor (EF-Tu) to the ribosome, where by matching the anticodon of the tRNA with the triplet codon on the mRNA, the correct amino acid is added into the elongating polypeptide chain. For further details refer to S F Ataide and M Ibba*.

particular amino acid at the 3'-end, and uses an anticodon (complementary to the triplet codon) to load the amino acid at a ribosomal site in response to the correct codon (*Figure* 1). Such codon–anticodon pairing allows proteins to be correctly synthesized according to the genetic information stored in the DNA genome.

In the early 1960s, biochemical and genetic work began to uncover the mechanistic details of protein synthesis. After Crick and Brenner [3] established that the genetic code was a non-overlapping triplet code, intense interest focused on assigning the 64 codons to the 20 canonical amino acids. This led to an inspired competition between the groups of Marshall Nirenberg [4], Severo Ochoa [5] and Har Gobind Khorana [6][1]. Elucidation of the genetic code required different techniques to eventually establish the nucleotide sequence of all codons. This was aided by the simultaneous advances in tRNA purification, acylation by the aminoacyl-tRNA synthetases, and sequence analysis. Many of these methods were developed in the laboratory of Robert Holley, who determined the first RNA sequence, which was of the yeast alanine tRNA [7]. For these achievements on genetic encoding Nirenberg, Khorana and Holley shared the 1968 Nobel Prize in Physiology or Medicine.

[1] See also Khorana Nobel Lecture: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1968/khorana-lecture.pdf.

After Crick and Brenner [3] established that the genetic code was a non-overlapping triplet code, intense interest focused on assigning the 64 codons to the 20 canonical amino acids.

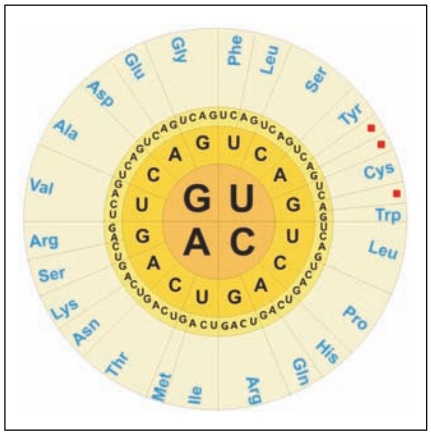


Figure 2. The standard genetic code translates into 20 canonical amino acids. One or more triplet codons are assigned to each amino acid in the standard code. For instance, AUA encodes methionine (Met), and CUN (N denotes A, U, G or C) codons are responsible for insertion of leucine (Leu) into proteins. The phenomenon that multiple codons are assigned to a single amino acid is called degeneracy. UAG, UGA, and UAA (stop codons) are not assigned to any amino acid in the standard code, and they signal a termination during protein synthesis. For further details refer to: A Ambrogelly, S Palioura, and D Söll, Natural expansion of the genetic code. *Nat. Chem. Biol.*, Vol.3, pp.29–35, 2007.
Courtesy: A Ambrogelly, S Palioura, and D Söll.

The combination of 4 different bases (A, U, G and C) gives rise to a total of 64 triplet codons (61 sense codons and 3 stop codons [UAA, UAG, and UGA]); yet they normally decode only the 20 canonical amino acids found in Nature (*Figure* 2). One codon, AUG, has a double role; it specifies initiation of protein synthesis in addition to being a normal Met codon (a codon for the amino acid methionine). As *Figure* 2 also shows, the genetic code is degenerate. Some amino acids are specified by 6, 4, 2, or 1 codon(s). The decoding process rests on the pairing rules of codon and anticodon; this was theoretically considered by the Wobble Hypothesis[2] [8]. Experimental proof of the anticodon [9] and of the fact that certain single aminoacyl-tRNA species can bind, and thus translate, multiple codons was also established in Khorana's laboratory [10]. The question of the evolution of the genetic code was vexing. Evolvability of the genetic code was believed to be constrained by the fact that changes to the code and the resulting mis-translated protein is detrimental to the cell. Thus, at the time of its elucidation, the genetic code was thought to be a frozen accident incapable of further evolution, and one expected that all living organisms use exactly the same set of codons (the standard genetic code) [11].

[2] The Wobble Hypothesis states that the third (wobble) position of the codon is relaxed and can allow formation of non-Watson-Crick pairs, such as a G:U pair.

Experimental proof of the anticodon [9] and of the fact that certain single aminoacyl-tRNA species can bind, and thus translate, multiple codons was also established in Khorana's laboratory

Today: Understanding Evolution of the Genetic Code

The stunning advances in our understanding of the mechanism of protein synthesis (highlighted by the 2009 Chemistry Nobel

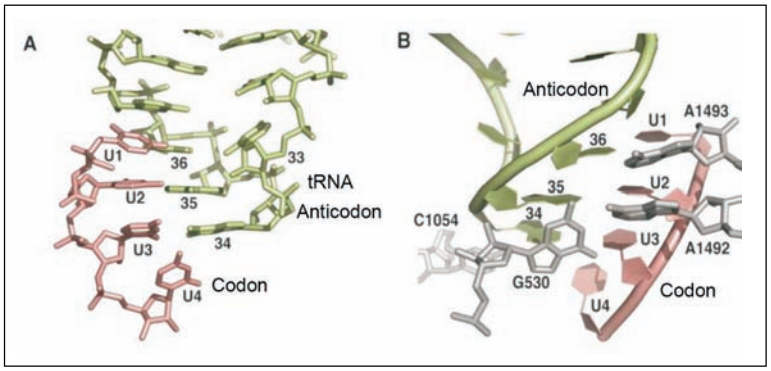


Figure 3. Codon recognition by tRNA on the ribosome. In the ribosomal decoding center, three bases in the tRNA anticodon loop (ASL, green) recognize the triplet codon on the mRNA (pink). In A and B, U1 and U2 form Watson–Crick pairs with A36 and A35 of the tRNA, respectively, and U3 recognizes G34 in a more relaxed wobble pair [8]. Wobbling and RNA modifications [13] at position 34 of the tRNA are primarily responsible for the genetic code degeneracy.

Courtesy: This figure is reprinted from reference [20] with permission (© 2007, Cold Spring Harbor Laboratory Press).

Prize[3] shared by V Ramakrishnan, T A Steitz and A E Yonath [12]) now allow us to visualize how the genetic code is deciphered by aminoacyl-tRNA on the ribosome. *Figure* 3 shows the structural context of the interaction of mRNA codon, with the tRNA anticodon on the ribosome. However, the genetic code is not static; it is interpreted (see *Figure* 1) by the translation machinery including tRNAs, aminoacyl-tRNA synthetases, elongation factors, and tRNA modification enzymes [13]. Even before the advent of genome sequencing, it was evident that the genetic code changed during evolution [14]. In currently known organisms, over 20 examples are known where individual organisms have reassigned the meaning of certain codons away from the standard one (*Figure* 4) [15]. A major deviation, for instance, is the use of the stop codon UGA for tryptophan in mitochondria. Recoding of stop codons was also Nature's way to encode two additional novel amino acids (selenocysteine encoded by UGA, and pyrrolysine by UAG). Such natural expansion of the genetic code allows synthesis of proteins with novel functionality (e.g., selenocysteine-containing proteins are highly sensitive to oxidation–reduction), and might offer modern organisms some competitive advantage during the course of evolution. Despite these

[3] See Laasya Samhita and Umesh Varshney, The Ribosome and the 2009 Nobel Prize in Chemistry, *Resonance*, Vol.15, No.6, pp.526–537, 2010.

In currently known organisms, over 20 examples are known where individual organisms have reassigned the meaning of certain codons away from the standard one.

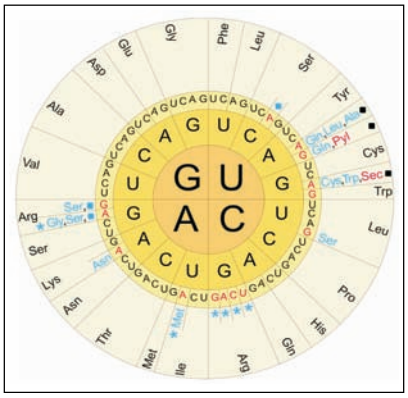


Figure 4. Evolution of the genetic code in Nature. The genetic code was thought to be frozen at the time of its elucidation. It was later found that the code has been continuously evolving. For instance, the UGA stop codon is reassigned to selenocys-teine (Sec) in certain bacteria, archaea and eukaryotes (including humans); and UAG decodes pyrrolysine (Pyl) in certain archaea and bacteria. In mitochondria (essential organelles in eukaryotes), UGA is recognized as a tryptophan (Trp) codon, and AUA is reassigned to Met. These codon reassignment events are examples of the natural expansion of the genetic code. For further details refer to [15].

Courtesy: A Ambrogelly, S Palioura, and D Söll.

extensions of the genetic code, a large degree of 'code universality' is still maintained in the living world. Thus, the genetic code still provides a communal language across the different kingdoms in the living world; this allows the evolutionarily important mechanism of horizontal gene transfer[4] (e.g., acquiring DNA from an unrelated organism and using its information) to work effectively.

[4] See Saurabh Dhawan and Tomás J Ryan, The Bacterium That Got Infected by a Cow! – Horizontal Gene Transfer and Evolution, *Resonance,* Vol.12, No.1, pp.49–59, 2007.

Tomorrow: Genetic Code Engineering and Synthetic Biology

Since Crick's frozen accident hypothesis[5], it was assumed that cellular fitness depends on highly accurate protein synthesis. However, in the last decade it became clear that organisms can tolerate sizable amounts of mis-translated proteins (e.g., 10% amino acid mis-incorporation in *E. coli*) and still maintain fitness. Genetic code engineering is therefore possible, and a certain level of inaccurate proteins synthesized as side products of orthogonal translation systems will be tolerated by the organism. This evolvability and natural expansion of the genetic code have inspired biologists and chemists to engineer the genetic code for interesting noncanonical amino acids. To date, over 70 such amino acids have been successfully inserted into proteins using laboratory-evolved orthogonal aminoacyl-tRNA synthetase variants [16]. This powerful approach is useful for introducing novel functional groups into proteins and studying post-translational modifications [17, 18]. Changes in the genetic code by reducing

[5] The Frozen Accident Hypothesis states that "the code is universal because at the present time any change would be lethal, or at least very strongly selected against" [11].

To date, over 70 such amino acids have been successfully inserted into proteins using laboratory-evolved orthogonal aminoacyl-tRNA synthetase variants.

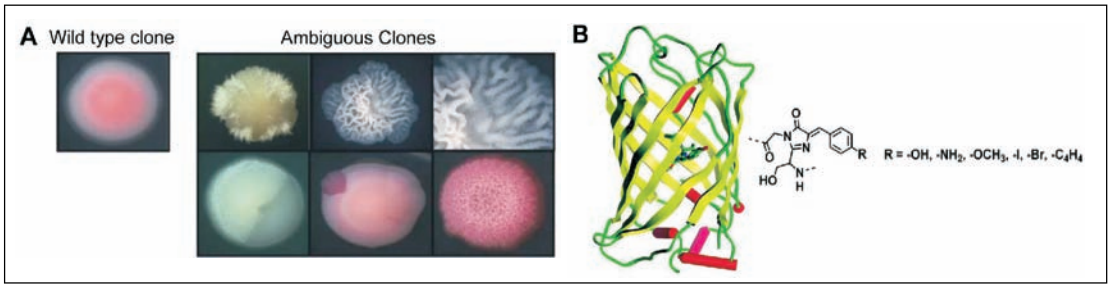


Figure 5. Phenotypic and functional diversity resulting from genetic code expansion.

(A) Genetic code ambiguity results in diverse phenotypes in *Candida albicans* (a fungus species). The wild-type strain forms smooth colonies, while increasing genetic code ambiguity causes cells to form colonies with rough surfaces. For further details refer to I Miranda *et al*., A genetic code alteration is a phenotype diversity generator in the human pathogen *Candida albicans*. *PLoS One*, Vol.2, e996, 2007. Courtesy: I Miranda *et al.*

(B) Insertion of non-canonical amino acids into the active site of green fluorescent protein (GFP) changes its absorption and emission patterns. A tyrosine (Tyr) at position 68 of GFP is replaced with various non-canonical amino acids shown in the figure. This is done by mutating the corresponding Tyr codon to UAG, and using an engineered AARS to insert the non-canonical amino acid at the UAG position. For further details refer to L Wang, J Xie, AA Deniz, and PG Schultz.

Courtesy: Reproduced with permission from L Wang, J Xie, A A Deniz, and P G Schultz, Unnatural amino acid mutagenesis of green fluorescent protein, *J. Org. Chem.*, Vol.68, pp.174–176, 2003. © 2003 American Chemical Society.

code degeneracy and reassigning a significant number of codons to new amino acid meanings could result in diverse phenotypes at the organism level and multi-functionality in proteins (*Figure* 5). Such proteins with novel or improved properties may contain multiple additional amino acids that may lead to new materials and new pharmaceuticals. In addition, as a result of post-translational modification, over 300 different amino acids exist in proteins in Nature [19]. In human cells, modified proteins play key roles in signaling, gene expression and disease. For synthetic biologists, who work on rewiring organisms by rearranging or creating new cellular networks, genetic code engineering provides fascinating opportunities to speed up natural evolution and create enzymes, biochemical pathways and organisms with improved properties that are containable. Continued evolution of the genetic code will benefit broad areas of medicine, materials science and biotechnology. By pushing forward the boundaries of genetic encoding to create synthetic proteins and biopolymers, we will discover more about the plasticity of life.

For synthetic biologists, genetic code engineering provides fascinating opportunities to speed up natural evolution and create enzymes, biochemical pathways and organisms with improved properties that are containable.

Suggested Reading

[1] D Söll and U L RajBhandary, The genetic code – thawing the 'frozen accident', *J. Biosci.*, Vol.31, No.4, pp.459–463, 2006.

[2] F H Crick, On protein synthesis, *Symp. Soc. Exp. Biol.*, Vol.12, pp.138–163, 1958.

[3] F H Crick *et al*, General nature of the genetic code for proteins, *Nature*, Vol.192, pp.1227–1232, 1961.

[4] M Nirenberg *et al*, The RNA code and protein synthesis, *Cold Spring Harb. Symp. Quant. Biol.*, Vol.31, pp.11–24, 1966.

[5] P Lengyel, Memories of a senior scientist: on passing the fiftieth anniversary of the beginning of deciphering the genetic code, *An. Rev. Microbiol.*, Vol.66, pp.27–38, 2012.

[6] D Söll *et al*, Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleo-tides and a survey of codon assignments for 20 amino acids, *Proc. Natl. Acad. Sci.*, USA, Vol.54, No.5, pp.1378–1385, 1965.

[7] R W Holley *et al*, Structure of a ribonucleic acid, *Science*, Vol.147, pp.1462–1465, No.3664, 1965.

[8] F H Crick, Codon–anticodon pairing: the wobble hypothesis, *J. Mol. Biol.*, Vol.19, No.2, p.548–555, 1966.

[9] D Söll and U L RajBhandary, Studies on polynucleotides. LXXVI. Specificity of transfer RNA for codon recognition as studied by amino acid incorporation, *J. Mol. Biol.*, Vol.29, No.1, pp.113–124, 1967.

[10] D Söll *et al*, Specificity of sRNA for recognition of codons as studied by the ribosomal binding technique, *J. Mol. Biol.*, Vol.19, No.2, pp.556–573, 1966.

[11] F H Crick, The origin of the genetic code, *J. Mol. Biol.*, Vol.38, No.3, pp.367–379, 1968.

[12] L Samhita, U Varshney, The ribosome and the 2009 Nobel Prize in Chemistry, *Resonance*, pp.526–537, 15 June 2010.

[13] E M Gustilo, F A Vendeix and P F Agris, tRNA's modifications bring order to gene expression, *Curr. Opin. Microbiol.*, Vol.11, No.2, pp.134–140, 2008.

[14] T H Jukes and S Osawa, Evolutionary changes in the genetic code, *Comp. Biochem. Physiol. B.*, Vol.106, No.3, p.489–494, 1993.

[15] A Ambrogelly, S Palioura and D Söll, Natural expansion of the genetic code, *Nat. Chem. Biol.*, Vol.3, No.1, pp.29–35, 2007.

[16] C C Liu and P G Schultz, Adding new chemistries to the genetic code, *Annu. Rev. Biochem.*, Vol.79, pp.413–444, 2010.

[17] L Davis and J W Chin, Designer proteins: applications of genetic code expansion in cell biology, *Nat. Rev. Mol. Cell. Biol.*, Vol.13, No.3, pp.168–182, 2012.

[18] H S Park *et al*, Expanding the genetic code of *Escherichia coli* with phosphoserine, *Science*, Vol.333, No.6046, p.1151–1154, 2011.

[19] Y Zhao and O N Jensen, Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques, *Proteomics*, Vol.9, No.20, pp.4632–41, 2009.

[20] C M Dunham *et al*, Structures of tRNAs with an expanded anticodon loop in the decoding center of the 30S ribosomal subunit, *RNA*, Vol.13, No.6, pp.817–823, 2007.

Address for Correspondence
Jiqiang Ling[1]
Department of Molecular Biophysics and Biochemistry
and
Dieter Söll[2]
Department of Chemistry
Yale University
New Haven, CT 06520, USA
Email:
[1]jiqiang.ling@yale.edu
[2]dieter.soll@yale.edu