Acta Biochimica Polonica

Vol. 54 No. 1/2007, 51–54

on-line at: www.actabp.pl

Minireview

The genetic code — 40 years on

Maciej Szymański and Jan Barciszewski✉

Institute of Bioorganic Chemistry of the Polish Academy of Sciences, Poznań, Poland

Received: 20 February, 2007; revised: 05 March, 2007; accepted: 09 March, 2007
available on-line: 20 March, 2007

The genetic code discovered 40 years ago, consists of 64 triplets (codons) of nucleotides. The genetic code is almost universal. The same codons are assigned to the same amino acids and to the same START and STOP signals in the vast majority of genes in animals, plants, and microorganisms. Each codon encodes for one of the 20 amino acids used in the synthesis of proteins. That produces some redundancy in the code and most of the amino acids being encoded by more than one codon. The two cases have been found where selenocysteine or pyrrolysine, that are not one of the standard 20 is inserted by a tRNA into the growing polypeptide.

Keywords: genetic code, tRNA, aminoacyl-tRNA synthetase

INTRODUCTION

There are few achievements in the short history of molecular biology that had a profound impact on the advancement of science, as well as are strongly imprinted in the public perception. Undoubtedly, one of such events was deciphering of the genetic code.

In the 1940s, it had been demonstrated for the first time that, contrary to previous beliefs, deoxyribonucleic acid (DNA) and not proteins is responsible for the transmission of genetic information through the generations. The nature of that process became apparent after the discovery of the DNA structure in 1953 (Watson & Crick, 1953), which, on the other hand, brought new challenges. The key question ahead was how the sequence of four nucleotides in DNA is translated into sequences of twenty amino acids in proteins and subsequently into functional protein folds (Woese, 2001). Much of the theoretical considerations concerning the nature of the genetic code were due to the activity of George Gamov, Francis Crick, Leslie Orgel, James Watson, Alexander Rich and other members of the exclusive *Tie Club* (Rich *et al.*, 2004).

However, the breakthrough came in 1961, when in the first experiments in *Escherichia coli* cell free system the poly-U programmed synthesis of polyphenylalanine was demonstrated. It has been earlier assumed that the genetic code was composed of nucleotide triplets. Thus, the first word of the code, UUU, had been deciphered as encoding phenylalanine (Nirenberg *et al.*, 1962). The code was cracked open. Subsequent work, over a period of about 5 years (1961–1966), led to assigning of all triplets to particular amino acids (Nirenberg *et al.*, 1966, Nirenberg, 2004). The genetic code is nearly universal across all life forms and, with a few exceptions, unambiguous. Its importance has been compared to that of the periodic table of the elements for chemistry and the scientists who contributed to its deciphering were awarded the Nobel Prize in 1968.

EXCEPTIONS

Initially, it was believed that the genetic code used by all present day organisms is universal and invariant. Any change in the meaning of codons would result in erroneous protein sequences. That concept of a ‘frozen accident’ was later revised by discoveries of alternative genetic codes which show slight differences from the established standard (Knight *et al.*, 2001). These deviations are limited to

✉Corresponding author: Prof. dr hab. Jan Barciszewski, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego 12/14, 61-704 Poznań, Poland; tel.: (48 61) 852 8503 ext. 132; fax: (48 61) 852 0532; e-mail: Jan.Barciszewski@ibch.poznan.pl

nuclear codes in certain taxonomic groups and to mitochondria (Fig. 1). Interestingly, the reassignments of codons are often recurrent in different groups of organisms. Mitochondria, which have much of their own protein translation machinery decode AUA as methionine rather than isoleucine. In *Mycoplasma*, the stop codon UAG is decoded as tryptophan as well as the usual UGG. The reassignment of termination codons UAA and UAG to encode glutamine is found in divergent groups including diplomonads, ciliates and some green algae (Gesteland & Atkins, 1996).

In organisms which utilize the standard genetic code there are also cases of alternative codon assignments (recoding) which may be due to several mechanisms: nonsense or missense suppression, ribosomal frameshifts, bypassing and natural suppression. In the last case the genetic code can be redefined as insertion of non-canonical amino acids at stop codons described below. While exceptions to the basic rules of the genetic code are of considerable biological interest, the universal features of the code indicate that a core of protein biosynthesis apparatus evolved before the divergence of the three kingdoms of life (Knight *et al.*, 2001; Miranda *et al.*, 2006).

EXPANSION

The complete set of all different amino acids found in natural proteins is currently estimated at 140. However, the canonical set comprises only 20 amino acids which have their corresponding triplets in the genetic code, and are incorporated into proteins during translation. This set is invariant in all organisms, irrespective of their evolutionary complexity and environment in which they live. The limits of the coding capacity were probably set at the very early stages of genetic code evolution. It is assumed that the present day three-letter triplet code evolved from a "two-letter triplet" code, in which only the first two nucleotides were in fact used for coding (Szathamary 1999; Travers, 2006). This property is somehow present in the contemporary genetic code in which most of the amino acids are encoded by groups of codons differing only at the third position (Fig. 1). Such degeneracy or redundancy of triplet combinations seems to have been preserved to minimize the deleterious effect of point mutations.

The traces of the evolutionary process of extending the coding capability of the genetic code can be observed in some of the present-day organisms. In certain species, there are no asparaginyl- or glutaminyl-tRNA synthetases, but they possess specific tRNAs which incorporate these amino acids into proteins. The synthesis of Asn-tRNA^{Asn} and Gln-tRNA^{Gln} is accomplished in a two-step process which involves aminoacylation of tRNA^{Asn} and tRNA^{Gln} with aspartate and glutamate, respectively. In the second step the acids are converted into amides by amidotransferases (Ibba & Soll, 2001).

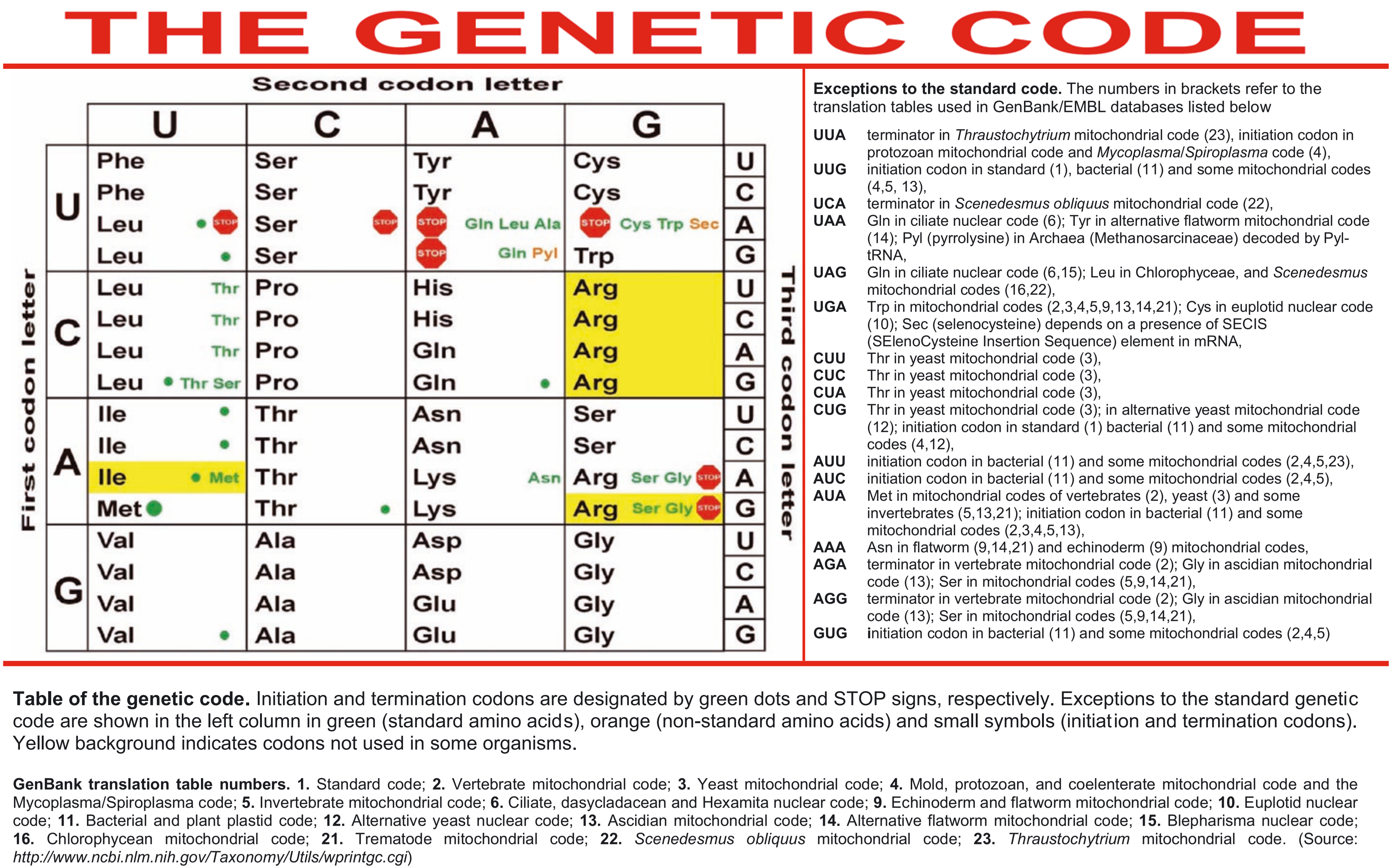
There are, however, two examples of the expansion of the genetic code beyond the standard set of 20 amino acids. The two non-canonical amino acids, selenocysteine (Sec) and pyrrolysine (Pyl), can be incorporated cotranslationally into proteins at positions specified by codons UGA and UAG, respectively, which are normally termination codons (Fig. 1).

The selenocysteinyl tRNA$^{[Ser]Sec}$ is first aminoacylated with serine which is then used as a substrate for selenocysteine synthesis. The aminoacylation step is performed by serine specific aminoacyl-tRNA synthetase (Ambrogelly *et al.*, 2007). In the case of pyrrolysine, the process is much simpler and involves direct aminoacylation of suppressor tRNA^{Pyl} with pyrrolysine (Srinivasan *et al.*, 2002). Thus in the two cases the expansion of coding capacity is achieved by different strategies.

The incorporation of selenocysteine and pyrrolysine into protein sequences can be treated as an expansion of the 'standard' genetic code because for both amino acids there are specific cognate tRNAs which recognize specific codons. However, unlike other codons, whether in the standard or alternative genetic codes, the presence of these codons is not sufficient for their decoding as Sec or Pyl. In this respect the two cases resemble suppression or a readthrough at termination codons, which is common both in eukaryotes and prokaryotes (Cornish *et al.*, 1995; Beier & Grimm, 2001). Selenocysteine and pyrrolysine insertions require additional structural features in the mRNA known as SECIS (selenocysteine insertion sequence) and PYLIS (pyrrolysine insertion sequence) which provide an appropriate context for the termination codon to be suppressed (Mix *et al.*, 2007; Longstaff *et al.*, 2007). Additionally, the insertion of selenocysteine strictly depends on the activity of specific translation elongation factors (Fagegaltier *et al.*, 2000).

PERSPECTIVES

The genetic code is the basis for the template-instructed synthesis of proteins, but it cannot be considered without the components of decoding machinery. As early as 1955, in an unpublished note: *On degenerate templates and adaptor hypothesis* written for the 'Tie Club', Francis Crick predicted the existence of adaptor molecules (tRNAs) which link particular codons with specific amino acids. Thus, the decoding process, which depends on the complementary interactions of the codons and tRNA mol-



THE GENETIC CODE

First codon letter	Second codon letter				Third codon letter
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu ● STOP	Ser STOP	STOP Gln Leu Ala	STOP Cys Trp Sec	A
	Leu ●	Ser	STOP Gln Pyl	Trp	G
C	Leu Thr	Pro	His	Arg	U
	Leu Thr	Pro	His	Arg	C
	Leu Thr	Pro	Gln	Arg	A
	Leu ● Thr Ser	Pro	Gln ●	Arg	G
A	Ile ●	Thr	Asn	Ser	U
	Ile ●	Thr	Asn	Ser	C
	Ile ● Met	Thr	Lys Asn	Arg Ser Gly STOP	A
	Met ●	Thr ●	Lys	Arg Ser Gly STOP	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val ●	Ala	Glu	Gly	G

Exceptions to the standard code. The numbers in brackets refer to the translation tables used in GenBank/EMBL databases listed below

UUA terminator in *Thraustochytrium* mitochondrial code (23), initiation codon in protozoan mitochondrial code and *Mycoplasma*/*Spiroplasma* code (4),

UUG initiation codon in standard (1), bacterial (11) and some mitochondrial codes (4,5, 13),

UCA terminator in *Scenedesmus obliquus* mitochondrial code (22),

UAA Gln in ciliate nuclear code (6); Tyr in alternative flatworm mitochondrial code (14); Pyl (pyrrolysine) in Archaea (Methanosarcinaceae) decoded by Pyl-tRNA,

UAG Gln in ciliate nuclear code (6,15); Leu in Chlorophyceae, and *Scenedesmus* mitochondrial codes (16,22),

UGA Trp in mitochondrial codes (2,3,4,5,9,13,14,21); Cys in euplotid nuclear code (10); Sec (selenocysteine) depends on a presence of SECIS (SElenoCysteine Insertion Sequence) element in mRNA,

CUU Thr in yeast mitochondrial code (3),

CUC Thr in yeast mitochondrial code (3),

CUA Thr in yeast mitochondrial code (3),

CUG Thr in yeast mitochondrial code (3); in alternative yeast mitochondrial code (12); initiation codon in standard (1) bacterial (11) and some mitochondrial codes (4,12),

AUU initiation codon in bacterial (11) and some mitochondrial codes (2,4,5,23),

AUC initiation codon in bacterial (11) and some mitochondrial codes (2,4,5),

AUA Met in mitochondrial codes of vertebrates (2), yeast (3) and some invertebrates (5,13,21); initiation codon in bacterial (11) and some mitochondrial codes (2,3,4,5,13),

AAA Asn in flatworm (9,14,21) and echinoderm (9) mitochondrial codes,

AGA terminator in vertebrate mitochondrial code (2); Gly in ascidian mitochondrial code (13); Ser in mitochondrial codes (5,9,14,21),

AGG terminator in vertebrate mitochondrial code (2); Gly in ascidian mitochondrial code (13); Ser in mitochondrial codes (5,9,14,21),

GUG initiation codon in bacterial (11) and some mitochondrial codes (2,4,5)

Table of the genetic code. Initiation and termination codons are designated by green dots and STOP signs, respectively. Exceptions to the standard genetic code are shown in the left column in green (standard amino acids), orange (non-standard amino acids) and small symbols (initiation and termination codons). Yellow background indicates codons not used in some organisms.

GenBank translation table numbers. 1. Standard code; **2.** Vertebrate mitochondrial code; **3.** Yeast mitochondrial code; **4.** Mold, protozoan, and coelenterate mitochondrial code and the Mycoplasma/Spiroplasma code; **5.** Invertebrate mitochondrial code; **6.** Ciliate, dasycladacean and Hexamita nuclear code; **9.** Echinoderm and flatworm mitochondrial code; **10.** Euplotid nuclear code; **11.** Bacterial and plant plastid code; **12.** Alternative yeast nuclear code; **13.** Ascidian mitochondrial code; **14.** Alternative flatworm mitochondrial code; **15.** Blepharisma nuclear code; **16.** Chlorophycean mitochondrial code; **21.** Trematode mitochondrial code; **22.** *Scenedesmus obliquus* mitochondrial code; **23.** *Thraustochytrium* mitochondrial code. (Source: *http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi*)

Figure 1. Table of the genetic code.

ecules requires a highly specific mechanism which links amino acids and their cognate tRNAs. This task is accomplished by aminoacyl-tRNA synthetases (AARS) which catalyse specific charging of their cognate tRNAs with the corresponding amino acid. This reaction establishes the direct link between the anticodon and the activated amino acid attached to the 3′ end of the tRNA. Therefore, aminoacyl-tRNA synthetases are considered to be the real interpreters (translators) of the genetic code. The determinants of the specificity of recognition between tRNAs and aminoacyl tRNA synthetases have been often referred to as the second genetic code (Beuning & Musier-Forsyth, 1999).

On top of the genetic code there is also an epigenetic code which constitutes a part of the complicated gene regulation system (Turner, 2007). It seems that deciphering the epigenetic features and their influence on the patterns of gene expression will have an impact on our understanding of molecular systems comparable with that of the genetic code forty years ago.

Acknowledgements

This work was supported with the MNISW grant to J.B.

REFERENCES

Ambrogelly A, Palioura S, Soll D (2007) Natural expansion of the genetic code. *Nat Chem Biol* **3**: 29–35.

Beier H, Grimm M (2001) Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res* **29**: 4767–4782.

Beuning PJ, Musier-Forsyth K (1999) Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers* **52**: 1–28.

Cornish VW, Mendel D, Schultz PG (1995) Probing protein structure and function with an expanded genetic code. *Angew Chem Int Ed Engl*. **34**: 621–633.

Fagegaltier D, Hubert N, Yamada K, Mizutani T, Carbon P, Krol A (2000) Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J* **19**: 4796–4805.

Gesteland RF, Atkins JF (1996) Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* **65**: 741–768.

Ibba M, Söll D (2001) The renaissance of aminoacyl-tRNA synthesis. *EMBO Rep* **2**: 382–387.

Knight RD, Freeland SJ, Landweber LF (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet* **2**: 49–58.

Longstaff DG, Blight SK, Zhang L, Green-Church KB, Krzycki JA (2007) *In vivo* contextual requirements for UAG translation as pyrrolysine. *Mol Microbiol* **63**: 229–241.

Miranda I, Silva R, Santos MA (2006) Evolution of genetic code in yeast. *Yeast* **23**: 203–213.

Mix H, Lobanov AV, Gladyshev VN (2007) SECIS elements in the coding regions of selenoprotein transcripts are functional in higher eukaryotes. *Nucleic Acids Res* **35**: 414–423.

Nirenberg M (2004) Historical review: deciphering the genetic code — a personal account. *Trends Biochem Sci* **29**: 46–54.

Nirenberg MW, Matthae JH, Jones OW (1962) An intermediate in the biosynthesis of polyphenylalanine directed by synthetic template RNA. *Proc Natl Acad Sci USA* **48**: 104–109.

Nirenberg M, Caskey T, Marshall R, Brimacombe R, Kellogg D, Doctor B, Hatfield D, Levin J, Rottman F, Pestka S, Wilcox M, Anderson F (1966) The RNA code and protein synthesis. *Cold Spring Harb Symp Quant Biol* **31**: 11–24.

Rich A (2004) The excitement of discovery. *Annu Rev Biochem* **73**: 1–37.

Srinivasan G, James CM, Krzycki JA (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**: 1459–1462.

Szathmary E (1999) The origin of the genetic code. *Trends Genet* **15**: 223–229.

Travers A (2006) The evolution of the genetic code revisited. *Orig Life Evol Biosph* **36**: 549–555.

Turner BM (2007) Defining an epigenetic code. *Nat Cell Biol* **9**: 2–6.

Watson JD, Crick FHC (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171**: 737–738.

Woese CR (2001) Translation: in retrospect and prospect, *RNA* **7**: 1055–1067.