

# Universal genetic code theory alternatives



## Introduction

The existence of a 'genetic code' was established through genetic and biochemical experiments that revealed an association between the sequence of nucleotides in DNA and the sequence of amino acids in polypeptides (Singer and Berg 1991). In 1961, Nirenberg and colleagues conducted *in vitro* studies using synthetic oligo- and polyribonucleotides as messenger RNAs (mRNAs). They demonstrated that a cell-free *Escherichia coli* system was able to catalyse the synthesis of polyphenylalanine in the presence of polyuridylic acid (polyU). Even though the reaction mix contained other amino acids, only phenylalanine was incorporated into the protein when polyU was used as the mRNA, leading to the conclusion that UUU must be the codon for phenylalanine (Nirenberg and Matthaei 1961). Similar experiments allowed codons to be identified for all other amino acids (Nirenberg et al. 1966).

The genetic code comprises a total of 64 triplet codons; of these, 61 code for specific amino acids and 3 are stop codons which terminate translation. All except two of the 20 amino acids used in protein synthesis (i. e. methionine and tryptophan) are encoded by more than one codon (Szymański and Barciszewski 2007). This degeneracy or redundancy was first recognised by Crick in 1966, who observed that while the first two nucleotides of the codon and the last two nucleotides of the anticodon follow Watson and Crick's canonical A: U or G: C base pairing rules (Watson and Crick 1953), rules for pairing between the third nucleotide of the codon and the first base of the anticodon are more relaxed and demonstrate 'wobble'. This led to Crick's 'Wobble Hypothesis' in which he proposed that the third nucleotide was less

important during the process of translation than the first two nucleotides of the codon (Crick 1966).

It has been suggested that the current three-letter triplet code evolved from a 'two-letter triplet' code in which only the first two nucleotides were employed (Szanathamary 1999; Travers 2006). Conserving this approach may minimise the deleterious effect of point mutations (Szymański and Barciszewski 2007). Studies have shown that in most organisms, certain codons are preferred over others encoding the same amino acid, a phenomenon known as codon bias (Hershberg and Petrov 2008). Population genetics experiments demonstrated that these preferred codons may be translated more accurately and/or with greater efficiency, and are thus the major forces driving selection, although other forces may also play a part (Hershberg and Petrov 2008).

The genetic code was initially thought to be universal in that a particular codon was believed to code for the same amino acid in all living organisms, since the code also applied to tobacco mosaic virus and vertebrates as well as *E. coli* (Osawa et al. 1992). Just over ten years after the discovery of the genetic code, researchers reported the first alternative [*Client: I have used 'alternative' rather than 'exception' throughout the brief – you may prefer the alternative*] to the universal code (Barrell et al. 1979). Since then, a number of other alternatives have been reported. This paper discusses the early research and theories that support the concept of a universal genetic code, examines the findings that have challenged these theories, and explores current hypotheses relating to why the genetic code is not universal.

## The frozen-accident theory

The early findings of Nyrenberg and others led Crick to propose the 'frozen-accident theory' in 1968, based on the earlier arguments of Hinegardner and Engelberg (1963). Crick's theory stated that: *"the code is universal because at the present time any change would be lethal, or at least very strongly selected against. This is because in all organisms (with the possible exception of certain viruses), the code determines (by reading the mRNA) the amino acid sequences of so many highly evolved protein molecules that any change to these would be highly disadvantageous unless accompanied by many simultaneous mutations to correct the 'mistakes' produced by altering the code. To account for it being the same in all organisms, one must assume that all evolved from a single organism (more strictly, from a single, closely interbreeding population). In its extreme form, the theory implies that the allocation of codes to amino acids at this point was entirely a matter of 'chance'."* There was always resistance to this hypothesis with scientists asking the question, 'If the genetic code is simply a 'frozen accident, why does it have such useful features?', hence a number of other hypotheses were also proposed to explain the early evolution of the code (Sonneborn 1965; Woese 1965; Wong 1975).

## Alternatives to the 'universal' genetic code

Just over ten years after the frozen-accident theory was proposed, the first alternative to the genetic code was reported. In 1979, Barrell et al. showed that in the cytochrome oxidase subunit II gene of human mitochondrial DNA, UGA was read as a tryptophan codon and not a stop codon, while AUA was

read as methionine, rather than isoleucine (Barrell 1979). It was initially thought that these changes could be tolerated in mitochondria due to the small size of their genome. However, alternatives in nuclear codes were also discovered. In 1985, Yamao et al. reported that in the bacterium, *Mycoplasma capricolum*, UGA was also read as tryptophan and not a stop codon, which this codon commonly reads throughout prokaryotes and eukaryotes. Over 20 alternatives to the universal code have now been reported (those up to 2007 are summarised in Table 1) in bacteria, archaea, eukaryotic nuclear genomes and organellar genomes (Szymański and Barciszewski 2007). Each of these alternatives show minimal differences from the standard code, which is thought to be indicative of a common ancestry (Cavalcanti and Landweber 2004) and many of the same codons are reassigned in independent lineages.

### *Theories of codon reassignment*

Three main theories have been proposed to explain the code changes observed over the past almost twenty years. These theories are not thought to be mutually exclusive (Massey et al. 2003). The codon capture theory proposed by Osawa and Jukes (1989) is based on the finding that a codon can disappear from the genome (particularly small genomes) as a result of directional GC/AT mutation pressure on the genome, then reappear with a different recognition, often due to random genetic drift. It is suggested that this theory is neutral and that codon reassignment would occur without the production of deleterious or non-functional proteins. The 'ambiguous intermediate' theory proposes that codon reassignment involves an intermediate stage during which the codon is read by both cognate and a

<https://assignbuster.com/universal-genetic-code-theory-alternatives/>

mutant transfer RNA (tRNA) (Schultz and Yarus 1994). As a result of competition between these two tRNAs, the gene coding for the cognate tRNA may eventually be eliminated. The final theory, 'genome streamlining', proposed by Andersson and Kurland (1995) suggests that selective pressure to reduce the size of mitochondrial genomes leads to the reassignment of specific codons, in particular stop codons.

### Hypotheses on the evolution of the genetic code

Many historical, chemical and selection hypotheses have been proposed to explain the origin and evolution of the genetic code (Knight et al. 1999).

Three of the main theories: the coevolution theory, the stereochemical theory, and the adaptive theory are discussed below.

#### *Coevolution theory*

Crick first suggested that the modern genetic code evolved from a simpler, primordial version that encoded fewer amino acids than the current form (Crick 1968). Later, this hypothesis formed the basis of the coevolution theory (Wong 1975), which proposed that the genetic code coevolved with biosynthetic pathways for new amino acids. It was thought that the early genetic code used only a small number of amino acids (Wong and Bronskill 1979), and evolved to its present form by incorporating new derivatives of the early, primordial amino acids as metabolic pathways also evolved. A substantial body of literature has been published that supports and builds on this coevolution model (reviewed in Di Giulio 1998; Szathmary 1999; Davis 1999; Di Giulio 2004). In particular, evidence for the existence of some

molecular fossils which are representative of biosynthetic pathways which, according to the theory, were involved in coevolution (Di Giulio 2004).

In 2000, Ronneberg and co-workers investigated the fundamental biochemical assumptions of the coevolution model and tested the validity of its statistical support. Their findings showed that the statistical significance demonstrated by the coevolution theory was actually calculated using incorrect assumptions. Based on what they believed to be correct assumptions, recalculations showed a 62% probability that chance alone could explain the pattern of codon assignments in the genetic code. It was therefore concluded that the coevolution theory could not adequately explain genetic code structure. Recently published literature also provides support for the inadequacy of this theory (Di Giulio 2008). DiGiulio argues that the coevolution theory is unable to adequately explain the earliest origins of the genetic code, since it does not assign a role to biosynthetic relationships between early amino acids, and proposes an extension to this theory which takes account of these early relationships (Di Giulio 2008).

### *Stereochemical theory*

This theory posits that a chemical interaction takes place between amino acids and their cognate codons/anticodons and this influences codon assignment. According to this theory, the structure of the genetic code is not accidental, rather planned. Experiments to study the binding of RNA aptamers (short RNA molecules) to amino acids demonstrated that those aptamers selected from random sequence mixtures by amino acid binding were significantly enriched with cognate codons for the respective amino

acids (Knight and Landweber 1998). The probability of such correlations occurring by chance was very low and out of 8 amino acids tested, only glutamine showed no correlation (Yarus et al. 2005). This theory has a number of shortcomings, most notably that: (1) affinities between codon/anticodon and amino acids are often only weak and (2) it fails to account for the assignment of those codons which do not show any chemical interactions with their cognate amino acids.

### *Adaptive theory*

It is hypothesised that natural selection has led to codon assignments that minimise the deleterious effects of single point mutations (Sonneborn 1965; Zuckerkandl and Pauling 1965) or translational errors (Woese 1967).

Quantitative evidence supporting this so-called 'error minimisation' hypothesis was first provided by Haig and Hurst (1991). However, there are a number of inherent problems with this theory and more recently, it has been suggested that the robustness of the universal code may actually be a product of evolution and selection pressures that are in no way associated with error minimisation (Stoltzfus and Yampolsky 2007).

### Conclusions

Despite many textbooks still referring to the genetic code as 'universal', the substantial body of evidence published over the past twenty years on the existence of a number of alternative codes shows that these are not merely 'exceptions' and that the genetic code should no longer be considered as universal. What still remains to be elucidated are the reasons for these code changes, since no one theory proposed to date provides an adequate



explanation for codon reassignment. Further, forty years after the publication of Crick's frozen-accident hypothesis, little progress has been made towards defining the evolution of the genetic code. While it appears highly likely that evolution of the code did involve a combination of some sort of frozen accident together with selection for error minimisation, other factors yet to be identified may also have played a key role.

## References

- Andersson, S. G. & Kurland, C. G. 1995, , Genomic evolution drives the evolution of the translation system', *Biochemistry and Cell Biology* , vol. 73, pp. 775-87.
- Barrell, B. G., Bankier, A. T., Drouin, J. 1979, ' A different genetic code in human mitochondria', *Nature* , vol. 282, no. 5735, pp. 189-94.
- Cavalcanti, A. R. O. & Landweber, L. F. 2004, ' Genetic code', *Current Biology* , vol. 14, no. 4, pp. R147.
- Crick, F. H. C. 1966, ' Codon-anticodon pairing: the wobble hypothesis', *Journal of Molecular Biology* , vol. 19, pp. 548-55.
- Crick, F. H. C. 1968, ' The origin of the genetic code', *Journal of Molecular Biology* , vol. 38, pp. 367-79.
- Davis, B. K. 1999, ' Evolution of the genetic code', *Progress in Biophysics and Molecular Biology* , vol. 72, pp. 157-243.
- Di Giulio, M. 1998, ' Reflections on the origin of the. genetic code: a hypothesis', *Journal of Theoretical Biology* , vol. 191, no. 2, pp. 191-6.

Di Giulio, M. 2004, ' The coevolution theory of the origin of the genetic code', *Physics of Life Reviews* , vol. 1, no. 2, pp. 128–37.

Di Giulio, M. 2008, ' An extension of the coevolution theory of the origin of the genetic code', *Biology Direct* , vol. 3, pp. 37.

Haig, D. & Hurst, L. D. 1991, ' A quantitative measure of error minimization in the genetic code', *Journal of Molecular Evolution* , vol. 33, pp. 412–7.

Hershberg, R. & Petrov, D. A. 2008, ' Selection on codon bias', *Annual Review of Genetics* , vol. 42, pp. 287–99.

Hinegardner, R. T. & Engelberg, J. 1963, ' Rationale for a universal genetic code', *Science* , vol. 142, pp. 1083–55.

Knight, R. D. & Landweber, L. F. 1998, ' Rhyme or reason: RNA-arginine interactions and the genetic code', *Chemistry and Biology* , vol. 5, pp. 215–20.

Knight, R. D., Freeland, S. J., Landweber, L. F. 1999, ' The 3 faces of the genetic code', *Trends in Biochemical Sciences* , vol. 24, no. 6, pp. 241–7.

Massey, S. E. G., Moura, P., Beltrao, R., Almeida, J. R., Garey, M. F., Tuite, M. F., Santos, M. A. S. 2003, ' Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp.', *Genome Research* , vol. 13, pp. 544–57.

Nirenberg, M. W. & Matthaei, J. H. 1961, ' The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides',

*Proceedings of the National Academy of Sciences USA* , vol. 47, pp. 1588–602.

Nirenberg, M. W., 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 Harbour Symposium on Quantitative Biology* , vol. 31, pp. 11–24.

Osawa, S. & Jukes, T. H. 1989, ' Codon reassignment (codon capture) in evolution', *Journal of Molecular Evolution* , vol. 28, pp. 271–8.

Osawa, S., Jukes, T. H., Watanabe, K., Muto, A. 1992, ' Recent evidence for evolution of the genetic code', *Microbiological Reviews* , vol. 56, no. 1, pp. 229–64.

Ronneberg, T. A., Landweber, L. F., Freeland, S. J. 2000, ' Testing a biosynthetic theory of the genetic code: fact or artefact', *Proceedings of the National Academy of Sciences USA* , vol. 97, no. 25, pp. 13690–5.

Schultz, D. W. & Yarus, M. 1994, ' Transfer RNA mutation and the malleability of the genetic code', *Journal of Molecular Biology* , vol. 235, pp. 1377–80.

Singer, M. & Berg, P. 1991, *Genes and genomes: a changing perspective* , University Science Books, Mill Valley, California.

Sonneborn, T. M. 1965, ' Degeneracy of the genetic code: extent, nature, and genetic implications. In *Evolving Genes and Proteins* (Bryson, V. & Vogel, H. J. Eds), pp. 377–97, Academic Press, New York.

Stoltzfus, A. & Yampolsky, L. Y. 2007, ' Amino acid exchangeability and the adaptive code', *Journal of Molecular Evolution* , vol. 65, pp. 456-62.

Szathmary, E. 1999, ' The origin of the genetic code', *Trends in Genetics* , vol. 15, pp. 223-9.

Szymański, M. & Barciszewski, J. 2007, ' The genetic code – 40 years on', *Acta Biochimica Polonica* , vol. 54, no. 1, pp. 51-4.

Travers, A. 2006, ' The evolution of the genetic code revisited', *Origins of Life and Evolution of Biospheres* , vol. 36, pp. 549-55.

Watson, J. D. & Crick, F. H. 1953, ' Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid', *Nature*, vol. 171, no. 4356, pp. 737-8.

Woese, C. R. 1965, ' On the evolution of the genetic code', *Proceedings of the National Academy of Sciences USA*, vol. 54, pp. 1546-52.

Woese, C. R. 1967, *The Genetic Code: the Molecular Basis for Genetic Expression* . Harper and Row, New York.

Wong, J. T-F. 1975, ' Title needed', *Proceedings of the National Academy of Sciences USA* , vol. 72, pp. 1909-12.

Wong, J. T-F. & Bronskill, P. M. 1979, ' Inadequacy of prebiotic synthesis as origin of proteinous amino acids', *Journal of Molecular Evolution* , vol. 13, pp. 115-25.

Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y., Osawa, S. 1985, ' UGA is read as tryptophan in *Mycoplasma capricolum* ', *Proceedings of the National Academy of Sciences USA* , vol. 82, pp. 2306–9.

Yarus, M., Caporaso, J. G., Knight, R. 2005, ' Origins of the genetic code: the escaped triplet theory', *Annual review of Biochemistry* , vol. 74, pp. 179–98.

Zuckerlandl, E. & Pauling, L. 1965, ' Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins* (Bryson, V. & Vogel, H. J. Eds), pp. 97–166, Academic Press, New York.

Table 1. Alternatives to the universal genetic code (from Szymański and Barciszewski 2007, p. 53).

Codon	Universal code	Exceptions
UUA	Leu	Stop codon in <i>Thraustochytrium</i> mitochondrial code Initiation codon in protozoan mitochondrial codes and <i>Mycoplasma</i> and <i>Spiroplasma</i> codes
UUG	Leu	Initiation codon in standard, bacterial and some mitochondrial codes
UCA	Ser	Stop codon in <i>Scenedesmus obliquus</i> mitochondrial code
UAA	Stop	Gln codon in ciliate nuclear code Tyr in alternative flatworm mitochondrial code Pyl in Archaea (Methanosarcinaceae) decoded by Pyl tRNA
UAG	Stop	Gln codon in ciliate nuclear code Leu codon in chlorophyceae and <i>Scenedesmus</i> mitochondrial codes

		Trp codon in mitochondrial codes Cys codon in euplotid nuclear code
UGA	Stop	codon (depends on presence of SElenoCysteine Insertion Sequence element in mRNA)
CUU	Leu	Thr codon in yeast mitochondrial code
CUC	Leu	Thr codon in yeast mitochondrial code
CUA	Leu	Thr codon in yeast mitochondrial code
CUG	Leu	Thr codon in yeast mitochondrial code and alternative yeast mitochondrial code Initiation codon in standard, bacterial and some mitochondrial codes
AUU	Ile	Initiation codon in bacterial and some mitochondrial codes
AUC	Ile	Initiation codon in bacterial and some mitochondrial codes
AUA	Ile	Met codon in mitochondrial codes of vertebrates, yeast and some invertebrates Initiation codon in bacterial and some mitochondrial codes
AAA	Lys	Asn codon in flatworm and echinoderm mitochondrial codes
AGA	Arg	Stop codon in vertebrate mitochondrial code Gly codon in ascidian mitochondrial code Ser codon in mitochondrial codes
AGG	Arg	Stop codon in vertebrate mitochondrial code Gly codon in ascidian mitochondrial code Ser codon in mitochondrial codes
GUG	Val	Initiation codon in bacterial and some mitochondrial codes