Genetics / Génétique

Degeneracy in the genetic code and its symmetries by base substitutions

Jean-Luc Jestin

Unité de chimie organique, département de biologie structurale et chimie, Institut Pasteur, 28, rue du Docteur-Roux, 75724 Paris cedex 15, France

Abstract

Degeneracy in the genetic code is known to minimise the deleterious effects of the most frequent base substitutions: transitions at the third base of codons are generally synonymous substitutions. Transversions that alter degeneracy were reported by Rumer. Here the other transversions are shown to leave invariant degeneracy when applied to the first base of codons. As a summary, degeneracy is considered with respect to all three types of base substitutions, the transitions and the two types of transversions. The symmetries of degeneracy by base substitutions are independent of the representation of the genetic code and discussed with respect to the quasi-universality of the genetic code. *To cite this article: J.-L. Jestin, C. R. Biologies 329 (2006).*
© 2006 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Résumé

La dégénérescence du code génétique et ses symétries par substitutions de base. La dégénérescence du code génétique minimise les effets délétères des substitutions les plus fréquentes : les transitions à la troisième base de codons sont généralement synonymes. Des transversions qui altèrent la dégénérescence ont été rapportées par Rumer. Les autres transversions appliquées à la première base des codons laissent invariantes la dégénérescence. En résumé, la dégénérescence est considérée par rapport à l'ensemble des trois types de substitutions, les deux types de transversions et les transitions. Les symétries de la dégénérescence par substitutions de base sont indépendantes de la représentation du code génétique et sont discutées par rapport à la quasi-universalité de ce dernier. *Pour citer cet article : J.-L. Jestin, C. R. Biologies 329 (2006).*
© 2006 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Keywords: Molecular evolution; Transition; Transversion; Mutation; Symmetry; Codon; Genetic code

Mots-clés : Évolution moléculaire ; Transition ; Transversion ; Mutation ; Symétrie ; Codon ; Code génétique

Living systems require the cooperation between nucleic acids and proteins, as shown by RNA-directed protein synthesis and by polymerase-catalysed nucleic acid amplification [1,2]. Reducing the effects of biochemical errors in the corresponding primordial systems has often been suggested as the major force shaping the genetic code [3–7].

Further progress on the understanding of why the genetic code is the way it is may result from the consideration of empirical rules that fit with models to be found.

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC		TCC		TAC		TGC	
TTA	Leu	TCA		TAA	stop	TGA	stop
TTG		TCG		TAG		TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC		CCC		CAC		CGC	
CTA		CCA		CAA	Gln	CGA	
CTG		CCG		CAG		CGG	
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC		ACC		AAC		AGC	
ATA		ACA		AAA	Lys	AGA	Arg
ATG	Met	ACG		AAG		AGG	
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC		GCC		GAC		GGC	
GTA		GCA		GAA	Glu	GGA	
GTG		GCG		GAG		GGG	

Fig. 1. A common representation of the standard genetic code.

Rumer identified, for example, transversions that, when applied to all three bases of codons, exchange sets of four codons for which the third base does not have to be defined so as to specify an amino acid into sets of four codons for which the third base is necessary to identify unambiguously the codons' assignments [8,9]. More precisely, the set of 64 codons can be divided into a set of 32 codons noted group (IV), whose third base does not have to be specified so as to define an amino acid. For the other 32 codons, noted group (II), three bases have to be specified so as to define unambiguously an amino acid or a stop signal. Rumer observed that there exists a unique symmetry exchanging group (IV) into group (II) that substitutes G into T, T into G, C into A or A into C and which is applied to all three codon bases (Figs. 1–3). These transversions are represented by the symbol **o** and by its corresponding central symmetry in the standard representation of the genetic code (Fig. 2). These same transversions altering degeneracy are represented by a horizontal line in Rumer's representation of the genetic code (Fig. 3) [8].

Another unique symmetry is described here: it exchanges each group into itself by substitution of G into C, C into G, A into T or T into A and is applied to the first codon base. In the standard representation of the genetic code, it is represented as an axial symmetry: the lines noted (α) exchanges sets of four codons of lines (A) into sets of four codons of lines (C); the lines noted (β) exchanges sets of four codons of lines (B) into sets of four codons of lines (D) (Fig. 2). This symmetry is represented by oblique lines in Rumer's representation of the genetic code (Fig. 3). These transversions,

together with the transversions noted by Rumer, represent all possible transversions.

The transitions are known to be synonymous substitutions when applied to the third codon base in the vertebrate mitochondrial genetic code for example (Fig. 2). The two types of transversions together with the transitions represent the complete set of possible base substitutions (Fig. 4). Each type of substitutions either leaves invariant or alters degeneracy depending on whether they are applied to the first base of codons, to the third base of codons or to all three bases of codons (Fig. 4).

These symmetries are clearly independent of the representations of the genetic code (Figs. 2 and 3) and can therefore be considered as intrinsic properties of the genetic code. Given a codon assigned to a group ((IV) or (II)), the three types of base substitutions (i.e. the transitions applied to the third base of codons and the two types of transversions applied to the first base of codons or to the three codon bases as indicated above) define three other codons whose group assignments are then predicted by the symmetries.

In the standard genetic code, three amino acids Leu, Arg and Ser are encoded by six codons, which can be considered each as two sets of synonymous codons, one belonging to group (IV) and one belonging to group (II). The vertebrate mitochondrial genetic code has been chosen here as an example (Figs. 2–4), as neither amino acids nor stop signals are coded by one codon or by three codons. This is not the case of the standard genetic code, as tryptophan and methionine are encoded by a single codon and as stop signals are encoded by three codons (Fig. 1). These differences in degeneracy for a few amino acids and for the stop codons between a

(A)	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
	TTC		TCC		TAC		TGC	
	TTA	Leu	TCA		TAA	stop	TGA	Trp
	TTG		TCG		TAG		TGG	

(B)	CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
	CTC		CCC		CAC		CGC	
	CTA		CCA		CAA	Gln	CGA	
	CTG		CCG		CAG		CGG	

α

(C)	ATT	Ile	ACT	Thr **o**	AAT	Asn	AGT	Ser
	ATC		ACC		AAC		AGC	
	ATA	Met	ACA		AAA	Lys	AGA	stop
	ATG		ACG		AAG		AGG	

β

(D)	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
	GTC		GCC		GAC		GGC	
	GTA		GCA		GAA	Glu	GGA	
	GTG		GCG		GAG		GGG	

Fig. 2. A representation of the vertebrate mitochondrial genetic code. Rumer's transformation is the central symmetry for this representation of the genetic code and is noted **o**: it exchanges a set of four codons into another set of four codons by substitutions of A into C, C into A, T into G and G into T applied to the three codon bases. If a four-codon set codes for a single amino acid, then the symmetrical four-codon set requires the last codon base to be given so as to define unambiguously an amino acid or a stop signal. Conversely, if a four-codon set requires the last base to be given so as to define unambiguously an amino acid or a stop signal, then the symmetrical four-codon set encodes a single amino acid. The transformation described here is for this representation of the genetic code an axial symmetry with respect to the line (α) for the four-codon groups in lines (A) and (C) and an axial symmetry with respect to the line (β) for the four-codon groups in lines (B) and (D). It corresponds to the transversion substituting G into C, C into G, A into T and T into A and is applied to the first base of codons. If a four-codon set requires the third base of codons to be given so as to define unambiguously an amino acid or a stop signal, then the symmetrical four-codon set requires also the third base of codons to be given to define unambiguously the assignments. If the four-codon set encodes a single amino acid, then the symmetrical four-codon set encodes also a single amino acid.

Gly	Arg	Ala	Pro	Thr	Ser	Val	Leu
GGG	CGG	GCG	CCG	ACG	TCG	GTG	CTG
GGA	CGA	GCA	CCA	ACA	TCA	GTA	CTA
GGC	CGC	GCC	CCC	ACC	TCC	GTC	CTC
GGT	CGT	GCT	CCT	ACT	TCT	GTT	CTT

Phe	Leu	Ile	Met	Tyr	stop	Asn	Lys	His	Gln	Asp	Glu	Cys	Trp	Ser	stop
TTT	TTA	ATT	ATA	TAT	TAA	AAT	AAA	CAT	CAA	GAT	GAA	TGT	TGA	AGT	AGA
TTC	TTG	ATC	ATG	TAC	TAG	AAC	AAG	CAC	CAG	GAC	GAG	TGC	TGG	AGC	AGG

Fig. 3. A representation of the vertebrate mitochondrial genetic code highlighting the symmetries of degeneracy by base substitutions. Rumer's transformation is an axial symmetry for this representation of the genetic code and is represented by the horizontal line: it exchanges a set of four codons into another set of four codons by substitutions of A into C, C into A, T into G and G into T applied to the three-codon bases. The transformation described here is for this representation of the genetic code an axial symmetry with respect to the oblique lines. It corresponds to the transversion substituting G into C, C into G, A into T and T into A and is applied to the first base of codons.

mitochondrial code and the standard genetic code do not change the symmetries as they depend only on whether the third codon base has to be defined for unambiguous assignment of an amino acid or of the stop signal.

The number of amino acids coded within the genetic code is quasi-universal and varies among the various known genetic codes: their diversity has been described by evolutionary models, which also account for the introduction of further amino acids such as selenocysteine within the genetic code [10,11]. Whether mitochondrial codes should be considered as ancestral codes or as codes that evolved during symbiosis remains unclear.
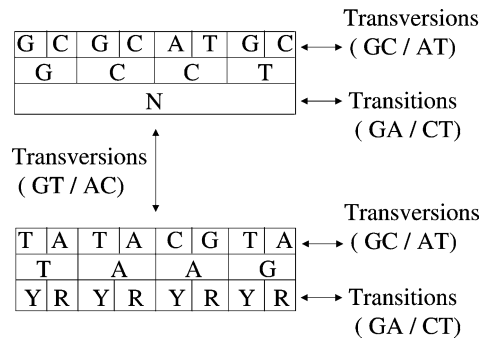
Fig. 4. A representation of the sets of synonymous codons highlighting the symmetries of degeneracy by the three types of base substitutions, transitions noted (GA/CT) and transversions noted (GC/AT: it is applied to the first base of codons) and (GT/AC: it is applied to the three bases of codons). The set of 64 codons of the genetic code is dissected into two groups depending on whether the third base of codons has to be given or not to define unambiguously an amino acid or the stop signal.

This diversity of codes is often associated to start signals as well as to stop signals. The codon assignment of stop signals was shown to optimise the tolerance of polymerase-induced frameshift mutations, i.e. to minimise the deleterious effects of single-base deletions catalysed by nucleic acid polymerases [12]. Accordingly, it should be of interest to verify experimentally whether the two different doublets coding for the stop signal in the vertebrate mitochondrial genetic code are sequence 'hot-spots' for single-base deletions catalysed by DNA polymerases in vertebrate mitochondria. A diversity of genetic codes was also created experimentally [13,14]. In this case, the diversity of genetic codes is not necessarily associated to start or to stop codons [13]. If we consider only the natural diversity of genetic codes, then the symmetries of degeneracy by base substitutions are almost universal. Whether the symmetries can be related to universal biochemical biases such as those found in genes, in genomes or in proteins [15,16] remains to be established. The absence of obvious links between biochemical properties and the transversions altering or leaving invariant degeneracy in the genetic code suggests that these symmetries may derive from physical constraints linked to intrinsic properties of the code: a model accounting for the symmetries by base substitutions of degeneracy in the genetic code is investigated [17].

References

[1] M. Eigen, Selforganisation of matter and the evolution of biological macromolecules, Naturwissenschaften 58 (1971) 465–523.

[2] J. Maynard-Smith, E. Szathmary, The Major Transitions in Evolution, Freeman, Oxford, 1995.

[3] T.M. Sonneborn, Degeneracy of the genetic code: extent, nature and genetic implications, in: V. Bryson, H.J. Vogel (Eds.), Evolving Genes and Proteins, Academic Press, New York, 1965, pp. 377–397.

[4] C.R. Woese, On the evolution of the genetic code, Proc. Natl Acad. Sci. USA 54 (1965) 1546–1552.

[5] A.L. Goldberg, R.E. Wittes, Genetic code: aspects of organization, Science 153 (1966) 420–424.

[6] S.J. Freeland, R.D. Knight, L.F. Landweber, L.D. Hurst, Early fixation of an optimal genetic code, Mol. Biol. Evol. 17 (2000) 511–518.

[7] G. Sella, D.H. Ardell, The impact of message mutation on the fitness of a genetic code, J. Mol. Evol. 54 (2002) 638–651.

[8] Y.B. Rumer, About the codon's systematization in the genetic code, Proc. Acad. Sci. USSR 167 (1966) 1393–1394 (in Russian).

[9] V.I. Shcherbak, Rumer's rule and transformation in the context of the co-operative symmetry of the genetic code, J. Theor. Biol. 139 (1989) 271–276.

[10] J. Ninio, The revised genetic code, Orig. Life Evol. Biosphere 20 (1990) 167–171.

[11] T.H. Jukes, S. Osawa, Evolutionary changes in the genetic code, Comp. Biochem. Physiol. 106B (1993) 489–494.

[12] J.L. Jestin, A. Kempf, Chain-termination codons and polymerase-induced frameshift mutations, FEBS Lett. 419 (1997) 153–156.

[13] V. Döring, H.D. Mootz, L.A. Nangle, T.L. Hendrickson, V. De Crécy-Lagard, P. Schimmel, P. Marlière, Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway, Science 292 (2001) 501–504.

[14] J.W. Chin, T.A. Cropp, J.C. Anderson, M. Mukherji, Z. Zhang, P.G. Schultz, An expanded eukaryotic genetic code, Science 301 (2003) 964–967.

[15] E.P.C. Rocha, A. Danchin, A. Viari, Universal replication biases in bacteria, Mol. Microbiol. 32 (1999) 11–16.

[16] G. Pascal, C. Médigue, A. Danchin, Universal biases in protein composition of model prokaryotes, Proteins: Struct. Funct. Bioinfor. 60 (2005) 27–35.

[17] J.-L. Jestin, Degeneracy in the genetic code, http://arXiv.org/abs/physics/0303045.