

Genetic Code Evolution Started with the Incorporation of Glycine, Followed by Other Small Hydrophilic Amino Acids

Harold S. Bernhardt · Wayne M. Patrick

Received: 25 April 2014 / Accepted: 23 May 2014 / Published online: 12 June 2014
© Springer Science+Business Media New York 2014

Abstract We propose that glycine was the first amino acid to be incorporated into the genetic code, followed by serine, aspartic and/or glutamic acid—small hydrophilic amino acids that all have codons in the bottom right-hand corner of the standard genetic code table. Because primordial ribosomal synthesis is presumed to have been rudimentary, this stage would have been characterized by the synthesis of short, water-soluble peptides, the first of which would have comprised polyglycine. Evolution of the code is proposed to have occurred by the duplication and mutation of tRNA sequences, which produced a radiation of codon assignment outwards from the bottom right-hand corner. As a result of this expansion, we propose a trend from small hydrophilic to hydrophobic amino acids, with selection for longer polypeptides requiring a hydrophobic core for folding and stability driving the incorporation of hydrophobic amino acids into the code.

Keywords Genetic code · Polyglycine · Water-soluble peptides · Transfer RNA · Transition mutation

Many theories have been proposed to explain the evolution of the genetic code. In a recent issue of *J. Mol. Evol.*, Brian Francis hypothesized that the “improved or altered protein function afforded by the changes in amino acid function provided the selective advantage underlying the expansion of the genetic code” (Francis 2013). While we agree that improved protein function was a primary driver of the code’s evolution, we believe that other aspects also played

important roles, such as the manner in which coding triplets are likely to have undergone mutation.

Francis’s model followed those of Eigen and Winkler–Oswatitsch (1981) and Higgs (2009), who proposed that the original genetic code consisted only of the four GNC triplets, coding for Gly, Asp/Glu, Ala and Val from the bottom row of the genetic code table. Higgs argued that these were the first encoded amino acids because they are found in the highest quantities in a range of abiotic environments (including meteorites and hydrothermal vents), and as the products of Miller-type early-Earth simulation experiments (Bada 2013). In addition, they are not the products of long biosynthetic pathways.

Previously, we proposed that the first tRNA to evolve was glycine tRNA, and that glycine was, therefore, the first genetically encoded amino acid. This was based on work by Di Giulio (1992), supported by Dick and Schamel (1995) and Widmann et al. (2005), suggesting that the first tRNA resulted from the duplication and ligation of a hairpin approximately half its length. If the original hairpin possessed a 3’CCA terminus—the universally conserved tRNA aminoacylation sequence—tandem ligation would have produced two CCA sequences, one of which was proposed to have formed part of the nascent NCC anticodon in the middle of the molecule (Bernhardt and Tate 2008). We proposed that the stable interaction between NCC and GGN (the universally conserved glycine codon sequence) marked the genesis of coded protein synthesis (Bernhardt and Tate 2008, 2010).

Francis (2013) outlined the potential utility of glycine-containing peptides in a variety of roles, including anion-binding motifs (of which phosphate-binding variants may have been particularly important in an RNA world). We have gone further, proposing that the first coded peptides comprised polyglycine alone (Bernhardt and Tate 2010).

H. S. Bernhardt (✉) · W. M. Patrick
Department of Biochemistry, University of Otago,
PO Box 56, Dunedin 9054, New Zealand
e-mail: harold.bernhardt@otago.ac.nz

		2nd codon base			
		U	C	A	G
1st codon base	U	UUN Phe Leu	UCN Ser	UAN Tyr stop	UGN Cys stop Trp
	C	CUN Leu	CCN Pro	CAN His Gln	CGN Arg
	A	AUN Ile Met	ACN Thr	AAN Asn Lys	AGN Ser Arg
	G	GUN Val	GCN Ala	GAN Asp Glu	GGN Gly

Fig. 1 The genetic code table. Codon assignment is proposed to have begun in the *bottom right-hand corner* with the GGN (Gly) codon box (*dark grey*), followed by AGN (Ser) and GAN (Asp/Glu) codon boxes (*medium grey*), followed by the AAN codon box (*light grey*), and then radiated outwards. The earliest amino acids incorporated into the code (with their codons) are shown in *white type*. The most hydrophobic amino acids are in the *left-most column* and *top row* of the table, furthest in sequence space from the *bottom right-hand corner*. *Thick lines* divide the table into quadrants, between which a transversion mutation is required to alter the amino acid that is encoded; *thin lines* separate codon boxes, between which a (more frequent) transition mutation is required to alter coding

Such peptides would have possessed useful attributes, such as being able to form cavities or ‘nest’-like structures (Milner-White and Russell 2008; see also Jabs et al. 1999), and to coordinate a variety of metal ions (Rabenstein and Libich 1972), including some with catalytic ability (Pogni et al. 1999), sufficient for the positive selection—and further evolution—of the nascent system for genetic coding. Starting with glycine tRNA, we propose that the genetic code evolved through a process of tRNA duplication and mutation similar to that which occurs in modern gene evolution: through transition and transversion mutations of the glycine tRNA anticodon sequence (Saks et al. 1998) and acceptor stem sequence (Schimmel et al. 1993) which play a major role in determining aminoacylation specificity, as well as in the codon sequences of the early mRNAs. In modern DNA genomes, transition mutations (pyrimidine→pyrimidine; purine→purine) occur much more frequently than transversion mutations (pyrimidine→purine and vice versa) (Collins and Jukes 1994; Ebersberger et al. 2002). A major contributing factor is that spontaneous deamination of cytosine yields uracil.

Therefore, we propose that codon assignment radiated outwards from the glycine codons in the bottom right-hand corner of the genetic code table (Fig. 1). The mutation of glycine anticodons to those for either Ser or Asp/Glu involved a single transition mutation. Although Arg shares a codon box with Ser in the modern code, we consider it

unlikely that Arg was an early codon assignment, as it is not produced in early-Earth simulation experiments and has a complex biosynthetic pathway. The AAN codons differ from Ser and Asp/Glu codons by a second transition mutation, and would have been the next codons assigned. However, the original amino acids these codons were assigned to are unclear, as the two amino acids in the modern code—Asn and Lys—are also likely to have been later codon assignments, and for the same reasons as for Arg. These three amino acids were probably incorporated into the code either through the later reassignment of Ser and Asp/Glu codons, or by stop codon takeover (Lehman and Jukes 1988). Consequently, we propose that the next amino acids to be incorporated following glycine were similarly small and hydrophilic. As a result, the peptides produced would have been water-soluble, and also short, because most codons at this stage still coded for chain termination, and because the ribosomal machinery would still have been rudimentary, increasing the likelihood of premature termination events. Such short water-soluble peptides could have possessed a number of selectable functions, including those outlined by Francis (2013). However, emerging within a fully functioning RNA world, it would not have been necessary for such peptides to possess all the functionalities of modern proteins.

Over time, as the proportion of sense codons increased and the ribosomal machinery became more refined, longer polypeptides could be synthesized reproducibly. Many of these would have been under strong positive selection due to their improved characteristics as, for example, enzymes or structural proteins. Once such polypeptides attained a critical size, the presence of a hydrophobic core (comprising the side-chains of hydrophobic amino acids) would have been required for folding and stability. Therefore, selection for longer polypeptides would have driven the incorporation of hydrophobic amino acids into the code; Fournier and Gogarten (2007) and Cleaves (2010) have put forward similar arguments. In support of our hypothesis, the eight most hydrophobic amino acids (Phe, Leu, Ile, Met, Val, Tyr, Cys and Trp) have codons that group together in the left-most column and top row of the genetic code table, equidistant in sequence space from the glycine codon box (Fig. 1).

The co-evolutionary hypothesis (Wong 1975, 2005; Di Giulio 2008) and the anticodon-flipping model of Crick et al. (1976) both proposed a similar order of early amino acid incorporation to our model. However, statistical (Amirnovin 1997; Di Giulio 1999; Ronneberg et al. 2000) and logical (Higgs 2009) support for the co-evolutionary hypothesis is problematic, while experimental support for an anticodon-flipping mechanism from tRNA structural studies is missing. In contrast, our hypothesis for duplication and mutation of an original glycine tRNA provides an

explanation for the evolution of the genetic code as captured in the organization of the modern genetic code table.

Acknowledgments This work was supported by a Grant from the Marsden Fund to HSB, and by a Rutherford Discovery Fellowship to WMP.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Amirnovin R (1997) An analysis of the metabolic theory of the origin of the genetic code. *J Mol Evol* 44:473–476
- Bada JL (2013) New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chem Soc Rev* 42:2186–2196. doi:10.1039/c3cs35433d
- Bernhardt HS, Tate WP (2008) Evidence from glycine transfer RNA of a frozen accident at the dawn of the genetic code. *Biol Direct* 3:53. doi:10.1186/1745-6150-3-53
- Bernhardt HS, Tate WP (2010) The transition from noncoded to coded protein synthesis: did coding mRNAs arise from stability-enhancing binding partners to tRNA? *Biol Direct* 5:16. doi:10.1186/1745-6150-5-16
- Cleaves HJ (2010) The origin of the biologically coded amino acids. *J Theor Biol* 263:490–498
- Collins DW, Jukes TH (1994) Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics* 20:386–396
- Crick FH, Brenner S, Klug A, Piecznik G (1976) A speculation on the origin of protein synthesis. *Orig Life* 7:389–397
- Di Giulio M (1992) On the origin of the transfer RNA molecule. *J Theor Biol* 159:199–214
- Di Giulio M (1999) The coevolution theory of the origin of the genetic code. *J Mol Evol* 48:253–255 (includes reply by Amirnovin and Miller)
- Di Giulio M (2008) An extension of the coevolution theory of the origin of the genetic code. *Biol Direct* 3:37. doi:10.1186/1745-6150-3-37
- Dick TP, Schamel WA (1995) Molecular evolution of transfer RNA from two precursor hairpins: implications for the origin of protein synthesis. *J Mol Evol* 41:1–9
- Ebersberger I, Metzler D, Schwarz C, Pääbo S (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet* 70:1490–1497
- Eigen M, Winkler-Oswatitsch R (1981) Transfer-RNA, an early gene? *Naturwissenschaften* 68:217–228
- Fournier GP, Gogarten JP (2007) Signature of a primitive genetic code in ancient protein lineages. *J Mol Evol* 65:425–436
- Francis BR (2013) Evolution of the genetic code by incorporation of amino acids that improved or changed protein function. *J Mol Evol* 77:134–158
- Higgs PG (2009) A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. *Biol Direct* 4:16. doi:10.1186/1745-6150-4-16
- Jabs A, Weiss MS, Hilgenfeld R (1999) Non-proline cis peptide bonds in proteins. *J Mol Biol* 286:291–304
- Lehman N, Jukes TH (1988) Genetic code development by stop codon takeover. *J Theor Biol* 135:203–214
- Milner-White EJ, Russell MJ (2008) Predicting the conformations of peptides and proteins in early evolution. *Biol Direct* 3:3. doi:10.1186/1745-6150-3-3
- Pogni R, Baratto MC, Busi E, Basosi R (1999) EPR and O₂-scavenger activity: Cu(II)-peptide complexes as superoxide dismutase models. *J Inorg Biochem* 73:157–165
- Rabenstein DL, Libich S (1972) Nuclear magnetic resonance studies of the solution chemistry of metal complexes. V. Cadmium, zinc, and lead complexes of polyglycine peptides. *Inorg Chem* 11:2960–2967
- Ronneberg TA, Landweber LF, Freeland SJ (2000) Testing a biosynthetic theory of the genetic code: fact or artifact? *Proc Natl Acad Sci USA* 97:13690–13695
- Saks ME, Sampson JR, Abelson J (1998) Evolution of a transfer RNA gene through a point mutation in the anticodon. *Science* 279:1665–1670
- Schimmel P, Giegé R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc Natl Acad Sci USA* 90:8763–8768
- Widmann J, Di Giulio M, Yarus M, Knight R (2005) tRNA creation by hairpin duplication. *J Mol Evol* 61:524–530
- Wong JT (1975) A co-evolution theory of the genetic code. *Proc Natl Acad Sci USA* 72:1909–1912
- Wong JT (2005) Coevolution theory of the genetic code at age thirty. *BioEssays* 27:416–425