

Question 6: Coevolution Theory of the Genetic Code: A Proven Theory

Jeffrey Tze-Fei Wong

Received: 22 December 2006 / Accepted: 12 February 2007 /
Published online: 5 July 2007
© Springer Science + Business Media B.V. 2007

Abstract The coevolution theory proposes that primordial proteins consisted only of those amino acids readily obtainable from the prebiotic environment, representing about half the twenty encoded amino acids of today, and the missing amino acids entered the system as the code expanded along with pathways of amino acid biosynthesis. The isolation of genetic code mutants, and the antiquity of pretran synthesis revealed by the comparative genomics of tRNAs and aminoacyl-tRNA synthetases, have combined to provide a rigorous proof of the four fundamental tenets of the theory, thus solving the riddle of the structure of the universal genetic code.

Keywords Genetic code · Coevolution theory · Amino acid biosynthesis · Pretran synthesis

Introduction

Is there anything solid known about the origin of the genetic code? To answer this question posed at the Erice Conference (2006), the present study examines the nature of the proof for the coevolution theory of the genetic code, or CET (Wong 2005a).

It is generally supposed that the development of life forms on Earth or elsewhere requires energy, information and catalysis. The need for energy is prescribed by thermodynamics, and evolution entails competition between replicating information systems, but is catalysis essential? This question is answered in the affirmative by recognizing the finite chemical stability of heteropolymeric templates/genes, which cannot multiply if one or more breaks in their structures occur during one replication cycle. The *stability theorem* follows:

$$kLT < 1$$

Presented at: *International School of Complexity – 4th Course: Basic Questions on the Origins of Life*, “Ettore Majorana” Foundation and Centre for Scientific Culture, Erice, Italy, 1–6 October 2006.

J. T.-F. Wong (✉)

Department of Biochemistry and Applied Genomics Laboratory, Hong Kong University of Science & Technology, Clear Water Bay, Hong Kong, China
e-mail: bejtw@ust.hk

where k is the rate of gene scission, L the number of inter-monomeric bonds in the genes and T the replication time. For a primitive ribo-organism consisting of three RNA genes each 50 nucleotide long, according to the stability theorem T must be less than 8.6 years. No biotic system utilizing RNA genes could acquire such a fast replication rate without catalysis. Enhancement of catalytic efficiency thus had to be a foremost evolutionary incentive at life's origin (Wong and Xue 2002).

Since ribozymes often display a low k_{cat} , thereby causing a low k_{cat}/K_m , or catalytic efficiency (Wong and Xue 2002), the transition from an RNA-like World to a Protein World is postulated to begin with the addition of amino acids and peptides to ribozymes to improve catalytic efficiency (Wong 1991; Szathmary 1993), which is supported by the observed ribozymic synthesis of peptide-RNA conjugates (Zhang and Cech 1997) and peptide activation of ribozymic function (Robertson et al. 2004).

Tenets of Coevolution Theory

The Asp-family amino acids produced from Asp as biosynthetic precursor occupy three codon boxes across the ANN row of the code. Ile and Met, sharing the AUN codon box, are sibling amino acids derived from Asp through homoserine. Cys and Trp, sharing the UGN codon box, are both biosynthetic products of Ser. Based on such biosynthesis-codon allocation relationships, CET proposes that the code at first encoded not all 20 proteinous amino acids, but only about 10 Phase 1 amino acids that were readily supplied by prebiotic synthesis. Subsequently, these brought into the code through primitive biosynthesis the Phase 2 amino acids, which vastly enhanced the catalytic and specificity performance of proteins. Still later, enhancement was continued with the entry of Phase 3 amino acids through post-translational modifications. *Pretran synthesis*, whereby a precursor amino acid while bonded to its tRNA is converted to a product amino acid, furnished an important mechanism for the entry of some Phase 2 amino acids into the code. The advantage of pretran synthesis is that the nascent product immediately receives the anticodons on the tRNA (Wong 1975a, 1981, 2005a; Di Giulio 2004). Accordingly, the basic tenets of CET are:

- Tenet 1 The prebiotic environment did not supply all 20 proteinous amino acids at life's origin, but had to be complemented by sourcing through inventive biosynthesis.
- Tenet 2 Pretran synthesis provided mechanisms for the encoding of some Phase 2 amino acids.
- Tenet 3 Biosynthetic relationships between amino acids were an important determinant of codon allocations.
- Tenet 4 The amino acid ensemble encoded by the genetic code is mutable, allowing early code expansion to admit the Phase 2 amino acids.

Genetic Code Mutation

All extant organisms use the same 20 encoded amino acids. In the face of this 3-billion year invariance, there is only one way to prove Tenet 4, which is to mutate the code. This was first achieved by the isolation of genetic code mutants of *Bacillus subtilis* where 4-fluoroTrp effectively replaces Trp as an encoded amino acid for indefinite cell growth, in

some mutants even displacing Trp entirely, with Trp being reduced to the status of an inhibitory analogue (Wong 1983, 2005a). More recently, 5-fluoroTrp and 6-fluoroTrp have also become genetically encoded amino acids fully capable of supporting indefinite growth (Mat et al. 2005). The addition of genetically encoded amino acids also has been extended to *E. coli*, yeast, mammalian cells, and to over 30 unnatural amino acids (Doring et al. 2001; Bacher and Ellington 2003; Bacher et al. 2004; Kohrer et al. 2004; Xie and Schultz 2005; Budisa 2006). As well, besides the top-down proteome-wide approach of code mutation employed in the displacement of Trp by 4-fluoroTrp, which throws direct light on code evolution, the low reactivity of some aaRS with tRNAs from another biological domain (Kwok and Wong 1980) has enabled a bottom-up position-specific approach for the genetic encoding of unnatural amino acids.

Active code evolution followed by a 3-billion year freeze is at first glance surprising, but it finds a ready parallel in human languages, where alphabets evolved to arrive at an adequate representation of the 40 different basic sounds of the human voice, and froze. Different alphabets froze with a different number of letters – Hebrew with 22, Latin 23, English 26, Cyrillic 33, and archaic Hungarian runan 39, but once the usage of any alphabet is established, it resists further evolutionary change. In this light, the Phase 1 amino acids from the prebiotic environment could launch life, but not allow the construction of high performance polypeptides. Therefore the Phase 2 code expansion was the Protein World's search for excellence. The dynamics of the coevolution process are such that evolution of the encoded amino acid ensemble, constantly enhancing the catalytic and specificity capabilities of proteins, never ceases until it arrives at a collection of amino acid side-chains with sufficient chemical versatility to ensure an extremely low error rate in the translation machinery, whereupon the code freezes because further revision would create an unacceptable level of noise in the context of low-noise translation and thus become an over-burdensome selective disadvantage (Wong 1976). The versatility of the 20-member amino acid code has withstood the test of time immemorial, underwriting such singular accomplishments of the Protein World as enzymes with diffusion-controlled kinetics (Wong 1975b), multicellular life and human intelligence.

The proving of Tenet 4 has opened up the genetic code to modifications and expansions to deepen understanding of protein structure and function, and broaden the scope of genetic engineering. Evolution is no longer confined merely to the endless sequence permutations of 20 standard amino acids. Instead, from now on in a sequel to life, both amino acid sequences and the amino acids themselves can be varied (Cohen 2000). Since the 20-amino acid code is so fundamental an attribute of life, the new genetic code mutants employing a different encoded amino acid ensemble in effect represent new types of life (Hesman 2000).

Primordial Pretran Synthesis

The lack of an efficient prebiotic synthesis for all 20 standard amino acids (Wong 1988, 2005a), and the chemical instability of some amino acids (Wong and Bronskill 1979; Wong 1984) support Tenet 1. The evident correlations between biosynthesis and codon allocations support Tenets 1–3. However, strong as these lines of evidence are, they fall short of a rigorous proof. Instead, rigorous proof has to be derived as follows.

Gln-tRNA is produced using GlnRS in a *direct pathway* in some organisms, but from Glu-tRNA using pretran synthesis in an *indirect pathway* in other organisms. The question is, which of these two alternate pathways is primordial, and which is modern? The same

question arises for the synthesis of Asn-tRNA and Cys-tRNA, where both direct and indirect pathways are known. Because Tenet 2 postulates that some Phase 2 amino acids entered the primitive expanding code through pretran synthesis, it is disproven if pretran synthesis is strictly a modern invention. If pretran synthesis is primordial, proving all of Tenets 1–3 becomes straightforward. Three lines of evidence are germane in this regard:

- (a) The genetic distances between alloacceptor tRNAs accepting dissimilar amino acids indicate that tRNA evolution began with sequences closely clustered in sequence space, which became dispersed in time. *Methanopyrus*, with the lowest alloacceptor tRNA distances, represents the slowest evolver that stands closest to the last universal common ancestor, or LUCA (Xue et al. 2003). Anticodon usages (Tong and Wong 2004), as well as sequence homologies between potentially paralogous aaRS pairs (Xue et al. 2005) have provided independent evidence for a *Methanopyrus*-proximal LUCA. On this basis, the absence of GlnRS, AsnRS and CysRS from *Methanopyrus* (Sauerwald et al. 2005; Wong 2005a, b) establishes that LUCA lacked these three aaRS and employed pretran synthesis for the encoding of Gln, Asn and Cys.
- (b) Comparative phylogenetics indicate that the indirect pathways using pretran synthesis to produce Gln-tRNA and Asn-tRNA are primordial, whereas both direct and indirect pathways are equally ancient for Cys-tRNA (O'Donoghue et al. 2005, Sauerwald et al. 2005).
- (c) Selenocysteine, or Sec, enters proteins through pretran synthesis from Ser-tRNA via either the SelA or the PSTK/SepSecS pathway. Since no SecRS is known, only pretran synthesis is employed for Sec encoding in organisms. Comparative phylogenetics indicates that pretran synthesis of Sec is primordial, and was utilized by LUCA (Yuan et al. 2006).

These lines of evidence converge to the conclusion that pretran synthesis is not a modern invention like much of secondary metabolism, but a primordial occurrence that brought Gln, Asn, Sec, and likely Cys and some other Phase 2 amino acids into the code, thereby proving Tenet 2. The entry of Gln, Asn and Sec into the biotic system through pretran synthesis proves Tenet 1. The pretran synthesis origins of Gln and Asn is corroborated by the thermal instabilities of Gln and Asn which are such that Gln could not exceed 3.7×10^{-12} M and Asn could not exceed 2.4×10^{-8} M in the prebiotic environment (Wong and Bronskill 1979): Gln and Asn were simply unavailable at the start of life. The UV-instabilities of Cys, Met, Trp, His, Tyr and Phe (Wong 1984) also favor the bulk of these amino acids being supplied to the pre-LUCA biotic system by primitive biosynthesis. The pretran synthesis origins of Sec and Cys constitute a remarkable validation of CET's suggestion that the UGN codon box was originally a Ser-box that connects the UCN and AGY codons into a contiguous Ser-domain with single-base separations between codons.

Proving Tenet 2 is tantamount to also proving Tenet 3. The reason is, when Gln, Asn, Sec or Cys is formed *in situ* on tRNA through pretran synthesis, it immediately acquires the anticodon on the tRNA to which it is bonded. Consequently, CAA and CAG are allocated to Gln, AAU and AAC to Asn, UGA to Sec, and UGU and UGC to Cys because in each instance the allocated codons belonged to the pretran synthesis precursor. In these instances physicochemical attributes such as hydrophobicity or molecular volume could make only a minor contribution to codon allocation by influencing which among the precursor's codons were to be assigned to the product, and potential stereochemical interactions between Gln, Asn, Sec and Cys with their cognate codons/anticodons had little role to play.

Conclusion

CET suggested that amino acid biosynthesis was the predominant, but not the sole, determinant of codon allocations (Wong 1975a). Recent estimates of the strengths of three different determinants of codon allocations have made possible a quantitative assessment of their relative contributions to the selection of the universal genetic code (Wong 2005a) as:

$$[\text{Amino Acid Biosynthesis}] : [\text{Error Minimization}] : [\text{Stereo chemical Interaction}] \\ = 40,000,000 : 400 : 1$$

Thus the contribution of amino acid biosynthesis relative to other factors in shaping the code turns out to be far more overwhelming than could have been surmised. In *Chance and Necessity*, Monod (1972) identified three frontiers that represent the foremost challenge of biology: the problem of life's origins, the riddle of the code's origins, and the central nervous system. The proving of CET reveals that amino acid biosynthesis is the key to deciphering the riddle posed by the structure of the genetic code. Just as the map of a country so often tells the story of its history, the structure of the universal genetic code is a lasting inscription of the history of its coevolution with the primordial pathways of amino acid biosynthesis.

Acknowledgments I thank Drs. Henri Grosjean, Ka-Lok Tong and Hannah Hong Xue for valuable discussion.

References

- Bacher JM, Ellington AD (2003) The directed evolution of organismic chemistry: unnatural amino acid incorporation. In: Lapointe J, Brakier-Gingras L (eds) Translation mechanisms. Landes Bioscience, Georgetown, Texas, pp 80–84
- Bacher JM, Hughes RA, Wong JT, Ellington AD (2004) Evolving new genetic codes. Trends Ecol Evol 19:69–75
- Budisa N (2006) Engineering the genetic code. Wiley-VCH, Weinheim, pp 1–296
- Cohen P (2000) Life the sequel. New Sci 167:32–36
- Di Giulio M (2004) The coevolution theory of the origin of the genetic code. Physics of Life Reviews 1:128–137
- Doring V, Mootz HD, Nangle LA, Hendrickson TL, de Crecy-Lagard V, Schimmel P, Marliere P (2001) Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. Science 292:501–504
- Erice Conference Booklet (2006) Basic questions about the origins of life. In: Luisi PL, Pietronero L (eds), pp 7–10
- Hesman T (2000) Code breakers. Scientists are altering bacteria in a most fundamental way. Sci News 157:360–362
- Kohrer C, Sullivan EL, RajBhandary UL (2004) Complete set of orthogonal 21st aminoacyl-tRNA synthetase-amber, ochre and opal suppressor tRNA pairs: concomitant suppression of three different termination codons in an mRNA in mammalian cells. Nucleic Acids Res 32:6200–6211
- Kwok Y, Wong JT (1980) Evolutionary relationships between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. Can J Biochem 58:213–218
- Mat FWK, Xue H, Wong JT (2005) Genetic encoding of 4-, 5-, and 6-fluorotryptophans: role of oligogenic barriers. ISSOL 2005 Abstracts, pp 97–98
- Monod J (1972) Chance and necessity. Vintage Books, New York, pp 138–148
- O'Donoghue P, Sethi A, Woese CR, Luthy-Schulten ZA (2005) The evolutionary history of Cys-tRNA^{Cys} formation. Proc Natl Acad Sci USA 102:19003–19008
- Robertson MP, Knudsen SM, Ellington AD (2004) In vitro selection of ribozymes dependent on peptides for activity. RNA 10:114–127

- Sauerwald A, Zhu W, Major TA, Roy H, Palioura S, Jahn D, Whitman WB, Yates JR 3rd, Ibba M, Soll D (2005) RNA-dependent cysteine biosynthesis in archaea. *Science* 307:1969–1972
- Szathmari E (1993) Coding coenzyme handles: a hypothesis for the origin of the genetic code. *Proc Natl Acad Sci USA* 90:9916–9920
- Tong KL, Wong JT (2004) Anticodon and wobble evolution. *Gene* 333:169–177
- Wong JT (1975a) A co-evolution theory of the genetic code. *Proc Natl Acad Sci USA* 72:1909–1912
- Wong JT (1975b) Kinetics of enzyme mechanisms. Academic Press, London, pp 200–201
- Wong JT (1976) The evolution of a universal genetic code. *Proc Natl Acad Sci USA* 73:2336–2340
- Wong JT (1981) Co-evolution of the genetic code and amino acid biosynthesis. *Trends Biochem Sci* 6:33–36
- Wong JT (1983) Membership mutation of the genetic code: loss of fitness by tryptophan. *Proc Natl Acad Sci USA* 80:6303–6306
- Wong JT (1984) Evolution and mutation of the amino acid code. In: Ricard J, Cornish-Bowden A (eds) *Dynamics of biochemical systems*. Plenum, New York, pp 247–257
- Wong JT (1988) Evolution of the genetic code. *Microbiol Sci* 5:174–181
- Wong JT (1991) Origin of genetically encoded protein synthesis: a model based on selection for RNA peptidation. *Orig Life Evol Biosph* 21:165–176
- Wong JT (2005a) Coevolution theory of the genetic code at age thirty. *BioEssays* 27:416–425
- Wong JT (2005b) On the formation of asp-tRNA^{Asn} by aspartyl-tRNA synthetases. *BioEssays* 27:1309
- Wong JT, Bronskill PM (1979) Inadequacy of prebiotic synthesis as origin of proteinous amino acids. *J Mol Evol* 13:115–125
- Wong JT, Xue H (2002) Self-perfecting evolution of heteropolymer building blocks and sequences as the basis of life. In: Palyi G, Zucchi C, Caglioti L (eds) *Fundamentals of life*. Elsevier, Paris, pp 473–494
- Xie J, Schultz PG (2005) Adding amino acids to the genetic repertoire. *Curr Opin Chem Biol* 9:548–554
- Xue H, Tong KL, Marck C, Grosjean H, Wong JT (2003) Transfer RNA paralogs: evidence for genetic code-amino acid biosynthesis coevolution and an archaeal root of life. *Gene* 310:59–66
- Xue H, Ng SK, Tong KL, Wong JT (2005) Congruence of evidence for a *Methanopyrus*-proximal root of life based on transfer RNA and aminoacyl-tRNA synthetase genes. *Gene* 360:120–130
- Yuan J, Palioura S, Salazar JC, Su D, O'Donoghue P, Hohn MJ, Cardoso AM, Whitman WB, Soll D (2006) RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proc Natl Acad Sci USA* 103:18923–18927
- Zhang B, Cech TR (1997) Peptide bond formation by *in vitro* selected ribozymes. *Nature* 390:96–100