Coevolution theory of the genetic code at age thirty

J. Tze-Fei Wong

Summary

The coevolution theory of the genetic code, which postulates that prebiotic synthesis was an inadequate source of all twenty protein amino acids, and therefore some of them had to be derived from the coevolving pathways of amino acid biosynthesis, has been assessed in the light of the discoveries of the past three decades. Its four fundamental tenets regarding the essentiality of amino acid biosynthesis, role of pretran synthesis, biosynthetic imprint on codon allocations and mutability of the encoded amino acids are proven by the new knowledge. Of the factors that guided the evolutionary selection of the universal code, the relative contributions of *Amino Acid Biosynthesis: Error Minimization: Stereochemical Interaction* are estimated to first approximation as 40,000,000:400:1, which suggests that amino acid biosynthesis represents the dominant factor shaping the code. The utility of the coevolution theory is demonstrated by its opening up experimental expansions of the code and providing a basis for locating the root of life. *BioEssays* 27:416–425, 2005. © 2005 Wiley Periodicals, Inc.

Introduction

The basic universality of the genetic code suggests that the code was established in its present form prior to the earliest branchings of extant organisms, at or before the emergence of LUCA, the Last Universal Common Ancestor. Shaped by pre-LUCA events, the code is a unique record of those events in the first eons of life's history on Earth. The frozen accident theory proposes that the structure of the code is accidental rather than rationally analyzable.[1] It would represent by default the preferred description of code origin and structure if no adequate rational explanation can be found. At least four broad classes of rational explanations, however, have been suggested.

Applied Genomics Laboratory and Department of Biochemistry, Hong Kong University of Science & Technology, Hong Kong, China.
E-mail: bcjtw@ust.hk
DOI 10.1002/bies.20208
Published online in Wiley InterScience (www.interscience.wiley.com).

Abbreviations: aaRS, aminoacyl-tRNA synthetase; CET, coevolution theory of the genetic code; LUCA, last universal common ancestor; Pyl, pyrrolysine; Sec, selenocysteine.

Error minimization. Placement of the codons for physically similar amino acids at neighboring positions (one base removed from each other) in the code would minimize the damage due to mutations, and errors of transcription and translation. The genetic code evolved to maximize this advantage.[2–14]

Stereochemical interaction. Stereochemical interactions between an encoded amino acid and its anticodons or codons determined their specific associations in the genetic code.[15–22]

Amino acid biosynthesis. The formation of amino acids by biosynthetic pathways guided the development of the genetic code.[23–42]

Expanding codons. Initially not all the triplet codons were utilized, and the number of effective codons increased from a smaller number to the present day 64.[43–48]

These four classes are not mutually exclusive, and each of them includes a range of formulations to explain the origin and structure of the genetic code. Together they furnish a multiplicity of elements useful to a reconstruction of the events that gave birth to the modern code. Among the *Amino Acid Biosynthesis* formulations, the important suggestion was made that codon allocations to the amino acids could have been guided by potential conversions between the amino acids, although the postulated conversions were based on structural relatedness of the amino acids and contradicted many of the actual metabolic conversions occurring in organisms.[15,24] The observation that Arg could have replaced its biosynthetic precursor ornithine if it competed successfully against ornithine for attachment to the latter's tRNA, however, furnished an unambiguous, albeit isolated, example of plausible biosynthetic impact on the code.[23] A systematic formulation of genetic code evolution based on amino acid biosynthesis was given by the coevolution theory (CET), which postulates that because prebiotic amino acid synthesis was an inadequate supplier of all 20 present-day protein amino acids, novel amino acid biosynthetic pathways had to supply some of them, thereby propelling a coevolution of the code and leaving a biosynthetic imprint on the code.[25–28] Since the proposal of CET thirty years ago, a great deal has been learned regarding genetic coding, and the sequencing of many genomes has transformed much of biology into an open book. It becomes important to assess the fundamental tenets and wider aspects of CET in the light of this new knowledge.

Tenet 1: Essentiality of amino acid biosynthesis

Since genetic systems cannot multiply if one or more breaks occur in their genes within one replication cycle, they need to replicate faster than the rate of chemical scission of their genes.[49] Catalysis was thus essential at all stages of the development of life. For enzymes, efficient catalysis requires a versatile ensemble of amino acids. The production of amino acids by exposing mixtures of primitive gases to electric discharge and other energy inputs, and the presence of amino acids on carbonaceous meteorites both point to the feasibility of prebiotic synthesis, but gaseous synthesis has failed to yield all 20 protein amino acids.[28,49,50] CET suggests that there are three phases of amino acid entry into proteins. Phase 1 amino acids came from prebiotic synthesis, and phase 2 ones from biosynthesis. Phase 3 amino acids are introduced into proteins via post-translational modifications without direct genetic encoding.[27]

The primitive atmosphere on Earth might only be moderately reducing, under which it would be difficult to synthesize amino acids by electric discharge. However, irradiation with high-energy protons of a mildly reducing mixture of carbon monoxide and nitrogen over water yielded a range of amino acids,[51,52] which included all of the phase 1 but none of the phase 2 amino acids. The perfect agreement between the results of irradiated synthesis and the phases 1 and 2 classification suggested by CET (Table 1) validates both the prebiotic synthesis model employed, and the phase 1–phase 2 division of CET. Also, a survey of forty different criteria relating to the possible order of appearance of different amino acids in proteins[53] resulted in a partition between 'old' and 'new' amino acids that coincides with the phase 1–phase 2 division, with the single exception that Ile, a phase 1 amino acid of CET, is considered a 'new' amino acid.

Phase 2 amino acids are not only difficult to synthesize prebiotically, but also easily degraded. Gln and Asn are thermally unstable,[50] and Cys, Met, Trp, His, Tyr and Phe are quickly destroyed by ultraviolet irradiation.[54] Even under the extremely optimistic scenario of every UV photon of <260 nm reaching primitive Earth being available for prebiotic synthesis, such that the primitive oceans accumulated to 20 M in amino acids in 1 Gyr, the steady-state concentrations of Gln and Asn would not exceed 3.7×10^{-12} M and 2.4×10^{-8} M respectively.[50] These two amides were clearly unavailable from the prebiotic environment, and therefore had to be biosynthesized.

Tenet 2. Role of pretran synthesis

To compete and survive, organisms have come to excel as metabolic inventors, bringing about the syntheses of a host of pigments, alkaloids, antibiotics and other secondary metabolites. This has been referred to as *inventive biosynthesis*,[27] which represents an important source of new metabolites including some phase 2 amino acids. CET postulates that an important source is also to be found in pretranslational modification, or *pretran synthesis*, which is a subset of inventive biosynthesis where the synthesis transforms the amino acid moiety in an aminoacyl-tRNA compound into a novel amino acid or metabolite, e.g. in the transformation of Met-tRNA into formyl-Met-tRNA. When a novel amino acid is produced by ordinary inventive biosynthesis, it has to find a tRNA to attach to in order to receive genetic encoding. When it is produced by pretran synthesis, there is no need to do so, for it is produced pre-attached to a precursor's tRNA.

Thirty years ago, pretran synthesis was something of a curiosity known to be employed only for fMet incorporation, and Gln incorporation in Gram-positive bacteria. Today it is known to be widespread in occurrence, accounting for the incorporation of Gln and Asn, and of selenocysteine (Sec),[55] into the proteins of many organisms (Table 2). Its participation in pyrrolysine (Pyl) incorporation is also not excluded.[56,57] The derivation of Asn synthetase and AsnRS from the AspRS gene suggests that pretran synthesis of Asn-tRNA preceded the direct incorporation of free Asn into proteins via AsnRS.[58,59] That pretran syntheses employing aminoacyl-tRNA substrates are also utilised in porphyrin synthesis, cell wall peptide synthesis, protein N-modifications and the ubiquitin pathway[30,60] attests further to the evolutionary importance of this subclass of inventive biosynthesis.

Table 1. Comparison of phases 1 and 2 amino acids and amino acids produced by high energy proton irradiation of a carbon monoxide–nitrogen–water mixture

	Gly	Ala	Ser	Asp	Glu	Val	Leu	Ile	Pro	Thr	Phe	Tyr	Arg	His	Trp	Asn	Gln	Lys	Cys	Met
Phase of entry[a]	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
Irradiated synthesis[b]	+	+	+	+	+	+	+	+	+	+	0	0	0	0	0	0	0	0	n	n

[a]Phase 1–phase 2 entries into the code are as described (27), where Pro and Thr are regarded as marginal phase 1, and Phe, Tyr and Cys marginal phase 2 amino acids.

[b]Observed irradiated synthesis (51,52) is indicated by +, and lack of synthesis by 0. The comparison is not applicable to Cys and Met (n) because there was no sulfur in the irradiated synthesis.

Table 2. Encoding of amino acids by pretran synthesis[28,34,49,108]

Pretran synthesis	Nature of evidence	Organisms
Met-tRNA → fMet tRNA	Lack of fMetRS Presence of formylase	Bacteria
Glu-tRNA → Gln-tRNA	Lack of GlnRS Presence of amidotransferase	Archaea Bacteria
Asp-tRNA → Asn-tRNA	Lack of AsnRS Presence of amidotransferase	Archaea Bacteria
Ser-tRNA → Sec-tRNA	Lack of SecRS Elucidation of pretran pathway	All three domains

Tenet 3. Biosynthetic imprint on codon allocations

CET is supported by or consistent with a range of studies,[30–42,44,45,59,61,62] but there are also two reappraisals of CET regarding its supposition of a significant biosynthetic imprint on the code.[63,64] The first reappraisal[63] yielded a random probability of 0.001 for codon-biosynthesis correlations within the code, which was weaker evidence for an imprint than the CET estimate of 0.0002[25] but still highly significant. This was further weakened to a still significant 0.036 by removing two of the precursor–product pairs from assessment, and ultimately to an insignificant 0.34 when the biosynthetic relationships between the amino acids suggested by CET are subjected to alterations that are difficult to justify.[65,66] The conclusion drawn by the study that the codon-biosynthesis correlations are only statistical in nature is no more than a truism.

The second reappraisal[64] likewise altered the biosynthetic relationships described by CET, and obtained a still highly significant random probability of 0.0062. Reducing the number of sense codons in the code from 61 to 45 decreased the number of possible alternative codes drastically, which together with changes in the method for probability assessment abolished all significant correlation. The validity of some of the basic postulates of this study is found to be unsupportable.[67] Moreover, cellular metabolism commonly employs opposing sets of different enzymes to achieve energy-dependent interconversions. For instance, fructose-6-phosphate is converted to fructose-1,6-bisphosphate by phosphofructokinase, but the reversal goes through a different enzyme, in this case fructose-1,6-bisphosphatase. Similarly, pyruvate is formed from phosphoenolpyruvate by pyruvate kinase, but the reversal is brought about by pyruvate carboxylase and phosphoenolpyruvate carboxykinase. Gln synthetase and glutaminase, and Asn synthetase and asparaginase are other examples of such opposing enzymes. Yet the study stipulated that the conversion of homoserine to Thr and its potential reversal must be catalysed by the same enzyme. Imposition of such a same-enzyme stipulation would block many metabolic conversions. Gluconeognesis would be gravely impeded if it had to reverse glycolysis exactly and produce fructose-6-phosphate from fructose-1,6-bisphosphate through phosphofructokinase accompanied by production of ATP. Gln hydrolysis would also become insignificant if it had to go through a reversal of Gln synthetase accompanied by production of ATP. The suggestion made by the study that pretran synthesis might not be a remnant of early biosynthetic expansion is also refuted by the species distributions of GlnRS and AsnRS, which establish the absence of GlnRS and AsnRS and dependence of Gln and Asn incorporation into proteins on pretran synthesis at the root of life.[68–70] It is noteworthy, however, that this group has now come around to the realization that their "claims for an adaptive canonical code might be spurious" on account of "potentially flawed" evidence, and brought in amino acid biosynthesis to help explain the code.[71]

In the genetic code for 20 + 2 amino acids (Fig. 1), all the amino acids synthesized from Asp are spread out across the left-center of the ANN row. Those synthesized from Ser or Glu also display same row or column neighbourliness, and the same applies to the Phe–Tyr and Val–Leu precursor–product pairs. Thus the imprints of biosynthesis on codon allocations are recognizable by either inspection or using statistical analysis.

Fig. 1 demonstrates as well the enrichment of sibling amino acids in the shared codon boxes. Among the 231 pairings of the 20 + 2 amino acids, based on the precursor–product relationships postulated by CET[25] plus that of Ser–Sec, only 10 of them or 4.3% are sibling pairs: Asn–Lys, Asn–Thr and Thr–Lys as siblings from Asp, Ile–Met as siblings from Thr (or its precursor homoserine), Gln–Arg, Gln–Pro and Arg–Pro as siblings from Glu, and Cys–Trp, Cys–Sec and Trp–Sec as siblings from Ser. Yet out of the 10 amino acids pairs that share a four-codon box, 5 of them or 50%, viz. Ile–Met, Asn–Lys, Cys–Trp, Cys–Sec and Trp–Sec are sibling pairs. Therefore the enrichment is more than ten times. In this regard, the sharing of the same codon box by Cys and Trp, the most physically dissimilar pair of encoded amino acids, is particularly striking: it is readily explained by their being siblings, but defies any explanation based on *Error Minimization*.

How was the biosynthetic imprint on the code brought about? In pretran synthesis, a product acquired an isoacceptor tRNA of its precursor and in so doing gained neighborliness with the remaining precursor codons. This is the case with Gln acquiring codons from Glu, Asn from Asp, Sec from Ser, and likely the case with Cys and Trp acquiring codons from Ser. Homoserine/Thr could have given rise to Met through pretran synthesis via cystathionine-tRNA and homocysteine-tRNA as intermediates. Homoserine, cystathionine and homocysteine might find utility as interim encoded amino acids prior to the

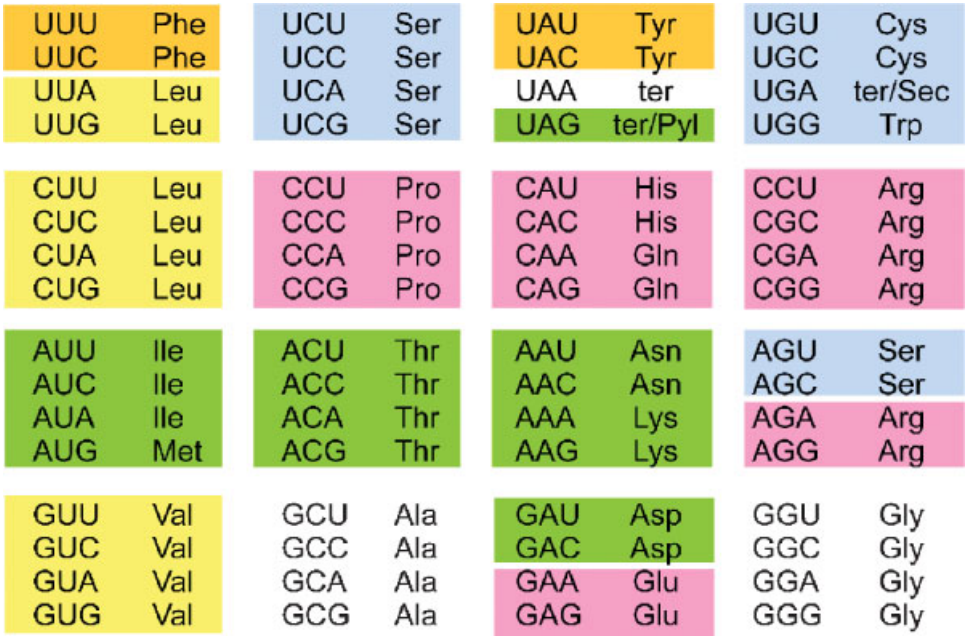

Figure 1. Genetic code showing amino acids belonging to the biosynthetic families of Asp (green), Glu (red), Phe (orange), Ser (blue) and Val (yellow). The partial assignment of in-frame UAG to Pyl is known so far only in species of *Methanosarcina, Desulfitobacterium* and *Methanococcoides*.[56]

introduction of Thr and Met into the code. Codon box sharing by Gln–His might be favored by the contribution of an N-atom by Gln-tRNA to form the imidazole ring, with ring closure catalysed by a primitive imidazole glycerol phosphate synthase. Pro-tRNA might form from Glu-tRNA by pretran synthesis via a Glu-5-semialdehyde-tRNA. The conversion of Glu to Arg could also be facilitated by pretran synthesis, especially in the transformation of ornithine to citrulline, and citrulline to Arg. Both ornithine and citrulline might be interim encoded amino acids.[23]

Where a product resembles its precursor physically, its competitive attachment to one or more isoacceptor tRNAs of the precursor could help it gain codons, e.g. Leu once formed might compete against Val. Since Tyr might be pretran-synthesized from Phe, and Tyr also resembles Phe physically, either pretran synthesis or ordinary inventive biosynthesis might suffice to ensure neighborliness of Phe and Tyr codons. For biosyntheses dependent on a thermolabile intermediate, such as acetyl-Glu-semialdehyde in Arg synthesis from Glu, α-ketobutyrate in Ile synthesis from Thr, or Asp-semialdehyde in Lys, homoserine, Thr and Met synthesis from Asp, metabolic channeling might favor attachment of a product amino acid to the tRNAs of its precursor.[35,37] Since metabolic channeling is known to be important for the preservation of labile intermediates in a hyperthermophilic environment,[72,73] its role in pre-LUCA evolution might be accentuated by the probably hyperthermophilic environment.[70]

The pretran syntheses of Gln, Asn and Sec are well understood. The pretran synthesis of Glu-semialdehyde from Glu-tRNA is still employed in *Methanopyrus kandleri* (Mka) close to LUCA, catalysed by Glu-tRNA reductase.[72] That Cys and Trp share UGN codons because they were both derived from Ser is supported by homology between biosynthetic enzymes for Cys and Trp.[74] In addition, it is known that tRNA-dependent pretran syntheses could come to be replaced in time by evolved tRNA-independent reactions, e.g. the synthesis of Gln by Gln synthetase, and Asn by Asn synthetase. The pretran synthesis of Glu-1-semialdehyde as precursor to 5-aminolevulinic acid and tetrapyrroles is also replaced in animals, fungi and some bacteria by 5-aminolevulinic acid synthase in a one-step condensation of succinyl-CoA and glycine.[72] Therefore some original pretran syntheses of amino acids could have been replaced back in pre-LUCA times, and vanished from all extant organisms.

Further research will be needed to decipher the yet uncertain biosynthetic pathways responsible for introducing

a number of novel amino acids into the pre-LUCA code. A potential contribution by chance occurrence to any observable codon-biosynthesis correlation in the code also cannot be ruled out.

Tenet 4. Mutability of encoded amino acids

A fundamental challenge confronting CET at the time of its proposal was that CET predicts and requires the code to be mutable regarding its ensemble of encoded amino acids, for which there has been no evidence throughout the course of biological evolution. While amino acid analogues may be incorporated into proteins replacing an amino acid over the short term, even successfully adding to growth yield,[75] no analogue could support indefinite cell growth. The only way to test the mutability of the code in this regard was to try to mutate the present-day code. Accordingly experiments were carried out that showed that *Bacillus subtilis* could be mutated to replace its Trp by 4-fluoroTrp, and even further to displace Trp entirely as a competent amino acid capable of supporting indefinite cell growth.[76,77] In the replacement mutants, either Trp or 4-fluoroTrp can support the indefinite growth of a Trp-auxotroph. In the displacement mutants, 4-fluorTrp can support indefinite growth, but Trp itself has lost this capability and turned into an inhibitory analogue (Fig. 2). Genetic adaptation to growth on 4-fluoroTrp has been extended to *E. coli* and phage, with identification in the latter instance of some of the mutations underlying the adaptation.[78–80] Unlike mutations of codon assignments,[81] which leave the nature of the proteome unaffected, such mutations relating to the encoded amino acids alter the building blocks and therefore the fundamental nature of the proteome.

Proof of code mutability relating to the encoded amino acids has turned the code into an open target for directed evolution employing both a 'top down' approach through mutations of the organism, or a 'bottom up' approach through orthogonal tRNA-aaRS pairs that make use of suppressor tRNAs and the lack of cross reactions between the tRNAs and aaRS from organisms belonging to different biological domains.[82–85] Genetic code expansion has become a fast growing field, making it possible to explore completely new proteins that were once inconceivable,[86] create new dimensions of protein engineering, and verify CET, ''The broader finding that the amino acids of the genetic code are mutable, as predicted by Wong, has wider implications for our understandings of the evolution of this aspect of protein synthesis''.[87] These developments have raised the question of whether extant life employing the 20 standard amino acids represents merely the first half of life's history, with a sequel of new life forms built on expanded amino acid repertoires ready to begin.[88,89] So after a 2–3 Gyr interlude of quiescence, the genetic code resumes its increment of encoded amino acids, now with humans instead of nature guiding its explorations.

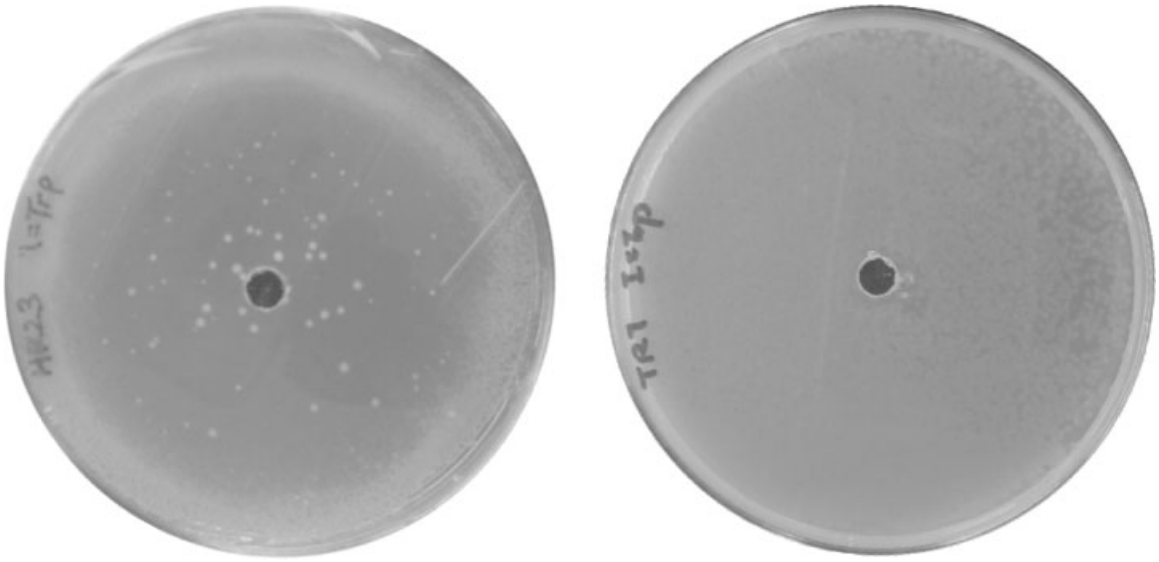

Figure 2. Trp inhibition of growth of *Bacillus subtilis* strain HR23, a faster growing variant of HR16, which grows on 4-fuoroTrp but not on Trp.[76] **Left:** Inhibition zone in HR23 growth layer caused by Trp placed in center well, showing within the zone revertant colonies resistant to Trp inhibition. **Right:** Loss of Trp inhibition against revertant TR7 of HR23 that had regained the ability to grow on Trp (FW Mat and JT Wong, unpublished observations).

Code universality

The transformation of the code from early active expansion to eons of quiescence signals a process of exceptional dynamics. CET estimates that there are $N = 2 \times 10^{19}$ alternative codes if code evolution were to begin with either one or twenty founder amino acids. If it began with 5–10 founder amino acids, and went through codon domain subdivisions to reach the final twenty, codon allocations could be permutated only within each of the separate founder domains, and N is reduced by a factor of 10^{-11} down to 2×10^{8}.[26] Enormous N reduction is needed for the emergence of a universal code. There are only 6×10^{16} seconds in 2 Gyr, so even if LUCA armed with its universal code eliminated one competitor organism bearing an alternative code every second, there was insufficient time in 2 Gyr to eliminate 2×10^{19} alternative codes.

Starting from an RNA or RNA-like World, the advantage of enhanced catalysis might propel the covalent attachment of aminoacyl and peptidyl cofactors to the ribozymes,[49,90] or the use of free amino acids and peptides as cofactors.[91] Aminoacyl- or peptidyl-oligonulceotides also might be joined together to form aminoacylated or peptidylated ribozymes.[22] As the peptide moieties on the ribozymes lengthened and surpassed the RNA backbone in catalytic performance, especially with regard to the k_{cat} parameter, encoding of peptide sequences rather than catalysis became the primary function of the RNA. Observations on tRNA aminoacylation by the aaRS-like YadB at a site outside the 3′ terminal hydroxy groups[92] suggest how the machineries for attaching amino acids and peptides to ribozymes might specialize to become aaRS, attaching amino acids exclusively to the 3′-ends of tRNA. The genomic tag hypothesis[93] also points to possible predecessor roles for tRNA prior to its utilization in protein synthesis.

The phase 1 code was developed prior to the biosynthetic addition of phase 2 amino acids. For its development, *Stereochemical Interaction*[15–22] between amino acids and their codons or anticodons could play a significant role in establishing codon assignments; tRNAs bearing complementary anticodons might also interact with each other to help introduce new species of tRNA into the coding system.[48] *Expanding Codons* also might add effective codons in stages to bring about a gradual development of the code.[43–48]. GNN codons could be favored initially,[94] while the full use of all 61 sense codons might have to await appropriate nucleoside modifications in the tRNA anticodon-loop to equalize the hydrogen-bonding strengths of different codon–anticodon pairs. The utility of *Amino Acid Biosynthesis* and CET would come into prominence in the development of the phase 2 code, while *Error Minimization* might assist code development in both phases 1 and 2.

In addition, advances in aaRS research have shown that the majority of aaRS recognize the anticodon on their cognate tRNAs as an identity element, with only SerRS and AlaRS not yet observed to employ any anticodon base for this purpose.[95] *B. subtilis* TrpRS, for instance, uses the anticodon and the discriminator base as major identity elements.[96,97] The row- and column-centered codon domains of the Asp, Glu, Phe, Ser and Val biosynthetic families (Fig. 1) suggest that the partition of the phase 1 code between these families could be facilitated by the *Anticodon Identity Elements* of primitive aaRS, e.g. a primitive AspRS preferring NNU anticodons could help to secure ANN codons for the Asp family, and a primitive ValRS preferring NAN anticodons secure NUN codons for the Val family. As importantly, *Anticodon Identity Elements* also provides an alternative to direct *Stereochemical Interaction*[4,7,16–20] for partnering amino acids and anticodons/codons of specified stereochemical properties. The reason is that use of the anticodon as an identity element requires not only a binding site on the aaRS for the amino acid substrate, but also a binding site for the anticodon. Consequently, the stereochemical properties of these two binding sites on the aaRS could determine directly what kinds of amino acids came to be encoded by what kinds of anticodons/codons, in terms of their respective stereochemical properties.

Early code expansion required the recruitment of new aaRS to activate some of the novel amino acids, including the adaptation of a preexistent aaRS belonging to an older amino acid. This may account for the similarity of TyrRS–TrpRS, and ValRS–LeuRS–IleRS sequences.[98] Recruitment of an aaRS for a new amino acid also might be facilitated by the docking of Class I and Class II aaRS to opposite sides of the tRNA, allowing them to recognize dissimilar identity elements on similar tRNA sequences accompanied by limited cross-aminoacylations. The usefulness of such double docking would also stabilize the existence of two universal classes of aaRS, which is otherwise difficult to explain. The pairing of ancestral aaRS suggests a reduced number of amino acids and tRNA identities in the early code,[99] in accord with the postulation by Tenet 1 of an evolutionary increment of encoded amino acids.

The finding that *Error Minimization* brought about a significant but partial optimization of the code[100] has been confirmed extensively.[11,13,14,31–34] It has been estimated, albeit accompanied by the unrealistic assumption that all 20 protein amino acids including Gln and Asn were available right from the start, that *Error Minimization* could yield a one-in-a-million, viz. 10^{-6}, selection against alternative codes.[9] The contribution by *Stereochemical Interaction* has been estimated from aptamer–amino acid binding to provide up to 0.04%, or 4×10^{-4} selection.[21] The concept of *Stereochemical Interaction* between amino acids and codons or anticodons helping to guide early code formation is an attractive one,[4,7,15–22] but the use of aptamer–amino acid interactions for such estimation might not be supportable.[101]

Based on the given estimates for their respective N-reducing powers of 10^{-11}, 10^{-6} and 4×10^{-4}, the relative

contributions made by *Amino Acid Biosynthesis: Error Minimization: Stereochemical Interaction* toward the selection of a universal code are 40,000,000:400:1. These estimates refer to evolutionary contributions made three billion years ago, and are by necessity a first approximation. Nonetheless, they suggest that *Amino Acid Biosynthesis* was the dominant factor shaping the universal code. *Error Minimization* and *Stereochemical Interaction* would play tiny but useful roles. Working together with the factors of *Identity Elements* and *Expanding Codons*, they might furnish the $1/(2 \times 10^8)$ selection needed to complement codon domain subdivision to accomplish the remarkable selection of a one-in-twenty-billion-billion code.

The partial assignments of some codons to Sec and Pyl underline the question of why only 20 amino acids receive full codon assignments. According to CET,[26,49] the early code expanded because the addition of novel amino acids enhanced the versatility of proteins, thereby paving the way to catalytic rate perfection in some enzymes in the form of diffusion-controlled kinetics,[102] and reduction of the combined transcriptional and translational errors to the <0.0003 range. Against this background of diminishing errors, the noise generated by the insertion of yet another amino acid with full codon assignment into multiple sites across the proteome eventually would represent too great a selective disadvantage, far outweighing the advantage offered by a new sidechain to the amino acid repertoire. Code expansion to recruit additional amino acids given full codon assignments had to come to a halt.[26] Post-translational modifications took over from code expansion as the operating mechanism to increase the amino acid repertoire, and started to bring the many phase 3 amino acids such as hydroxy-Pro, adenylyl-Tyr, His-flavin, Cys-heme, pyro-Glu, glycosyl-Asn, desmosine into proteins. Because the phase 3 amino acids are only placed into a limited number of residues in the proteome, the benefit/noise ratio arising from the placement becomes much more favorable in the context of low background noise than the across the proteome placements caused by the full assignment of codons to a novel phase 2 constituent.

Root of Life

Biologists have long searched for the root of life. DNA and rRNA cannot be used for rooting because they lack paralogs. The fast evolution and horizontal gene transfers of proteins have also rendered difficult rooting based on protein sequences.[103] CET suggests that tRNAs might be enriched in paralogs, generated when a phase 2 amino acid acquired one of its precursor's isoacceptor tRNAs through pretran synthesis. Analysis of the tRNA genes of 60 Bacteria, Archaea and Eukarya genomes indicates that some of the alloacceptor tRNAs (accepting different amino acids) in the hyperthermophilic Archaea possess highly similar sequences. This suggests that early code evolution began with a group of closely

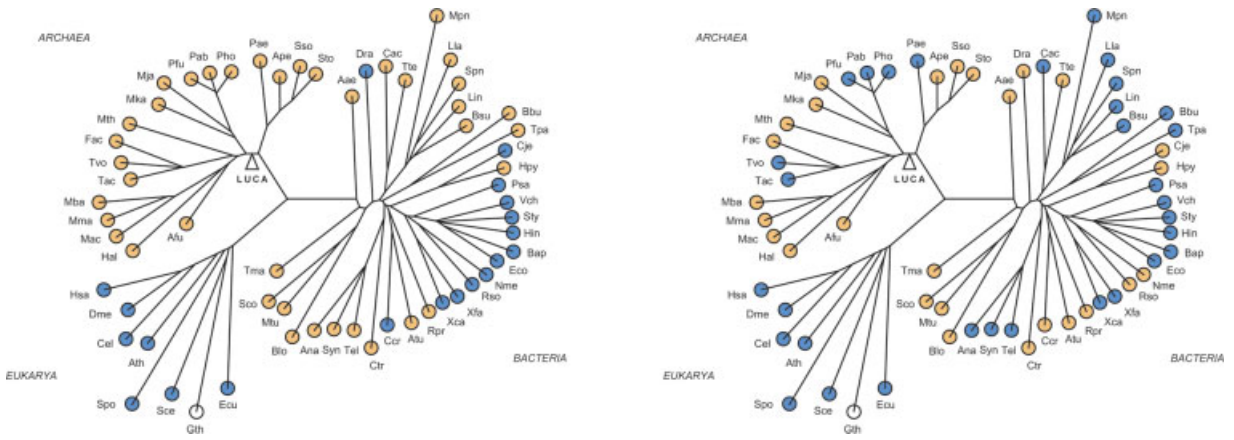

Figure 3. Universal tRNA tree showing distribution of pretran synthesis of Gln-tRNA and Asn-tRNA. Occurrence of pretran synthesis is indicated where an organism contains only a single aaRS gene for Glu and Gln, or for Asp and Asn. **Left:** Genome sequences[109] identified by BLASTP[110] to contain two or more (blue), a single (orange), or no (colorless) gene homologous to *E. coli* GluRS and GlnRS as query sequences. **Right:** Genome sequences identified by BLASTP to contain two or more (blue), a single (orange), or no (colorless) gene homologous to *E. coli* AspRS and AsnRS as query sequences. *Methanopyrus kandleri* (Mka), which is closest to LUCA in terms of tRNA genotypes (70) and therefore the best model for LUCA, contains also the GatA (Mka0238), GatB (Mko960), GatC (Mk0359) and GatE (Mk0129) indicative of the amidotransferase capablity required for pretran synthesis of Gln-tRNA and Asn-tRNA, and the SelD (Mk1369) and SelA (Mk0620) required for pretran synthesis of Sec-Trna.[69] Abbreviations for species as given[70] include *Aeropyrum pernix* (Ape), *Pyrococcus horikoshii* (Pho), *Pyrobaculum aerophilum* (Pae), *Thermoplasma acidophilum* (Tac), *Aquifex aeolicus* (Aae), *E. coli* (Eco), *B. subtilis* (Bsu), *Deinococcus radiodurans* (Dra), *Clostridium acetobutylicum* (Cac), *Anabaena sp.* (Ana), *Guillardia theta* (Gth), yeast (Sce) and humans (Hsa).

related tRNAs that underwent gene duplications, coevolution with amino acid biosynthesis, and dispersion in sequence space through evolutionary change. The extent of tRNA sequence clustering within a genome measured by the alloacceptor distance D_{allo} thus provides a useful index of its antiquity, which places LUCA on the universal tRNA tree closest to the hyperthermophilic archaeon *Methanopyrus kandleri*.[70] The uniform anticodon usages of the Archaea, and their departures from the Crick wobble rules[104] also support a *Methanopyrus*-like LUCA.

Fig. 3 shows the position of LUCA on the universal tRNA tree, and the distribution of species with separate GluRS and GlnRS, or separate AspRS and AsnRS. The suggestion that GlnRS entered the Bacteria from the Eukarya through horizontal gene transfer at the level of Dra[105] is consistent with the fairly deep branching position of Dra in the Bacteria domain.

Finding LUCA is important to CET. First, the enrichment of tRNA pairs with $D_{allo} < 0.2$ among biosynthetically related amino acid pairs[70] supports CET, and confirms earlier studies in this regard.[106,107] Secondly, that Mka, closest to LUCA, lacks GlnRS and AsnRS and accordingly must rely on pretran synthesis to incorporate Gln and Asn into proteins indicates that Gln-tRNA and Asn-tRNA were derived from pretran synthesis rather than prebiotic synthesis during genetic code evolution, as postulated by Tenets 1 and 2.[28,34,49,108] Thirdly, the finding of a single LUCA root readily explains the universality of the code on the basis that this code was employed by LUCA. In contrast, universality would be difficult to explain if there are multiple roots in the tree of life.

Conclusion

There have been vigorous advances during the past three decades relating to fundamental aspects of genetic coding. Although a range of important questions, e.g. how tRNA-mediated translation evolved from an RNA-like world, or why the primitive tRNAs consisted of closely clustered sequences, remain unanswered, the new knowledge has confirmed the basic tenets of CET. CET has also demonstrated its utility in reopening genetic code expansion after prolonged quiescence, and providing an approach for locating the root of life based on tRNA genes.

The dependence of *Methanopyrus*, and by implication likely also LUCA, on pretran synthesis to produce and encode Gln, Asn and Sec proves the validity of Tenet 1 regarding the essentiality of biosynthesis in supplying some of the amino acids to the pre-LUCA code, Tenet 2 regarding the usefulness of pretran synthesis for codon acquisition by some novel amino acids, and Tenet 3 regarding biosynthesis leaving a detectible imprint on the code, for the allocations of CAR to Gln, AAY to Asn and UGA partially to Sec are all direct consequences of the pretran synthesis of Gln-tRNA, Asn-tRNA and Sec-tRNA. These three tenets will stand proven unless there emerges compelling evidence to relocate LUCA on the tree of life away from the proximity of *Methanopyrus* to the midst of organisms that possess both GlnRS and AsnRS and do not catalyse pretran synthesis of Sec. For Tenet 4, code mutability regarding the encoded amino acids is proven by the isolation of the 4-fluoroTrp mutants of *Bacillus subtilis*, and the ongoing experimental expansion of encoded amino acids that is opening up exciting vistas for the future.

In conclusion, the coevolution theory is more strongly supported to-day than it could have been imagined thirty years ago. It unveils a magnificent convergence of prebiotic synthesis, inventive biosynthesis, pretran synthesis, tRNA gene duplications, codon domain subdivisions, anticodon identity elements, and physicochemical factors to construct the genetic code, and establish its universality through a *Methanopyrus*-like LUCA.

Acknowledgments

The valuable discussion of Dr. Hong Xue, Dr. Ka-Lok Tong, Flora Mat and Siu-Kin Ng, and support by the Research Grants Council of Hong Kong are gratefully acknowledged.

References

1. Crick FHC. 1968. The origin of the genetic code. J Mol Biol 38:367–379.
2. Sonneborn TM. 1965. Degeneracy of the genetic code, extent, nature and genetic implications. In: Bryson V, Vogel HJ. (ed) Evolving Genes and Proteins. New York: Academic Press. p 379–397.
3. Woese CR, Dugre DH, Dugre SA, Kondo M, Saxinger WC. 1966. On the fundamental nature and evolution of the genetic code. Cold Spring Harbour Symp Quant Biol 31:723–736.
4. Jungck JR. 1978. The genetic code as a periodic table. J Mol Evol 11: 211–224.
5. Figureau A. 1989. Optimization and the genetic code. Orig Life Evol Biosph 19:57–67.
6. Luo LF. 1989. The distribution of amino acids in the genetic code. Orig Life Evol Biosph 19:621–631.
7. Lacey JC Jr, Wickramasinghe NSMD, Cook GW. 1992. Experimental studies on the origin of the genetic code and the process of protein synthesis: a review update. Orig Life Evol Biosp 22:243–275.
8. Goldman N. 1993. Further results on error minimization in the genetic code. J Mol Evol 37:662–664.
9. Freeland SJ, Hurst LD. 1998. The genetic code is one in a million. J Mol Evol 47:238–248.
10. Freeland SJ, Wu T, Keulman N. 2003. The case for an error minimizing standard genetic code. Orig Life Evol Biosph 33:457–477.
11. Judson OP, Haydon D. 1999. The genetic code: what is it good for? An analysis of the effects of selection pressure on genetic codes. J Mol Evol 49:539–550.
12. Chiusano ML, Alvarez-Valin F, Di Giulio M, D'Onofrio G, Ammirato G, et al. 2000. Second codon positions of genes and the secondary structures of proteins. Relationships and implications for the origin of the genetic code. Gene 261:63–69.
13. Ardell DH, Sella G. 2002. No accident: genetic codes freeze in error-correcting patterns of the standard genetic code. Phil Trans R Soc London B 357:1625–1642.
14. Zhu CT, Zeng XB, Huang WD. 2003. Codon usage decreases the error minimization within the genetic code. J Mol Evol 57:533–537.
15. Pelc SR. 1965. Correlation between coding triplets and amino acids. Nature 207:597–599.
16. Dunnill P. 1966. Triplet nucleotide-amino acid pairing, a stereochemical basis for the division between protein and non-protein amino acids. Nature 215:355–359.
17. Melcher G. 1974. Stereospecificity of the genetic code. J Mol Evol 3: 121–140.

18. Hendry LB, Bransome ED Jr, Hutson MS, Campbell LK. 1981. First approximation of a stereochemical rationale for the genetic code based on the topography and physicochemical properties of ''cavities'' constructed from models of DNA. Proc Natl Acad Sci USA 78:7440–7444.
19. Shimizu M. 1982. Molecular basis for the genetic code. J Mol Evol 18: 297–303.
20. Yarus M. 1988. A specific amino acid binding site composed of RNA. Science 240:1751–1758.
21. Knight R, Landweber XX, Yarus M. 2003. Tests of a stereochemical genetic code. In Lapointe J, Brakier-Gingras L. (ed); Translation Mechanisms. Georgetown, Texas: Landes Bioscience. p 115–129.
22. Seligmann H, Amzallag GN. 2002. Chemical interactions between amino acid and RNA: multiplicity of the levels of specificity explains origin of the genetic code. Naturwissenschaften 89:542–551.
23. Jukes TH. 1973. Arginine as an evolutionary intruder into protein synthesis. Biochem Biophys Res Com 53:709–714.
24. Dillon LS. 1973. The origins of the genetic code. Bot Rev 39:301–345.
25. Wong JT. 1975. A coevolution theory of the genetic code. Proc Natl Acad Sci USA 72:1909–1912.
26. Wong JT. 1976. The evolution of a universal genetic code. Proc Natl Acad Sci USA 73:2336–2340.
27. Wong JT. 1981. Coevolution of the genetic code and amino acid biosynthesis. Trends Biochem Sci 6:33–35.
28. Wong JT. 1988. Evolution of the genetic code. Microbiol Sci 5:174–181.
29. McClendon JH. 1986. The relationship between the biosynthetic paths to the amino acids and their coding. I. the aliphatic amino acids and proline. Orig Life 17:401–417.
30. Danchin A. 1989. Homeotropic transformation and the origin of translation. Prog Biophys Mol Biol 54:81–86.
31. Di Giulio M. 1989. Some aspects of the organization and evolution of the genetic code. J Mol Evol 29:191–201.
32. Di Giulio M. 1996. The β-sheets of proteins, the biosynthetic relations between amino acids, and the origin of the genetic code. Origins Life Evol Biosph 26:589–609.
33. Di Giulio M, Medugno M. 1998. The historical factor: the biosynthetic relationships between amino acids and their physicochemical properties in the origin of the genetic code. J Mol Evol 46:615–621.
34. Di Giulio M. 2004. The coevolution theory of the origin of the genetic code. Phyics Life Rev 1:128–137.
35. Wachtershauser G. 1988. Before enzymes and templates: theory of surface metabolism. Microbiol Rev 52:452–484.
36. Taylor FJR, Coates D. 1989. The code within the codons. BioSystems 22:177–187.
37. Edwards MR. 1996. Metabolite channeling in the origin of life. J Theor Biol 179:313–322.
38. Budisa N, Moroder L, Huber R. 1999. Structure and evolution of the genetic code viewed from the perspective of the experimentally expanded amino acid repertoire in vivo. CMLS Cell Mol Life Sci 55: 1626–1635.
39. Bermudez CI, Daza EE, Andrade E. 1999. Characterization and comparison of *Escherichia coli* transfer RNAs by graph theory based on secondary structure. J Theoret Biol 197:193–205.
40. Stevenson DS. 2002. Co-evolution of the genetic code and ribozyme replication. J Theor Biol 217:235–253.
41. Cavalcanti AR, Leite ES, Neto BB, Ferreira R. 2004. On the classes of aminoacyl-tRNA synthetases, amino acids and the genetic code. Orig Life Evol Biosph 34:407–420.
42. Klipcan L, Safro MM. 2004. Amino acid biogenesis, evolution of the genetic code and aminoacyl-tRNA synthetases. J Theo Biol 228:389–396.
43. Eigen M, Schuster P. 1978. The hypercycle. A principle of natural self-organization. Naturwissenschaften 65:341–369.
44. Jurka J, Smith TF. 1987. β-turn driven early evolution: the genetic code and biosynthetic pathways. J Mol Evol 25:15–19.
45. Delarue M. 1995. Partition of aminoacyl-tRNA synthetases in two different structural classes dating back to early metabolism: implications for the origin of the genetic code and the nature of protein sequences. J Mol Evol 41:703–711.
46. Davis BK. 1999. Evolution of the genetic code. Prog Biophys Mol Biol 72:157–243.
47. Yang CM. 2004. Molecular versus atomic information logic behind the genetic coding contents constrained by two evolutionary axes and the Fibonacci-Lucas sequence. J Biol Systems 12:21–44.
48. Guimaraes RC, Moreira CHC. 2004. Genetic code: a self-referential and functional model. In Palyi G, Zucchi C, Caglioti L. (ed); Progress in Biological Chirality. Paris: Elsevier. p 83–118.
49. 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. Paris: Elsevier. p 473–494.
50. Wong JT, Bronskill PM. 1979. Inadequacy of prebiotic synthesis as origin of proteinous amino acids. J Mol Evol 13:115–125.
51. Kobayashi K, Tsuchiya M, Oshima T, Yanagawa H. 1990. Abiotic synthesis of amino acids and imidazole by proton irradiation of simulated primitive earth atmospheres. Orig Life Evol Biosph 20:99–109.
52. Kobayashi K, Kaneko T, Saito T, Oshima T. 1998. Amino acid formation in gas mixtures by high energy particle irradiation. Orig Life Evol Biosph 28:155–165.
53. Trifonov EN. 2000. Consensus temporal order of amino acids and evolution of the triplet code. Gene 261:139–151.
54. Wong JT. 1984. Evolution and mutation of the aminoacid code. In: Ricard J, Cornish-Bowden A. (eds). Dynamics of Biochemical Systems. New York: Plenum Press, p 247–257.
55. Commans S, Bock A. 1999. Selenocysteine inserting tRNAs: an overview. FEMS Microb Rev 23:335–351.
56. Krzycki JA. 2004. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. Curr Opinion Chem Biol 8:484–491.
57. Polycarpo C, Ambrogelly A, Berube A, Winbush SM, McCloskey JA, et al. 2004. An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. Proc Natl Acad Sci USA 101:12450–12444.
58. Roy H, Becker HD, Reinbolt J, Kern D. 2003. When contemporary aminoacyl-tRNA synthetases invent their cognate amino acid metabolism. Proc Natl Acad Sci USA 100:9837–9842.
59. Francklyn C. 2003. tRNA synthetase paralogs: evolutionary links in the transition from tRNA-dependent amino acid biosynthesis to de novo synthesis. Proc Natl Acad Sci USA 100:9650–9652.
60. Ibba M, Curnow AW, Soll D. 1997. Aminoacyl-tRNA synthesis: divergent routes to a common goal. Trends Biochem Sci 22:39–42.
61. Di Giulio M, Medugno M. 1999. Physicochemical optimization in the properties in the genetic code origin as the number of codified amino acids increases. J Mol Evol 49:1–10.
62. Di Giulio M, Medugno M. 2001. The level and landscape of optimization in the origin of the genetic code. J Mol Evol 52:372–382.
63. Amirnovin R. 1997. An analysis of the metabolic theory of the origin of the genetic code. J Mol Evol 44:473–476.
64. Ronneberg TA, Landweber LF, Freeland SL. 2000. Testing a biosynthetic theory of the genetioc code: fact or artifact? Proc Natl Acad Sci USA 97:13690–13695.
65. Di Giulio M. 1999. The coevolution theory of the genetic code. J Mol Evol 48:253–255.
66. Di Giulio M, Medugno M. 2000. The robust statistical basis of the coevolution theory of genetic code origin. J Mol Evol 50:258–263.
67. Di Giulio M. 2001. A blind empiricism against the coevolution theory of the origin of the genetic code. J Mol Evol 53:724–732.
68. Siatecka M, Rozek M, Barciszewski J, Mirande M. 1998. Modular evolution of the Glx-tRNA synthetase family. Eur J Biochem 256:80–87.
69. Tumbula D, Vothknecht UC, Kim HS, Ibba M, Min B, et al. 1999. Archaeal aminoacyl-tRNA synthesis: diversity replaces dogma. Genetics 152: 1269–1276.
70. 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.
71. Freeland SJ, Knight RD, Landweber LF, Hurst LD. 2000. Early fixation of an optimal genetic code. Mol Biol Evol 17:511–518.
72. Moser J, Schubert W-D, Beier V, Bringemeier I, Jahn D, et al. 2001. V-shaped structure of glutamyl-tRNA reductase, the first enzyme of tRNA-dependent tetrapyrrole biosynthesis. EMBO J 23: 6583–6590.

73. Manssant J, Verstreken P, Durbecq V, Kholti A, Legrain C, et al. 2002. Metabolic channeling of carbamoyl phosphate, a thermolabile intermediate. J Biol Chem 277:18517–18522.
74. Levy S, Danchin A. 1988. Phylogeny of metabolic pathways: O-acetylserine sulphydrylase A is homologous to the tryptophan synthase beta subunit. Molec Microb 2:777–783.
75. Pezo V, Metzgar D, Hendrickson TL, Waas WF, Hazebrouck S, et al. 2004. Artificially ambiguous genetic code confers growth yield advantage. Proc Natl Acad Sci USA 101:8593–8597.
76. Wong JT. 1983. Membership mutation of the genetic code: loss of fitness by tryptophan. Proc Natl Acad Sci USA 80:6303–6306.
77. Bronskill PM, Wong JT. 1988. Supression of fluorescence of tryptophan residues in proteins by replacement with 4-fluorotryptophan. Biochem J 249:305–308.
78. Bacher JM, Ellington AD. 2001. Selection and characterization of *Escherichia coli* variants capable of growth on an otherwise toxic tryptophan analogue. J Bacteriol 183:5414–5425.
79. Bacher JM, Bull JJ, Ellington AD. 2003. Evolution of phage with a chemically ambiguous proteome. BMC Evol Biol 3:24.
80. Bacher JM, Hughes RA, Wong JT, Ellington AD. 2004. Evolving new genetic codes. Trends Ecol Evol 19:69–75.
81. Barrell BG, Bankier AT, Drouin J. 1979. A different genetic code in human mitochondria. Nature 282:189–194.
82. Bacher JM, Ellington AD. 2003. The directed evolution of organismic chemistry: unnatural amino acid incorporation. In Lapointe J, Brakier-Gingras L ed; Translation Mechanisms. Georgetown Texas: Landes Bioscience. p 80–94.
83. Kowal AK, Kohrer C, RajBhandary UL. 2001. Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. Proc Natl Acad Sci 98:2268–2273.
84. Magliery TJ, Pastrnak M, Anderson C, Santoro SW, Herberich B, et al. 2003. In vitro tools and in vivo engineering: incorporation of unnatural amino acids into proteins. In Lapointe J, Brakier-Gingras L. (ed); Translation Mechanisms. Georgetown Texas: Landes Bioscience. p 95–114.
85. Kwok Y, Wong JT. 1980. Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. Can J Biochem 58:213–218.
86. Bock A. 2001. Invading the genetic code. Science 292:453–454.
87. Ibba M, Stathopoulos C, Soll D. 2001. Protein synthesis: twenty three amino acids and counting. Curr Biol 11:R563–565.
88. Cohen P. 2000. Life the sequel. New Scientist 167:32–36.
89. Hesman T. 2000. Code breakers. Scientists are altering bacteria in a most fundamental way. Science News 157:360–362.
90. Wong JT. 1991. Origin of genetically encoded protein synthesis: a model based on selection for RNA peptidation. Orig Life Evol Biosph 21:165–176.
91. Szathmary E. 1999. The origin of the genetic code: amino acids as cofactors in an RNA world. Trends Genet. 15:223–229.
92. Ibba M, Francklyn C. 2004. Turning tRNA upside down: when aminoacylation is not a prerequisite to protein synthesis. Proc Natl Acad Sci USA 101:7493–7494.
93. Maizels N, Weiner AM. 1999. The genomic tag hypothesis: what molecular fossils tell us about the evolution of tRNA. In Gesteland RF, Cech TR, Atkins JF. (ed); The RNA World. Cold Spinr Harbor, NY: Cold Spring Harbor Laboratory Press, p 79–111.
94. Wong JT. 1981. The origin of the genetic code. Trends Biochem Sci 6:XIII.
95. Giege R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res 22:5017–5035.
96. Xue H, Shen W, Giege R, Wong JT. 1993. Identity elements of $tRNA^{Trp}$. Identification and evolutionary conservation. J Biol Chem 268:9316–9322.
97. Guo Q, Gong Q, Tong KL, Vestergaard B, Costa A, et al. 2002. Recognition by tryptophanyl-tRNA synthetases of discriminator base on $tRNA^{Trp}$ from three biological domains. J Biol Chem 277:14343–14349.
98. Nagel GM, Doolittle RF. 1995. Phylogenetic analysis of the aminoacyl-tRNA synthetases. J Mol Evol 40:487–498.
99. Ribas de Pouplana L, Schimmel P. 2001. Aminoacyl-tRNA synthetases: potential markers of genetic code development. Trends Biochem Sci 26:591–596.
100. Wong JT. 1980. Role of minimization of chemical distances between amino acids in the evolution of the genetic code. Proc Natl Acad Sci USA 77:1083–1086.
101. Ellington AD, Khrapov M, Shaw CA. 2000. The scene of a frozen accident. RNA 6:485–498.
102. Wong JT. 1975. Kinetics of Enzyme Mechanisms. London: Academic Press. 201p.
103. Wolf YI, Rogozin IB, Grishin NV, Koonin EV. 2002. Genome trees and the tree of life. Trends in Genet 18:472–479.
104. Tong KL, Wong JT. 2004. Anticodon and wobble evolution. Gene 333:169–177.
105. Handy J, Doolittle RF. 1999. An attempt to pinpoint the phylogenetic introduction of glutaminyl-tRNA synthetase among bacteria. J Ml Evol 49:709–715.
106. Di Giulio M. 1995. The phylogeny of tRNAs seems to confirm the predictions of the coevolution theory of the origin of the genetic code. Orig Life Evol Biosph 25:549–564.
107. Chaley MB, Korotkov EV, Phoenix DA. Relationships among isoacceptor tRNAs seem to support the coevolution theory of the origin of the genetic code. J Mol Evol 48:168–177.
108. Di Giulio M. 2002. Genetic code origin: are the pathways of the type $Glu\text{-}tRNA^{Gln} \leftarrow Gln\text{-}tRNA^{Gln}$ molecular fossils or not? J Mol Evol 55:616–622.
109. Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, et al. 2004. Database resources of the National Center for Biotechnology Information: update. Nucl. Acid Res 32:D35–D40.
110. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acid Res 25:3389–3402.