

- 31 Ouyang, L., Chen, X. and Bieker, J. J. (1998) *J. Biol. Chem.* 273, 23019–23025
- 32 Tanese, N. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 13611–13616
- 33 Ryu, S., Zhou, S., Ladurner, A. G. and Tjian, R. (1999) *Nature* 397, 446–450
- 34 Schaeper, U. *et al.* (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 10467–10471
- 35 Nibu, Y. *et al.* (1998) *EMBO J.* 17, 7009–7020
- 36 Koritschoner, N. P. *et al.* (1997) *J. Biol. Chem.* 272, 9573–9580
- 37 Sogawa, K. *et al.* (1993) *Nucleic Acids Res.* 21, 1527–1532
- 38 Yet, S. F. *et al.* (1998) *J. Biol. Chem.* 273, 1026–1031
- 39 Subramaniam, M. *et al.* (1995) *Nucleic Acids Res.* 23, 4907–4912
- 40 Imataka, H. *et al.* (1992) *EMBO J.* 11, 3663–3671
- 41 Wimmer, E. A. *et al.* (1993) *Nature* 366, 690–694

Selection, history and chemistry: the three faces of the genetic code

Robin D. Knight, Stephen J. Freeland
and Laura F. Landweber

The genetic code might be a historical accident that was fixed in the last common ancestor of modern organisms. ‘Adaptive’, ‘historical’ and ‘chemical’ arguments, however, challenge such a ‘frozen accident’ model. These arguments propose that the current code is somehow optimal, reflects the expansion of a more primitive code to include more amino acids, or is a consequence of direct chemical interactions between RNA and amino acids, respectively. Such models are not mutually exclusive, however. They can be reconciled by an evolutionary model whereby stereochemical interactions shaped the initial code, which subsequently expanded through biosynthetic modification of encoded amino acids and, finally, was optimized through codon reassignment. Alternatively, all three forces might have acted in concert to assign the 20 ‘natural’ amino acids to their present positions in the genetic code.

THE GENETIC CODE remains an enigma, even though the full codon catalog was deciphered over 30 years ago. Although we know which base triplets encode which amino acids, and even how these assignments vary among taxa, we do not know why the specific codon assignments take their actual form¹. Why, for instance, does the AUU triplet encode isoleucine rather than some other amino acid? Why do some amino acids have more codons than others? And why do amino acids that have similar chemical properties tend to have similar codons (Fig. 1)?

The simplest answer is that codon assignments were historical accidents that became fixed in the last common

ancestor of all modern organisms and that, therefore, their pattern requires no further explanation². This ‘frozen accident’ hypothesis is a useful null model against which other models can be tested, but does not predict the observed order in the genetic code. The model has also been criticized because we now know that the code is not universal, and thus variant codes might have existed before the last common ancestor, as well as at present.

There are three main challenges to the frozen-accident model, which are based on ‘adaptive’, ‘historical’ and ‘chemical’ arguments. All three deal only with the genetic code present in the last universal ancestor and might not apply to more-recent changes. The ‘adaptive’ challenge suggests that the pattern of codon assignments is an adaptation that optimizes some function, such as minimization of errors caused by mutation or mistranslation. The ‘historical’ challenge suggests that

the genetic code accumulated amino acids over a long period of time and that codon assignments reflect this pattern of incremental expansion. The ‘chemical’ challenge suggests that certain codon assignments were directly influenced by favorable chemical interactions between particular amino acids and short nucleic acid sequences, whereas lack of such interactions excluded other amino acids from proteins entirely. Here, we evaluate the evidence for these three views and suggest how they might be combined into a coherent synthesis of code evolution.

Adaptation – the best of all possible codes?

The earliest explanations for the observed order in the genetic code, such as Crick’s ingenious commaless code³, assumed that natural selection somehow optimized the codon catalog. Given that more changes to a protein are deleterious than beneficial, the genetic code should reduce the impact of errors: the pattern of degeneracy, which groups together codons for the same amino acid, certainly has this effect (Fig. 1). The ‘lethal mutation’ model⁴ proposed that the genetic code reduces the effects of point mutation, whereas the ‘translation error’ model⁵ proposed that the code structure instead reduces the effects of errors during translation.

The principal evidence that supported these early models came from inspection of the genetic code itself: (1) codons for the same amino acid typically vary only at the third position; (2) amino acids that have U at the second position of their codon are hydrophobic, whereas those that have A at the second position are hydrophilic; and (3) the genetic code initially appeared to be universal⁵. This evidence is neither compelling nor unequivocal. Crick’s wobble hypothesis⁶ explained much of the degeneracy of the code in terms of simple chemical considerations: a single tRNA anticodon can recognize multiple codons by nonstandard base pairing. The association between second-position base and amino acid hydrophobicity holds only for two of the four bases

R. D. Knight, S. J. Freeland and L. F. Landweber are at the Dept of Ecology and Evolutionary Biology, Guyot Hall, Princeton University, Princeton, NJ 08544-1003, USA.
Email: lfl@princeton.edu

	U	C	A	G
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys
	UUA Leu	UCA Ser	UAA TER	UGA TER
	UUG Leu	UCG Ser	UAG TER	UGG Trp
C	CUU Leu	CCU Pro	CAU His	CGU Arg
	CUC Leu	CCC Pro	CAC His	CGC Arg
	CUA Leu	CCA Pro	CAA Gln	CGA Arg
	CUG Leu	CCG Pro	CAG Gln	CGG Arg
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser
	AUC Ile	ACC Thr	AAC Asn	AGC Ser
	AUA Ile	ACA Thr	AAA Lys	AGA Arg
	AUG Met	ACG Thr	AAG Lys	AGG Arg
G	GUU Val	GCU Ala	GAU Asp	GGU Gly
	GUC Val	GCC Ala	GAC Asp	GGC Gly
	GUA Val	GCA Ala	GAA Glu	GGA Gly
	GUG Val	GCG Ala	GAG Glu	GGG Gly

 Acidic	 Amide	 Hydroxyl containing
 Alkyl	 Aromatic	 Sulfur containing
 Alkyl	 Basic	 STOP

Figure 1

The ‘universal’ genetic code. Shading indicates polar requirement (PR)¹: lighter shades (black text), PR < 6 (hydrophobic); medium shades (yellow text), PR = 6–8 (medium); darker shades (white text) PR > 8 (hydrophilic). Amino acids whose codons have U at the second position tend to be unusually hydrophobic; those whose codons have A at the second position tend to be hydrophilic. Amino acids that share structural similarity tend to share codon sets connected by single point mutations: for instance, the basic amino acids arginine, lysine and histidine are connected. Ter, termination codon.

(Fig. 1). Finally, if code optimization had actually occurred, then the present genetic code must have been selected from a large pool of alternative genetic codes (a problem when the code was thought to be absolutely invariant). These shortcomings, given the choice of the frozen-accident theory as an alternative, probably account for the decline of adaptive explanations towards the end of the 1960s.

A variety of criteria have been used to assess whether the genetic code is in some sense optimal. These analyses fall into two main classes: ‘statistical’ and ‘engineering’. The statistical approaches^{7–11} compare the natural code with many randomly generated alternative codes and typically have concluded that the genetic code conserves amino

acid properties far better than would a random code. In contrast, the engineering approaches^{12–16} compare the natural code with only the best possible alternative (i.e. the code that formally minimizes the change in amino acid properties following an average single point mutation), and conclude that the genetic code is still far from optimal.

The statistical approach provides a more realistic representation of the variability available to selection than does the engineering approach. Because the engineering approach measures optimality on a linear scale as a fraction of the distance between the mean and optimal codes, it ignores the distribution of possible codes. This distribution is roughly Gaussian: increasingly optimal codes are increasingly rare, and the dif-

ference between successively more optimal codes decreases as optimality increases. Consequently, the globally optimal code might be unattainable, whereas the most optimal code accessible by point mutations is still closer to optimal than almost all alternatives. In fact, our unpublished results indicate that the canonical genetic code is closer to optimal than practically all alternatives, and this conclusion holds for differences in both measurement of optimality and distribution of possible codes. However, the evolutionary plasticity of the code might have been limited by unknown chemical or historical constraints.

The principal objection to optimization theories has been that a change in the genetic code causes mutations in every protein, most of which are likely to be deleterious. Consequently, once cells relied on a particular genetic code to any appreciable extent, the further changes required by the optimization process would have become increasingly unlikely². The ability of the genetic code to change is a prerequisite for theories that involve optimization through a stepwise evolutionary process. The discovery that the genetic code is not invariant¹⁷ removed this objection: if the genetic code recently has changed in apparently nonadaptive ways, then similar changes might have facilitated adaptation in the past. Actual changes in the nuclear genomes of eukaryotes (Fig. 2a) indicate that, even in metabolically complex organisms, the code is far from frozen.

Two mechanisms account for the codon swapping evident in a variety of species, and in both nuclear and mitochondrial genomes (Fig. 2). In the Osawa–Jukes mechanism¹⁸, particular codons vanish from the genome because of mutational pressure on the genome for changes in A.T or G.C composition, and the corresponding tRNAs are lost. When the mutational pressure later reverses, codons that lack cognate tRNAs inhibit translation. Consequently, any mutation that allows translation of these codons is advantageous. Such a mutation can occur through duplication of an existing tRNA gene and subsequent mutation of the anticodon to recognize a different codon. If the mutated tRNA still retains its original aminoacyl-tRNA synthetase specificity, the codon will encode an amino acid that differs from that used by the canonical code.

The Schultz–Yarus mechanism¹⁹ is similar but does not require the complete disappearance of a codon from the

genome before the transfer takes place. Instead, a mutation in a duplicated tRNA that generates either a new anticodon or a new aminoacyl-charging specificity leads to ambiguous translation of one or more codons. If this new specificity confers an advantage, selection will fix the new codon set. The fact that certain *Candida* species have ambiguous translation – depending on the circumstances, CUG will encode either serine or leucine – supports the model²⁰.

History – searching for footprints of the code's ancestors

Historical theories propose that the present code evolved from a simpler ancestral form: proteins produced by the initial, limited, set of amino acids synthesized new amino acids that could in turn be incorporated into the code. Recently introduced amino acids presumably would take over codons from their metabolic precursors; this could happen only if the resulting changes in protein structure were not widely deleterious². Consequently, historical theories often predict that similar amino acids would be assigned to similar codons even without explicit selection for error minimization.

The principal evidence for coevolution of amino acids and the code through stepwise expansion comes from cases in which dissimilar amino acids from related biosynthetic pathways also share similar codons (Fig. 3). Several authors argue that a disproportionate number of biosynthetically related amino acids have codons connected by single point mutations^{14,16,21,22}; however, because many amino acids are interconvertible, even randomized codes show similar associations between biosynthetically related amino acids and single base changes in codons²³.

One intriguing suggestion is that the first- and second-position bases have different functions: the second-position bases connect amino acids that have similar properties; and the first-position bases connect amino acids from the same biosynthetic pathway²⁴. Codons of the form GNN correspond to amino acids thought to be most primitive for several reasons²⁴; this might suggest that UNN, CNN and ANN codons were transferred to novel amino acids as their synthesis became possible. This hypothesis constrains the set of possible codes considerably, but does not explain the near optimality of the code¹¹.

Another approach looks at the phylogenies of tRNAs and of aminoacyl-tRNA

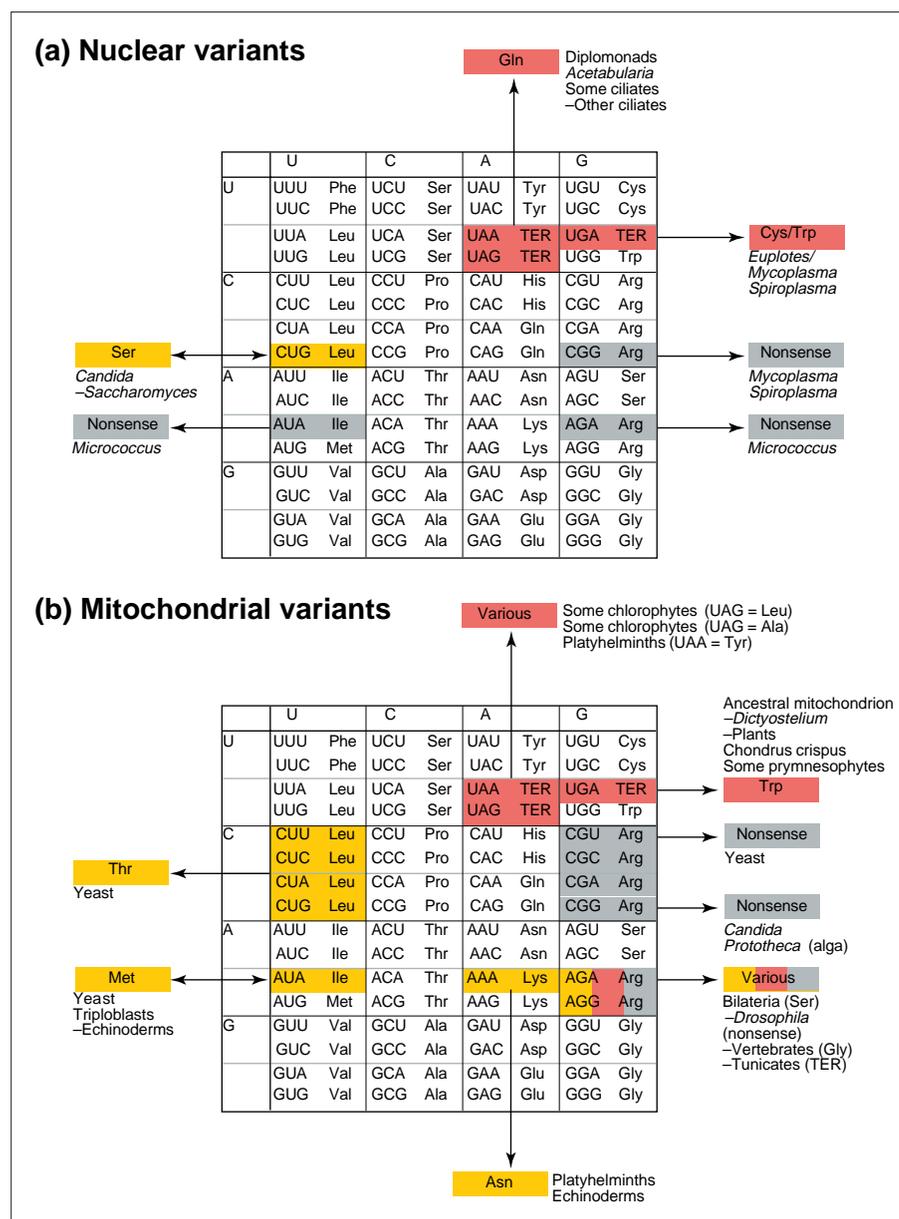


Figure 2

Naturally occurring variants of the canonical genetic code. **(a)** Nuclear variants (including changes effective within bacterial genomes)^{34,48,49}. **(b)** Mitochondrial variants^{48,50,51} (yeast variants are from <http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c>). Missense changes are shown in yellow; nonsense changes are shown in gray; changes in termination codons are shown in red. ‘–’ indicates a reversal of a change in a particular lineage.

synthetases (the enzymes that specifically link amino acids to their cognate tRNAs). If amino acids were added sequentially to the code, then tRNA and aminoacyl-tRNA synthetase phylogenies should be congruent; this would reflect duplication and divergence of a tRNA and its cognate synthetase as each amino acid was added. Unfortunately, most studies that examined tRNA phylogenies^{25–27} have assumed that trees derived from the set of tRNAs in different species are congruent, which is not the case²⁸. Because tRNAs can change either their anticodons or their amino acid specificity remarkably easily²⁹,

modern tRNA phylogenies are unlikely to reveal anything about the phylogeny of tRNAs in the last common ancestor. Furthermore, tRNA phylogenies are likely to become increasingly unstable as more sequences are added: this apparent tRNA flexibility is consistent with the requirement of the adaptive theories that the code be able to change.

Phylogenies of aminoacyl-tRNA synthetases prove slightly more revealing. Aminoacyl-tRNA synthetases fall into two main classes. Some of those for related amino acids cluster together³⁰, and phylogenies are similar among widely separated taxa³¹. Interestingly,

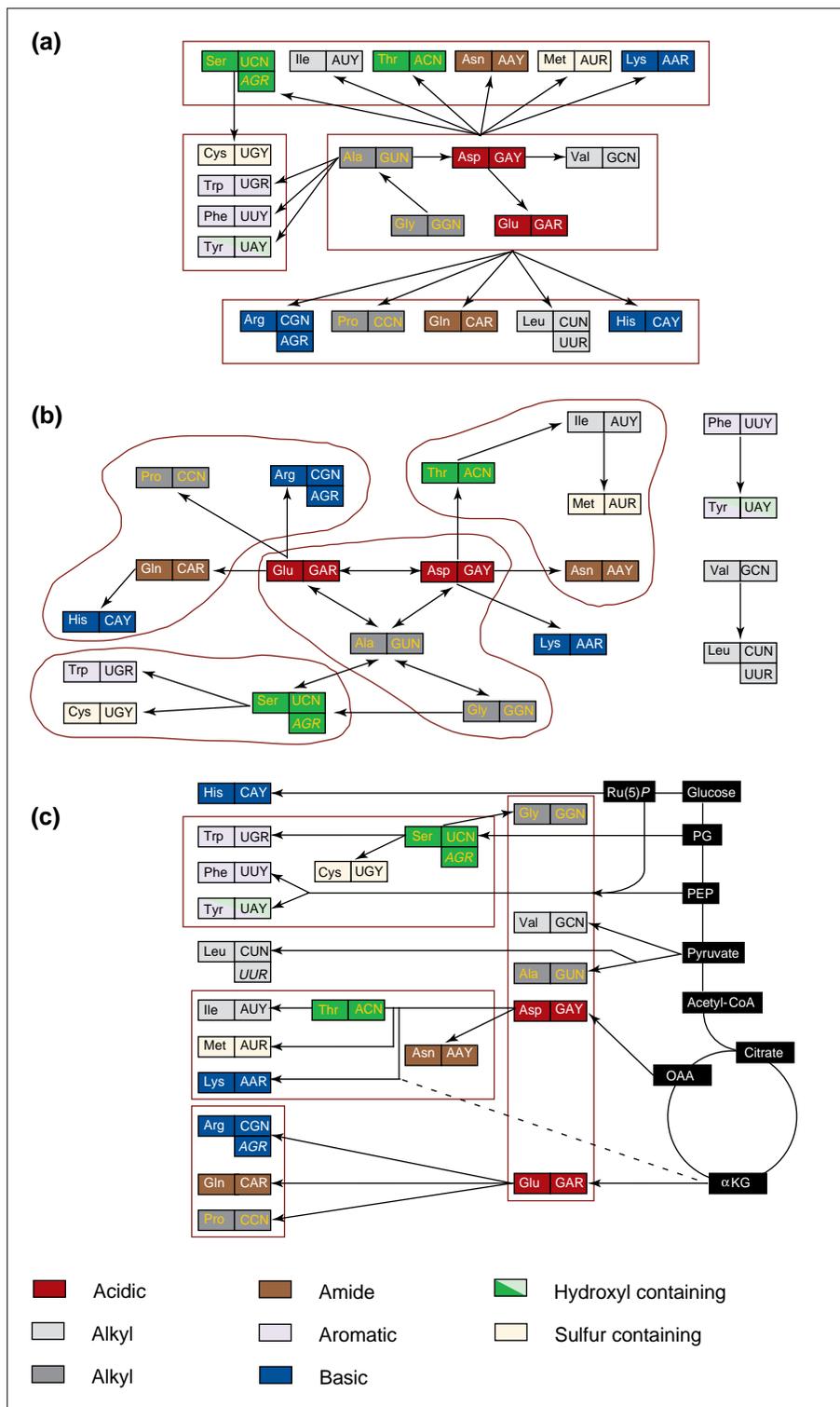


Figure 3

Biosynthetic pathways and code assignments. **(a)** Primitive sulfur-metabolizing bacteria (hypothetical)⁴⁷. **(b)** Generalized prokaryotes²¹. **(c)** *Escherichia coli*²⁴. Shading indicates polar requirement (PR)¹: lighter shades (black text), PR < 6 (hydrophobic); medium shades (yellow text), PR = 6–8 (medium); darker shades (white text) PR > 8 (hydrophilic). Bounded areas highlight codons that share the same first base identity. αKG, α-ketoglutarate; OAA, oxaloacetic acid; PEP, phosphoenolpyruvate; PG, phosphoglycerate; Ru(5)P, ribulose 5-phosphate.

although most organisms have a class II lysyl-tRNA synthetase, some archaea and spirochetes have a class I lysyl-tRNA synthetase³². Given that the class I lysyl-tRNA synthetases are monophyletic

and cluster within the other type I synthetases³¹, the last common ancestor of all organisms probably contained both types of synthetase, and all lineages probably lost one or the other at a later

stage³³. However, because the complete set of tRNA synthetases and tRNAs was present in the last common ancestor, phylogenetic analysis alone cannot discriminate between stepwise introduction of amino acids into translation and stepwise takeover of aminoacylation by protein aminoacyl-tRNA synthetases from more-primitive catalysts. Although congruence between tRNA and synthetase phylogenies would have provided striking evidence for sequential amino acid incorporation, the lack of such congruence provides evidence against expansion of the code during synthetase evolution. The present synthetases might have usurped the roles of earlier ribozymes that had the same functions, erasing the information in the original synthetases about the order in which amino acids were added to the code.

Stereochemistry – does it fit the evidence?

Stereochemical theories propose that amino acids are assigned to particular codons because of direct chemical interactions between RNA and amino acids. If these interactions follow consistent patterns, similar amino acids should bind to similar short RNA motifs and should therefore have similar codons. Although the resulting pattern of codon assignments might be adaptive, relative to randomized codes (because a point mutation would tend to substitute a relatively similar amino acid), it need not have been explicitly selected for this effect. Thus, the rules that constrain the set of chemically plausible codes might also lead to apparent error minimization.

The fact that the genetic code initially appeared to be universal provided the strongest support for stereochemical theories, because it suggested that the actual code is the only possible code. However, the known variations in the code do not disprove the stereochemical theories. All deviations from the canonical code appeared recently in comparison with the last common ancestor: the first surviving change probably appeared in the lineage leading to diplomonads³⁴, and most are much more recent. Furthermore, no known code differs by more than a few amino acids from the standard code. Because translation pairs codons with amino acids through a tRNA adaptor, the mechanisms that allowed recent changes in the genetic code might be entirely different from those that generated the code initially. All stereochemical theories have dealt only with the canonical code found in the last common ancestor,

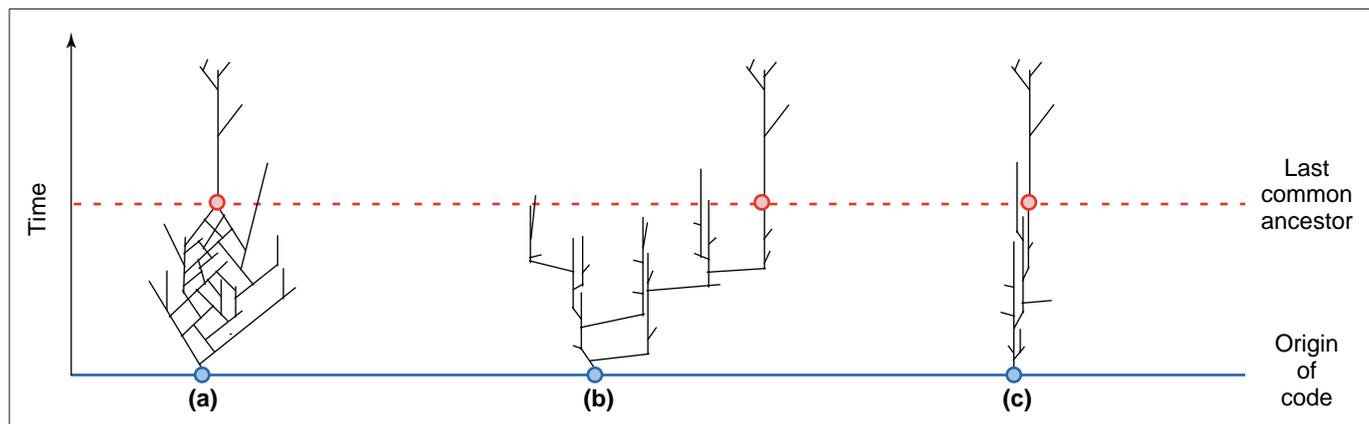


Figure 4

Three models of early code evolution. The ‘universal’ genetic code found in the last common ancestor (pink circle) might or might not be similar to the first genetic code that evolved (blue circle). **(a)** The primordial genetic code is maintained by lineage merging in a reticulate network: there is little competition between lineages, and lineages that share the majority genetic code have the advantage of using novel proteins from other lineages when protocells merge. **(b)** Strong selection for increased code efficiency among lineages drives the code in the last common ancestor far from the primordial code. Most lineages with variant codes become extinct, but a few successfully reach new local optima. **(c)** Despite competition among lineages, the chemical factors leading to the establishment of the original genetic code are much the same as the factors that influence the error in a given amino acid substitution; therefore the final code remains similar to the initial code. Aptamer experiments can distinguish (b) from (a) and (c) by providing evidence for a primordial code that might or might not be similar to the code in the last common ancestor.

because later changes probably were unaffected by stereochemical constraints.

The first stereochemical theories about the origin of the code relied on chemical models. These provided weak

support for a variety of possible pairing mechanisms: amino acids might bind to their cognate codons³⁵, anticodons³⁶, reversed codons³⁷, codon–anticodon double helices³⁸ or a complex of four

nucleotides containing the anticodon at the end of the acceptor stem³⁹. Unfortunately, the diversity of results reduces their significance: the apparent freedom inherent in the building and

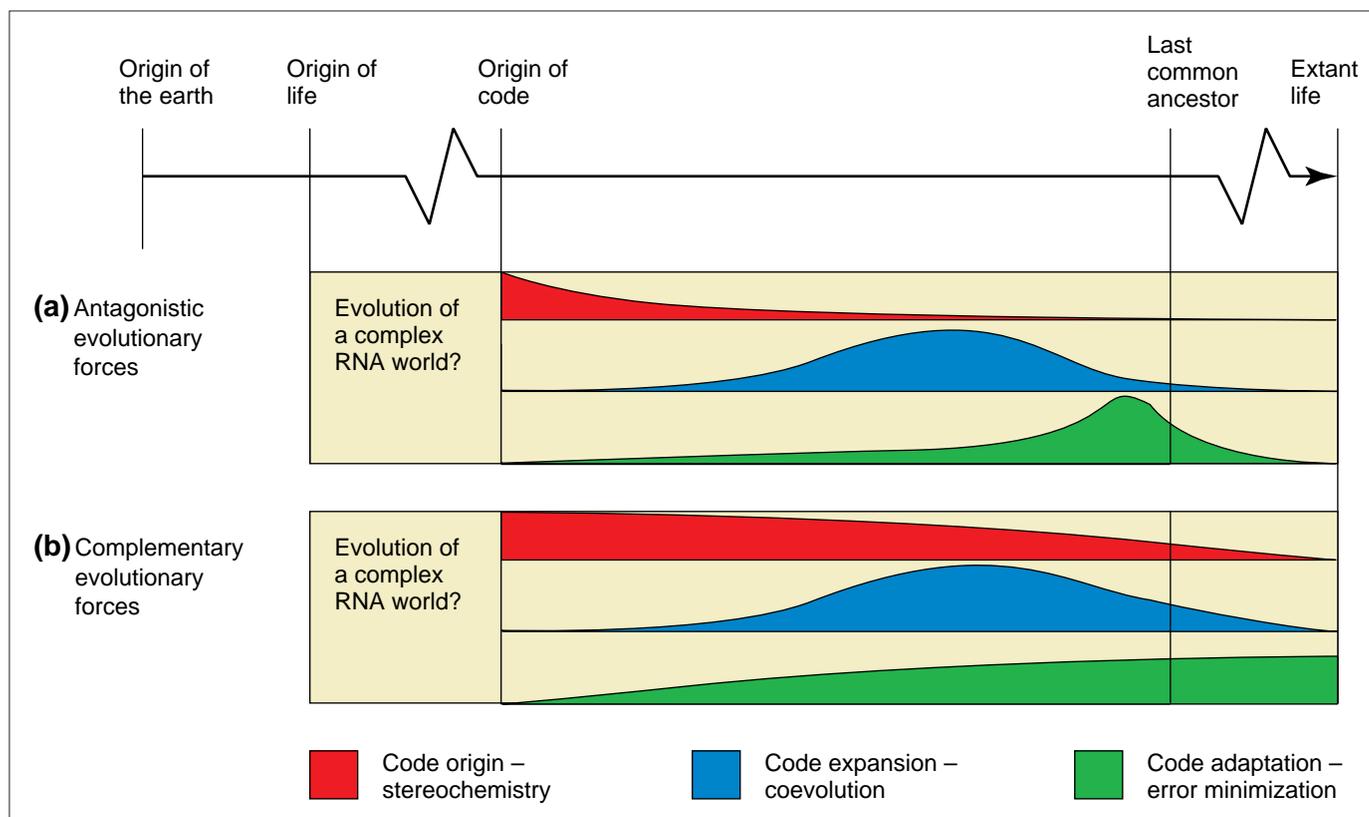


Figure 5

Three facets of code evolution. The genetic code probably originated through stereochemical interactions and, then, underwent a period of expansion in which new amino acids were incorporated. The evolution of the tRNA system, which separated codons from direct interaction with amino acids, then allowed reassignment of codons and, therefore, adaptive evolution. Traditionally, these forces have been assumed to be antagonistic **(a)**, but they might actually have been complementary **(b)**; for example, current codon assignments might assign biosynthetically similar amino acids to similar codons, which would meet both stereochemical and adaptive criteria.

interpretation of these models has undermined the significance of any particular model, especially in the absence of empirical predictions.

Another approach has been to examine interactions between amino acids and individual bases or nucleotides. Early studies showed that 'polar requirement,' a partitioning coefficient of a water-pyridine system that reflects hydrophobicity, varies among second-position bases¹. Other approaches included tests for the following: (1) correlations between the hydrophobicity of an amino acid and particular nucleotides or dinucleotides; (2) correlations between the partitioning coefficients of amino acids and nucleotides on various surfaces; and (3) differential effects of particular amino acids on nucleotide solubility. These studies tend to show weak associations between anticodons and amino acids⁴⁰.

The most direct test of RNA-amino acid interactions is to determine the precise RNA sequences that bind most strongly to each amino acid. *In vitro* selection, which isolates nucleic acid molecules that bind to a particular target by selective amplification over several generations⁴¹, has generated aptamers (RNA ligands) for several amino acids. Interactions between arginine and RNA have been studied in most detail: several laboratories have selected and characterized the binding of arginine to arginine aptamers. The set of codons assigned to arginine occurs far more often at arginine-binding sites than would be expected by chance: arginine anticodons, and the codon sets assigned to other amino acids, do not show this association⁴². We propose that this is also the case for at least some other amino acids and their codons, and that arginine interacts with its codons in other contexts, such as in RNA-binding proteins. Such intrinsic affinities between codons and amino acids might have influenced early codon assignments. Information about RNA molecules that bind to other amino acids will test the generality of this hypothesis. The first isoleucine aptamers seem to have critical isoleucine codons at their binding sites, although the first valine aptamers do not⁴³.

The RNA world: the milieu of code evolution?

Translation presents a 'chicken or egg' problem: given that many crucial components of the translation apparatus (including aminoacyl-tRNA synthetases, release factors and much of the ribosome) are made of protein, how

could translation ever have evolved? The RNA-world hypothesis⁴⁴ avoids this problem by suggesting that RNA preceded DNA and protein and acted as both genetic material and catalyst. The structure of the genetic code might contain information about the chemical environment in which the code evolved.

Two plausible pathways explain how a genetic code arose in an RNA world. First, RNA catalysts might have built specific peptides residue by residue, much in the way that short peptides are now constructed by specific enzymes. Once a general translation system evolved, it would have supplanted these early peptide-synthesis pathways. Second, some ribozymes might have used amino acids, and later peptides, as cofactors⁴⁵. As peptide synthesis became more feasible, the peptide parts of the hybrid catalysts would increasingly have replaced the RNA components; the final result was a protein world in which a few essential nucleotide cofactors remained as molecular fossils. In either case, specific interactions between RNA and amino acids would have been necessary to establish the initial coding system.

Compelling evidence (see above) supports the idea that arginine, and perhaps isoleucine, interacts with its codons in RNA aptamers and that the genetic code is highly optimal with respect to error minimization. When sequences for aptamers for more amino acids are available, we will be able to test whether chemical factors influenced the choice of amino acids and their codon assignments in the canonical genetic code. Assuming that each amino acid was originally assigned those codons for which it has greatest chemical affinity, it would be possible to reconstruct this primordial genetic code. The divergence between this primordial code and the code found in the last common ancestor of all life could test models of early code evolution (Fig. 4).

We envisage a series of definite, although perhaps overlapping, stages in the evolution of the code (Fig. 5). At first, in the RNA world, stereochemical interactions would have largely determined the correspondence between certain RNA-sequence tags and amino acids. Such early peptides, generated by direct templating⁴³ or similar mechanisms, need not have had catalytic function: for instance, short positively charged arginine repeats might have neutralized the phosphate backbones of RNA molecules, potentially allowing uptake of the latter through membranes⁴⁶

and/or their refolding into active structures. As amino acid and peptide cofactors, and eventually catalysts, became more prevalent at the onset of the RNA-protein world, coevolution of the code and the amino acid set might have led to expansion of the code on the basis of metabolic relatedness⁴⁷. This expansion would also have preserved the rules initially established by stereochemical interactions in order to continue making the original templated protein or proteins. Finally, after the evolution of the mRNA-tRNA-aminoacyl-tRNA-synthetase system removed direct interaction between amino acids and codons, codon swapping in different lineages would have permitted some degree of code optimization by codon reassignment.

Code optimization, however, need not be limited to this late stage: error minimization might have acted in concert both with stereochemical considerations and with biosynthetically driven code expansion to produce the canonical code (Fig. 5b). Recent evidence that suggests that the code has a highly optimized structure⁷⁻¹¹ highlights the crucial gap in our understanding of its evolution: the pattern of chemical interactions between the 64 codons and 20 amino acids remains largely unknown. Only when these interactions are known will we be able to understand the relative importance of selection, history and chemistry in code evolution.

References

- 1 Woese, C. R., Dugre, D. H., Saxinger, W. C. and Dugre, S. A. (1966) *Proc. Natl. Acad. Sci. U. S. A.* 55, 966-974
- 2 Crick, F. H. C. (1968) *J. Mol. Biol.* 38, 367-379
- 3 Crick, F. H. C. (1957) *Biochem. Soc. Symp.* 14, 25-26
- 4 Sonneborn, T. M. (1965) in *Evolving Genes and Proteins* (Bryson, V. and Vogel, H. J., eds), pp. 377-297, Academic Press
- 5 Woese, C. R. (1967) *The Genetic Code: The Molecular Basis for Genetic Expression*, Harper and Row
- 6 Crick, F. H. (1966) *J. Mol. Biol.* 19, 548-555
- 7 Alff-Steinberger, C. (1969) *Proc. Natl. Acad. Sci. U. S. A.* 64, 584-591
- 8 Haig, D. and Hurst, L. D. (1991) *J. Mol. Evol.* 33, 412-417
- 9 Ardell, D. H. (1998) *J. Mol. Evol.* 47, 1-13
- 10 Freeland, S. J. and Hurst, L. D. (1998) *J. Mol. Evol.* 47, 238-248
- 11 Freeland, S. J. and Hurst, L. D. (1998) *Proc. R. Soc. London Ser. B* 265, 2111-2119
- 12 Wong, J. T. (1980) *Proc. Natl. Acad. Sci. U. S. A.* 77, 1083-1086
- 13 Di Giulio, M. (1989) *J. Mol. Evol.* 29, 288-293
- 14 Di Giulio, M. (1991) *Z. Naturforsch.* 46c, 305-312
- 15 Di Giulio, M., Capobianco, M. R. and Medugno, M. (1994) *J. Theor. Biol.* 168, 43-51
- 16 Di Giulio, M. (1998) *J. Mol. Evol.* 46, 615-621
- 17 Barrell, B. G., Bankier, A. T. and Drouin, J. (1979) *Nature* 282, 189-194

- 18 Osawa, S. and Jukes, T. H. (1988) *Trends Genet.* 4, 191–198
- 19 Schultz, D. W. and Yarus, M. (1994) *J. Mol. Biol.* 235, 1377–1380
- 20 Yarus, M. and Schultz, D. W. (1997) *J. Mol. Evol.* 45, 1–8
- 21 Wong, J. T-F. (1975) *Proc. Natl. Acad. Sci. U. S. A.* 72, 1909–1912
- 22 Miseta, A. (1989) *Physiol. Chem. Phys. Med. NMR* 21, 237–242
- 23 Amirnovin, R. (1997) *J. Mol. Evol.* 44, 473–476
- 24 Taylor, F. J. R. and Coates, D. (1989) *Biosystems* 22, 177–187
- 25 Eigen, M. and Winkler-Oswatitsch, R. (1981) *Naturwissenschaften* 68, 282–292
- 26 Fitch, W. M. and Upper, K. (1987) *Cold Spring Harbor Symp. Quant. Biol.* 52, 759–767
- 27 Eigen, M. et al. (1989) *Science* 244, 673–679
- 28 Saks, M. E. and Sampson, J. R. (1995) *J. Mol. Evol.* 40, 509–518
- 29 Saks, M. E., Sampson, J. R. and Abelson, J. (1998) *Science* 279, 1665–1670
- 30 Nagel, G. M. and Doolittle, R. F. (1995) *J. Mol. Evol.* 40, 487–498
- 31 Ribas de Pouplana, L., Turner, R. J., Steer, B. A. and Schimmel, P. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 11295–11300
- 32 Ibba, M., Bono, J. L., Rosa, P. A. and Soll, D. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 14383–14388
- 33 Landweber, L. F. and Katz, L. A. (1998) *Trends Ecol. Evol.* 13, 93–94
- 34 Keeling, P. J. and Doolittle, W. F. (1997) *Mol. Biol. Evol.* 14, 895–901
- 35 Pelc, S. R. and Welton, M. G. E. (1966) *Nature* 209, 868–872
- 36 Dunnill, P. (1966) *Nature* 210, 1267–1268
- 37 Root-Bernstein, R. S. (1982) *J. Theor. Biol.* 94, 895–904
- 38 Hendry, L. B. and Whitham, F. H. (1979) *Perspect. Biol. Med.* 22, 333–345
- 39 Shimizu, M. (1982) *J. Mol. Evol.* 18, 297–303
- 40 Lacey, J. C., Jr (1992) *Orig. Life Evol. Biosph.* 22, 243–275
- 41 Landweber, L. F., Simon, P. J. and Wagner, T. A. (1998) *BioScience* 48, 94–103
- 42 Knight, R. D. and Landweber, L. F. (1998) *Chem. Biol.* 5, R215–R220
- 43 Yarus, M. (1998) *J. Mol. Evol.* 47, 109–117
- 44 Gilbert, W. (1986) *Nature* 319, 618
- 45 Szathmáry, E. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 9916–9920
- 46 Jay, D. G. and Gilbert, W. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 1978–1980
- 47 Dillon, L. S. (1973) *Bot. Rev.* 39, 301–345
- 48 Osawa, S. (1995) *Evolution of the Genetic Code*, Oxford University Press
- 49 Tourancheau, A. B. et al. (1995) *EMBO J.* 14, 3262–3267
- 50 Hayashi-Ishimaru, Y. et al. (1996) *Curr. Genet.* 30, 29–33
- 51 Hayashi-Ishimaru, Y., Ehara, M., Inagaki, Y. and Ohama, T. (1997) *Curr. Genet.* 32, 296–299

Back to Camelot: defining the specific role of tRNA in protein synthesis

After graduating from Medical School in 1961, I went to work in Seymour Benzer's laboratory at Purdue University, where I was privileged to participate in a series of exciting experiments on the then emergent genetic code. One study that received some notoriety was a critical test of the 'adaptor hypothesis' proposed by Francis Crick in 1958. Crick had postulated that a small oligonucleotide, possibly soluble RNA (sRNA, as it was then known; tRNA as it is known today), functions as an adaptor for the incorporation of amino acids into protein¹. Thus, it followed that once an amino acid is attached to sRNA, the specificity with which it is incorporated into protein resides solely in the sRNA adaptor to which it is attached.

The Raney-nickel experiment² (Fig. 1), as it came to be called, is often cited as the critical experiment that proved the adaptor hypothesis. It allowed us to demonstrate that, in an *in vitro* protein-synthesis system, alanine from alanyl-tRNA^{Cys} is incorporated into protein at positions normally occupied by cysteine rather than at those occupied by alanine. The Raney-nickel experiment, however, was only one of a series of experiments that confirmed the adaptor hypothesis. As interesting as the results

of the experiments themselves was the way in which these experiments came to be done and what followed.

Genetic studies of allele-specific suppression that led to the Raney-nickel experiment

This trail of research did not start out as an attempt to test the adaptor hypothesis, but developed from allele-specific (i.e. mutant-specific) genetic-suppressor studies of the phage-T4 *rII* system by Benzer and Champe, which began around 1959 (Ref. 3). While these studies were in progress, a series of papers from Francois Gros' laboratory^{4–6} revealed that *Escherichia coli* grown in the presence of 5-fluorouracil (5-FU) made abnormal proteins. For example, alkaline phosphatase and β -galactosidase were shown to have altered amino acid compositions and altered thermostability, but conserved antigenicity. We now believe that 5-FU exerts its suppressor activity because, although it is incorporated into mRNA as uracil, it base pairs with guanine in aminoacyl-tRNA anticodons (i.e. it exhibits the incorporation specificity of cytosine). The observation that the amounts of proline and tyrosine incorporated into total protein, as well as into tRNA-containing fractions, were markedly

increased, suggested that the effect was informationally specific. In parallel with the 5-FU studies, yet a third system, *E. coli* tryptophan synthetase, provided insight into allele-specific suppression. Yanofsky and St Lawrence, in a review entitled 'Gene Action'⁷, suggested that some forms of allele-specific suppression that they had seen in their studies might be caused by alterations in the specificity with which amino acids are incorporated into protein.

Members of the Benzer lab decided to attempt to explain allele-specific suppression of *rII* mutants of phage T4. The fluorouracil effect suggested a biochemical variable that they could include in their studies. The most striking aspect of the fluorouracil effect was its high degree of specificity: it restored enzyme activity very effectively for some *rII* mutants but not at all for others. This suggested that there was a relationship between the fluorouracil effect and the apparently altered specificity of amino acid incorporation reported by the Gros and Yanofsky labs^{4–7}.

By 1960, five years of intensive genetic mapping by the Benzer lab had saturated the *rII* region with mutations to a degree unprecedented in any other genetic system. We therefore hoped that the patterns of suppression by 5-FU at specific sites might be correlated with the wealth of detailed information about those sites. The problem of allele-specific suppression became even more interesting when it was noted that certain strains of *E. coli* K carry genetic suppressors (which eventually turned out to be mutant tRNA genes, as predicted) whose action mimics the phenotypic-suppressor activity of 5-FU.