The puzzling origin of the genetic code

Robert Cedergren and Pedro Miramontes

Recent results add to the mystery of the origin of the genetic code. In spite of early doubts, RNA can discriminate between hydrophobic amino acids under certain contexts. Moreover, codon reassignment, which has taken place in several organisms and mitochondria, is not a random process. Finally, phylogenies of some aminoacyl-tRNA synthetases suggest that the entire code was not completely assigned at the time of the divergence of bacteria from nucleated cells.

THE IDEA OF a genetic code that related the nucleotide sequences of genes to the amino acid sequences of proteins generated great excitement when it was first realized some 35 years ago. Ever since, there has been speculation as to the origin of this code: how were amino acids assigned to codons? In 1967, Carl Woese[1] suggested that there was a direct structural complementarity between an amino acid and a codon: a 'hand in glove' recognition. However, this idea was soon challenged by the 'adaptor hypothesis' of Francis Crick (for example, see Ref. 2). Even though Crick never ruled out a stereochemical relationship between the codon or RNA and its amino acid, the postulated existence of an 'adaptor' molecule that could simultaneously recognize a codon and an amino acid circumvented the need for a direct structural link between the two.

The presence of this adaptor thus suggested that the origin of the genetic code could be the result of a 'frozen accident' (so-called 'frozen' because once established, any change would be lethal), whereby the assignment of amino acids to codons was owing to a unique random event or series of events; it has been the dominant theory on the origin of the genetic code to date. By contrast, the idea that structural complementarity is the primordial origin of the genetic code suggests a more organized evolutionary approach, and although testable, it has lacked experimental verification.

Robert Cedergren and **Pedro Miramontes** are at the Département de Biochimie, Université de Montréal, Montréal, Québec, Canada H3C 3J7.

Recently, results from investigations dealing with selective binding of amino acids by RNA, alternative genetic codes and aminoacyl-tRNA synthetase phylogenies indicate that the time has come to re-evaluate our beliefs. That Crick's adaptor was later identified as transfer RNA (tRNA) does not rule out a primordial origin based on the stereochemical hypothesis.

Selectivity of ribonucleotides

One drawback of the stereochemical theory is that it suggests that ribonucleotides can distinguish between amino acids. But, how could an oligonucleotide possibly distinguish between two hydrophobic amino acids such as valine and isoleucine? Experiments in the Yarus laboratory addressed this question head on and with great elegance. Using *in vitro* selection from random RNA libraries, the Yarus group found RNA molecules having a highly conserved, asymmetric internal loop and that bind to L-valine 20 times better than to isoleucine, and 15 times better than to D-valine[4]. More surprisingly, RNAs that bind arginine contain the arginine codons in an internal loop region that is probably the attachment site for the amino acid[5].

Furthermore, in a rudimentary aminoacyl-exchange assay, an RNA molecule composed of a codon (for Gly, Ala, Val, Trp, Ser or Phe) at the 5′-terminus of a small stem–loop structure is surprisingly aminoacylated only in the presence of the cognate aminoacyl-adenylate[6]. These findings suggest that codons in the correct structural context can be selective binding agents of amino acids; whereas isolated codons are not efficient in amino acid recognition[7].

Alternative genetic codes

The universality of the genetic code was well accepted until 1983 when alternative codon assignments were discovered in mitochondria[8]. More recently, variant codes have been found in a variety of eukaryotes and the eubacterial *Mycoplasma*[9]. As the above organisms (and mitochondria) share ancestors with organisms that use the dominant (universal) code, from which all alternative codes were derived, then this dominant code must be ancestral. How, then, is it possible that UAA and UAG code for glutamine (dominant codons: UUA, UUG) in such widely diverging organisms as the ciliates, algae and diplomonads[10]? Furthermore, independent, but parallel codon changes must also explain the fact that UGA means tryptophan (dominant codon: UGG) in the extremely divergent mitochondria and *Mycoplasma* lineages (Fig. 1). In these cases, independent changes in the genetic code have resulted in the assignment of the same amino acid to the same new codons in distant phyla, suggesting that codon reassignments are anything but random.

Aminoacyl-tRNA synthetases

The present-day genetic code is not simply a relationship between codons and amino acids, as aminoacyl-tRNA synthetases aminoacylate individual tRNAs with a specific amino acid[11]. Codon assignments are therefore a result of the ternary interaction between tRNA, aminoacyl-tRNA synthetases and amino acids. It follows that if the code was determined before the divergence of eukaryotes and eubacteria, synthetases and tRNAs for a given amino acid should be related through a common ancestor that predates the separation, and indeed two commonly cited anomalies involving selenocysteine and glutamine follow this expected pattern: (1) selenocysteine is unusual as it has no codon for itself. However, the mechanism by which selenocysteine is introduced into a protein sequence is so similar in eukaryotes and eubacteria to leave little doubt that they shared a common ancestor before the divergence of the two cell types (before point A in Fig. 1)[12,13]. (2) In eukaryotes, individual glutamine and glutamic acid tRNAs and aminoacyl-tRNA synthetases are present and functional; however, in most eubacteria, as well as mitochondria, chloroplasts and possibly archaea, glutamyl-tRNA synthetase charges glutamic acid on both $tRNA^{Glu}$ and $tRNA^{Gln}$.

PII: S0968-0004(96)20015-5

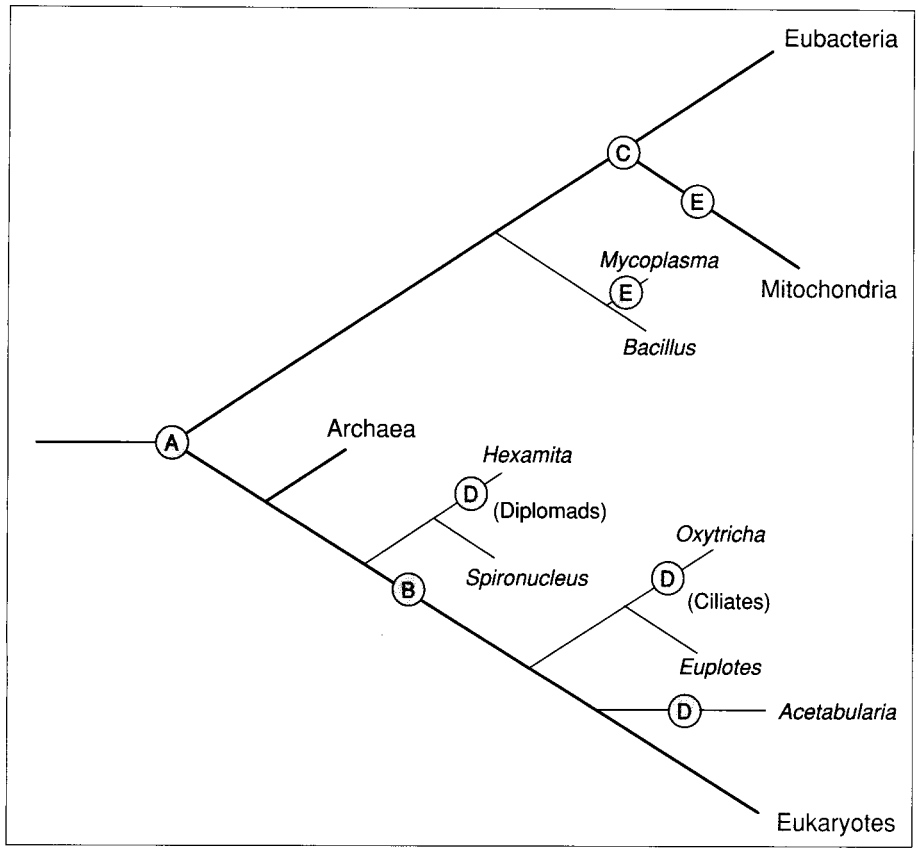

Figure 1

Phylogenetic distribution of organisms with alternative genetic codes. Point A represents the divergence between the bacterial and archaeal/eukaryote branches; point B represents the introduction of the mitochondria in eukaryotic cells; point C, the origin of mitochondria; point D, the conversion of the stop codons UAA and UAG to glutamine codons; point E, the conversion of the stop codon UGA to tryptophan.

Glutamyl-tRNA^{Gln} is then converted to glutaminyl-tRNA^{Gln} by a specific transaminase[14,15]. The γ-purple bacteria such as *Escherichia coli* are an exception and behave as eukaryotes owing to a proposed lateral transfer of glutaminyl-tRNA synthetase from eukaryotes[14]. These data predict that the Glu-tRNA synthetase, which could aminoacylate either tRNA^{Glu} or tRNA^{Gln}, existed at the divergence of eukaryotes and prokaryotes. A subsequent duplication of the synthetase gene allowed the emergence of a glutaminyl-tRNA synthetase in eukaryotes. Meanwhile, in the prokaryotes, glutamine insertion into proteins remained transaminase-dependent.

By contrast to these above examples, data on the tryptophanyl- and tyrosyl-tRNA synthetases recently reported by Ribas de Pouplana *et al.* clearly are more provocative[11]. Studies on the similarity of the two enzymes in eukaryotes and eubacteria led to the conclusion that the two synthetases are more related to each other within their respective superkingdoms than they are to their homologue in the other kingdom. Comparison of tRNA sequences, albeit less robust, shows the same tendency of independent divergences in eukaryotes and eubacteria, leading to tRNA^{Tyr} and tRNA^{Trp}. The most obvious explanation for these facts require that the codons for tyrosine and tryptophan were assigned independently, but in an identical manner, subsequent to the divergence of eukaryotes (after point A in Fig. 1). The authors suggest an alternative scenario with some reservations: that the current version of one of these proteins in eukaryotes originated as a gene duplication of the other[11]. However, even this possibility requires a mechanism whereby two similar, if not identical proteins, would differentiate between tRNA^{Tyr} and tRNA^{Trp}, otherwise the code would change. In addition, this explanation of the data does not deal with the tRNA phylogeny.

Conclusions

Even though, individually, none of the above observations would warrant overturning the 'frozen accident' theory of the origins of the correspondence between codons and amino acids – indeed a number of explanations could be offered for each of the examples cited – considering them together suggests that our ideas on how codons were assigned should be re-evaluated. In the RNA world, the possibility that small RNA molecules could bind amino acids specifically[4] and act as primitive aminoacyl-tRNA synthetases[16] has been demonstrated. We suggest that perhaps some of the above conflicting data could be resolved if these primitive aminoacyl-tRNA synthetases survived much later in protein synthesis than suspected up until now.

Acknowledgements

R. C. is a Fellow of the Evolutionary Biology Group of the Canadian Institute for Advanced Research. P. M. is on leave from the Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, México D. F. México. This work was supported by NSERC of Canada.

References

1 Woese, C. R. (1967) *The Genetic Code*, Harper & Row
2 Crick, F. H. C. (1968) *J. Mol. Biol.* 38, 367–379
3 Labuda, D. and Grosjean, H. (1981) *Biochimie* 63, 77–81
4 Majerfeld, I. and Yarus, M. (1994) *Nat. Struct. Biol.* 1, 287–292
5 Yarus, M. (1993) *The RNA World* (Gesteland, R. F. and Atkins, J. F., eds), pp. 205–217, Cold Spring Harbor Press
6 Shimizu, M. (1995) *J. Biochem.* 117, 23–26
7 Labuda, D. and Grosjean, H. (1981) *Biochimie* 63, 77–81
8 Anderson, S. *et al.* (1983) *Nature* 290, 457–465
9 Watanabe, K. and Osawa, S. (1995) in *tRNA Structure, Biosynthesis and Function* (Söll, D. and RajBhandary, U. L., eds), pp. 225–250, ASM Press
10 Keeling, P. J. and Doolittle, W. F. *EMBO J.* (in press)
11 Ribas de Pouplana, L., Frugier, M., Quinn, C. and Schimmel, P. (1995) *Proc. Natl Acad. Sci. USA* 93, 166–170
12 Baron, C. and Böck, A. (1995) in *tRNA Structure, Biosynthesis and Function* (Söll, D. and RajBhandary, U. L., eds), pp. 529–544, ASM Press
13 Low, S. C. and Berry, M. J. (1996) *Trends Biochem. Sci.* 21, 208–213
14 Lamour, V. *et al.* (1994) *Proc. Natl Acad. Sci. USA* 91, 8670–8674
15 Gagnon, Y., Lacoste, L., Champagne, N. and Lapointe, J. *J. Biol. Chem.* (in press)
16 Illangasekare, M., Sanchez, G., Nickles, T. and Yarus, M. (1995) *Science* 267, 643–647