INDICATIONS OF AN EVOLUTIONARY PATHWAY IN THE AMINO ACID CODE

Thomas H. Jukes

Space Sciences Laboratory
University of California
Berkeley, California 94720*

Received May 5, 1967

In the amino acid code, shown below, the numbers of codons for the various amino acids range from one to six. It is proposed in this communication that this pattern of differences in numbers provides evidence of four progressive stages in evolution of the code, when viewed in terms of the "wobble" hypothesis (1).

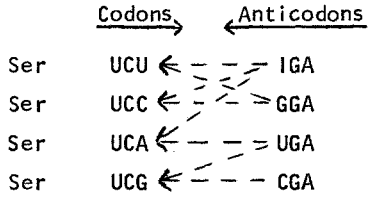
UUU Phe	CUU Leu	AUU Ile	GUU Val
UUC "	CUC "	AUC "	GUC "
UUA Leu	CUA "	AUA "	GUA "
UUG "	CUG "	AUG Met	GUG "
UCU Ser	CCU Pro	ACU Thr	GCU Ala
UCC "	CCC "	ACC "	GCC "
UCA "	CCA "	ACA "	GCA "
UCG "	CCG "	ACG "	GCG "
UAU Tyr	CAU His	AAU Asn	GAU Asp
UAC "	CAC "	AAC "	GAC "
UAA Chain Termn.	CAA Gln	AAA Lys	GAA Glu
UAG " "	CAG "	AAG "	GAG "
UGU Cys	CGU Arg	AGU Ser	GGU Gly
UGC "	CGC "	AGC "	GGC "
UGA Not translated	CGA "	AGA Arg	GGA "
UGG Trp	CGG "	AGG "	GGG "

It is assumed that evolution takes place through changes in DNA, and specifically that a major part in the evolution of the code has been played by mutations in the cistrons that are transcribed into the various transfer RNA (tRNA) molecules (2). It is also assumed that the number of amino acids involved in protein synthesis has increased during evolution

*Supported by grant NsG 479 by the National Aeronautics and Space Administration to the University of California, Berkeley, California.

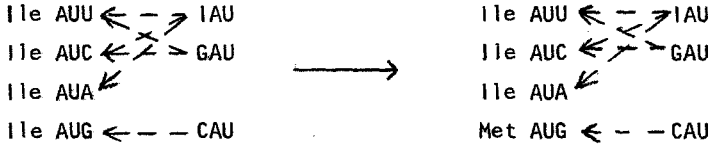
(3, 4, 5). Mutations in a tRNA molecule, taking place in its anticodon or in the region which recognizes the specific activating enzyme (amino acyl synthetase) that charges the tRNA with an amino acid, can enable the tRNA to pair with a different codon or to combine with a different amino acid. Another possibility is that a mutation taking place in a tRNA molecule could change an essential region so that the tRNA would become non-functional. As a consequence, the corresponding cistron might not be conserved, and the tRNA would then disappear. By mutations of these types, and by the additional phenomenon of duplication followed by differentiation, the tRNA molecules can bring about evolutionary changes in the code. The active role of tRNA may be contrasted with the comparatively passive effect of changes in messenger RNA; which is dependent upon tRNAs for its translation.

Seven amino acids each have four codons in which the third base is synonymously U, C, A or G. For each of these amino acids, there are four corresponding anticodons which participate in codon-anticodon pairing in terms of the wobble hypothesis (1), as shown below for serine.

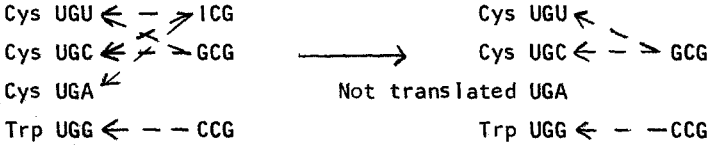


Pairing is indicated by dotted lines. It is not known whether all four anticodons are present in the tRNAs of every organism, but for translation of all four codons, it is necessary that at least IGA + CGA, IGA + UGA, or GGA + UGA be present, as pointed out by Crick (1). It is assumed that these minimum requirements for utilization of amino acids coded by all four bases in the third position of the codon are fulfilled in all normal organisms. It is inferred that the seven amino acids with this coding pattern (termed stage 1) have retained it from an earlier version of the code (3, 4). Leucine (CUX codes; X or Y = U, C, A or G), proline, arginine (CGX codes), threonine, glycine, valine and alanine are the other examples of stage 1.

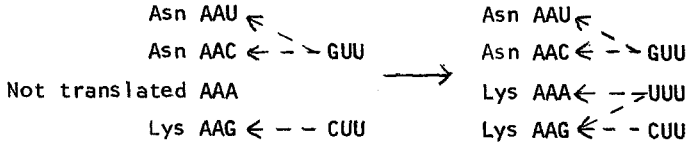
It is postulated that the introduction of a new amino acid into the group used in protein synthesis takes place as follows: the tRNA bearing the anticodon UXY becomes non-functional or is deleted. This loss does not interfere with the translation of all four codons provided that the anticodons IXY and CXY are still present. The tRNA with the anticodon CXY is then changed by mutation in the region that recognizes the activating enzyme so that it becomes charged with the new amino acid. For this crucial step to take place, the evolution of a new activating enzyme is necessary, which will transfer the new amino acid to ester linkage with the tRNA in question. The illustrative example of this step (stage 2) is the case of isoleucine and methionine. It is assumed that the codon AUG was changed in its assignment from Ile to Met:



The next step is to set free a second codon for assignment to the new amino acid. It is proposed that this takes place by deletion of the tRNA having the anticodon IXY, as in the following example (stage 3):



The unassigned codon is then "captured" by the new amino acid. No change in the activating enzyme is necessary. The tRNA for the new amino acid, carrying the CXY anticodon, becomes duplicated. This is followed by a single-base mutation of the anticodon CXY to UXY in one of the two tRNAs that are formed by the cistronic duplication. This is typified by the case of asparagine and lysine (stage 4):



The other examples of stage 4 are the pairs aspartic and glutamic acids; phenylalanine and leucine (UUX codons); histidine and glutamine; and serine and arginine (AGX codons). It is inferred that tRNAs with IXY anticodons, formed from AXY by deamination of adenine in the first position of the triplet, are excluded from stage 4, because such tRNAs would complement with the codons for two different amino acids, as noted by Crick (1).

The above proposals imply that the codons UAA and UAG may have acquired the function of polypeptide chain termination by the disappearance of tyrosine tRNAs having the anticodons IUA, UUA and CUA. This would lead to non-translation of UAA and UAG. Alternatively, tRNAs carrying IUA, UUA and CUA might never have existed, and in this case UAA and UAG would have been "interval" codons ab initio. A third possibility is that tRNAs exist with the anticodons UUA and CUA, and that these tRNAs carry a polypeptide-chain-terminating mechanism.

It is possible that the complexity and specialization of the numerous proteins found currently in living organisms "freezes" the code in its present state, and that evolution of the code has ceased. However, mutations that change certain aspects of protein synthesis still occur in tRNAs, such as suppressor mutations; and duplication of a serine tRNA has been discovered in yeast (6).

The above proposals imply that amino acids whose codons terminate only with purines are the more recent additions to the roster. The amino acids having codons that terminate with pyrimidines are hence the group which functioned in protein synthesis during a preceding ("intermediate" (5)) era. These are phenylalanine, serine, tyrosine, cysteine, leucine, proline, histidine, arginine, isoleucine, threonine, asparagine, valine, alanine, aspartic acid and glycine. This list is slightly different from one suggested previously (4).

To postulate a still earlier step, the problem is involved of further reducing the number of amino acids to eight, each with eight codons.

Such a proposal implies a reduction in codon:anticodon pairing specificity (5). For example, all 8 codons beginning with CA and CC might be translated as proline, etc.

The amino acid composition of the two clostridial ferredoxins whose sequences have been completely determined (7, 8) may be of significance in this context. It has been suggested that these are of very early evolutionary origin and are vestigial substances that originated before the terrestrial atmosphere contained oxygen (9, 10). It is conjectured that their amino acid content is predominantly of "primitive" amino acids that have been conserved and that their composition has been disturbed only slightly by intrusions of "new" amino acids inserted by point mutations. Their combined compositions are Cys_{16} Ala_{15} Val_{12} Asp_{10} Gly_{9} Ile_{9} Ser_{8} Asn_{7} Pro_{6} Gln_{5} Glu_{4} Thr_{4} Phe_{3} Lys_{1} Tyr_{1}. Arginine, histidine, leucine, methionine and tryptophan are absent, and Cl. butyricum ferredoxin contains no lysine or tyrosine. Only 10 of the 110 amino acid residues have codons that must terminate in a purine. If the 15 amino acids in the group postulated for the "intermediate" era are arranged in pairs so that there is a single-base difference between the codons in each pair as shown below (a), one member (b) of each pair may be selected as being the more primitive of the two on the basis of its greater abundance in the two clostridial ferredoxins:

(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
His CAX		Asp GAX		Cys UGX		Ile AUX	
	Pro		Asp		Cys		Ile
Pro CCX		Asn AAX		Arg CGX		Phe UUX	
Tyr UAX		Ala GCX		Gly GGX		Val GUX	
	Ser		Ala		Gly		Val
Ser UCX		Thr ACX		Ser AGX		Leu CUX	

The eight amino acids in columns (b) comprise 77% of the amino acid content of the two clostridial ferredoxins.

In summary, the pattern of the amino acid code is interpreted with the aid of the wobble hypothesis as indicating that some amino acids are of

more recent inclusion than others in the protein-synthesizing mechanism. These five "newer" amino acids are coded exclusively by codons terminating in a purine. The other 15 amino acids are postulated to be the group that were involved in protein synthesis in an earlier era. It is also suggested that the amino-acid composition of two clostridial ferredoxins supports this proposal and provides some indications as to which of these 15 amino acids are even more "primitive" in terms of protein synthesis.

References

1. Crick, F. H. C., J. Mol. Biol., 19, 548 (1966).
2. Jukes, T. H., Biochem. Biophys. Res. Comm., 24, 744 (1966).
3. Jukes, T. H., Biochem. Biophys. Res. Comm., 19, 391 (1965).
4. Jukes, T. H., Molecules and Evolution, (Columbia University Press, N. Y., 1966), p. 68.
5. Crick, F. H. C., Nature, 213, 119 (1967).
6. Zachau, H. G., Dütting, D. and Feldman, H., Angew. Chemie, Inter. Ed., 5, 422 (1966).
7. Tanaka, M., Nakashima, T., Benson, A., Mower, H. F. and Yasunobu, K. T., Biochem. Biophys. Res. Comm., 16, 422 (1964).
8. Benson, A. M., Mower, H. F. and Yasunobu, K. T., Proc. Natl. Acad. Sci., 55, 1532 (1967).
9. Eck, R. V. and Dayhoff, M. O., Science, 152, 363 (1966).
10. Jukes, T. H., loc. cit., p. 227.