J Mol Evol (1981) 18:15–17

Journal of Molecular Evolution

Amino Acid Codes in Mitochondria as Possible Clues to Primitive Codes

Thomas H. Jukes

Space Sciences Laboratory, University of California, Berkeley, California 94720, USA

Summary. Differences between mitochondrial codes and the universal code indicate that an evolutionary simplification has taken place, rather than a return to a more primitive code. However, these differences make it evident that the universal code is not the only code possible, and therefore earlier codes may have differed markedly from the previous code. The present universal code is probably a "frozen accident." The change in CUN codons from leucine to threonine (*Neurospora* vs. yeast mitochondria) indicates that neutral or near-neutral changes occurred in the corresponding proteins when this code change took place, caused presumably by a mutation in a tRNA gene.

Key words: Amino acid code – Evolution – Primitive codes – Mitochondria

The concept of a "doublet code," in which two nucleotides in the three per codon specify each amino acid, was first proposed by Roberts (1962a, 1962b). This was based on the early indication of degeneracy in the code. Roberts noted that a doublet code must contain apparently ambiguous words, because there are 20 amino acids and only 16 possible doublets. To resolve this problem partially, I proposed (Jukes 1965, 1966) that an archetypal doublet (two-out-of-three) code once existed, consisting of 16 quartets of triplets in which the third base was used synonymously and interchangeably, and that this code recognized only 15 amino acids. The proposal was encouraged by the circumstance that the present universal code still contains eight such quartets. The proposal included the suggestion that five amino acids were added subsequently, to complete the present universal code, by dividing several codon quartets into two pairs, each of which pair would code for a different amino acid. For example, the quartet CAN, suggested as being originally for histidine, would be divided into CAY, histidine, and CAR, glutamine.

The "two-out-of-three" proposal was revived by Mitra and co-workers (1977), who found that under conditions of in vitro protein synthesis, three *E. coli* valine tRNAs each recognized all four valine codons. Each of the three had a different anticodon. More recently, Lagerkvist (1981) has presented a generalized explanation of the pattern of the genetic code based on strong and weak interactions in pairs formed with anticodons by the bases in the first two positions of codons. This says there are three types of interactions – strong (two GC pairs), intermediate (one GC, one AU) and weak (two AU). All the quartets of type 1 (CCN, GCN, CGN, and GGN) are fourfold codons, which means that each quartet codes for one amino acid, and there are no fourfold codons of type 3 (UUN, AUN, UAN, and AAN). So far, so good, but the intermediate set does not follow any consistent pattern. For example, CUN is fourfold, but CAN codes for two amino acids. Also, ACN is fourfold, but AGN codes for two amino acids.

The tRNAs of mitochondria (human, yeast and *Neurospora*) are fewer in number than those in prokaryotes and in eukaryotic cytoplasmic systems. The allocation of tRNAs in the mitochondrial code is such that one tRNA translates all four codons of each of the eight fourfold sets. This is not a "two-out-of-three" pairing, but represents pairing of all three codon bases with an anti-codon, so that U in the first anti-codon position pairs loosely ("wobbles") with G, U and C (and pairs conventionally with A) in the third position of codons (Barrell et al. 1980; Heckman et al. 1980) (Tables 1 and 2).

CUN codons specify leucine in the universal code and also in mammalian and *Neurospora* mitochondrial codes,

1 N = A, G, C or U; Y = U or C; R = A or G

0022–2844/81/0018/0015/$ 01.00

Table 1. Codons, anti-codons and amino acids in the human mitochondrial code (after Barrell et al. 1980)

UUY	GAA	Phe				UAY	GUA	Tyr	UGY	GCA	Cys
			UCN	UGA	Ser						
UUR	UAA	Leu				UAR		CT	UGR	UCA	Trp
						CAY	GUG	His			
CUN	UAG	Leu	CCN	UGG	Pro				CGN	UCG	Arg
						CAR	UUG	Gln			
AUY	GAU	Ile				AAY	GUU	Asn	AGY	GCU	Ser
			ACN	UGU	Thr						
AUR*	UAU	Met				AAR	UUU	Lys	AGR		CT
						GAY	GUC	Asp			
GUN	UAC	Val	GCN	UGC	Ala				GGN	UCC	Gly
						GAR	UUC	Glu			

*AUG, CAU, Initiator Met

Table 2. Differences from human mitochondrial code in other mitochondrial codes

Species	Codons	Anti-codons	Amino acid
Yeast	CUN	UAG	Thr
Yeast	AUA	GAU	Ile
Yeast	AUG	CAU	Met
Yeast, *Neurospora*	AGR	UCU	Arg
Maize	CGG	CCG	Trp

but specify threonine in the yeast mitochondrial code. The "acquisition" of CUN by threonine in the yeast mitochondrial code is therefore an evolutionary event that is comparatively recent, because yeast and *Neurospora* belong to the same taxon, Ascomycetes, of eukaryotic fungi. The changed assignment of CUN to threonine must have taken place after yeast and *Neurospora* diverged from a common Ascomycete ancestor. Moreover, yeast mitochondria contain the gene for cytochrome oxidase which has protein sequence homology ranging up to 70% with human mitochondrial cytochrome oxidase (Bonitz et al. 1980a).

The fact that there are differences between mitochondrial codes (Table 1 and 2) and the universal code suggests that similar changes could quite probably have taken place in earlier codes, before the universal code was fixed. Such a conclusion weighs against the idea that the universal code is the "best of all possible codes," and came into being as a result of rigorous natural selection. The opposing viewpoint, that the universal code is a "frozen accident" is therefore favored.

The consensus is that differences from the universal code that are found in mitochondrial codes are not a return to an earlier, simpler and more primitive code. Instead, the differences represent an evolutionary simplification in which a minimum number of tRNAs has been conserved without compromising functional efficiency (Bonitz et al. 1980a). This simplification has evolved in order to lessen the amount of DNA assigned to genes for tRNAs, and thus to help minimize the size of the mitochondrial genome (Barrell et al. 1980). Therefore, the mitochondrial code is not a primitive survivor of a precursor of the universal code. These conclusions have been well summarized by Heckman et al. (1980).

It is also possible that some of the simplifications in the mitochondrial codes do indeed correspond to conditions that existed in an earlier coding system. I suggested (Jukes 1978) that AUA may have been an earlier codon for methionine. A specialized biochemical mechanism is needed to prevent the anti-codon UAU from pairing with AUG. A similar mechanism would be needed if UGA were a cysteine codon, to prevent UCA from recognizing UGG, but UGA became a stop codon.

Another feature of mitochondrial codes that appears primitive is the use of one tRNA for each codon quartet. Each such quartet pairs with a single anti-codon starting with U. The use of a single tRNA to translate four codons fits with a model of an earlier code containing fewer amino acids, each with four codons. Moreover, in Table 1, the first anticodon base is U for all the fourfold codons and for all twofold codons terminating with purines. Also, the only other first anti-codon base in Table 1 is G. This is a simpler system than the universal code in which I and C occur in the first position of many anti-codons. The universal code makes use of procedures that supply redundancy. For example, the valine codons GUU and GUC can each pair with two anti-codons, GAC and IAC. The codon GUA can pair with UAC and IAC, and GUG with UAC and CAC. All four of these anti-codons have been identified in tRNA molecules. In mitochondria, all four valine codons pair only with UAC, a simplification that is suggestive of an earlier and more primitive translation system.

The fidelity of codon/anti-codon pairing is enhanced by the finding of contacts between codon and anti-codon loops that are maximized by bringing these two loops together in the ribosome. Yarus and co-worders (1981) have shown that this contact process involves the three nucleotides in tRNA that are next (distal, 3') to the anti-codon. Modifications in the first of these nucleotides (the one adjoining the anti-codon) were shown by Nishimura (1972) often to be related to the identity of the third nucleotide of the anti-codon. Such modifications occur in some mitochondrial tRNAs, for example, t_6A adjoins the third base (U) of the threonine anti-codon in *Neurospora* tRNA (Heckman et al. 1980), but the function of mitochondrial ribosomes in enhancing fidelity of translation has not been explored.

Differences between the universal and mitochondrial codes serve to remove the concept that any changes in the genetic code would be lethal and would hence not persist. This concept has been a barrier to proposals for the existence of other codes than the universal code. The use of CUN (while retaining the regular ACN threonine codons) for threonine by yeast mitochondria (Bonitz et al. 1980b) shows a major change can occur in the code. Why was this change tolerated? Evidently, there were a number of sites in mitochondrial proteins where threonine could replace leucine. Bonitz et al. (1980a) list 14 CUN codons for threonine in a total of 1,217 codons in five yeast mitochondrial genes. There are also 62 ACN codons for threonine. Changes from leucine to threonine do not occur in known mutations because such changes necessitate a minimum of two nucleotide substitutions per codon. We must conclude that the switch from leucine to threonine at CUN sites in yeast mitochondrial genes, to be tolerated, must have resulted in neutral or near-neutral changes.

The same five genes also contain 164 UUA and 2 UUG codons for leucine, corresponding to 14% leucine. Whether this followed a replacement of some of the CUN codons by UUR, following the change in the code, and representing a resistance against substitution of leucine by threonine, is unknown.

It is commonly assumed that the universal code remains constant because any change in it, in any living organism, would result in numerous alterations of protein structure, and hence in impairment of protein function. Such changes are therefore rejected, but they can obviously be engendered by mutations in tRNA genes, either in the anti-codons or in the nucleotides responsible for recognizing the cognate amino acid to which the tRNA is esterified. Indeed, there is evidence that certain tRNAs have undergone such changes. For example, lysine and arginine tRNAs in yeast have similarities in sequence indicating recent divergence from a common ancestor (Holmquist et al. 1973). We interpreted this as resulting from gene duplication, followed by withdrawal of one of the duplicate genes from use until its amino acid recognition site and anti-codon had both changed, after which its function was restored. This series of steps avoided a change in the code.

The differences from the universal code recorded in mitochondria show that changes in tRNA genes must have taken place that altered their coding functions. Whether such differences exist in other organisms is not known, and, no doubt, will now be investigated as a result of the discoveries in mitochondrial codes.

It is noteworthy that the genome of mitochondria is much smaller (16, 569 base pairs) than that of complete organisms (Anderson et al. 1981). Therefore, mitochondria could probably tolerate changes in the code that would be lethal to a larger and more complex system. By the same token, the small primitive organisms that existed at an early period of evolution perhaps could accommodate changes in their genetic codes, and so the code did not become frozen until a level of organismal complexity had been reached that precluded such changes. It is also noteworthy that, in contrast to mitochondria, complete organisms often contain duplicate tRNA genes with the same anti-codon. This might enable such organisms to survive codon-changing mutations in one tRNA gene.

Acknowledgement. This work was supported by grant NGR 05–003–460 from the National Aeronautics and Space Administration. The collaboration of Carol Fegté in the preparation of the manuscript is acknowledged.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Nature 290:457–470

Barrell BG, Anderson S, Bankier AT, de Bruijn MHL, Chen E, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1980) Proc Natl Acad Sci USA 77:3164–3166

Bonitz SG, Coruzzi G, Thalenfeld BE, Tzagoloff A, Macino G (1980a) J Biol Chem 255:11927–11941

Bonitz SB, Berlani R, Coruzzi G, Li M, Macino G, Nobrega FB, Nobrega MP, Thalenfeld BE, Tzagoloff A (1980b) Proc Natl Acad Sci USA 77:3167–3170

Heckman JE, Sarnoff J, Alzner-DeWeerd B, Yin S, RajBhandary UL (1980) Proc Natl Acad Sci USA 77:3159–3163

Holmquist R, Jukes TH, Pangborn S (1973) J Mol Biol 78:91–116

Jukes TH (1965) Biochem Biophys Res Commun 19:391–396

Jukes TH (1966) Molecules and Evolution, 68, Columbia Univ. Press New York, 285 pp

Jukes TH (1978) Advances in Enzymology 47:375–432

Lagerkvist U (1981) Cell 23:305–306

Mitra SK, Lusty F, Akesson B, Lagerkvist U, Strid L (1977) J Biol Chem 252:471–478

Nishimura S (1972) Progress in Nucleic Acid Research Molecular Biology 12:49–85

Roberts RB (1962a) Proc Natl Acad Sci USA 48:897–900

Roberts RB (1962b) ibid. 1245–1250

Yarus M, McMillan C, Cline S, Bradley D, Snyder M (1980) Proc Natl Acad Sci USA 77:5092–5096

Received May 21, 1981/Revised September 5, 1981