Proc. Natl. Acad. Sci. USA
Vol. 77, No. 6, pp. 3164–3166, June 1980
Biochemistry

Different pattern of codon recognition by mammalian mitochondrial tRNAs

(mitochondrial genetic code/codon recognition/two out of three reading/U·N wobble)

B. G. BARRELL*, S. ANDERSON*, A. T. BANKIER*, M. H. L. DE BRUIJN*, E. CHEN*†, A. R. COULSON*, J. DROUIN*‡, I. C. EPERON*, D. P. NIERLICH*§, B. A. ROE*†, F. SANGER*, P. H. SCHREIER*¶, A. J. H. SMITH*, R. STADEN*, AND I. G. YOUNG*‖

*Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England; †Chemistry Department, Kent State University, Kent, Ohio 44242; ‡Department of Biochemistry and Biophysics, School of Medicine, University of California, San Francisco, California 94143; §Department of Microbiology, College of Letters and Science, University of California, Los Angeles, California 90024; ¶Institut für Genetik der Universität, 5 Köln 41, Weyertal 121, West Germany; and ‖ Department of Biochemistry, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra City, A.C.T. 2601, Australia

Contributed by Frederick Sanger, March 5, 1980

ABSTRACT **Analysis of an almost complete mammalian mitochondrial DNA sequence has identified 23 possible tRNA genes and we speculate here that these are sufficient to translate all the codons of the mitochondrial genetic code. This number is much smaller than the minimum of 31 required by the wobble hypothesis. For each of the eight genetic code boxes with four codons for one amino acid we find a single specific tRNA gene with T in the first (wobble) position of the anticodon. We suggest that these tRNAs with U in the wobble position can recognize all four codons in these genetic code boxes either by a "two out of three" base interaction or by U·N wobble.**

Mitochondrial DNA codes for a number of tRNAs as well as ribosomal RNAs and several respiratory proteins. Nineteen tRNAs have been mapped (by hybridization) on the human mitochondrial genome (1, 2) and mitochondrial tRNAs specific for all amino acids except L-proline, glutamine, histidine, and asparagine have been identified by charging experiments (1). Similar studies on rat liver mitochondrial tRNAs have identified all species except those for glutamine and histidine (3). Assuming that all mammalian mitochondria code for the same complement of tRNAs, these studies bring the number of mammalian mitochondrial tRNAs so far identified to 21, which includes two serine tRNAs, two leucine tRNAs, a $tRNA_m^{Met}$, and a $tRNA_f^{Met}$. Studies on purified rat liver mitochondria and cytoplasmic tRNAs have failed to show any importation of cytoplasmic tRNAs into the mitochondrion (3). The mitochondrial tRNAs had different chromatographic properties on RPC-5 columns and polyacrylamide gel electrophoresis from their cytoplasmic counterparts, and no mitochondrial tRNAs hybridized to nuclear DNA.

The small number of mammalian mitochondrial tRNAs and the lack of any imported cytoplasmic tRNAs raises the problem of how these tRNAs recognize all possible codons in mitochondrial protein synthesis. On the basis of the wobble hypothesis (4), a minimum of 32 tRNAs are needed to read all possible codons of the "universal" genetic code. Recent studies have established that human and bovine mitochondria have a different genetic code in that UGA is tryptophan and not termination and AUA is methionine and not isoleucine (ref. 5; I.G.Y., unpublished results). Thus, a minimum of 31 tRNAs would be needed.

However, recently a number of *in vitro* experiments have indicated that the wobble hypothesis might not apply and that in some cases the third (wobble) position of the codon has no discriminatory function. Mitra *et al.* (6) studied the incorporation of valine in an *in vitro* protein-synthesizing system dependent upon added valine tRNA and found that valine tRNAs with anticodons VAC (V, 5-oxyacetic uridine), GAC, and IAC could each recognize all four valine codons. From these and other results, Lagerkvist (7) proposed an alternative method for codon reading called the "two out of three" method. Whereas this mechanism could operate with impunity in the genetic code boxes with four codons for one amino acid (a codon family), it would cause misreading in the genetic code boxes with four codons for two amino acids (non-family boxes). However, if it were selectively used in the four-codon family boxes, there being a mechanism to prevent misreading in the non-family boxes, then a minimum of 24 tRNAs would be needed to decode the mammalian mitochondrial genetic code. This number is consistent with the number of tRNAs found in mammalian mitochondria, and we discuss here the evidence for such a mechanism operating in human and bovine mitochondria.

RESULTS AND DISCUSSION

We have obtained extensive sequence information on human mitochondrial DNA (approximately 97% of total sequence) by using the primed synthesis chain-termination method (8). The sequence was compiled by using the computer programs of Staden (9). Fragments of human mitochondrial DNA were cloned in pBR322 (10) and single-stranded template DNA was generated by the exonuclease III technique (11). Similarly, fragments of bovine mitochondrial DNA were cloned and sequenced by using the bacteriophage M13 vector (12). These fragments span the areas in human mitochondrial DNA where we have no sequence data. Because of the high homology between human and bovine mitochondrial DNA (unpublished data), the two sets of data were combined to produce a virtually complete mammalian mitochondrial DNA sequence with few remaining ambiguous sequences. In this sequence we have identified 23 tRNA genes by visual inspection and by using the program TRNA (13) which searches for the base-paired cloverleaf structures of tRNAs. The identification of two of these genes, those for $tRNA_m^{Met}$ (AUA/G) and $tRNA^{Ser}$ (AGU/C), are considered tentative. This number confirms the 21 tRNAs identified by earlier studies (1–3) and identifies the tRNAs for glutamine and histidine which were previously not found. If additional tRNA genes lie in those areas of the sequence where there are ambiguities, then one or two more genes at the most are considered possible. The identification of several of these genes has been confirmed by sequence studies on the bovine mitochondrial tRNAs, and these studies also show that there are only 22–24 unique tRNA sequences found in bovine mitochondria (B.A.R. and E.C., unpublished results). The sequences of these tRNAs and a discussion of their unusual structures will be published elsewhere. The tRNAs have been identified by their anticodon sequences as given in Table 1.

Table 1. Anticodon sequences of human mitochondrial tRNA genes

Anticodon	Codon	Amino acid
Family boxes (codons ending in N):		
TAG	CUN	Leu
TAC	GUN	Val
TGA	UCN	Ser
TGG	CCN	Pro
TGT	ACN	Thr
TGC	GCN	Ala
TCG	CGN	Arg
TCC	GGN	Gly
Non-family boxes (codons ending in Y):		
GAA	UUY	Phe
GAT	AUY	Ile
GTA	UAY	Tyr
GTG	CAY	His
GTT	AAY	Asn
GTC	GAY	Asp
GCA	UGY	Cys
GCT*	AGY	Ser
Non-family boxes (codons ending in R):		
TAA	UUR	Leu
TAT*	AUR	Met_m
TTG	CAR	Gln
TTT	AAR	Lys
TTC	GAR	Glu
TCA	UGR	Trp
Initiation codon:		
CAT	AUG	Met_f

Sequences were deduced from the DNA sequence and assigned to specific amino acids and codons by using the modified human and bovine mitochondrial genetic code (ref. 5; I.G.Y., unpublished results). No tRNA gene for Arg (AGR) has been found. Abbreviations: Y, pyrimidine; R, purine; N, any base.

* The identification of the tRNA genes for Met_m (AUR) and Ser (AGY) is considered tentative.

If we consider the eight boxes in the genetic code where a mechanism such as "two out of three" reading can operate without causing misreading—i.e., in the four-codon families—then a striking pattern of codon/anticodon relationships is seen. Only one tRNA for each such box is found and in each case, ignoring any possible modification, the tRNAs have U in the first position of the anticodon—i.e., the wobble position. On the basis of G·U wobble, these tRNAs should only recognize the bottom two codons in these boxes. The wobble hypothesis predicts that these boxes would be decoded by using G·U wobble as follows: the top two codons in the valine box (GUU and GUC) are recognized by the anticodon GAC and the bottom two codons (GUA and GUG) are recognized by the anticodon UAC. No mitochondrial tRNAs are found having G in the first position of the anticodon to recognize the top two codons. They are found, however, in the non-family boxes where the top two codons are used for a different amino acid than the one for the bottom two. We have found tRNAs with the predicted G·U wobble anticodons for all these boxes except arginine AGA/G, in addition to an AUG specific formylmethionine initiator tRNA. Here it is interesting to note that the two changes in the human mitochondrial genetic code are brought about by G·U wobble tRNAs for methionine codons AUA and AUG and tryptophan codons UGA and UGG.

Analysis of the codon usage in two well-characterized genes in human mitochondrial DNA, those for cytochrome oxidase subunits II and III (ref. 5; unpublished results), show that all codons are used in these genes except for Ser UCG, Lys AAG, Arg CGG, AGA, and AGG (Fig. 1). Therefore, tRNAs must exist to recognize the top two codons in the family boxes. As previously discussed, it is unlikely that the extra eight tRNAs are imported from the cytoplasm or coded for by the mitochondrial DNA. Thus, we conclude that all four codons in the family boxes are read by a tRNA with first position U in the anticodon. The predicted pattern of decoding is shown in Fig. 1. A similar conclusion has been reached for *Neurospora* and yeast mitochondria by Heckman *et al.* (14) and Bonitz *et al.* (15).

The decoding of these family boxes could operate by a two out of three base codon/anticodon pairing or the U in the first position of the anticodon could recognize any base in the third position of the codon. Whatever mechanism is being used, it must not occur in the non-family boxes; otherwise, misreading would occur. A discrimination mechanism is suggested by Heckman *et al.* (14) who studied *Neurospora crassa* mitochondrial tRNAs. *Neurospora* also codes for a limited number of tRNAs—less than the 32 needed to recognize all codons by G·U wobble—although nothing is known about the codon usage or the possible importation of cytoplasmic tRNAs into *N. crassa* mitochondria. Sequence analysis of several *N. crassa* mitochondrial tRNAs by Heckman *et al.* (14) reveals that a codon recognition pattern similar to that of human mitochondrial tRNAs is found and that the tRNAs for the family boxes have an unmodified U in the wobble position and the tRNAs specific for the bottom two codons in the nonfamily boxes have an unknown modified U in this position. Thus, the modified U could serve to prevent misreading of the top two codons in the non-family boxes, whereas the unmodified U could allow recognition of all four codons in the family boxes. The pattern of modification is not yet known for mammalian mitochondrial tRNAs but it would seem likely that a pattern similar to that described by Heckman *et al.* (14) for *N. crassa* will be found.

In this connection it is interesting that, in all published tRNA sequences (16) with U in the wobble position, the U is always found to be modified with the exception of yeast $tRNA^{Leu}_{UAG}$ (17–19). This tRNA has an unusual codon recognition pattern in that it can somehow recognize all six leucine codons. However, the situation in prokaryotes and in the cytoplasm of eukaryotes must be more complex with a greater number of tRNA species being used, and it is difficult to draw any analogy between the patterns of modification of these tRNAs and those of the mitochondria.

Does the U in the first position of the anticodon of the mitochondrial tRNAs specific for the family boxes pair with all four bases or is there a strict two out of three base codon/anticodon interaction? Crick (4) in his wobble hypothesis noted that U·U and U·C base pairs were possible but discounted them on the basis that they were too close and would cause misreading in the non-family boxes. Grosjean *et al.* (20) have examined the stability of a number of anticodon/anticodon interactions and found that the only stable nonwobble pairs involved U (or 2-thio derivatives of U) pairing with U and C. They also found by model building that U·U and U·C pairs are possible. Thus, U is the only base that can possibly form a stable base pair with A, G, C, or U and it is significant that all the tRNAs for the codon families have U in the wobble position. This would seem to suggest that this U wobbles with all four bases in the third position of the codon and that there is not a two out of three base interaction with the U remaining unpaired.

Finally, we have strongly suggestive evidence that a further simplification of the mammalian mitochondrial genetic code has occurred. The pattern of codon recognition shown in Fig.

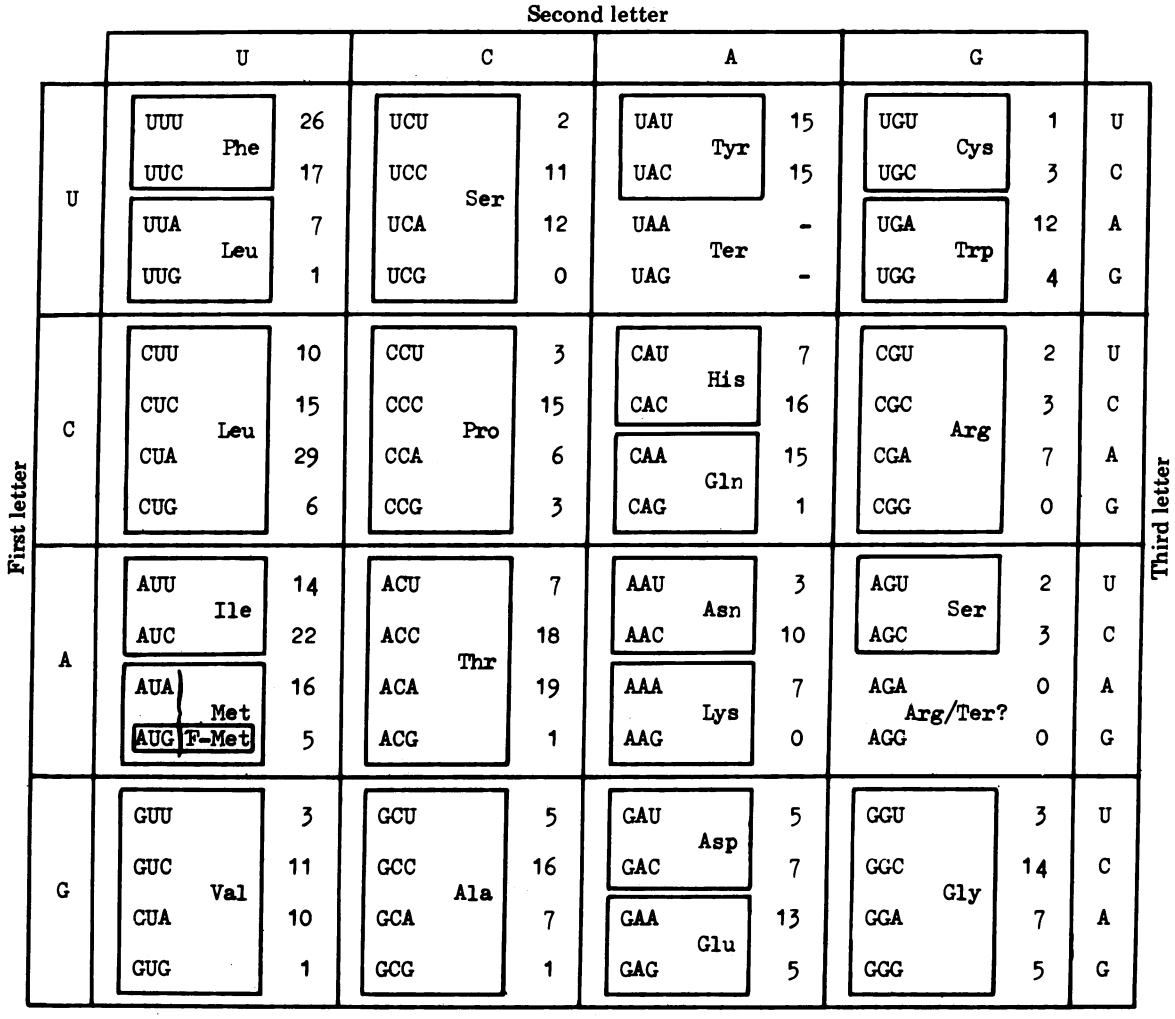

FIG. 1. Predicted pattern of codon recognition by mammalian mitochondrial tRNAs. Each internal box represents codons predicted to be recognized by a single tRNA using U·N wobble for the genetic code boxes with four codons for one amino acid and G·U wobble for the genetic code boxes with four codons for two amino acids. The number of codons found in the genes for cytochrome oxidase subunits II and III are listed alongside (ref. 5; unpublished results).

1 predicts that we should find a tRNA gene in the mitochondrial DNA sequence with a G·U wobble anticodon for the arginine codons AGA and AGG. No codons for arginine AGA and AGG are found in the genes for cytochrome oxidase subunits II and III, and an analysis of all significant reading frames in the rest of the sequence suggests that AGA and AGG are probably termination codons and not coding for arginine (unpublished results). Thus, mammalian mitochondria may only use 23 tRNAs to translate its genetic code.

We thank Dr. G. Attardi for a generous gift of HeLa cell mitochondria DNA. S.A. is a Research Fellow of the Cystic Fibrosis Foundation. B.A.R. was supported by a National Institutes of Health Research Career Development Award and P.H.S., by a European Molecular Biology Organization fellowship.

1. Lynch, D. C. & Attardi, G. (1976) *J. Mol. Biol.* **102**, 125–141.
2. Angener, L., Davidson, N., Murphy, W., Lynch, D. C. & Attardi, G. (1976) *Cell* **9**, 81–90.
3. Aujame, L. & Freeman, K. B. (1979) *Nucleic Acids Res.* **6**, 455–469.
4. Crick, F. H. C. (1966) *J. Mol. Biol.* **19**, 548–555.
5. Barrell, B. G., Bankier, A. T. & Drouin, J. (1979) *Nature (London)* **282**, 189–194.
6. Mitra, S. K., Lustig, F., Åkesson, B. & Lagerkvist, U. (1977) *J. Biol. Chem.* **252**, 471–478.
7. Lagerkvist, U. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1759–1762.
8. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
9. Staden, R. (1979) *Nucleic Acids Res.* **6**, 2601–2610.
10. Drouin, J. (1980) *J. Mol. Biol.*, in press.
11. Smith, A. J. H. (1979) *Nucleic Acids Res.* **6**, 831–848.
12. Schreier, P. H. & Cortese, R. (1979) *J. Mol. Biol.* **129**, 169–172.
13. Staden, R. (1980) *Nucleic Acids Res.* **8**, 817–825.
14. Heckman, J. E., Sarnoff, J., Alzner-DeWeerd, B., Yin, S. & RajBhandary, U. L. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3159–3163.
15. Bonitz, S. G., Berlini, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F. G., Nobrega, M. P., Thalenfeld, B. E. & Tzagoloff, A. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3167–3170.
16. Gaus, D. H., Grüter, F. & Sprinzl, M. (1979) *Nucleic Acids. Res.* **6**, Suppl. r1–r19.
17. Weissenbach, J., Dirheimer, G., Falcoff, R., Sanceau, J. & Falcoff, E. (1977) *FEBS Lett.* **82**, 71–76.
18. Randerath, K., Chia, L. S. Y., Gupta, R. C., Randerath, E., Hawkins, E. R., Brum, C. K. & Chang, S. H. (1975) *Biochem. Biophys. Res. Commun.* **63**, 157–163.
19. Randerath, E., Gupta, R. C., Chia, L. S. Y., Chang, S. H. & Randerath, K. (1979) *Eur. J. Biochem.* **93**, 79–94.
20. Grosjean, H. J., de Henau, S. & Crothers, D. M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 610–614.