Cell, Vol. 26, 167–180, October 1981 (Part 2), Copyright © 1981 by MIT

Sequence and Gene Organization of Mouse Mitochondrial DNA

Maureen J. Bibb, Richard A. Van Etten, Catharine T. Wright, Mark W. Walberg and David A. Clayton
Department of Pathology
Stanford University School of Medicine
Stanford, California 94305

Summary

The complete sequence of the 16,295 bp mouse L cell mitochondrial DNA genome has been determined. Genes for the 12S and 16S ribosomal RNAs; 22 tRNAs; cytochrome c oxidase subunits I, II and III; ATPase subunit 6; cytochrome b; and eight unidentified proteins have been located. The genome displays exceptional economy of organization, with tRNA genes interspersed between rRNA and protein-coding genes with zero or few noncoding nucleotides between coding sequences. Only two significant portions of the genome, the 879 nucleotide displacement-loop region containing the origin of heavy-strand replication and the 32 nucleotide origin of light-strand replication, do not encode a functional RNA species. All of the remaining nucleotide sequence serves a defined coding function, with the exception of 32 nucleotides, of which 18 occur at the 5′ ends of open reading frames. Mouse mitochondrial DNA is unique in that the translational start codon is AUN, with any of the four nucleotides in the third position, whereas the only translational stop codon is the orthodox UAA. The mouse mitochondrial DNA genome is highly homologous in overall sequence and in gene organization to human mitochondrial DNA, with the descending order of conserved regions being tRNA genes; origin of light-strand replication; rRNA genes; known protein-coding genes; unidentified protein-coding genes; displacement-loop region.

Introduction

The mammalian cell contains several thousand copies of a species-specific, closed circular mitochondrial genome that is autonomously replicated and transcribed within the organelle. Mitochondria maintain a complete protein-synthesizing system that is physically and genetically distinct from the cytoplasmic system. Several components of the system, including rRNA, tRNA and mRNA, are encoded by mitochondrial DNA, while others are nuclear gene products and must be imported into the organelle from the cytoplasm. The majority of proteins present in mitochondria are encoded by nuclear genes, but a small number of proteins are encoded by mitochondrial DNA and translated on mitochondrial ribosomes. Those mitochondrial translation products that have thus far been identified are components of enzyme complexes of the inner mitochondrial membrane that function in electron transport and oxidative phosphorylation.

The mouse L cell mitochondrial genome has been especially well characterized with respect to replication (Berk and Clayton, 1974, 1976; Bogenhagen and Clayton, 1977, 1978a, 1978b; Martens and Clayton, 1979; Gillum and Clayton, 1979; Bogenhagen et al., 1981) and transcription (Battey and Clayton, 1978; Van Etten et al., 1980, 1981). We report here the sequence of mouse L cell mitochondrial DNA as determined by the chemical sequencing method of Maxam and Gilbert (1980). An analysis of the sequence reveals several notable features unique to mouse mitochondrial DNA, as well as a striking conservation of overall gene organization as compared with the sequence of human mitochondrial DNA (Anderson et al., 1981).

Results and Discussion

Order of Functional Domains in the Genome

The 16,295 bp mouse mitochondrial DNA sequence is shown in Figure 1. Nucleotide 1 is the 5′ terminus of $tRNA^{Phe}$, and the numbering proceeds counterclockwise around the genome, as illustrated in Figure 2. The major coding strand of mouse mitochondrial DNA is the heavy (H) strand; most transcripts are therefore of light (L)- strand sequence and are complementary to the H strand. The sequence shown is L-strand sequence. The only sequences of significant length that do not code for RNA or protein are the displacement-loop region and the origin of L-strand synthesis, O_L (Figure 2). The remaining 94.4% of the genome reflects an essentially continuous coding function, with tRNA genes interspersed between the ribosomal RNA genes and protein-coding genes.

Displacement-Loop Origin of Heavy-Strand Synthesis

The 879 nucleotide displacement-loop region contains the origin of H-strand synthesis (O_H) and is defined as the sequence bounded by the genes for $tRNA^{Phe}$ (nucleotide 1) and $tRNA^{Pro}$ (nucleotide 15,416). The displacement-loop region is maintained as a triple-stranded structure by repeated synthesis and degradation of a family of short, single-stranded DNAs complementary to the L strand, with concomitant displacement of the parental H strand as a single-stranded loop. The 5′ and 3′ ends of these displacement-loop strands have been mapped (Gillum and Clayton, 1979; Doda et al., 1981), and their size distribution and relative stability are unusually complex for a single role in DNA replication (Bogenhagen and Clayton, 1978a). Because this region probably contains promoters for initiation of transcription of both the H and L strands, it is possible that displace-

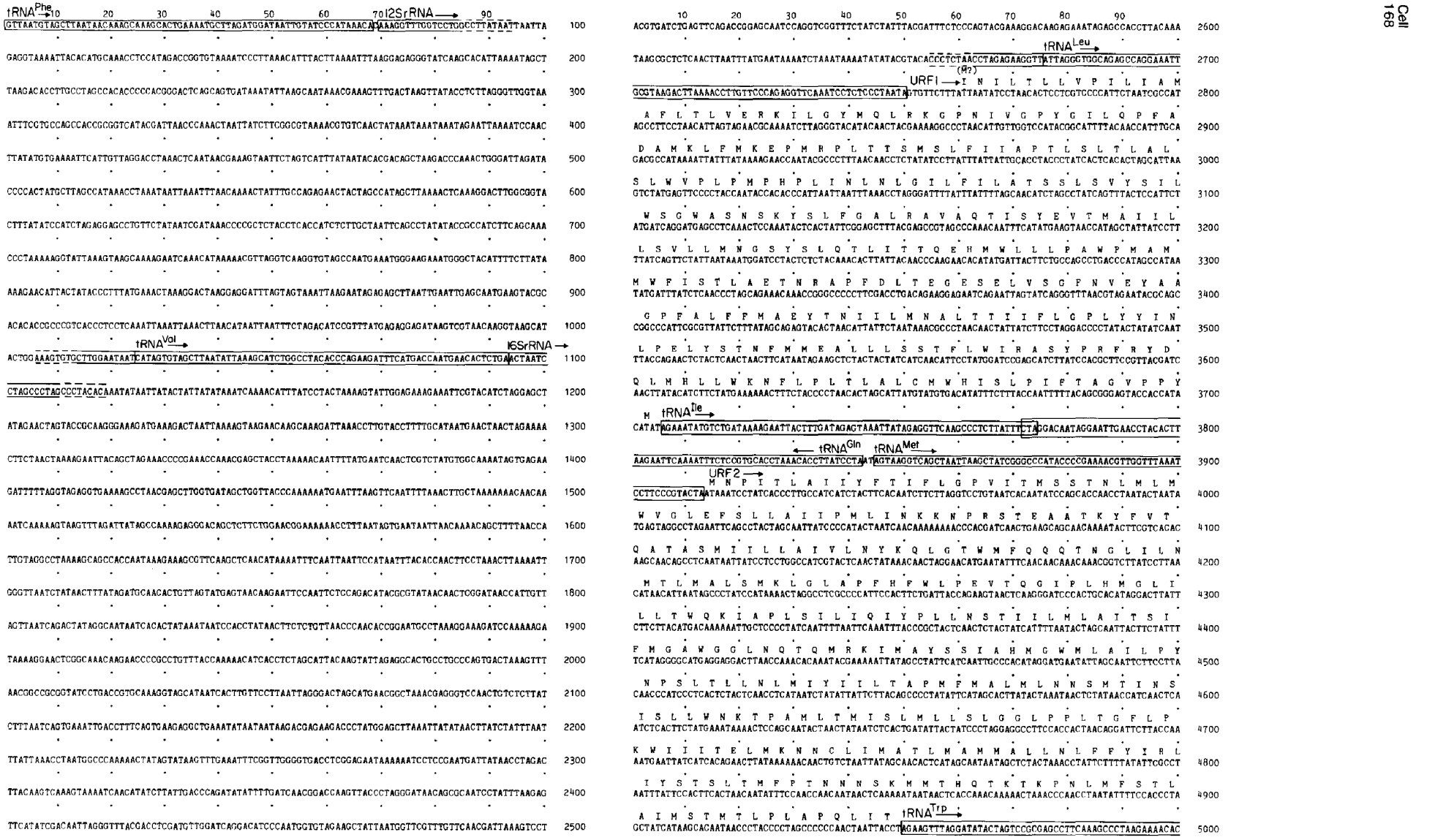

tRNA^Phe
12SrRNA
tRNA^Val
16SrRNA
tRNA^Leu
URF1
tRNA^Ile
tRNA^Gln
tRNA^Met
URF2
tRNA^Trp

Figure 1 (*continued overleaf*)


URF4
tRNA His
tRNA Ser
tRNA Leu
(M?) URF5
tRNA Glu
Cyt b
URF6

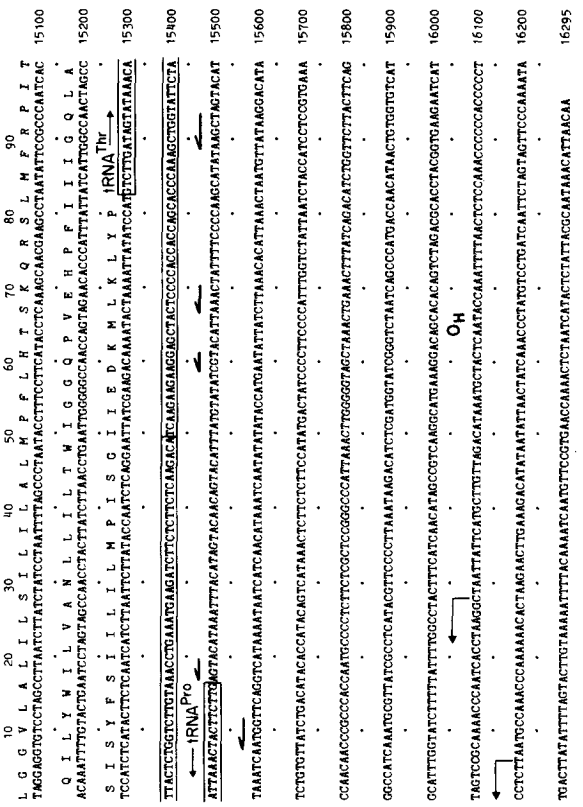


Figure 1. Complete DNA Sequence of Mouse L Cell Mitochondrial DNA

Sequence shown is that of the light (L) strand. Numbering begins at the first 5′ nucleotide of $tRNA^{Phe}$ and proceeds with increasing numbers through the rRNA genes. O_H and O_L: origins of heavy (H)- and L-strand mitochondrial DNA replication. Polarity of all genes is indicated by the arrows. 5′ and 3′ boundaries of tRNA genes have been designated such that the 5′ end of a gene begins with the first 5′ nucleotide in the 7 bp amino acid acceptor stem, while the 3′ end of a gene is the nucleotide immediately following the last paired 3′ nucleotide in this stem. tRNA genes are boxed, rRNA genes are boxed only at their 5′ and 3′ ends, and the identified and unidentified protein coding genes are translated according to the mitochondrial genetic code (see Table 1). The one-letter amino acid notation is used. ATT and ATC codons may code for methionine when used as initiation codons. TAA termination codons, when encoded in the DNA sequence, are indicated by END. In the displacement-loop region (nucleotides 15,417–16,295), 5′ ends of the major and the longest displacement-loop strand DNAs are indicated at positions 16,028 and 16,106, respectively (Gillum and Clayton, 1979). The various 3′ ends of the displacement-loop DNA species are likewise indicated by the half arrows at positions 15,417–15,419; 15,460–15,462; 15,467–15,470; 15,489–15,493; and 15,508–15,511 (Doda et al., 1981). The sequence of the rRNA gene region published previously (Van Etten et al., 1980) has been revised by deleting nucleotide 340 and by identifying the 3 undetermined nucleotides in that sequence. The complete L-strand sequence is 34.5% A, 28.7% T, 24.4% C and 12.3% G.

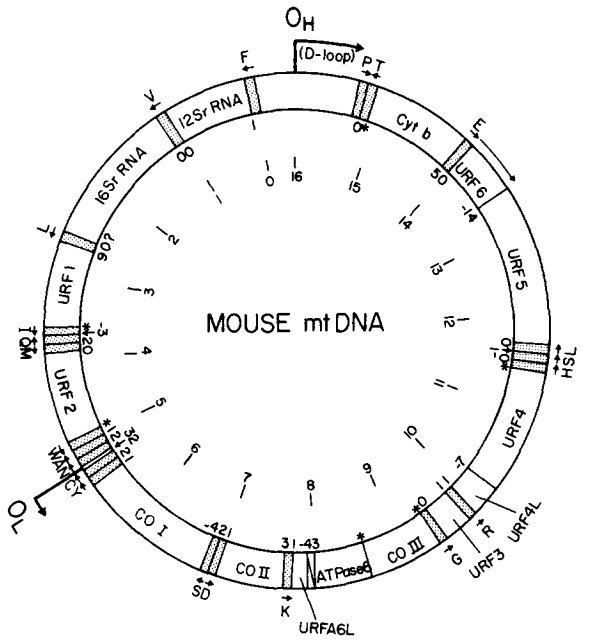
ment-loop metabolism is intimately involved in the control of gene expression. There are no open reading frames of significant size in the displacement-loop region, and it is therefore unlikely that this sequence serves a coding function. Consistent with this fact is the lack of any RNA species that can be identified as complementary to the region between $tRNA^{Phe}$ and $tRNA^{Pro}$ (Van Etten et al., 1981).

Origin of Light-Strand Synthesis

The origin of light-strand synthesis (O_L; nucleotides 5160–5191) is located among five tRNA genes one third of the way around the genome counterclockwise from the displacement loop (Figure 2). In contrast to O_H, O_L contains a simple major dyad symmetry with a perfect 12 bp stem and a 13 nucleotide loop at which the 5′ ends of nascent daughter L strands have been positioned (Martens and Clayton, 1979). This region of secondary structure is notable in that the stem portion is 58.3% G+C in a genome that is 36.7% G+C, while the loop portion contains 11 consecutive T bases in the template H strand. L-strand synthesis does not initiate at O_L until unidirectional expansion of a daughter H strand proceeds clockwise from O_H past this site on the genome (Berk and Clayton, 1974; Martens and Clayton, 1979). Thus the potential hairpin loop structure found at O_L may have biological significance, because the origin is not functional until exposed as a single-stranded template. The stem of the hairpin is immediately bounded by the genes for $tRNA^{Asn}$ and $tRNA^{Cys}$, which are encoded on the opposite L strand. In fact, the stem overlaps $tRNA^{Asn}$ and $tRNA^{Cys}$ by 1 and 4 nucleotides, respectively. Thus O_L represents a remarkable economy of DNA sequence utilization in a region otherwise devoted to tRNA coding. It is possible that the potential additional secondary structure that might be conferred by tDNA sequences is functionally important for the initiation of DNA synthesis at this origin.

Ribosomal RNA Genes

Mouse mitochondrial DNA codes for two rRNAs, 12S and 16S. No discrete equivalents of 5S or 5.8S rRNAs or genes for such rRNAs have been identified in mammalian mitochondria. The 5′ ends of the two rRNA genes have been precisely located by comparing the directly determined sequences of the rRNA termini to the mitochondrial DNA sequence (Van Etten et al., 1980). The 3′ terminus of the 12S rRNA yielded a sequence that was taken to indicate that the 3′-terminal nucleotide was a pyrimidine, corresponding to the T at position 1024 (Figure 1). Additional sequencing experiments have shown that these data were misinterpreted in that the majority of 12S rRNA terminates in GAAUAAUA–OH–3′ (J. W. Bird et al., manuscript in preparation). Since the next DNA template nucleotide following T is a C (position 1025), this suggests that the 3′-terminal A residue is added post-



MOUSE mt DNA

Figure 2. Organization of the Mouse Mitochondrial Genome

Distances (in kb) corresponding to the numbering of Figure 1 are shown on the inside of the circle. Shaded bars: tRNA genes, identified by the single-letter amino acid code, with polarity given by the arrows. All protein-coding genes are H-strand-encoded (with counterclockwise polarity), with the exception of URF6, which is L-strand-encoded, as indicated. COI, COII and COIII: cytochrome c oxidase subunits I, II and III. Cyt b: cytochrome b. URF: unidentified reading frame. The distance (in nucleotides) between two tRNA genes or between a tRNA gene and an initiation or termination codon is shown next to the junction of the genes just inside the circle; negative numbers indicate an overlap. Asterisk: no termination codon is encoded in the DNA for the preceding gene, and the termination codon is created by polyadenylation of the mRNA (see text). O_H and O_L: the H- and L-strand origins of DNA replication.

transcriptionally. The 3′ terminus of the 16S rRNA is the only mouse mitochondrial rRNA terminus that has not been directly sequenced. This gene boundary has been placed at either of two T residues just 5′ to the gene for $tRNA^{Leu}$ (Figure 1) based on a correspondence with the 3′-terminal nucleotide (U) of mouse 16S rRNA (Van Etten et al., 1980), and on S1 nuclease protection experiments that indicate that the space between the 16S rRNA gene and the $tRNA^{Leu}$ gene is at most a few nucleotides long (Nagley and Clayton, 1980). A comparison with the 3′-terminal sequence of the equivalent of 16S rRNA from hamster mitochondria is consistent with this assignment. However, the termini of a significant fraction of hamster 16S rRNAs are polyadenylated, and there is heterogeneity spanning a few nucleotides at the point in which the transition of encoded RNA sequence and poly(A) addition occurs (Dubin et al., 1981). Recent data indicate that at least some of the mouse 16S rRNA also contains 3′ A, but further work is required to establish the precise nature of the 3′ terminus. Although there are sequence homologies between mitochondrial rRNAs and procaryotic and eucaryotic rRNAs (Rubtsov et al., 1980), they do not approach the strikingly high level of 71%–74% homology between the rRNAs of corn chloroplast and E. coli (Schwarz and Kössel, 1980; Edwards and Kössel, 1981). Thus the sequence and structural homologies that do exist indicate that procaryotic, eucaryotic and mitochondrial genomes are related, but they do not provide a delineation of an evolutionary path for the mitochondrion.

Transfer RNA Genes

Twenty-two potential tRNA genes have been identified in the mouse mitochondrial DNA sequence by computer and visual search for characteristic dyad symmetries and certain conserved features of anticodon loops (Figure 1). Eight of the tRNA genes are located on the L strand of mouse mitochondrial DNA, while the other 14 are encoded on the H strand. The number, identity and map positions of the mouse mitochondrial tRNA genes are identical with the tRNA genes of human mitochondrial DNA. No other plausible tRNA genes can be found in the genome, and it is believed that the tRNAs transcribed from these 22 genes are sufficient to allow translation of all the mitochondrial-DNA-encoded proteins. With a single exception ($tRNA^{Ser}_{AGY}$), all the predicted mitochondrial tRNAs have the ability to form cloverleaf secondary structures. The tRNAs have a 7 bp amino acid acceptor stem, a 5 bp anticodon stem with an anticodon loop of 7, a "D" arm with a stem of 3–4 bp and a loop of 3–10, a variable arm of 4–5 nucleotides and a "TψC" arm with a stem of 4–5 bp and a loop of 4–8. The 3′-terminal oligonucleotide CCA, which has been found to be a common feature of all sequenced tRNAs, is not encoded in the DNA sequence and must be added posttranscriptionally.

The predicted sequences of the tRNAs are highly unusual in that they lack several features found in all other nonorganelle tRNAs. These deviations include the following (numbers refer to the yeast cytoplasmic $tRNA^{Phe}$ numbering system [Kim et al., 1974]): the D loop lacks the invariant $-G_{18}-G-$ sequence and usually has fewer nucleotides; the TψC loop lacks the universal sequence $-T_{54}-\psi-C-Pu-A-$, and varies in size from the standard 7 nucleotide length; there is no maintenance of constant nucleotides U_8, A_{14}, G_{15}, A_{21}, nor of the constant G–C pair adjacent to the TψC loop; and there is a higher proportion of mismatched bases in stem regions. The exceptions to the first three points are the mouse mitochondrial $tRNA^{Leu}_{UUR}$, $tRNA^{Ser}_{UCN}$ and $tRNA^{Gln}$, which have orthodox D and TψC loops and retain most of the conserved bases. Since these conserved features are generally bases that participate in noncovalent interactions between arms in the tertiary structure of conventional tRNAs, the mitochondrial tRNAs may not have as defined a three-dimensional conformation as do other tRNAs. The most striking example of this is $tRNA^{Ser}_{AGY}$, which is

smaller than the other tRNAs (59 nucleotides versus ~70) and, when positioned with the amino acid acceptor stem and anticodon arm in the usual configuration, is completely lacking a "D" arm. The same truncated $tRNA^{Ser}_{AGY}$ is found in human and bovine mitochondria, where it has been sequenced and shown to be chargeable with serine (Acari and Brownlee, 1980; de Bruijn et al., 1980).

Protein-Coding Genes

The gene for COII in mouse mitochondrial DNA has been identified by homology to the human gene (Barrell et al., 1979) and by comparison of the predicted mouse amino acid sequence with the known amino acid sequence of the bovine protein (Steffens and Buse, 1979). This comparison indicates that at least two changes in the standard genetic code have occurred in mouse mitochondrial DNA. The codon triplet UGA codes for tryptophan instead of functioning as a termination codon, and the codon AUA codes for methionine instead of isoleucine. Because these codon changes also exist in the human and bovine mitochondrial genetic codes (Anderson et al., 1981), they may be a common feature of mammalian mitochondrial systems.

Computer translation of both strands of the mouse mitochondrial genome in all possible phases has identified 13 potential open reading frames. All but one of these open reading frames is encoded on the H strand. Five of the potential protein-coding genes have been identified by homology with the corresponding human genes and by comparison with amino acid and nucleotide sequences for proteins from bovine and yeast mitochondria. The identified genes are for the three largest subunits of cytochrome c oxidase (COI, II and III), subunit 6 of the ATPase complex and cytochrome b (Cyt b) of the cytochrome bc_1 complex. The known genes are identified on the map in Figure 2; and corresponding genes exist in the same relative positions in human mitochondrial DNA (Anderson et al., 1981). The remaining eight open reading frames have not been identified on the basis of amino acid homology with any known protein, and we follow the terminology of Anderson et al. (1981) in designating them unidentified reading frames (URFs). These URFs are indicated in Figure 2, and the same eight URFs are found in the corresponding map positions in the human and bovine mitochondrial genomes. The maintenance of these conserved reading frames in intact form in all three species, in a genome that is apparently under great selective pressure to be as small as possible, suggests that these reading frames do represent genes for mitochondrial proteins. Consistent with this is the fact that polyadenylated mitochondrial transcripts corresponding to these URFs have been mapped to the mouse and human mitochondrial genomes (Battey and Clayton, 1978; Ojala et al., 1980; Van Etten et al., 1981).

Several of the URFs exist in the genome as overlapping reading frames. The 3′ end of the gene for URFA6L overlaps for 43 nucleotides with the gene for ATPase 6, which begins upstream from the URFA6L termination codon in a different reading frame on the same strand. The same is true for the URF4L and URF4 genes, except that the overlap is only 7 nucleotides. URF6 is the only non-tRNA gene encoded on the L strand of mouse mitochondrial DNA, and overlaps for 14 nucleotides at its 3′ end with the 3′ end of the URF5 gene, located on the opposite, H strand.

No other open reading frames longer than 70 amino acids are present anywhere in the mouse mitochondrial genome. Although several shorter open reading frames do exist, they are not considered significant because they are not conserved in the human mitochondrial genome.

Mechanism of Mitochondrial Translation

Table 1 shows the mouse mitochondrial genetic code with the pattern of codon utilization in the mitochondrial genome. According to the wobble hypothesis (Crick, 1966), the minimum number of tRNAs required to translate the classic genetic code is 32 using G:U or I:A/C/U wobble. In mouse mitochondrial DNA, as well as in the human and bovine mitochondrial genomes, the codons AGG and AGA, which code for arginine in the traditional genetic code, are never used in the middle of a reading frame, and no gene for a tRNA to decode these codons is found. Taking account of this and of the other changes in the mitochondrial genetic code reduces this theoretical number to 30, but no tRNA genes other than the 22 discussed above can be found in mouse mitochondrial DNA and no mammalian tRNAs appear to be imported from the cytoplasm to function in mitochondrial translation (Aujame and Freeman, 1979). Barrell et al. (1980) have proposed a mechanism by which the 22 human mitochondrial tRNAs are sufficient to translate all internal codons of mitochondrial reading frames, and it appears that the same mechanism is operating in mouse mitochondria. Under this proposal, each group of codons in the same box in Table 1 would be read by a single mitochondrial tRNA, whose anticodon is indicated (see Table 1).

Initiation Codons

The 5′ end regions of most of the mitochondrial protein-coding genes are shown in Figure 3. All the identified genes have an ATG initiation codon at the beginning of the reading frame, as do URFA6L, URF4L, URF4 and URF6. In the remaining URFs, the first in-phase ATG codon is many codons downstream in the potential reading frame, except for URF2 and URF3, which are completely lacking ATG codons. However, if the pattern of having an initiation codon within the first few nucleotides following a tRNA gene is maintained, then it is possible to write each of these

Table 1. Genetic Code and Codon Usage in Mouse Mitochondrial DNA

Phe	UUU	109	(GAA)	Ser	UCU	43	(UGA)	Tyr	UAU	65	(GUA)	Cys	UGU	11	(GCA)
	UUC	132			UCC	47			UAC	59			UGC	20	
Leu	UUA	131	(UAA)		UCA	148		TER	UAA	7		Trp	UGA	97	(UCA)
	UUG	16			UCG	4			UAG	0			UGG	7	
Leu	CUU	87	(UAG)	Pro	CCU	30	(UGG)	His	CAU	34	(GUG)	Arg	CGU	8	(UCG)
	CUC	65			CCC	34			CAC	63			CGC	18	
	CUA	266			CCA	139		Gln	CAA	79	(UUG)		CGA	36	
	CUG	26			CCG	2			CAG	3			CGG	3	
Ile	AUU	234	(GAU)	Thr	ACU	58	(UGU)	Asn	AAU	60	(GUU)	Ser	AGU	14	(GCU)
	AUC	140			ACC	80			AAC	108			AGC	35	
Met	AUA	218	(CAU)		ACA	157		Lys	AAA	100	(UUU)		AGA	0	
	AUG	31			ACG	6			AAG	2			AGG	0	
Val	GUU	53	(UAC)	Ala	GCU	47	(UGC)	Asp	GAU	31	(GUC)	Gly	GGU	36	(UCC)
	GUC	34			GCC	82			GAC	43			GGC	39	
	GUA	76			GCA	97		Glu	GAA	83	(UUC)		GGA	109	
	GUG	10			GCG	7			GAG	11			GGG	29	

The changes in the mouse mitochondrial genetic code from the universal code are that UGA codes for tryptophan and not for termination, and AUA codes for methionine and not isoleucine. AGG and AGA codons, which code for arginine in the universal genetic code, are never used in reading frames and no gene for a tRNA to decode them has been identified (see text). Each group of codons in the same box is read by a single tRNA whose anticodon, written 5′ to 3′, is in parentheses. Each four-codon box is read by a tRNA with a U in the first 5′ position of the anticodon, which occurs either by U:N wobble or by a "two-out-of-three" base-pairing mechanism (Lagerkvist, 1978). Two-codon boxes with codons ending in either U/C or A/G are read with G:U wobble by tRNAs, with G or U, respectively, in the first position of the anticodon. The number of $tRNA^{Met}$ species and their role in translation are not known and are discussed in the text. Total numbers of each type of codon found among the 13 open reading frames of mouse mtDNA are indicated; codons ending in A are used most frequently (45.5%), codons ending in C and U are used less frequently (26.2% and 24.1%, respectively) and codons ending in G are used rarely (4.1%).

unidentified reading frames as beginning with either an ATA, ATT or ATC initiation codon (underlined in Figure 3). Similar examples of putative ATA and ATT initiation codons are found in human and bovine mitochondrial DNA (Anderson et al., 1981), but no gene in either of these systems has an ATC initiator codon. Both URF3 and URF5 in human mitochondrial DNA have ATA codons immediately adjacent to 3′ ends of tRNA genes, exactly where the mouse genome has ATC codons. We therefore conclude that the generalized initiation codon in mouse mitochondria has the form ATN, with any of the four nucleotides permissible in the last position.

Since all the unorthodox initiation codons begin genes that code for unidentified proteins, it is not possible to determine from protein sequence data whether these proteins begin with methionine or isoleucine. However, two species of $tRNA^{Met}$ exist in mammalian mitochondria, methionyl-tRNA and N-formylmethionyl-tRNA (Lynch and Attardi, 1976; Aujame and Freeman, 1979), and the N terminus of the bovine COII protein has been shown to be N-formylmethionine (fMet) (Steffens and Buse, 1979). This suggests that fMet-tRNA may be the only aminoacyl-tRNA species used to initiate mitochondrial protein synthesis, and that the URF1, URF3 and URF5 polypeptides, if synthesized, probably begin with fMet. Only one $tRNA^{Met}$ gene has been identified in either mouse or human mitochondrial DNA. On the basis of its anticodon, CAT, and because of other structural features, Anderson et al. (1981) assigned the human $tRNA^{Met}$ gene as coding for $tRNA_f^{Met}$, the tRNA that participates in initiation of protein synthesis. All other eucaryotic initiator tRNAs have CAT anticodons and distinctive anticodon arms (Wrede et al., 1979), and the mouse mitochondrial $tRNA^{Met}$ shares these features with the human mitochondrial $tRNA^{Met}$. The CAT anticodon normally should be specific for the AUG codon, but because the initiation codon in mouse mitochondria is AUN, there is a requirement that the C in the anticodon interact with all four possible bases in the third position of the initiation codon. This would probably demand that this C residue in the $tRNA^{Met}$ be modified in a way that enables it to read all four initiation codons (Fukada and Abelson, 1980; Kuchino et al., 1980). From the genetic code pattern found in mammalian mitochondria, the $tRNA_m^{Met}$ species, which would code for a tRNA specific for internal AUG and AUA codons, would be expected to have UAU as an anticodon. Computer search of the mouse mitochondrial genome has failed to locate any convincing candidates for a $tRNA_m^{Met}$ gene. This suggests that there may be only one methionine tRNA gene, from which the two tRNA species are produced by differential modification of a primary transcript.

Termination Codons

The 3′ end regions of most of the mouse mitochondrial protein-coding genes are shown in Figure 4. Seven of these genes, COI, COII, URFA6L, URF3, URF4L, URF5 and URF6, have UAA termination codons encoded in the DNA sequence. The remaining six genes, URF1, URF2, ATPase 6, COIII, URF4 and Cyt b, do not have termination codons encoded in the DNA, but

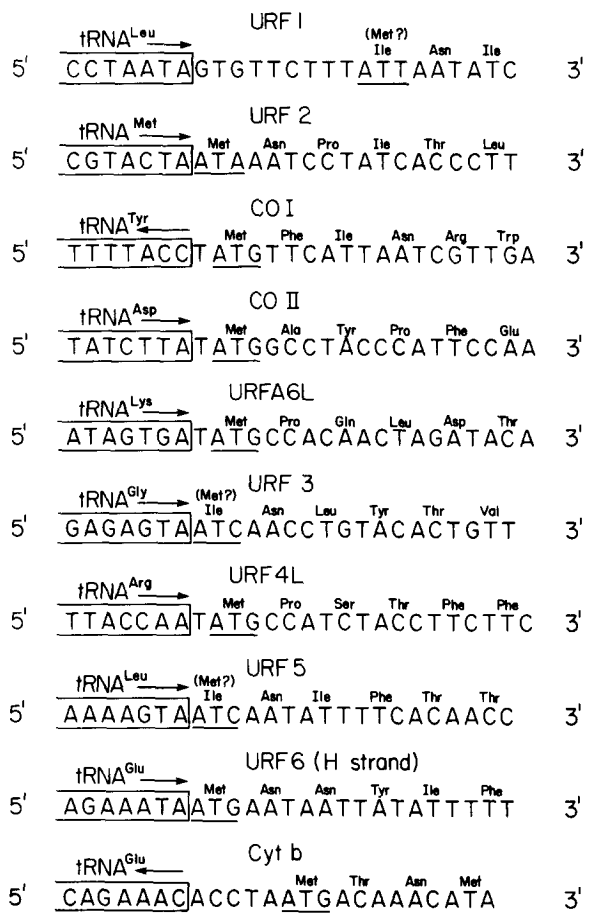


Figure 3. 5′ End Regions of Mouse Mitochondrial Protein-Coding Genes

DNA sequences in the regions of the 5′ ends of several of the mitochondrial open reading frames are shown. Putative initiation codons are underlined; ATT and ATC when used as initiation codons may code for methionine, not isoleucine (see text). All sequences are L-strand sequence, except for the URF6 5′ end, which is H-strand.

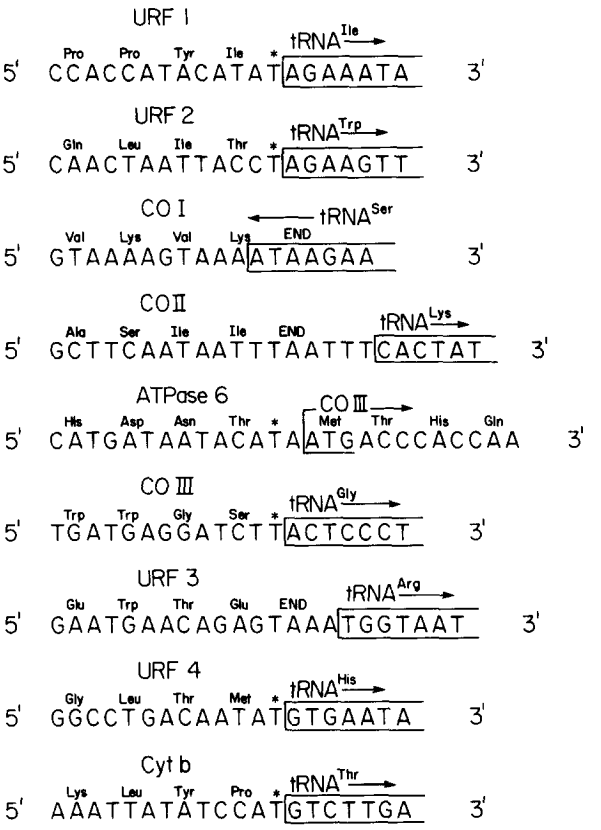


Figure 4. 3′ End Regions of Mouse Mitochondrial Protein-Coding Genes

DNA sequences in the regions of the 3′ ends of several of the mitochondrial open reading frames are shown. END: termination codons encoded in the DNA. Asterisk: polyadenylation of an RNA transcript at this nucleotide is required to create a termination codon. All sequence shown is L-strand sequence.

instead have a solitary T residue before the 5′ end of an adjacent tRNA gene or, in the case of ATPase 6, a TA before the 5′ end of the adjacent COIII gene. Since mouse mitochondrial mRNAs are polyadenylated (Battey and Clayton, 1978), cleavage of a primary transcript of one of these genes after the lone U followed by polyadenylation would create an in-phase termination codon for these transcripts. The same situation exists for several genes in human mitochondrial DNA (Anderson et al., 1981), and the polyadenylation model has been confirmed by direct sequence analysis of 3′ ends of human mitochondrial mRNAs (Ojala et al., 1981). In the human and bovine mitochondrial systems, AGG and AGA codons are found only at the end of reading frames and function as termination codons (Anderson et al., 1981), but these codons are never found at the end of reading frames in the mouse genome. The only termination codon utilized in mouse mitochondria is UAA.

Implications of Gene Organization for Transcription and Expression

Previous work detailing the organization of the rRNA region of the mouse mitochondrial genome showed exceptionally conservative utilization of DNA sequence in that the rRNA genes and associated tRNA genes were immediately adjacent to one another with zero or one intervening nucleotide (Van Etten et al., 1980). The same compact gene organization has now been shown to exist throughout the rest of the genome, with tRNA genes closely interspersed between structural genes.

Detailed transcriptional mapping of the mouse mitochondrial genome, by both S1 protection (Battey and Clayton, 1978) and RNA blotting (Van Etten et al., 1981) techniques, has demonstrated that every open reading frame on the H strand has a stable RNA transcript homolog, with no requirement or evidence for RNA splicing. The map positions of the transcripts are such that the lengths of the various RNAs, minus poly(A) tails, agree very well with the sizes of the reading frames. The three exceptions to this are the

transcripts for the overlapping reading frames of URFA6L/ATPase 6, URF4L/URF4 and URF5/URF6. The first two overlaps involve reading frames in different phases on the same strand of mitochondrial DNA, and in both cases the stable RNA that is observed begins at the 5′ end of the first reading frame and continues past the overlap region to end at the 3′ terminus of the second reading frame. The S1 mapping results clearly show that neither of these transcripts are spliced; there is therefore no small intervening sequence present in the DNA that is removed to create a single continuous reading frame in the RNA (Battey and Clayton, 1978). If both proteins of each pair of overlapping genes are produced, they must be synthesized from the same mRNA.

The other case of overlapping reading frames is between URF5, a reading frame of 607 codons on the H strand, and URF6, 172 codons on the L strand, whose 3′-terminal codons overlap each other by 14 nucleotides. In this case, a single transcript of 2350 nucleotides is found that spans from the 5′ end of URF5 through the overlap region and extends through the antisense sequences of URF6 up to the URF6 5′ terminus (Van Etten et al., 1981). No transcript corresponding to the URF6 coding region could be found, even in experiments that had the sensitivity to detect a transcript present at 1/100 the copy number of the H-strand-encoded species. The only stable L-strand-encoded transcripts that can be identified in mouse mitochondria are the eight tRNAs and a small, less abundant transcript that maps in the region of O_L, which may represent a precursor to the four clustered tRNA genes located there (Van Etten et al., 1981). Ojala et al. (1980) have mapped three relatively unstable L-strand-encoded transcripts in human mitochondria, all of which overlap the URF6 gene and could potentially serve as the mRNA for expression of URF6. No such transcripts have been detected in mouse mitochondria, although it may be that similar transcripts are present in very low quantities in the mouse system. It is possible that the URF6 gene is only expressed in certain developmental states or differentiated tissues, but there are no differences between the transcription pattern seen from mitochondria from cultured L cells and from mouse liver or kidney tissue (Van Etten et al., 1981).

Although the 5′ ends of the mouse mitochondrial mRNAs have not been determined directly, the transcript sizes are consistent with the notion of each RNA transcript beginning at or near the initiation codon of the gene. This is supported by the demonstration that in the human mitochondrial system most mRNAs have a 5′ end that maps at, or only a few nucleotides upstream from, the putative initiation codon (Montoya et al., 1981). Thus it appears that a general characteristic of mammalian mitochondrial mRNAs is the lack of any significant 5′ untranslated sequences. The longest potential 5′ leader sequence observed is only 9 nucleotides (URF1), which is probably insufficient to serve any role normally ascribed to such sequences. There is no feature common to the 5′ ends of the coding sequences that could be inferred as important to a translational recognition system or ribosome-binding site, nor are there any homologies to the 3′ end of the mitochondrial 12S small rRNA analogous to those found in procaryotic systems (Shine and Dalgarno, 1974; Steitz and Jakes, 1975). The mechanism of initiation of protein synthesis in mitochondria may involve a process whereby the small mitochondrial ribosomal subunit attaches to mRNAs by virtue of an affinity for 5′ ends, then moves downstream until the first AUN codon is encountered, at which point $tRNA_f^{Met}$ is bound and translation begins. However, this simple model cannot explain the process of initiation of synthesis of ATPase 6 or URF4 polypeptides, which must take place at internal AUG codons. It is possible that the secondary structure of these mRNAs is such as to permit utilization of only the correct AUG codon for internal initiation of translation; however, no simple conserved sequence or possible secondary structure is found in the region of the 5′ end of the internal reading frame in these messages. The fact that all identified genes in mouse, human and bovine mitochondrial DNA have orthodox ATG initiation codons is intriguing in view of the heterogeneity of start codons among the URFs. This raises the possibility that a translational control mechanism might operate at the level of selection of initiator codons for translation.

The most persuasive model of transcription of the H-strand genes of mouse mitochondrial DNA is one in which the entire coding region on the H strand is transcribed as a single polycistronic RNA beginning from a promoter located in the displacement-loop region. Individual rRNAs, tRNAs and mRNAs would then be generated by an RNA-processing activity that makes single endonucleolytic cleavages at the 5′ and 3′ borders of tRNA sequences, similar to the way the processing enzyme RNAase P generates the 5′ ends of tRNAs in E. coli (Lund and Dahlberg, 1977; Altman, 1978). Under this model, at least one cleavage (to generate the 5′ end of the Cyt b mRNA) would have to be made before or after an antisense tRNA sequence in the primary transcript. It is possible that the various sense and antisense tRNA sequences could fold into a tertiary structure capable of being recognized by the processing enzyme while still in the primary transcript, since it is known that RNAase P and other tRNA-processing enzymes in E. coli recognize features of secondary and tertiary structure rather than cutting at specific sequences (Altman, 1978). The boundary between the genes for ATPase 6 and COIII is the only one not demarcated by a tRNA gene, and it is not obvious how this specific cleavage

site is recognized by the processing enzyme(s). Although no tRNA-like dyad symmetries can be found in this region, some alternative secondary structure could exist that allows a processing activity to make the correct cleavage. In this regard, the junctional sequence does contain a dyad symmetry with the potential to form a structure with an 8 nucleotide stem and a 31 nucleotide loop, with the 5′ terminus of the COIII gene lying at the base of the stem on the 3′ side. It may be that the generation of the 3′ termini of the various mRNAs is closely coupled to the polyadenylation step, which is necessary to create termination codons for several of the RNAs. There is no obvious common sequence at or near the 3′ ends of the mRNAs that would represent a potential polyadenylation signal such as that found in nuclear-DNA-encoded mRNAs (Proudfoot and Brownlee, 1976).

Besides the overlapping reading frames discussed above, there are three other sequence overlaps in mouse mitochondrial DNA that involve tRNA genes. Two of these, a −3 overlap between genes for $tRNA^{Ile}$ and $tRNA^{Gln}$ and a −4 overlap between COI and $tRNA^{Ser}_{UCN}$, pose no conceptual difficulty because the two overlapping genes in each case are on opposite strands. The other overlap involves a single common nucleotide between the 3′ end of the $tRNA^{Ser}_{AGY}$ gene and the 5′ end of the $tRNA^{Leu}_{CUN}$ gene, which are on the same strand. A similar situation exists in human mitochondrial DNA for the genes for $tRNA^{Cys}$ and $tRNA^{Tyr}$. It is possible that this overlapping nucleotide is present in only one of the two mature tRNAs and not the other, or that a tRNA that lacks the nucleotide could be repaired posttranscriptionally. In the absence of repair, two transcripts of the region are required to produce one copy each of the functional tRNAs, a situation inconsistent with the otherwise striking economy of mitochondrial DNA. It is interesting that only 20% (49 of 242) of the serine codons in the open reading frames of mouse mitochondrial DNA require the $tRNA^{Ser}_{AGY}$ species to decode them; the other 80% are decoded by $tRNA^{Ser}_{UCN}$. The mitochondrion may therefore subsist with reduced molar amounts of $tRNA^{Ser}_{AGY}$ per transcription event, with the cleavage between the two tRNA sequences in the primary transcript occurring preferentially just 5′ to the A residue at nucleotide 11,665. This would yield a functional $tRNA^{Leu}_{CUN}$ and an incomplete $tRNA^{Ser}_{AGY}$.

Sequence Comparison to the Human Mitochondrial Genome

The mouse mitochondrial DNA sequence is directly comparable to the recently determined sequence of human mitochondrial DNA (Anderson et al., 1981). The overall order of gene organization is identical to that of human mitochondrial DNA, and the location of rRNA, tRNA and known protein-coding genes, as well as URFs, are at the same map positions in both systems. The human genome is 16,569 bp in length and the mouse is 16,295 bp. The 274 nucleotide difference in size is almost entirely due to the fact that the human displacement-loop region is 243 nucleotides larger than the mouse displacement-loop region, while the remaining 31 nucleotide difference is dispersed throughout the genome with no obvious functional implications.

The displacement-loop region in mouse and human mitochondrial DNA contains the origin of H-strand mitochondrial DNA synthesis and, most likely, the major promoters for both H- and L-strand transcription. The family of displacement-loop strands maintained in the D-loop structure is highly species-specific (Gillum and Clayton, 1978; Brown et al., 1978). This region of the mammalian mitochondrial genome is the most divergent in sequence, and the most striking lack of homology is at the 5′ and 3′ ends of the region. A comparison of the sequences of the displacement-loop regions of human, mouse and rat has revealed several blocks of conserved sequence located in the midregion of the displacement loop (Walberg and Clayton, 1981, Nucl. Acids Res. 9, in press). In addition, a short sequence in the L-strand template mapping at the 3′ ends of displacement-loop strands has been implicated as possibly serving a functional role in the arrest of displacement-loop-strand synthesis (Doda et al., 1981).

The origin of L-strand synthesis (O_L) was first identified in mouse mitochondrial DNA (Martens and Clayton, 1979), and was inferred to be in the same location in human mitochondrial DNA by the identification of an analogous and highly homologous potential hairpin structure in the human mitochondrial DNA sequence (nucleotides 5730–5763; Anderson et al., 1981). Direct mapping of the 5′ ends of human nascent daughter L strands has shown that this region is the origin (Tapper and Clayton, 1981). This impressive hairpin structure is a striking example of conservation of secondary structural properties with a 24% divergence in primary sequence. In contrast to O_H, it seems likely that O_L will prove to be highly conserved among mammalian mitochondrial genomes.

The 5′ ends of the human mitochondrial rRNA genes have been identified by the comparison of direct RNA sequence (Crews and Attardi, 1980) to the mitochondrial DNA sequence (Anderson et al., 1981). The 3′ ends of the genes have been assigned as immediately adjacent to tRNA genes by analogy to the mouse rRNA genes. The mouse and human 12S and 16S rRNAs are 75% and 74% homologous, respectively. There are domains of highly homologous sequences of lengths ranging between 10 and >100 nucleotides that are separated by regions of much lower homology. Stretches of mouse mitochondrial rRNA gene sequence that are longer than 20 nucleotides and that range between 20% and 46% homologous to human

mitochondrial rRNAs are nucleotides 128–179, 361–439, 797–816, 926–964, 1108–1144, 1232–1254, 1300–1329, 1765–1800, and 2181–2201 (Figure 1). The location of point mutations conferring resistance to chloramphenicol (Blanc et al., 1981; Kearsey and Craig, 1981) is within one of the most highly homologous stretches of the 16S rRNAs. On the 3′ side of this mutation site is the longest stretch of perfect homology between the mouse and human 16S rRNAs, the 52 nucleotides from 2473 to 2524 (Figure 1). Interestingly, the last 75 nucleotides of the mouse 16S rRNA gene are only 56% homologous to the human sequence.

The mouse mitochondrial tRNA genes are more homologous to their human counterparts than are the rRNA genes, the known protein-coding genes or the URFs. The average nucleotide conservation among the tRNAs is 80%, ranging from the $tRNA^{Gly}$ gene, which is least homologous at 68%, to the $tRNA^{Met}$ gene, which is most homologous at 90%. In terms of structural domains, the anticodon loop is the most strongly conserved, followed by the anticodon and amino acid acceptor stems and the stem to the "D" arm. Most divergent are the D loop, and especially the TψC arm and loop.

A comparison of the known and unidentified protein-coding genes is given in Table 2. Within the family of known genes, the amino acid conservation ranges from 71% (COII) to 90% (COI), while nucleotide conservation ranges from 71% (COII) to 77% (COI). Except for the COII gene, the percentage of nucleotide conservation is less than that for the amino acids, which reflects the degeneracy of the genetic code. The fraction of total nucleotide changes that do not change the amino acid inserted at that position varies from 43% (COII) to 74% (COI). Among the known genes, the fraction of base changes that are silent correlates with the level of amino acid conservation. Overall, the known protein-coding genes have 82% amino acid conservation and 74% nucleotide conservation between mouse and human mitochondrial DNA. The unidentified genes show less conservation of amino acid and nucleotide sequence, with 64% and 67% amino acid and nucleotide homology to human URFs, respectively (Table 2). The family of URFs does not follow an ordered pattern of increased amino acid conservation relative to nucleotide conservation and, in general, the fraction of nucleotide changes that are silent is significantly less than for the five known genes.

Because the mouse URFs are in general less conserved than the identified genes and utilize a different set of translational start codons, it might be argued that they represent a functionally different class of mitochondrial genes. However, it is very unlikely that they do not represent true genes. They do show significant homology to the corresponding URFs in human mitochondrial DNA (Table 2), and they have maintained a single open reading frame over their entire sequence. The overlap region of 43 nucleotides between URFA6L and ATPase 6 shows a significantly higher degree of nucleotide conservation between

Table 2. Homology of Protein-Coding Genes Between Mouse and Human Mitochondrial DNA

Gene	Amino Acid Homology (%) (Mouse/ Human)	Nucleotide Homology (%) (Mouse/ Human)	Silent Base Changes (%)	No. of Amino Acids (Mouse)	Predicted Molecular Weight (Mouse)	Initiation/Termination Codons[a]		
						Mouse	Human	Bovine
COI	90	77	74	514	56,900	ATG/TAA	ATG/AGA	ATG/TAA
COIII	86	76	67	261	29,900	ATG/ *	ATG/ *	ATG/ *
Cyt b	78	74	48	381	43,200	ATG/ *	ATG/ *	ATG/AGA
ATPase 6	75	72	48	226	25,100	ATG/ *	ATG/ *	ATG/ *
COII	71	71	43	227	26,000	ATG/TAA	ATG/TAG	ATG/TAA
URF1	78	71	49	315	35,600	ATT/ *	ATA/ *	ATG/ *
URF4L	67	69	39	97	10,500	ATG/TAA	ATG/TAA	ATG/TAA
URF4	67	68	38	459	51,900	ATG/ *	ATG/ *	ATG/ *
URF3	65	66	35	114	13,100	ATC/TAA	ATA/ *	ATA/ *
URF5	63	65	31	607	68,400	ATC/TAA	ATA/TAA	ATA/TAA
URF2	57	61	26	345	38,800	ATA/ *	ATT/ *	ATA/ *
URF6	52	63	19	172	18,600	ATG/TAA	ATG/AGG	ATG/TAA
URFA6L	46	65	14	67	7,800	ATG/TAA	ATG/TAG	ATG/TAA

Shown are percentages of amino acids and nucleotides conserved between the species, and the percentage of the total number of nucleotide changes that are "silent," in that they do not affect the amino acid that would be incorporated at that position. The comparisons are separated into two classes, known genes and URFs, and are ordered with respect to decreasing homology within each class.

[a] Data on human and bovine initiation and termination codons are from Anderson et al. (1981). Asterisk: a termination codon is created by polyadenylation.

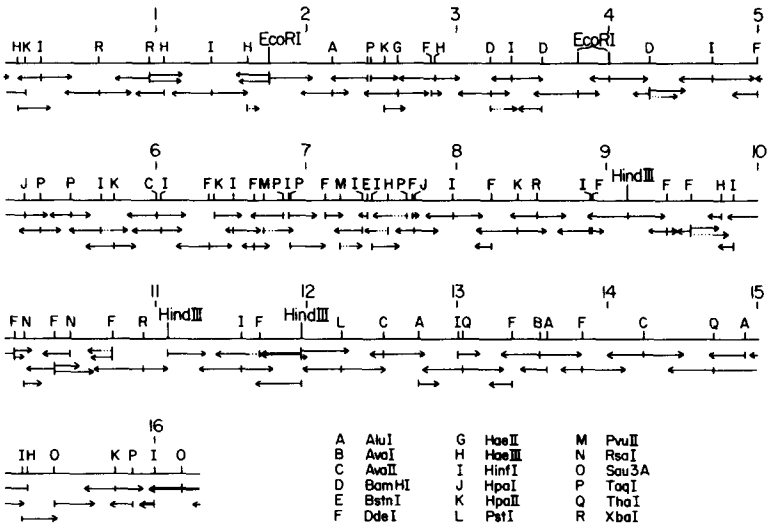


Figure 5. Sequencing Strategy for the Mouse Mitochondrial Genome

The restriction endonuclease cleavage sites used in determining the genomic sequence are shown; they do not represent a complete restriction map. Numbers: sequence length (in kilobases), corresponding to the numbering of Figures 1 and 2. Arrows: regions of determined sequence (vertical bar: labeled end). Dotted lines: portions of a particular sequence determination not analyzed. Heavy arrows: regions of the genome for which L cell mitochondrial DNA was used in sequence determination.

human and mouse than either of these genes do alone. This would be expected of a region that has overlapping genes in different reading frames producing two functional proteins, because such a region no longer has the freedom to acquire "silent" third-position changes in codons. Finally, although the overall nucleotide and amino acid homology between the mouse and human URFs is relatively low, the divergence between the two families of genes is decidedly nonrandom, in that numerous blocks of sequence can be identified, particularly in the middle of genes, that have a much higher degree of homology than that of the gene as a whole. The most extreme examples of this are found in URF5, where 38% of the 78 N-terminal and 26% of the 19 C-terminal amino acids are conserved while the remaining amino acids are 69% conserved, and in URF6, where 29% of the 21 N-terminal and 37% of the 88 C-terminal amino acids are conserved while 79% of the remaining amino acids are conserved. This pattern suggests that certain regions of the URF polypeptides, perhaps functional domains, are under strong selection, while the remaining regions are free to undergo more extensive divergence. Elucidation of the identities of these proteins will be a major step in the understanding of the biogenesis and function of mammalian mitochondria.

Experimental Procedures

Source of DNA

Recombinant plasmids containing Hind III–Eco RI fragments of mouse mitochondrial DNA were propagated under P2-EK1 conditions and isolated from the E. coli SF8-C600 host as described by Battey and Clayton (1978). Mitochondrial DNA was isolated from the LA9 derivative of mouse L cells as described by Bogenhagen and Clayton (1974).

DNA Sequencing

The chemical method of Maxam and Gilbert (1980) was used. Most restriction sites utilized for end labeling were overlapped by sequencing from a different site (Figure 5). Mouse tissue-culture mitochondrial DNA was sequenced in addition to cloned mitochondrial DNA in many instances, and no sequence differences between the L cell DNA and cloned DNA were found. Particular care was taken to identify all Eco RII restriction endonuclease recognition sites, at which the cloned DNA used has methylated cytosine residues that are inert to cleavage in the sequencing reaction (Maxam and Gilbert, 1980). Digestion of the cloned mitochondrial DNA with Bst NI, an isoschizomer of Eco RII that cuts methylated DNA, yielded only the fragments expected from the DNA sequence.

Acknowledgments

We thank J. N. Doda, D. P. Tapper, L. S. Kaguni, N. L. Michael and A. Brennicke for helpful discussions. We acknowledge the Stanford Molgen Project and the NIH SUMEX-AIM computer facility for use of facilities required for analysis of DNA sequences. This investigation was supported by grants from the NIH and American Cancer Society, Inc. R. A. V. and M. W. W. are Medical Scientist Training Program Trainees (NIH), and D. A. C. is a Faculty Research Awardee (ACS).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 27, 1981

References

Acari, P. and Brownlee, G. G. (1980). The nucleotide sequence of a small (3S) seryl-tRNA (anticodon GCU) from beef heart mitochondria. Nucl. Acids Res. *8*, 5207–5212.

Altman, S. (1978). Transfer RNA biosynthesis. In Biochemistry of Nucleic Acids II, *17*, B. F. Clark, ed. (Baltimore: University Park Press), pp. 19–44.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. and Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457–465.

Aujame, L. and Freeman, K. B. (1979). Mammalian mitochondrial transfer RNAs: chromatographic properties, size and origin. Nucl. Acids Res. *6*, 455–470.

Barrell, B. G., Bankier, A. T. and Drouin, J. (1979). A different genetic code in human mitochondria. Nature *282*, 189–194.

Barrell, B. G., Anderson, S., Bankier, A. T., de Bruijn, M. H. L., Chen, E., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. and Young,

I. G. (1980). Different pattern of codon recognition by mammalian mitochondrial tRNAs. Proc. Nat. Acad. Sci. USA *77*, 3164–3166.

Battey, J. and Clayton, D. A. (1978). The transcription map of mouse mitochondrial DNA. Cell *14*, 143–156.

Berk, A. J. and Clayton, D. A. (1974). Mechanism of mitochondrial DNA replication in mouse L-cells: asynchronous replication of strands, segregation of circular daughter molecules, aspects of topology and turnover of an initiation sequence. J. Mol. Biol. *86*, 801–824.

Berk, A. J. and Clayton, D. A. (1976). Mechanism of mitochondrial DNA replication in mouse L-cells: topology of circular daughter molecules and dynamics of catenated oligomer formation. J. Mol. Biol. *100*, 85–102.

Blanc, H., Wright, C. T., Bibb, M. J., Wallace, D. C. and Clayton, D. A. (1981). Mitochondrial DNA of chloramphenicol-resistant mouse cells contains a single nucleotide change in the region encoding the 3′ end of the large ribosomal RNA. Proc. Nat. Acad. Sci. USA *78*, 3789–3793.

Bogenhagen, D. and Clayton, D. A. (1974). The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. J. Biol. Chem. *249*, 7991–7995.

Bogenhagen, D. and Clayton, D. A. (1977). Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. Cell *11*, 719–727.

Bogenhagen, D. and Clayton, D. A. (1978a). Mechanism of mitochondrial DNA replication in mouse L-cells: kinetics of synthesis and turnover of the initiation sequence. J. Mol. Biol. *119*, 49–68.

Bogenhagen, D. and Clayton, D. A. (1978b). Mechanism of mitochondrial DNA replication in mouse L-cells: introduction of superhelical turns into newly replicated molecules. J. Mol. Biol. *119*, 69–81.

Bogenhagen, D., Lowell, C. and Clayton, D. A. (1981). Mechanism of mitochondrial DNA replication in mouse L-cells: replication of unicircular dimer molecules. J. Mol. Biol. *148*, 77–93.

Brown, W. M., Shine, J. and Goodman, H. M. (1978). Human mitochondrial DNA: analysis of 7S DNA from the origin of replication. Proc. Nat. Acad. Sci. USA *75*, 735–739.

Crick, F. H. C. (1966). Codon–anticodon pairing: the wobble hypothesis. J. Mol. Biol. *19*, 548–555.

Crews, S. and Attardi, G. (1980). The sequences of the small ribosomal RNA gene and the phenylalanine tRNA gene are joined end to end in human mitochondrial DNA. Cell *19*, 775–784.

de Bruijn, M. H. L., Schreier, P. H., Eperon, I. C. and Barrell, B. G. (1980). A mammalian mitochondrial serine transfer RNA lacking the "dihydrouridine" arm. Nucl. Acids Res. *8*, 5213–5222.

Doda, J. N., Wright, C. T. and Clayton, D. A. (1981). Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Nat. Acad. Sci. USA *78*, in press.

Dubin, D. T., Timko, K. D. and Baer, R. J. (1981). The 3′ terminus of the large ribosomal subunit ("17S") RNA from hamster mitochondria is ragged and oligoadenylated. Cell *23*, 271–278.

Edwards, K. and Kössel, H. (1981). The rRNA operon from Zea mays chloroplasts: nucleotide sequence of 23S rRNA and its homology with E. coli 23S rDNA. Nucl. Acids Res. *9*, 2853–2869.

Fukada, K. and Abelson, J. (1980). DNA sequence of a T4 transfer RNA gene cluster. J. Mol. Biol. *139*, 377–391.

Gillum, A. M. and Clayton, D. A. (1978). Displacement-loop replication initiation sequence in animal mitochondrial DNA exists as a family of discrete lengths. Proc. Nat. Acad. Sci. USA *75*, 677–681.

Gillum, A. M. and Clayton, D. A. (1979). Mechanism of mitochondrial DNA replication in mouse L-cells: RNA priming during the initiation of heavy-strand synthesis. J. Mol. Biol. *135*, 353–368.

Kearsey, S. E. and Craig, I. W. (1981). Altered ribosomal RNA genes in mitochondria from mammalian cells with chloramphenicol resistance. Nature *290*, 607–608.

Kim, S.-H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A., Seeman, N. C. and Rich, A. (1974). Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science *185*, 435–440.

Kuchino, Y., Watanabe, S., Harada, F. and Nishimura, S. (1980). Primary structure of AUA-specific isoleucine transfer ribonucleic acid from *Escherichia coli*. Biochemistry *19*, 2085–2089.

Lagerkvist, U. (1978). "Two out of three": an alternative method for codon reading. Proc. Nat. Acad. Sci. USA *75*, 1759–1762.

Lund, E. and Dahlberg, J. E. (1977). Spacer transfer RNAs in ribosomal RNA transcripts of E. coli: processing of 30S ribosomal RNA in vitro. Cell *11*, 247–262.

Lynch, D. C. and Attardi, G. (1976). Amino acid specificity of the transfer RNA species coded for by HeLa cell mitochondrial DNA. J. Mol. Biol. *102*, 125–141.

Martens, P. A. and Clayton, D. A. (1979). Mechanism of mitochondrial DNA replication in mouse L-cells: localization and sequence of the light-strand origin of replication. J. Mol. Biol. *135*, 327–351.

Maxam, A. M. and Gilbert, W. (1980). Sequencing end-labeled DNA with base specific chemical cleavages. Meth. Enzymol. *65*, 449–560.

Montoya, J., Ojala, D. and Attardi, G. (1981). Distinctive features of the 5′-terminal sequences of the human mitochondrial mRNAs. Nature *290*, 465–470.

Nagley, P. and Clayton, D. A. (1980). Transcriptional mapping of the ribosomal RNA region of mouse L-cell mitochondrial DNA. Nucl. Acids Res. *8*, 2947–2965.

Ojala, D., Merkel, C., Gelfand, R. and Attardi, G. (1980). The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell *22*, 393–403.

Ojala, D., Montoya, J. and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature *290*, 470–474.

Proudfoot, N. J. and Brownlee, G. G. (1976). 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature *263*, 211–214.

Rubtsov, P. M., Musakhanov, M. M., Zakharyev, V. M., Krayev, A. S., Skryabin, K. G. and Bayev, A. A. (1980). The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from Saccharomyces cerevisiae^{+}. Nucl. Acids Res. *8*, 5779–5794.

Schwarz, Zs. and Kössel, H. (1980). The primary structure of 16S rDNA from Zea mays chloroplast is homologous to E. coli 16S rRNA. Nature *283*, 739–743.

Shine, J. and Dalgarno L. (1974). The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Nat. Acad. Sci. USA *71*, 1342–1346.

Steffens, G. J. and Buse, G. (1979). Studies on cytochrome *c* oxidase. IV. Primary structure and function of subunit II. Hoppe-Seyler's Z. Physiol. Chem. *360*, 613–619.

Steitz, J. A. and Jakes, K. (1975). How ribosomes select initiator regions in mRNA: base pair formation between the 3′ terminus of 16S rRNA and mRNA during initiation of protein synthesis in *Escherichia coli*. Proc. Nat. Acad. Sci. USA *72*, 4734–4738.

Tapper, D. P. and Clayton, D. A. (1981). Mechanism of replication of human mitochondrial DNA. Localization of the 5′ ends of nascent daughter strands. J. Biol. Chem. *256*, 5109–5115.

Van Etten, R. A., Walberg, M. W. and Clayton, D. A. (1980). Precise localization and nucleotide sequence of the two mouse mitochondrial rRNA genes and three immediately adjacent novel tRNA genes. Cell *22*, 157–170.

Van Etten, R. A., Michael, N. L., Bibb, M. J., Brennicke, A. and Clayton D. A. (1981). Expression of the mouse mitochondrial DNA genome. In Mitochondrial Genes, G. Attardi, P. Borst and P. Slonimski, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), in press.

Wrede, P., Woo, N. H. and Rich, A. (1979). Initiator tRNAs have a unique anticodon loop conformation. Proc. Nat. Acad. Sci. USA *76*, 3289–3293.