J Mol Evol (1989) 28:497–516

Journal of
Molecular Evolution

The Complete Nucleotide Sequence of the *Rattus norvegicus* Mitochondrial Genome: Cryptic Signals Revealed by Comparative Analysis between Vertebrates

G. Gadaleta,[1] G. Pepe,[1] G. De Candia,[1] C. Quagliariello,[2] E. Sbisà,[1] and C. Saccone[1]

[1] Centro di Studio sui Mitocondri e Metabolismo Energetico, CNR Bari, Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Italy
[2] Dipartimento di Biologia Cellulare, Università della Calabria, Cosenza, Italy

Summary. This paper reports the nucleotide sequence of rat mitochondrial DNA, only the fourth mammalian mitochondrial genome to be completely sequenced. Extensive comparative studies performed with similar genomes from other organisms revealed a number of interesting features.

1) Messenger RNA genes: the codon strategy is mainly dictated by the base compositional constraints of the corresponding codogenic DNA strand. The usage of the initiation and termination codons follows well-established rules. In general the canonical initiator, ATG, and terminators, TAA and TAG (in rat, only TAA), are always present when there is gene overlapping or when the mRNAs possess untranslated nucleotides at the 5′ or 3′ ends.

2) Transfer RNA genes: a number of features suggest the peculiar evolutionary behavior of this class of genes and confirm their role in the duplication and rearrangement processes that took place in the evolution of the animal mitochondrial genome.

3) Ribosomal RNA genes: accurate sequence analysis revealed a number of significant examples of complementarity between ribosomal and messenger RNAs. This suggests that they might play an important role in the regulation of mitochondrial translation and transcription mechanisms.

The properties revealed by our work shed new light on the organization and evolution of the vertebrate mitochondrial genome and more importantly open up the way to clearly aimed experimental studies of the regulatory mechanisms in mitochondria.

Key words: Evolution of animal mitochondrial DNA — Complete genome sequence — Regulatory signals — Mitochondrial tRNA evolution — Initiation and termination codons

Introduction

The mitochondrial (mt)DNA of animal cells is a circular molecule ranging in size between 14 and 39 kb. This molecule has become a very useful tool for studies of evolutionary genetics because of its reduced size, high copy number, unisexual mode of inheritance, and evolutionary behavior (Wilson et al. 1985). For these reasons numerous laboratories are now engaged in comparative studies of animal mtDNA not only by using restriction endonuclease enzymes but also by determining nucleotide sequences.Vertebrate mtDNA has been completely sequenced in human, cow, mouse, and in the frog *Xenopus laevis*; a large part of the mitochondrial genome of rat was sequenced in three laboratories, including our own, between 1981 and 1983 (Gortz and Feldmann 1982; Koike et al. 1982; Pepe et al. 1983; Cantatore and Saccone 1987, for review). However, sequencings were performed by using different rat strains that are known to have polymorphic variants of mtDNA.

The opportunity for analyzing the entire primary structure of another mammalian mitochondrial ge-

Offprint requests to: C. Saccone

nome together with needing to know the complete sequence of a single mtDNA type caused us to complete the sequencing of the rat mitochondrial genome.

In this paper we report the complete nucleotide sequence of *Rattus norvegicus* mtDNA, variant A, 16,298 bp long. The genome properties that are similar to those previously described in other related organisms and that have been discussed in recent reviews (Attardi 1985; Cantatore and Saccone 1987) are not treated here. We report some peculiar features of the genes coding for messenger, transfer, and ribosomal RNAs that were revealed by extensive comparative studies. These features shed new light on both the genetic organization and the evolutionary behavior of the mitochondrial genome in animal cells.

For the structure and evolution of the D-loop-containing region we refer to our recent papers (Brown et al. 1986; Saccone et al. 1987).

Materials and Methods

Source and Sequencing of mtDNA. The complete rat mitochondrial genome (*Rattus norvegicus*, type A), digested with Eco RI and Hind III, was cloned in pSF2124 plasmid. Each inserted fragment was excised from the recombinant, purified by preparative gel electrophoresis, and eluted separately. After further digestions with useful restriction enzymes the fragments obtained were sequenced using various strategies already described (Saccone et al. 1981; Cantatore et al. 1982; Pepe et al. 1983). More recently, large fragments, lacking appropriate restriction sites, were treated with nuclease Bal 31 for various time intervals and ligated into pUC8 to obtain overlapping segments (Barnes et al. 1983). After transformation of the *Escherichia coli* TG1 strain, the recombinant DNAs were isolated by alkaline treatment according to Birnboim and Doly (1979). Following NaOH denaturation, sequencing experiments were performed by the dideoxynucleotide chain terminator method of Sanger et al. (1980), using both synthetic oligonucleotide primer and reverse primer.

Computer Analysis. The sequences analyzed were extracted from the GenBank collection using ACNUC software (Gouy et al. 1985). Analysis of the sequences was performed on a VAX 11/780(DEC) computer by using the software GLORIA (Attimonelli, Lanave, Liuni, and Pesole, unpublished).

Results and Discussion

Base Compositional Constraints and Codon Usage

The complete sequence of the *Rattus norvegicus* mtDNA, variant A, is presented in Fig. 1. The mitochondrial genome of rat as well as of other vertebrates has a G-C content ranging from 36 to 44%, but the base composition of the two strands is highly asymmetrical as reported in Table 1, which refers to the L-strand of the mtDNA from rat and other vertebrates. It must be noted that in rodents, as in cow and in *Xenopus*, the most represented base is A followed by T > C > G, with G definitely underrepresented. In humans, on the other hand, the order is A = C > T > G. If we consider the different RNA species separately, we find that this bias in base composition is present in all functional sequences. However, in the case of rRNA genes the values are more uniform than in the messenger and transfer RNA genes in all vertebrates owing to their high degree of conservation. The compositional constraints of the rat mitochondrial genome can also be observed in the codon usage of protein coding genes, shown in Table 2. The bias to highly avoid the base G, particularly in the third codon position (see also Pepe et al. 1983; Lanave et al. 1985) that reflects the generally low amount of this base in the total L-strand is stressed in such a way that some G-ending codons are not used at all (i.e., TAG and AGG). Only ATG is significantly more abundant because of its initiator function (see below). In Table 3 the use of the four bases in the three codon positions of the rat mitochondria is presented: in the first position the use of A greatly exceeds that of G; T is mostly used in the second position, and A and C are the most used in the third one. In the mitochondrial genome of the other vertebrates the codon strategy is similar.

A regular three-base periodicity in protein coding genes has been suggested by various authors. In particular, Trifonov (1987) has reported recently that in the protein coding sequences there is a universal three-base periodical pattern (G-nonG-N)*n* and has suggested that this pattern is responsible for the correct monitoring of the frame during translation, so that only the frame having the highest G content can be chosen. To confirm the existence of this mechanism he found several sites with a complementary (N-N-C)*n* motif that should appropriately bind mRNA in the *E. coli* 16S rRNA sequence. We inspected the codon usage of rat mt mRNAs in order to check for a triplet periodical pattern that could support this theory. The data reported in Table 3 clearly demonstrate that a G-nonG-N periodicity cannot be considered universal behavior and is certainly not applicable to mitochondrial genes. We found a very high frequency of A-Y-N codons (24% of total) of which more than 50% are A-T-N codons. Furthermore, in the frame of all the mRNAs there are (A-N-N)*n* stretches that in the genetic system studied by us could be the counterpart of the (G-nonG-N)*n* tracts. However, in our opinion, the preferential use of these triplets with A in the first position is not linked strictly to the necessity of monitoring the correct reading frame. It is dictated rather by two major constraints. The first is the base composition of the codogenic strand: the calculated frequency of A-N-N codons is equal roughly to the

Table 1. Length and base composition of vertebrate mitochondrial genomes

Species	Total length (bp)	Percent A	T	C	G
Rat	16.298	34	27	26	13
Mouse	16.295	35	29	24	12
Cow	16.338	33	27	26	14
Human	16.569	31	25	31	13
X. laevis	17.553	33	30	23	14

Table 2. Codon usage in rat mitochondrial genome

F	TTT 77		TCT 39	Y	TAT 50	C	TGT 8
	TTC 142	S	TCC 75		TAC 71		TGC 17
			TCA 117				
L	TTA 110		TCG 3	STOP	TAA 10	W	TGA 97
	TTG 6				TAG 0		TGG 1
				H	CAT 36		
	CTT 72		CCT 35		CAC 62		CGT 9
L	CTC 100	P	CCC 55			R	CGC 14
	CTA 266		CCA 105				CGA 40
	CTG 13		CCG 5	Q	CAA 82		CGG 2
					CAG 5		
I	ATT 178			N	AAT 58	S	AGT 6
	ATC 172		ACT 51		AAC 113		AGC 41
		T	ACC 94				
			ACA 163				
M	ATA 191		ACG 9	K	AAA 95	STOP	AGA 0
	ATG 24				AAG 4	?	AGG 0
				D	GAT 20		
	GTT 16		GCT 44		GAC 46		GGT 19
V	GTC 38	A	GCC 97			G	GGC 55
	GCA 98		GCA 81				GGA 95
	GTG 8		GCG 3	E	GAA 79		GGG 18
					GAG 7		

Table 3. Percentage of base frequency at the codon position and periodicity pattern in rat mitochondrial protein coding genes

Base	Codon position I	II	III	Periodicity pattern	Found	Expected
A	33	20	44	A-N-N	33	32
T	23	41	20	A-Y-N	24	18
C	25	27	33	A-T-N	16	9
G	19	12	3	A-nonG-N	32	29

value expected in all the possible frames considering the base distribution in the genome. The second is due to the type of products that such a genome must code for. In fact the abundance of the A-T-N codons, the only ones significantly more frequent than the expected ones ($\chi^2 = 5.5\%$), seems to be imposed by the necessity for the mitochondrial genome to code for proteins with a peculiar amino acid content. The frequency of each amino acid in the protein coding genes can be seen in Table 2. The most used is leucine (with T in the second position of the codon), but isoleucine and methionine (coded by A-T-N) are also very frequent, as well as are threonine and serine (coded by A-N-N). Moreover, we have searched carefully for a significant (N-N-T)*n* periodicity in rat mitochondrial small rRNA. Indeed we found many tracts with this feature, some of which are exposed in a loop structure, and are thus hypothetically accessible mRNA–zRNA contact sites, but again the calculated frequency is not significantly higher than the expected values. Thus, it seems unlikely that the mechanism proposed by Trifonov is at work in the animal mitochondrial genome: as we shall see later, only a few rRNA small regions (some of which were also indicated by Trifonov) are probably involved in messenger binding.

Unusual Genetic Code: Initiation and Termination Codons

The main peculiarities of the mitochondrial genome are the use of TGA as the tryptophan codon and the possibility that codons other than canonicals can act as initiators and terminators (Barrell et al. 1979). The latter suggestion is based on the fact that they are found at the 5′ or 3′ ends of mRNAs that do not have normal initiation or termination codons. On the basis of a detailed comparative analysis performed on the sequences of mtDNAs so far available, we propose a new interpretation of the use of the initiation and termination signals in the mitochondrial translation process.

It is well known that in vertebrate mtDNA the 22 tRNAs are interspersed between the ribosomal and protein coding genes, probably serving as signals (punctuation points) for the processing of the primary transcripts (Ojala et al. 1980, 1981). Few or no bases, in fact, separate mRNAs and tRNAs in the various genomes, and indeed sequencing of the 5′ end of mRNAs (Montoya et al. 1981) has revealed that the mature transcript starts from the base immediately after the tRNA of the same strand (see also below) and not from the initiation codon. In Fig. 2a the number of bases separating the mRNAs from the tRNAs and the nature of the initiation codons found in the mt mRNAs of vertebrates are shown. From these data a peculiar correlation emerges: when there are spacer bases between a tRNA and an mRNA or when there is gene overlapping (i.e., ATPase 6, ND4, see Fig. 1 for definition of abbreviations), we always find ATG as the initiator codon probably because a strong initiator is necessary to put the messenger in frame. On the other hand, when the reading frames start

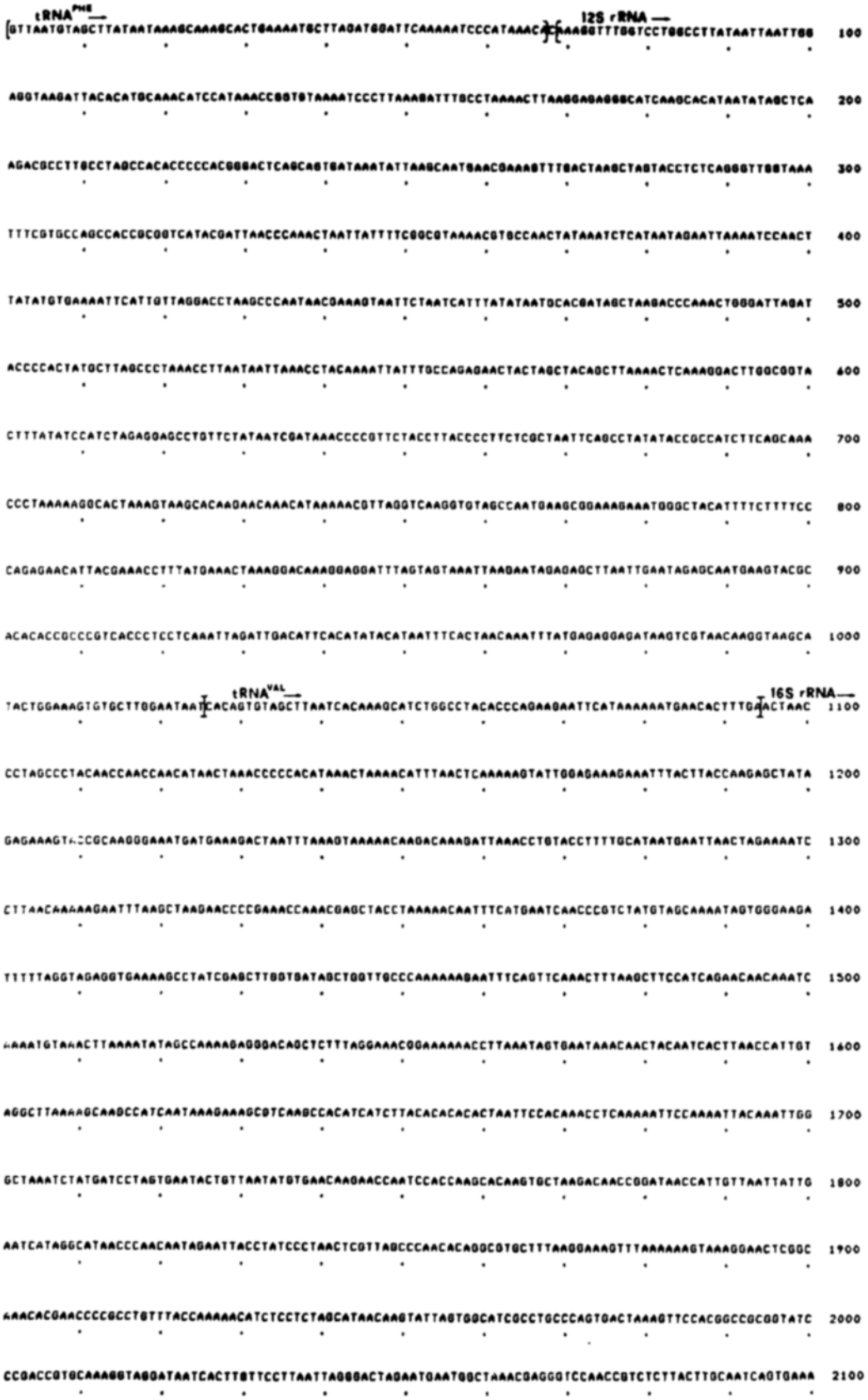

Fig. 1. Sequence of the *R. norvegicus* mtDNA (variant A). The sequence reported corresponds to the L-strand (5′ → 3′). The arrows indicate the direction of transcription of each gene. The predicted amino acid sequence (one-letter code) of each protein gene is shown above the sense strand. COI, II, III: cytochrome oxidase subunits; Cyt *b*: cytochrome *b*. ND1–6: NADH dehydrogenase subunits. Continued on pages 501–507.

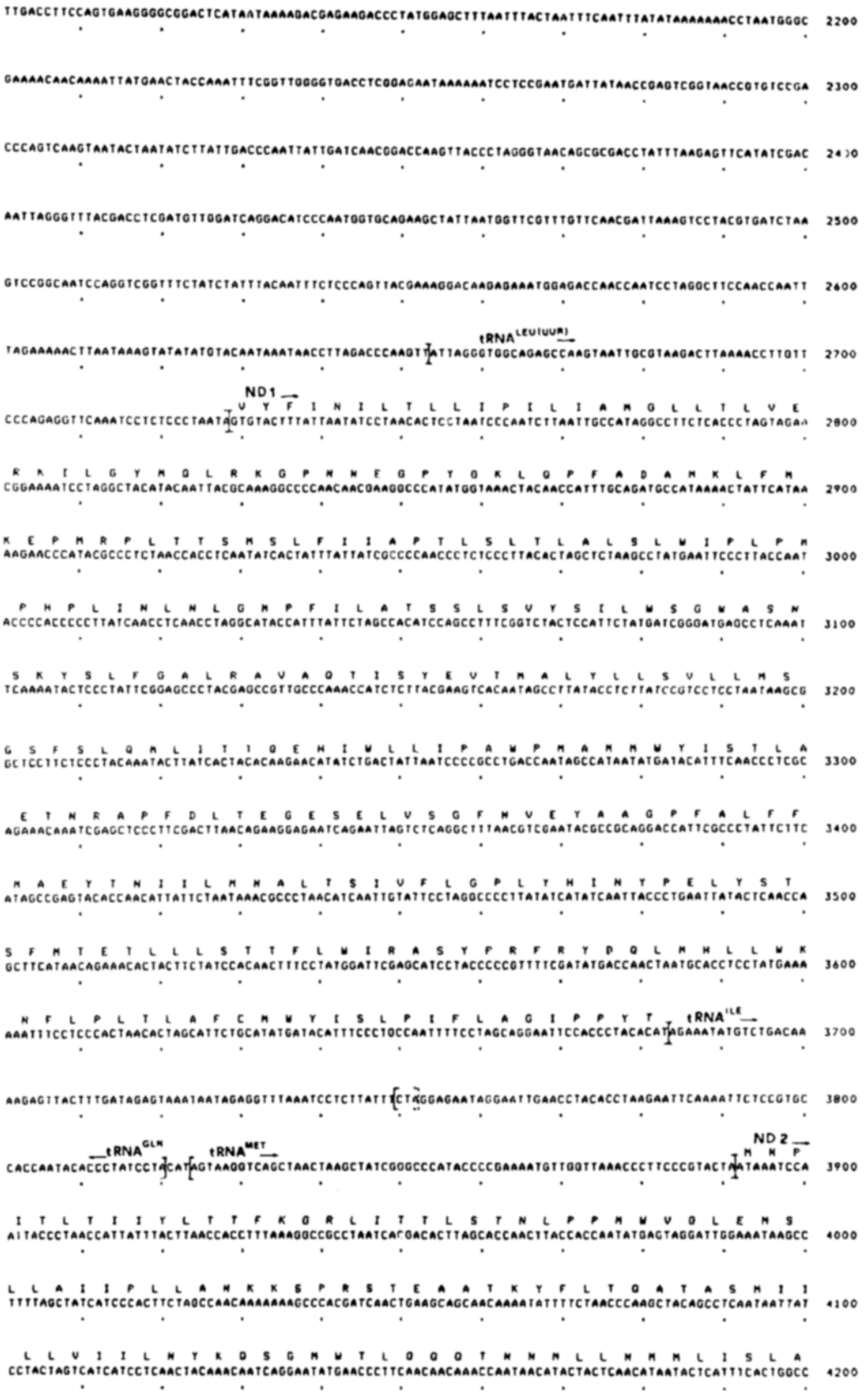

Fig 1. Continued

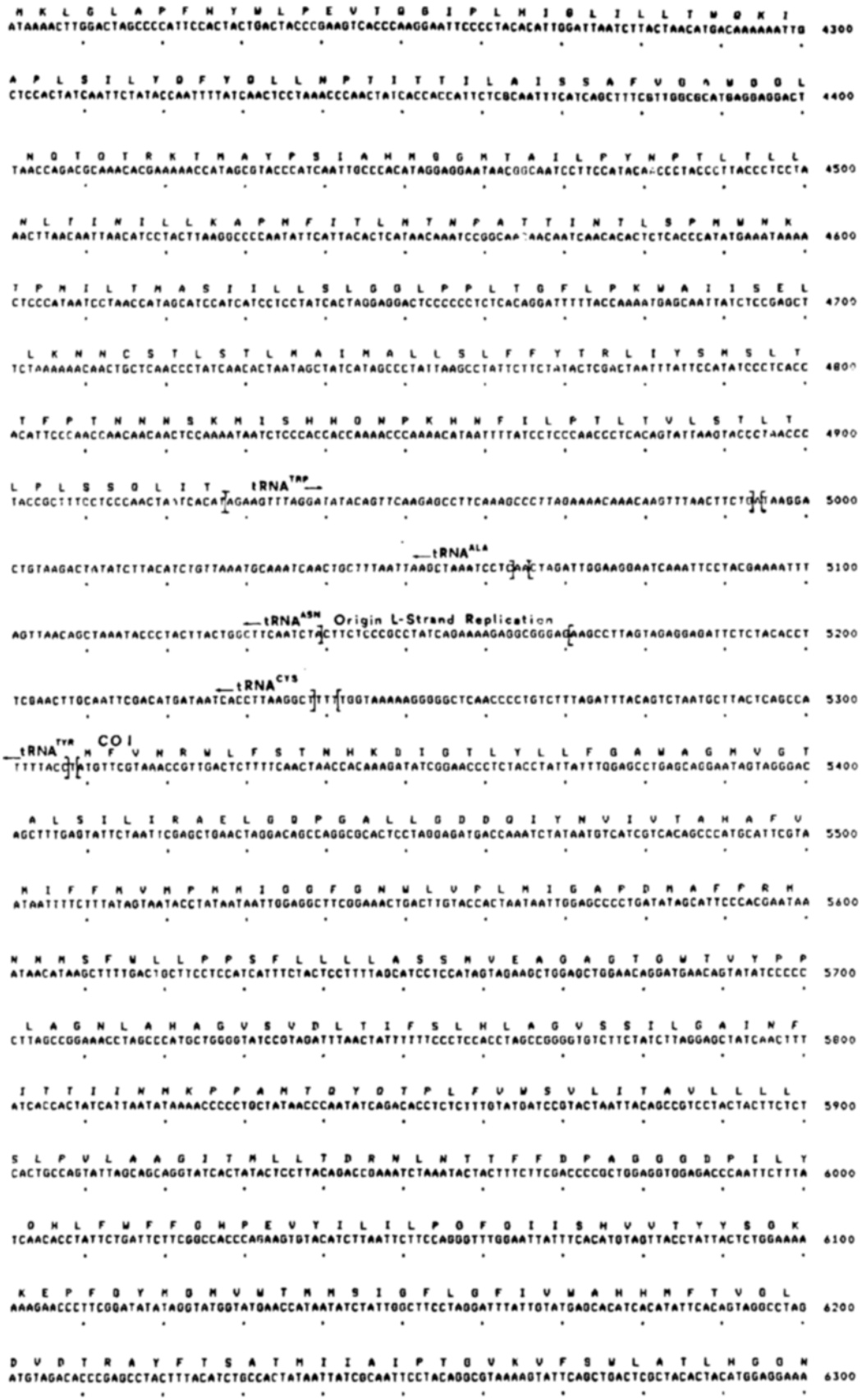

Fig. 1. Continued

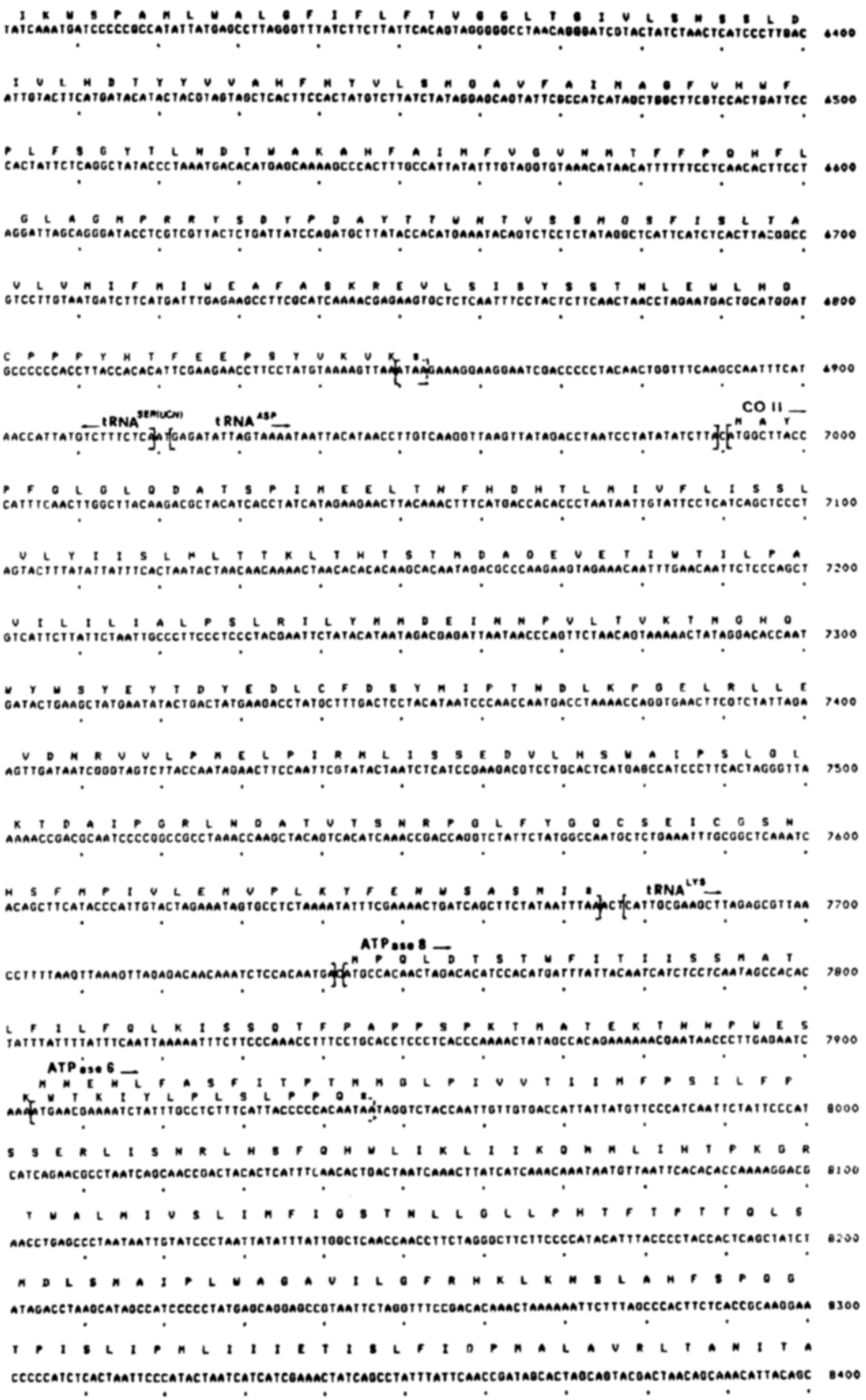

Fig. 1. Continued

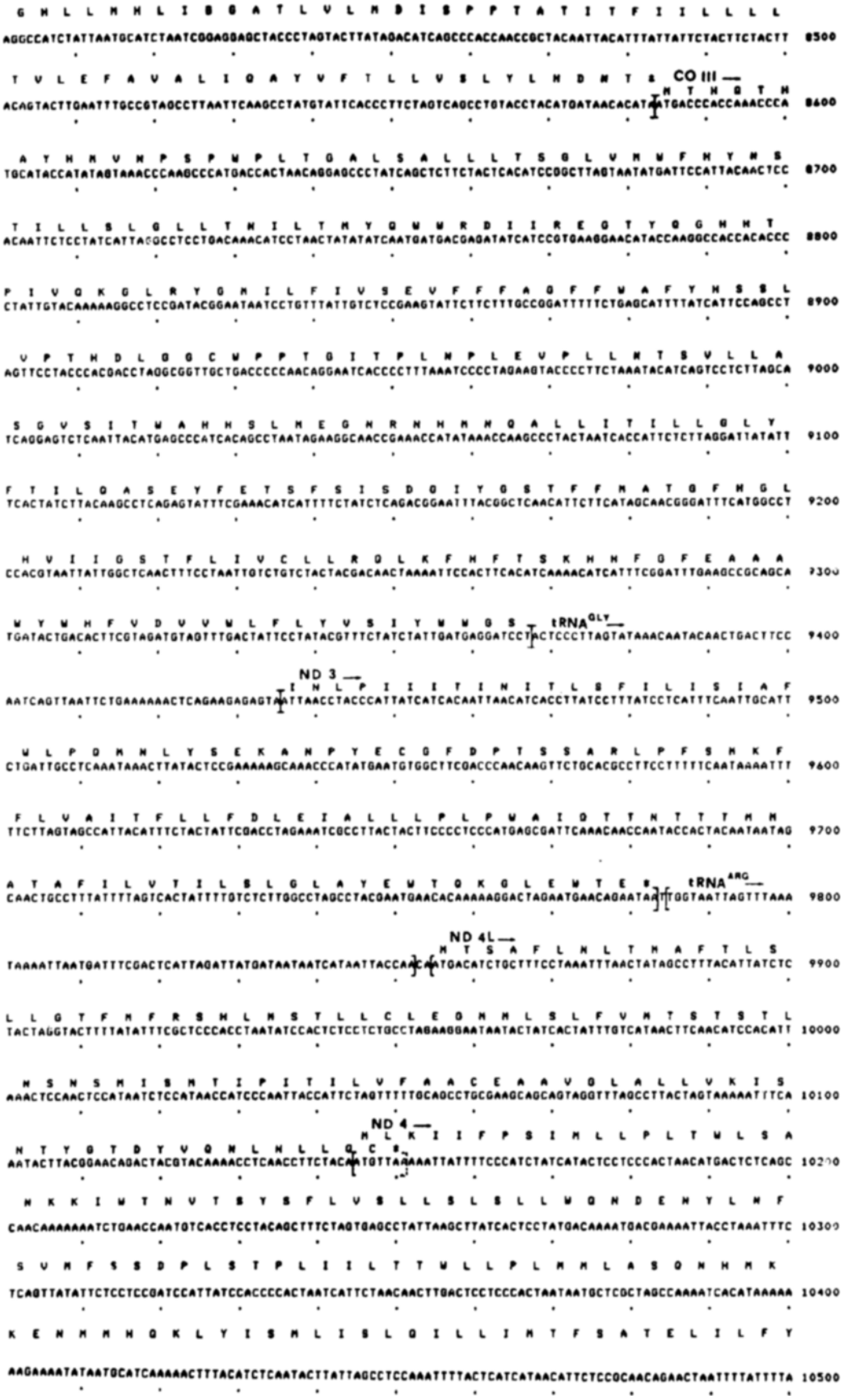

Fig. 1. Continued

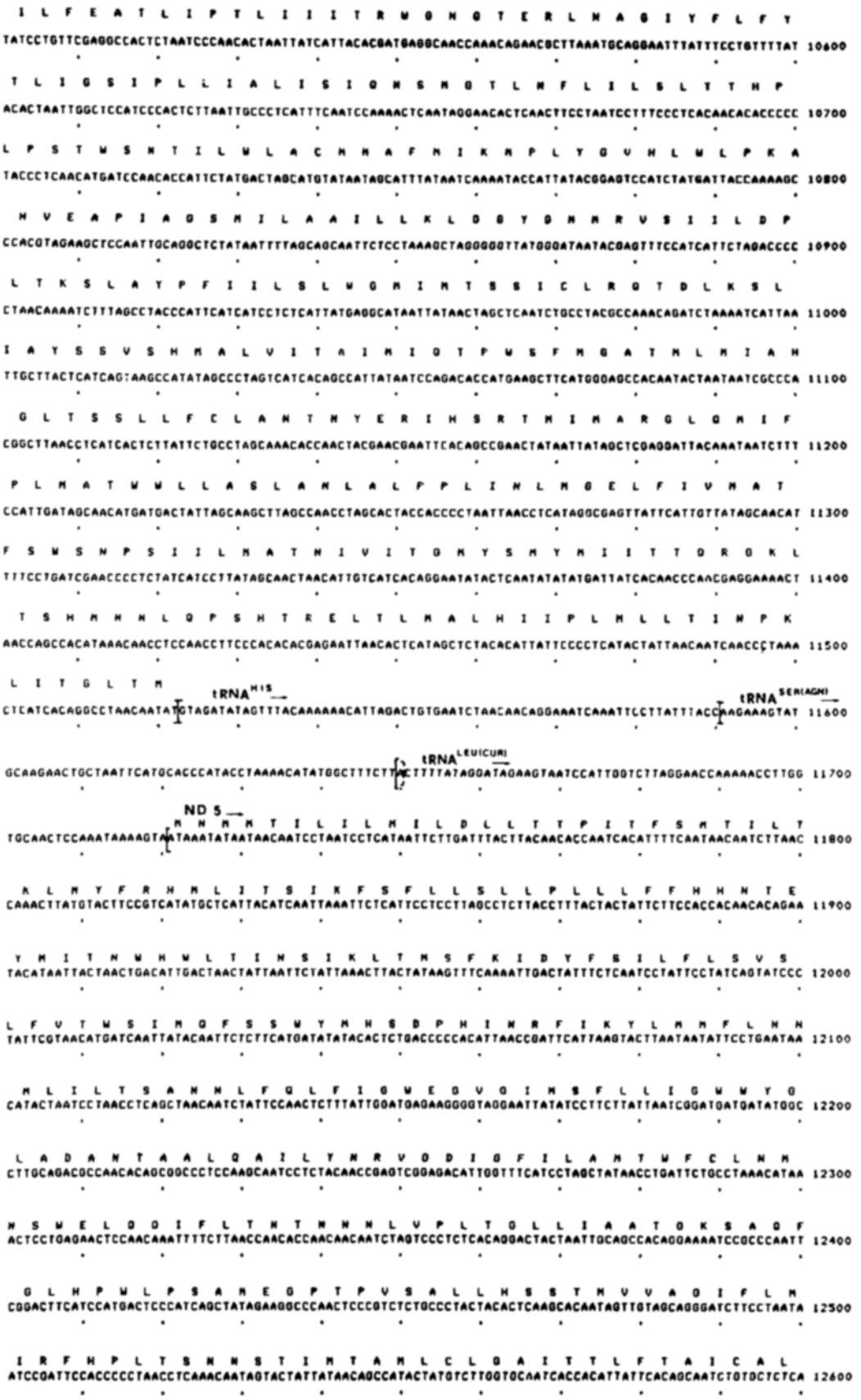

Fig. 1. Continued

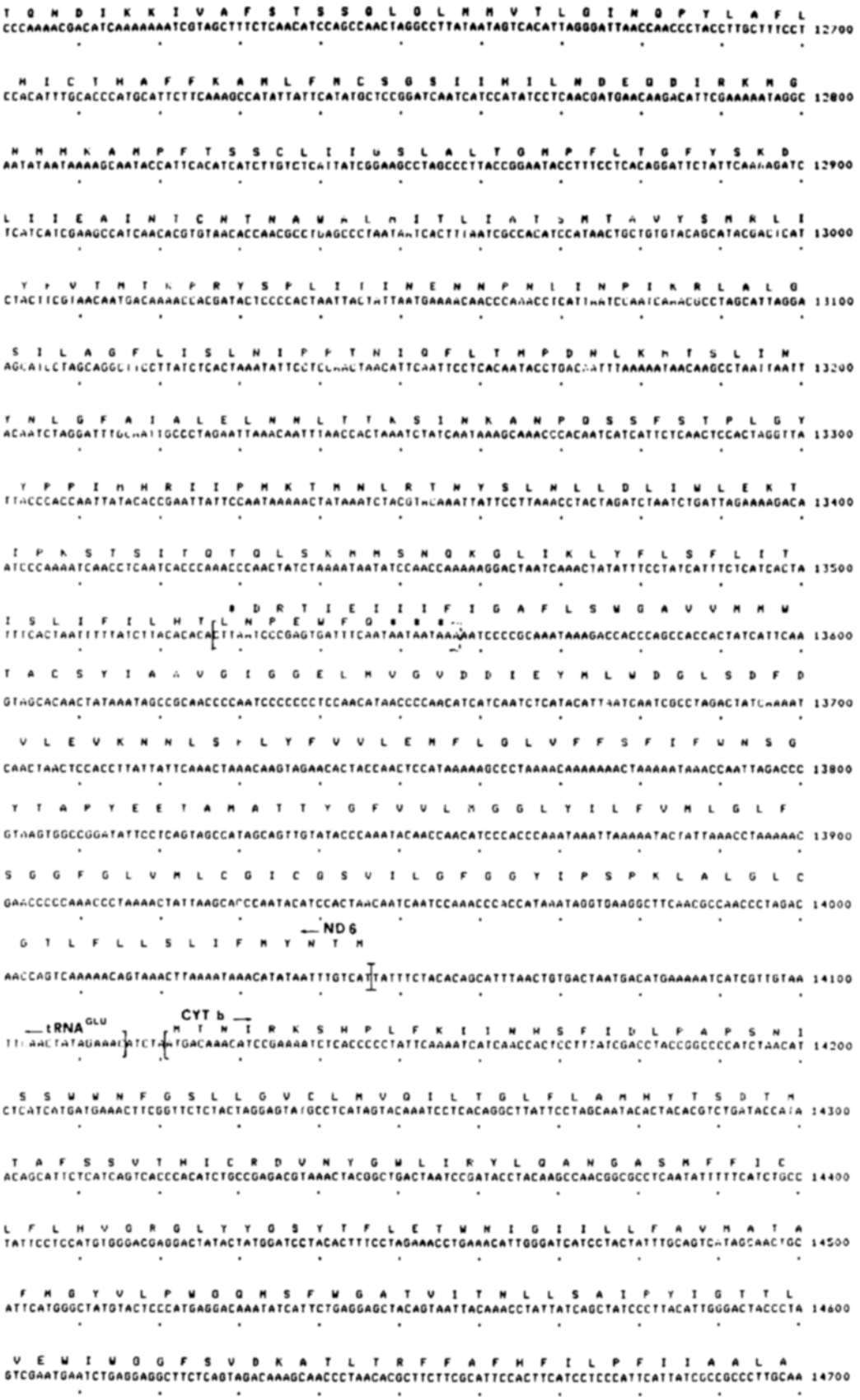

Fig. 1. Continued

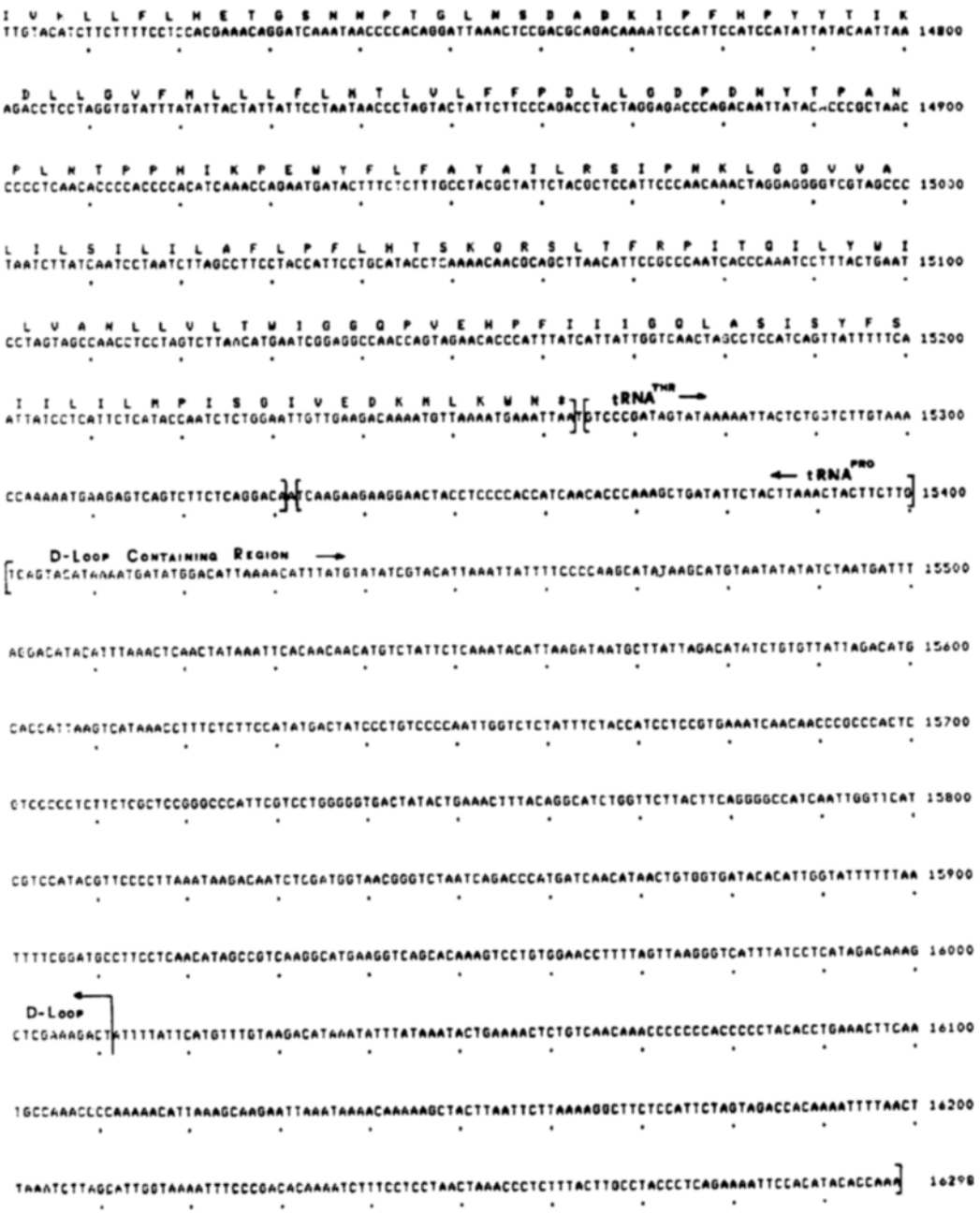

Fig. 1. Continued

from the first base of the messengers, any ATN codon and possibly also a GTG codon (see below) can act as initiator. The case of ND1 that could represent an exception requires some comment. In cow the ATG codon lies after two spacer bases, as expected according to our prediction. In humans the corresponding mRNA has been sequenced: it starts with two spacer bases followed by ATA and, after a triplet, by ATG that in any case would keep the messenger in frame. In mouse by considering ATT as the initiation codon the start has been fixed nine bases after the tRNA leu (uur) gene (Fig. 2b). However, more careful inspection of the alignment at the 5′ region reveals that if the ND1 mRNAs of mouse and rat initiate at the ATT codon, they end up shorter than the other similar genes by exactly nine bases [315 amino acids (aa) instead of 318 aa]. On the other hand, if the translation begins soon after the tRNA gene, the mRNAs of rodents will have the same length as those of the genes of other vertebrates (Fig. 2b). This implies that not only ATT or ATC, but also GTG, could act as initiator when there are no extra bases at the 5′ end of the messenger. It should be recalled that the codon GTG has been suggested as the initiator in the mitochondrial ND1 gene of mouse, in the mitochondrial ND5 gene of

a

	MAN		COW		MOUSE		RAT		XENOPUS	
	sb	ic	sb	ic	sb	ic	sb	ic	sb	ic
ND1	2	ATA	2	ATG	9	ATT ?	–	GTG	–	ATG
ND2	–	ATT	–	ATA	–	ATA	–	ATA	–	ATG
COI	12	ATG	1	ATG	1	ATG	1	ATG	2	ATG
COII	–	ATG	1	ATG	1	ATG	1	ATG	2	ATG
ATPase8	1	ATG	1	ATG	1	ATG	1	ATG	1	ATG
ATPase6	ov	ATG	ov	ATG	ov	ATG	ov	ATG	ov	ATG
COIII	–	ATG	–	ATG	–	ATG	–	ATG	–	ATG
ND3	–	ATA	–	ATA	–	ATC	–	ATT	–	ATG
ND4L	–	ATG	–	ATG	1	ATG	2	ATG	–	ATG
ND4	ov	ATG	ov	ATG	ov	ATG	ov	ATG	ov	ATG
ND5	–	ATA	–	ATA	–	ATC	–	ATA	–	ATG
ND6	–	ATG	–	ATG	–	ATG	–	ATG	–	ATG
Cyt b	4	ATG	4	ATG	5	ATG	5	ATG	2	ATG

b

```
MAN       Leu uur...TTAACAAC.ATA.CCC.ATG.GCC.AAC...ND1
COW          ...TTAACAAA.ATG.TTC.ATA.ATT.AAC...
MOUSE        ...TCCCTAATAgtg.TTC.TTT.ATT.AAT...
RAT          ...TCCCTAATAGTG.TAC.TTT.ATT.AAT...
XENOPUS      ...TCGCTAACTATG.TTA.ACT.ATT.ATT...
```

Fig. 2. Initiation codons of vertebrate mt mRNAs. **a** ic = predicted initiation codons; sb = untranslated spacer bases at 5′ end; ov = gene overlapping; – = absence of spacer bases; ? = cases discussed in the text. **b** Best alignment of the ND1 5′ end. ——: initiation codons as reported in literature. Lower cases: initiation codons as proposed by us.

Drosophila yakuba (Clary and Wolstenholme 1985), and in the Cyt *b* gene of sunflower mitochondria (Gallerani, personal communication), and that it has also been found to act as initiator in a few prokaryotic genes (Stormo et al. 1982). The feature we have described seems to be well confirmed in the genome of *X. laevis* (Roe et al. 1985), where we observe that all mRNA genes that start with a few spacer bases always contain the ATG codon. A constraint in the use of ATG versus ATA (or other initiation codons) is further corroborated by the fact that most of the mRNAs in mammals initiate with ATG, in spite of the higher probability of finding ATA instead of ATG, because of the low G content of the mtDNA L-strand and because of the higher use of A at the third codon position (see above).

We have also observed some peculiar features for the use of terminators. It has been reported (Clary and Wolstenholme 1985) that the presence or absence of complete termination codons at the 3′ end of the mRNA genes is a poorly conserved property in mammals or flies. However, even here some rules seem to be followed. In particular, only when an mRNA ends immediately before a tRNA might the termination codon be incomplete: in this case the transcript is processed immediately before the transfer RNA so that the poly A addition can create the termination codon TAA. When the mRNA contains, as in the human COII gene, a 3′ untranslated region, we always find a normal terminator such as TAG or TAA. The same of course holds for the cases in which there is gene overlapping (ATPase 8, ND4L, ND5, and ND6). This rule seems to be validated by the mtDNA of *Drosophila* that has a different gene organization. It is well known that in vertebrates AGA and AGG have been proposed as terminators. The two codons are never used within the open reading frames (ORFs) of vertebrates and have been found at the 3′ end of a few genes as shown in Fig. 3a. In other animals, like *Drosophila* (Clary and Wolstenholme 1985), echinoids (Cantatore et al. 1987b; Himeno et al. 1987), and nematode worms (Wolstenholme et al. 1987), AGG and AGA code for serine. In our opinion the cases in which the two codons have been found at the 3′ termini of mammalian mRNAs require comment (Fig. 3b). The COI gene, which is the best-conserved protein coding gene, ends in the human genome with AGA but this makes the gene one amino acid shorter than in all other vertebrates (513 aa instead of 514 aa). It is also remarkable that the codon following AGA is CAA in humans, whereas in the other mammals, rat included, it is TAA, and in *Xenopus* it is TCG. Likewise, in the Cyt *b* of cow, if the stop codon is AGA, the protein is one amino acid shorter. Moreover, in this case the following codon is CAG that differs from the normal terminator only by a transition in the first base, namely a C to T conversion. As regards ND6, it should be noted that the human gene has a regular TAG codon nine bases after the AGG codon and in *Xenopus* a normal TAA codon six bases after AGA. In the genomes of rat, mouse, and cow there is overlapping between the ND5 and ND6 mRNA genes. In humans, if ND6 terminates at the AGG, the two messengers become contiguous, whereas if the reading frame of ND6 reaches the regular terminator TAG, the genes become four amino acids longer but the overlapping is reestablished. Likewise, in *Xenopus* the reading of AGA (arginine in all similar mammalian genes) normalizes gene length.

Our observations cannot exclude the previous assignment of initiation and termination codons in mt mRNA genes. However, they emphasize the importance of multiple comparative analyses and the necessity of sequencing mt mRNAs and proteins in order to solve these important questions and to shed light on the mitochondrial translation process.

An unusual codon assignment is not unique to mitochondria. In ciliates TAA and TAG stop codons are found in ORFs, and TAA seems to code for glutamine (Horowitz and Gorovsky 1985). More recently, both in prokaryotes and in eukaryotes, sequencing has revealed that selenocysteine is encoded by TGA (Chambers and Harrison 1987). This raises the question as to how the cell distinguishes between two functions for the same codon. It has been suggested that, by analogy with the suppressor tRNA genes, the choice depends on the sequences within the mRNA in the vicinity of the termination codon (the context effect involving structural interactions at the ribosomal level).

Table 4. Conservation degree of different functional regions of the vertebrate mt tRNAs

	AA stem	D stem	D loop	D stem	AC stem	AC loop	AC stem	V loop	T stem	T loop	T stem	AA stem
Met	***	****	**	****	***	****	***	***	****	**	****	***
Leu (cun)	***	***	*	****	***	****	***	***	***	***	***	***
Ile	***	***	***	***	***	****	***	***	***	***	***	***
Ala	**	****	****	****	**	***	***	***	**	*	**	***
Arg	***	***	**	***	***	***	***	***	***	—	***	***
Tyr	***	***	**	***	**	***	**	**	**	—	****	***
His	***	****	**	****	***	***	***	**	**	—	**	***
Asn	***	***	*	***	****	****	****	*	***	—	**	**
Leu (uur)	*	****	**	***	**	***	**	*	***	***	***	**
Glu	**	****	**	****	***	**	***	**	**	*	**	**
Gln	**	***	—	***	**	**	**	**	***	***	**	**
Phe	**	****	*	****	**	****	**	***	*	—	*	***
Trp	**	**	*	***	***	****	****	**	***	—	**	***
Lys	*	****	*	***	***	****	***	***	*	—	*	**
Ser (ucn)	—	—	*	***	**	***	**	*	****	***	**	**
Thr	*	**	**	**	**	****	**	***	***	—	**	**
Gsp	**	***	**	***	***	***	***	**	—	—	*	**
Val	*	****	*	****	*	***	*	**	**	—	**	*
Gly	*	****	**	***	*	****	*	*	*	—	*	**
Cys	—	***	**	**	*	****	*	***	**	—	**	**
Pro	*	***	**	***	*	****	*	***	*	—	*	*
Ser (ggn)	*	**	...	...	****	****	*	**	—	—	**	**

% similarity: 1–25 (—), 26–45 (*), 46–70 (**), 71–90 (***), 91–100 (****), bases completely absent (···)

Transfer RNA Genes: Degree of Conservation among Vertebrates and Peculiar Similarity Features

The 22 tRNA species coded by the animal mitochondrial genomes have structural anomalies to such an extent that they may be considered a separate class of tRNA molecules. They are generally smaller than their cytoplasmic and prokaryotic counterparts and lack many of the invariable features (see Cantatore and Saccone 1987, for review). Besides having a structural role, the tRNAs of vertebrates, according to the punctuation model (Ojala et al. 1980, 1981), also act as recognition sites for processing RNase-P-like enzymes that cleave the transcripts at the junction between the mRNAs (or the rRNAs) and the tRNA genes. In view of this dual function of the mt tRNA genes, we have carried out sequence analysis by comparing similar tRNA genes in five vertebrate species and the 22 tRNA genes in the same organisms. Table 4 and Fig. 4 show the degree of conservation of the different functional regions of the tRNAs in five species, namely rat, mouse, cow, human, and *Xenopus*. The tRNAs are listed according to their overall conservation degree. It appears that (1) some of the most conserved species (leucine and isoleucine tRNAs) are those that decode the most used amino acids (see also Table 2); (2) the anticodon loop is the most conserved region followed by the D stem; and (3) the T loop and stem are the most variable regions. In general the degree of conservation of the 5′ half is higher than that of the 3′ half, suggesting that the 5′ region contains stronger functional constraints.

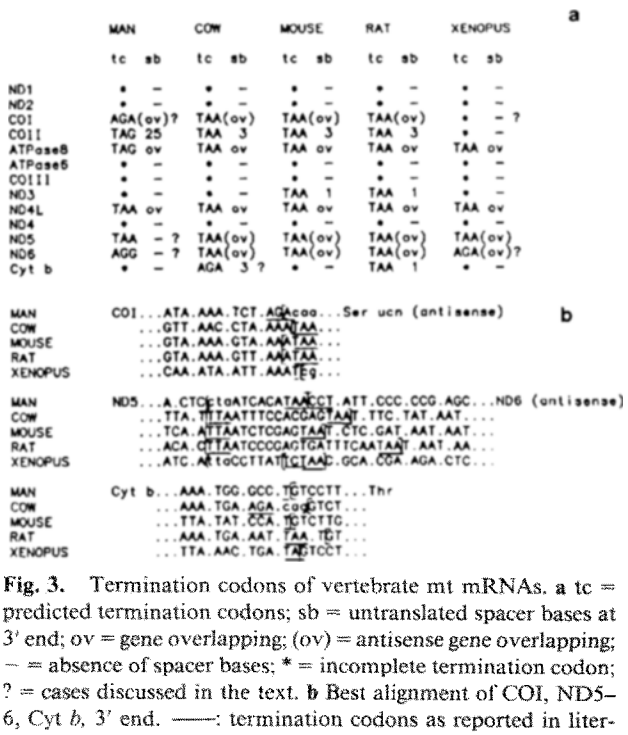

Fig. 3. Termination codons of vertebrate mt mRNAs. **a** tc = predicted termination codons; sb = untranslated spacer bases at 3′ end; ov = gene overlapping; (ov) = antisense gene overlapping; — = absence of spacer bases; * = incomplete termination codon; ? = cases discussed in the text. **b** Best alignment of COI, ND5–6, Cyt *b*, 3′ end. ——: termination codons as reported in literature. Lower cases: termination codons as proposed by us.

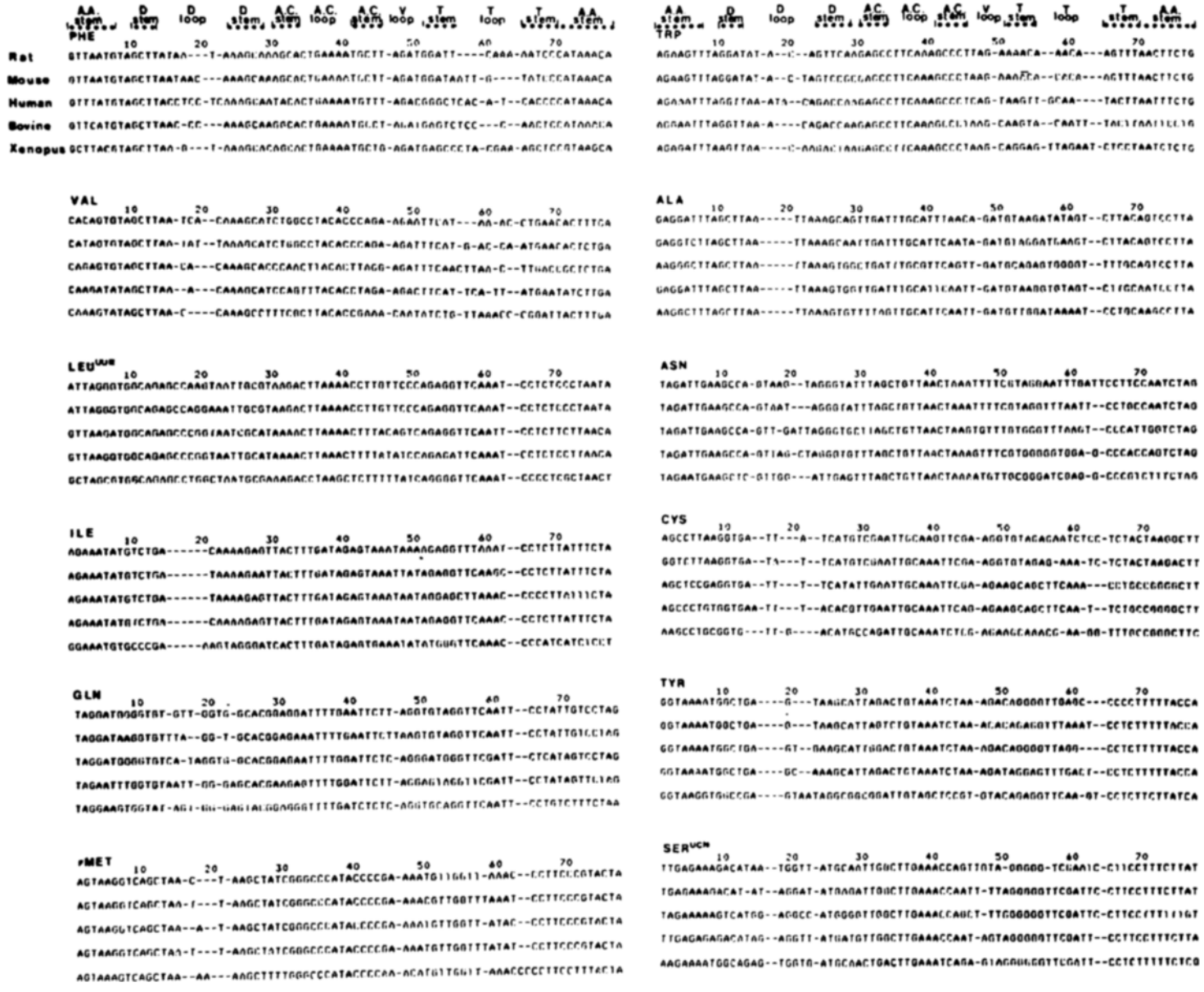

Fig. 4. Alignments of vertebrate mt tRNAs. On the top the different functional areas are indicated. Dots correspond to the spaces created in order to optimize the sequence similarity. Continued on next page.

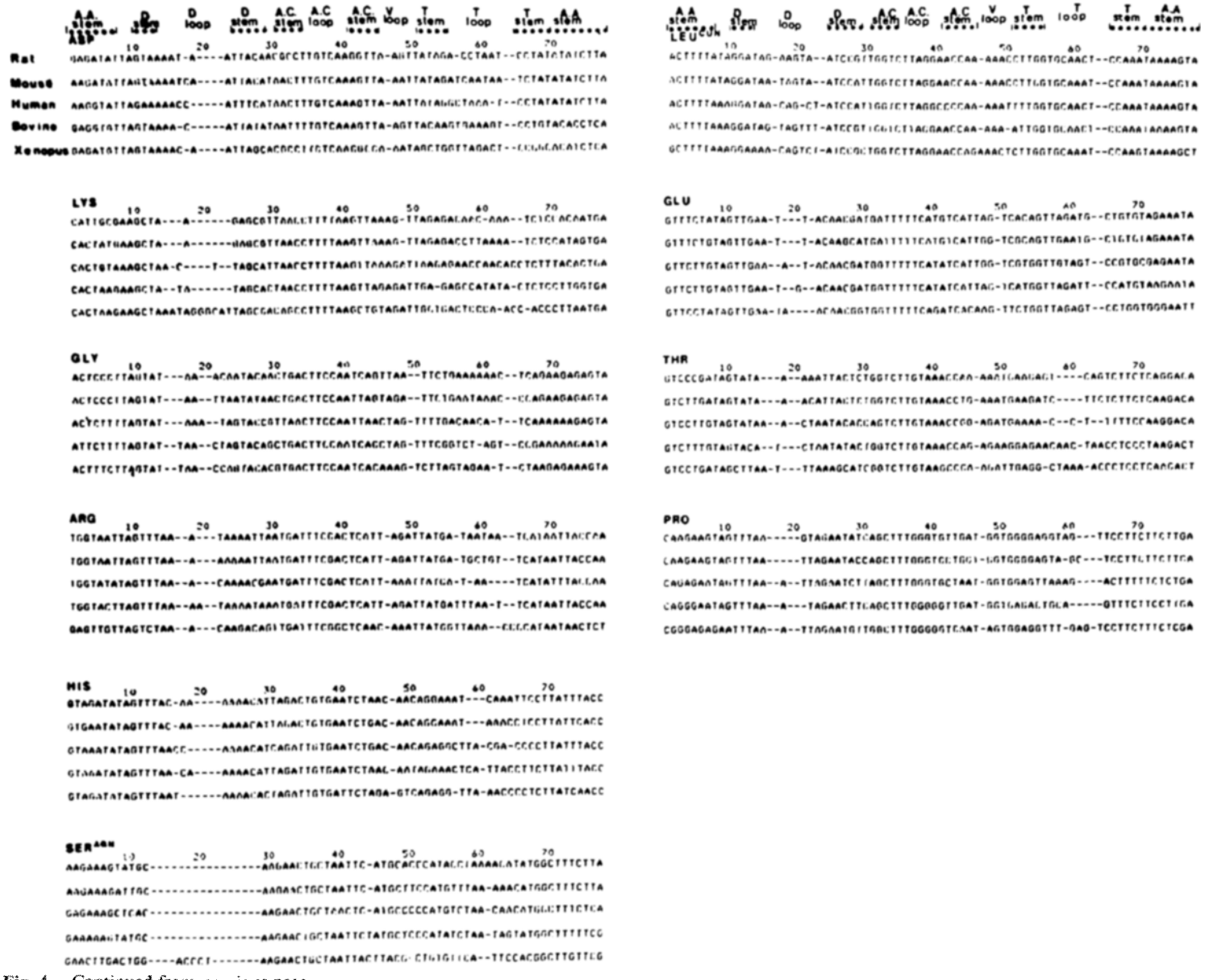

Fig. 4. Continued from previous page

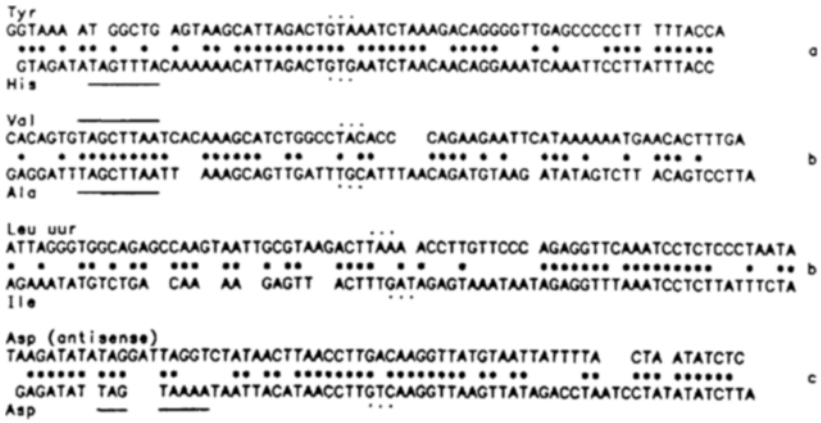

Fig. 5. Peculiar similarity features of rat mt tRNA. **a, b, c,** Examples of the three classes of similarity discussed in the text. ···: anticodon. ——: consensus sequence.

Comparative studies between tRNAs of the same organism revealed the existence of peculiar similarity features well conserved in mammals. In particular we observed:

1) The overall similarity between tyrosine and histidine tRNAs is significantly higher than that between any other tRNAs not only in rat (Fig. 5a) but also in all other vertebrates. In particular, in the central region there is a block of 19 identical bases with only one mismatch at the third anticodon base. Some of us have recently found (Cantatore et al., 1987a) that during the evolution of sea urchins the mt leucine tRNA gene for CUN codon lost its function and became part of a protein coding gene. The new CUN tRNA gene appears to have evolved by duplication and divergence from the UUR tRNA gene. This suggests that tRNAs could themselves act as a primer for mtDNA synthesis and sometimes fail to be removed from the newly synthesized strand, becoming a template in the next round of duplication. The strong similarity between heterologous tRNA genes of the same organism observed here seems to indicate that this mechanism of duplication and remoulding of tRNA genes observed in the sea urchin might be a more general evolutionary event that occurred also in the vertebrate lineage. It is to be noted further that in the case of the histidine–tyrosine pair, the two transfers belong to different strands and are both clustered with other tRNA genes. This reinforces the view that tRNA genes can move independently of other mitochondrial genes and stresses their importance in duplication and rearrangement mechanisms occurring during the evolution of animal mtDNA as indicated by several authors (Clary and Wolstenholme 1985; Brown, personal communication).

2) In each organism tRNA genes might be grouped in families on the basis of a significant degree of similarity in the 5′ or central or 3′ region (the different degree of similarity found in the 5′ or 3′ region is due to the high number of mismatches in the amino acid stem of mt tRNAs). Examples are shown in Fig. 5a and b. We are tempted to interpret this feature on the basis of the multiple roles played by tRNAs. We have already stressed the function of mt tRNAs in the processing of the transcripts and in other mitochondrial processes (Cantatore et al. 1987a). It is interesting to note that recent sequence analyses indicate that both the secondary and the primary structures present in tRNAs are involved in many functions, such as origins of replication, gene regulation, recombination, retroposon generation, and others (Saccone et al., 1987; see Weiner and Maizels 1988 for review). What is remarkable in this context is the recent finding that in the mitochondria of *Neurospora crassa* one protein component required for splicing mitochondrial introns is mt tyrosil tRNA synthetase that recognizes features resembling tRNA substrate in introns (Akins and Lambowitz 1987). All these data suggest that the various mt tRNAs may have evolved under different evolutionary pressures depending on their multiple roles.

3) Two tRNA species, namely arginine and aspartic acid, display an unusual high degree of self-complementarity. This feature has been tentatively interpreted in terms of both the function and the evolution of mt tRNA genes as is shown later. With regard to the property of mt tRNAs to act as a recognition signal for RNase enzymes, we recall that according to the data of Attardi's group (Montoya et al. 1981; Ojala et al. 1981), in humans the tRNAs are cut exactly at the 5′ and 3′ ends; however, it seems that the corresponding antisense tRNAs are not always processed with the same accuracy at their termini. This is demonstrated by the presence of 5′ and 3′ untranslated regions in COI mRNA corresponding to the antisense tyrosine and serine tRNAs and by the detection of an RNA species (RNA 6) precursor of RNA 9 that includes the antisense sequences of the tRNAs clustered at the Ori-L region (Gaines et al. 1987). Another example is the inaccurate processing of Cyt *b* mRNA at the 5′ end. Moreover, the three abundant long L-strand tran-

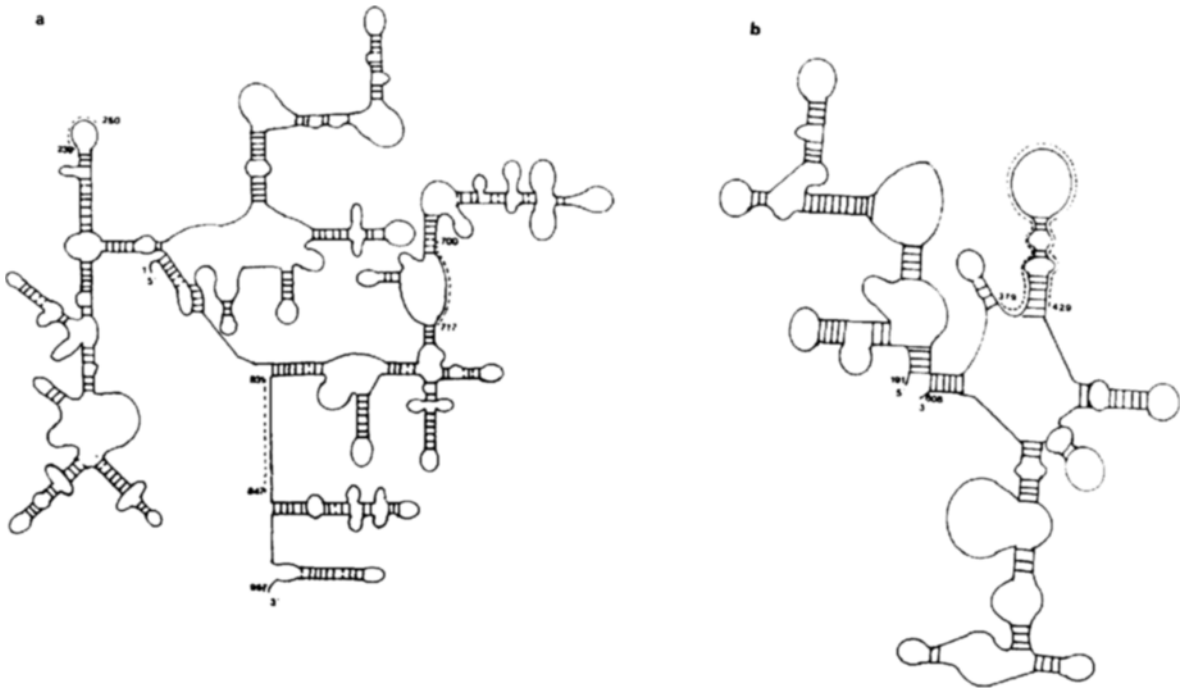

Fig. 6. Secondary structures of rat mt rRNAs. These structures are based on the model presented by Maly and Brimacombe (1983). **a** Small 12S rRNA. The binding sites to the mRNA proposed by us are underlined. **b** 3′ region of the large 16S rRNA structure containing the sequence complementary to the ND6 mRNA (underlined).

scripts identified in humans can have their termini which correspond in terms of position with those of the tRNAs coded by the same strand with the exception of RNA 3 (Ojala et al. 1980) that stops at the arginine tRNA antisense level. Thus, it seems that tRNA structures are neither necessary (the cutting between ATPase 6 and COIII does not involve tRNAs) nor sufficient (antisense RNAs are not always processed) for the maturation of mitochondrial transcripts. Our observation suggests that for the processing of the polycistronic transcripts both primary and secondary structural elements may be necessary as recognition signals. The self-complementarity of the arginine tRNA, which is particularly relevant in humans (71%), means that the same primary structure is present in the two DNA strands. This, in other terms, should explain why some antisense tRNAs are processed more efficiently than others and thus justifies the processing of RNA 3 at the level of antisense-arginine tRNA. Because the aspartic acid tRNA displays the same unusual self-complementarity (Fig. 5c), it could serve as the processing site for RNA 2, whose terminus has not been defined exactly. Obviously, our speculations need to be supported by further experimental analyses that these studies will hopefully stimulate.

In the attempt to identify common features, we found that many tRNAs possess the triplet TAG at positions 8, 9, and 10. A second TAG is present also in 11 genes at a fixed distance (20–27 bases) from the 3′ terminus. The occurrence of this triplet in both L- and H-strand-coded genes is significantly higher than expected. TAG at positions 8, 9, and 10 is also found at high frequency in prokaryotic and eukaryotic tRNAs (listed in Nucleic Acid Res suppl vol 13, 1985; Spritzl et al. 1987), although only T at position 8 is considered an invariant. In 12 mt tRNA genes we found, starting from position 8, the consensus sequence TAG(Y)*n*(R)*n* (Fig. 5). This sequence is at the start of the A block of the Pol III enzyme and is probably important for the evolution of the tRNAs in general. In the complete rat mitochondrial genome the sequence TAG-TAAA, present in aspartic tRNA, was found only at the end of 12S rRNA and at the 5′ terminus of COIII, which is the only messenger processed without a tRNA acting as a signal. We are tempted to speculate as to its involvement in the processing mechanism. A similar sequence is also present in the short interspersed nuclear elements (SINEs) containing tRNA-like structures of mammalian nuclear genomes, and this strongly argues for an important role in both tRNAs and SINEs (Daniels and Deininger 1985; Lawrence et al. 1985).

Ribosomal RNA Genes: Interactions between Ribosomal and Messenger RNAs

The 12S and 16S rRNA genes in rat mitochondria are 956 and 1558 bases long, respectively. These are

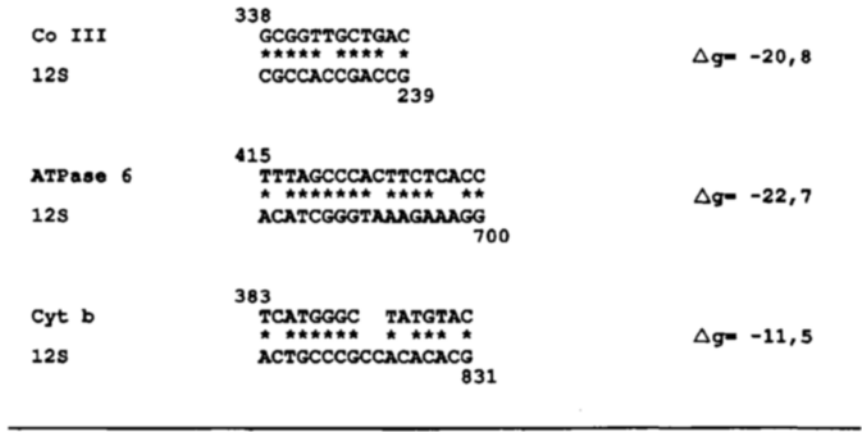

Fig. 7. Possible interactions between 12S small rRNA and mRNAs in rat. The distances from the 5′ end of the genes are indicated. Free energy (Δg) was calculated according to the method of Papanicolau et al. (1984). The statistical significance of the sequence complementarities, calculated according to the method of Smith et al. (1985), is 6 ± 1.52, 7 ± 13.1, and 7 ± 0.97, respectively, for the three sequences presented.

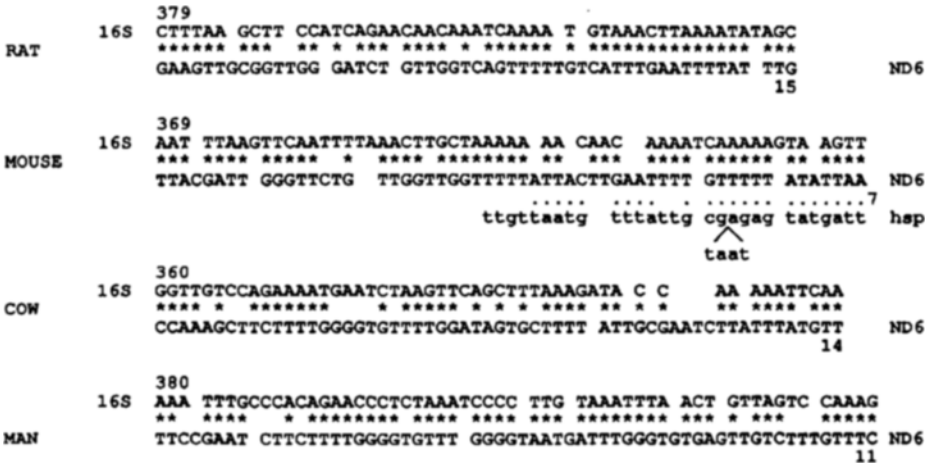

Fig. 8. Sequence complementarity between 16S rRNA and ND6 mRNA of mammalian mitochondria. The distances from the 5′ end of the genes are indicated. hsp = coding H-strand promoter (Chang and Clayton 1986).

the most conserved mitochondrial genes in all animals, but together with their counterparts from other sources, both mitochondrial (lower eukaryotes and plants) and nonmitochondrial (bacteria, chloroplasts, and cytoplasm) genes display similarity only at the level of secondary structure.

Our graphic representation of the folding patterns for the large and small rRNAs of rat mitochondria is essentially the same as those presented for human, mouse, and *Xenopus laevis* (Glotz et al. 1981; Kuntzel and Kochel 1981; Dunon-Bluteau and Brun 1986; Hixson and Brown 1986) with a few modifications (Fig. 6).

In the 12S rRNA gene we identified a number of interesting sites, three of which are illustrated in Fig. 7. At position 239 from the 5′ end, a sequence that is conserved almost perfectly in organisms ranging from *E. coli* to animal mitochondria is present. It is in an exposed loop and finds complementarity in some protein coding genes (COI, COIII, Cyt *b*, ND4, ND5). Another sequence, from base 700, which is conserved in small rRNA genes and in addition is localized in an unpaired region, is complementary to a sequence present in almost all mRNA genes (e.g., ATPase 6; see also Attimonelli et al. 1985; Saccone et al. 1985). Finally, a sequence of 17 bases starting from position 831 near the 3′ end is nearly completely identical in primary and secondary structures in all the organisms compared, ranging from *E. coli* to mammalian mitochondria. This site also displays a statistically significant complementarity to a sequence present in some reading frames (COI, Cyt *b*). The degree of complementarity of these sequences varies from gene to gene, but in any case the Δg values are much higher than those exhibited by the Shine–Dalgarno type of interaction (Shine and Dalgarno 1975). The evolutionary conservation and the high statistical significance of these interactions strongly suggest their important function in the regulation of the translation. Watson–Crick base pairing at sites within the mRNA molecules probably are necessary to ensure the correct protein elongation (as has also been suggested by Weiss et al. 1988), especially for mitochondrial systems that do not possess a Shine–Dalgarno type of interaction.

In the 16S rRNA we found a region complementary to ND6 mRNA, and again this feature is conserved evolutionarily in mammals in spite of a divergent primary sequence in both genes (Fig. 8). Part of the same region of ND6 in some organisms (e.g.,

mouse) is complementary also to the HSP, the sequence that acts as promoter in the transcription of the H-strand ($\Delta g = -9.9$ kcal). We recall that in mammalian mitochondria the transcription of the two strands is symmetric and that the transcription of rRNAs and mRNA is controlled probably by different mechanisms (Attardi 1985). We might envisage a regulatory mechanism via antisense RNA mediated by early products of mitochondrial transcription. In other words, the products of early transcription ND6 mRNA and 16S rRNA for the L- and H-strand, respectively, could modulate their own level by interacting with each other or with their promoter. The availability of cytoplasmic ribosomal proteins that use mt rRNA to form active ribosomes could represent a key element in the regulation of mitochondrial biogenesis.

Conclusion

Detailed comparative studies performed with the sequences of five vertebrate mitochondrial genomes have revealed new and interesting properties of mitochondrial gene products.

The high compactness and economic organization of animal mtDNA leaves little space for regulatory regions. The results reported in this paper suggest that the regulatory signals for expression of the mitochondrial genome are contained in the primary and higher structural elements of the products themselves. The validity of our hypothesis is based mainly upon the evolutionary conservation of peculiar structural features. More importantly, the data here reported offer a valid theoretical basis for future experiments on mitochondrial regulatory mechanisms.

Acknowledgments. We thank Prof. R. Gallerani's group for having provided us with some sequence data and Drs. M. Attimonelli and G. Pesole for help with computer analysis. This work has been supported partly by Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie, CNR, and partly by a Ministero Pubblica Istruzione grant.

References

Akins RA, Lambowitz AM (1987) A protein required for splicing group I introns in *Neurospora* mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. Cell 50:331–345

Attardi G (1985) Animal mitochondrial DNA: an extreme example of genetic economy. Int Rev Cytol 93:93–145

Attimonelli A, Lanave C, Sbisà E, Preparata G, Saccone C (1985) Multisequence comparisons in protein coding genes. Search for functional constraints. Cell Biophys 7:239–250

Barnes WM, Bevan M, Son PH (1983) Kilo-sequencing: creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13. Methods Enzymol 101: 98–122

Barrell BG, Bankier AT, Drouin J (1979) A different genetic code in human mitochondria. Nature 282:184–194

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Brown GG, Gadaleta G, Pepe G, Saccone C, Sbisá E (1986) Structural conservation and variation in the D-loop containing region of vertebrate mitochondrial DNA. J Mol Biol 192: 503–511

Cantatore P, Saccone C (1987) Organization, structure and evolution of mammalian mitochondrial genes. Int Rev Cytol 108: 149–206

Cantatore P, De Benedetto C, Gadaleta G, Gallerani R, Kroon AM, Holtrop M, Lanave C, Pepe G, Quagliariello C, Saccone C, Sbisà E (1982) The nucleotide sequences of several tRNA genes from rat mitochondria: common features and relatedness to homologous species. Nucleic Acids Res 10:3279–3289

Cantatore P, Gadaleta MN, Roberti M, Saccone C, Wilson AC (1987a) Duplication and remoulding of tRNA genes during evolutionary rearrangement of mitochondrial genomes. Nature 329:853–855

Cantatore P, Roberti M, Morisco P, Rainaldi G, Gadaleta MN, Saccone C (1987b) A novel gene order in the *Paracentrotus lividus* mitochondrial genome. Gene 53:41–54

Chambers I, Harrison PR (1987) A new puzzle in selenoprotein biosynthesis: selenocysteine seems to be encoded by the 'stop' codon UGA. Trends Biochem Sci 12:255–256

Chang DD, Clayton DA (1986) Precise assignment of the heavy-strand promoter of mouse mitochondrial DNA: cognate start sites are not required for transcriptional initiation. Mol Cell Biol 6:3262–3267

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecules of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. J Mol Evol 22:252–271

Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. Nature 317:819–822

Dunon-Bluteau D, Brun G (1986) The secondary structures of the *Xenopus laevis* and human mitochondrial small ribosomal subunit RNA are similar. FEBS Lett 198:333–338

Gaines G, Rossi C, Attardi G (1987) The excised leader of human cytochrome c oxidase subunit I mRNA which contains the origin of mitochondrial DNA light-strand synthesis accumulates in mitochondria and is polyadenylated. Mol Cell Biol 7:925–931

Glotz C, Zwieb C, Brimacombe R, Edwards K, Kossel H (1981) Secondary structure of the large subunit ribosomal RNA from *Escherichia coli, Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. Nucleic Acids Res 9:3287–3306

Gortz G, Feldmann H (1982) Nucleotide sequence of the cytochrome b gene and adjacent regions from rat liver mitochondrial DNA. Curr Genet 5:221–225

Gouy M, Gautier C, Attimonelli M, Lanave C, Di Paola G (1985) ACNUC—a portable retrieval system for nucleic acid sequence database: logical and physical designs and usage. CABIOS 1:166–172

Himeno H, Masaki H, Kawai T, Ohta T, Kumagai I, Miura K-I, Watanabe K (1987) Unusual genetic code and a novel gene structure for tRNA ser(AGY) in starfish mitochondrial DNA. Gene 56:219–230

Hixson JE, Brown WM (1986) A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution and phylogenetic implications. Mol Biol Evol 3:1–18

Horowitz S, Gorovsky MA (1985) An unusual genetic code in nuclear gene of *Tetrahymena*. Proc Natl Acad Sci USA 82: 2452–2455

Koike K, Kobayashi M, Taira M, Yoshida E, Imai M (1982) Nucleotide sequence and evolution of the cytochrome b gene

containing the ochre termination codon from rat mitochondria. Gene 20:177–185

Kuntzel H, Kochel G (1981) Evolution of rRNA and origin of mitochondria. Nature 293:751–755

Lanave C, Preparata G, Saccone C (1985) Mammalian genes as molecular clocks? J Mol Evol 21:346–350

Lawrence CB, McDonnell DP, Ramsey WJ (1985) Analysis of repetitive sequence elements containing tRNA-like sequences. Nucleic Acids Res 13:4239–4252

Maly P, Brimacombe R (1983) Refined secondary structure models for the 16S and 23S ribosomal RNA of *Escherichia coli*. Nucleic Acids Res 11:7263–7286

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

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

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

Papanicolau C, Gouy M, Ninio J (1984) An energy model that predicts the correct folding of both the tRNA and the 5S RNA molecules. Nucleic Acids Res 12:31–44

Pepe G, Holtrop M, Gadaleta G, Kroon AM, Cantatore P, Gallerani R, De Benedetto C, Quagliariello C, Sbisà E, Saccone C (1983) Nonrandom patterns of nucleotide substitutions and codon strategy in the mammalian genes coding for identified and unidentified reading frames. Biochem Int 6:553–563

Roe BA, Ma D-P, Wilson RK, Wong JF-H (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J Biol Chem 260:9759–9774

Saccone C, Cantatore P, Gadaleta G, Gallerani R, Lanave C, Pepe G, Kroon AM (1981) The nucleotide sequence of the large ribosomal RNA gene and adjacent tRNA genes from rat mitochondria. Nucleic Acids Res 9:4139–4147

Saccone C, Attimonelli M, Sbisà E (1985) Primary and higher order structural analysis of animal mitochondrial DNA. In: Quagliariello E (ed) Achievements and perspectives of mitochondrial research, vol II: biogenesis. Elsevier Science Publishers BV (Biomedical Division) Amsterdam, The Netherlands, p 37

Saccone C, Attimonelli M, Sbisà E (1987) Structural elements highly preserved during the evolution of the D-loop containing region in vertebrate mitochondrial DNA. J Mol Evol 26: 205–211

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA (1980) Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143:161–178

Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. Nature 254:34–38

Smith TF, Waterman MS, Burks C (1985) The statistical distribution of nucleic acid similarities. Nucleic Acids Res 13: 645–656

Spritzl M, Hartman T, Meissner F, Moll J, Vorderwulbecke T (1987) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 15:r53–r188

Stormo GD, Schneider TD, Gold LM (1982) Characterization of translational sites in *E. coli*. Nucleic Acids Res 10:2971–2996

Trifonov EN (1987) Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16S rRNA nucleotide sequences. J Mol Biol 194:643–652

Weiner AM, Maizels N (1988) tRNA-like structures tag the 3′ ends of genomic RNA molecules for replication: implication for the origin of protein synthesis. Proc Natl Acad Sci USA (in press)

Weiss RB, Dunn DM, Dahlberg AE, Atkins JF, Gesteland RF (1988) Reading frames switch caused by base-pair formation between the 3′ end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. EMBO J 7:1503–1507

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowki KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc 26:375-400

Wolstenholme DR, Macfarlane JL, Okimoto R, Clary DO, Wahleithner JA (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc Natl Acad Sci USA 84:1324–1328

Received May 19, 1988/Revised and accepted November 22, 1988