Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 5368–5371, June 1994
Evolution

A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria

(transcription termination signal/mitochondria/mitochondrial ribosomal RNA/motif conservation)

JOSÉ R. VALVERDE, ROBERTO MARCO, AND RAFAEL GARESSE*

Departamento de Bioquímica, Instituto de Investigaciones Biomédicas, Facultad de Medicina, Universidad Autónoma de Madrid, c/Arzobispo Morcillo 4, 28029 Madrid, Spain

Communicated by Arthur Kornberg, January 24, 1994

ABSTRACT **A search of sequence data bases for a tridecamer transcription termination signal, previously described in human mtDNA as being responsible for the accumulation of mitochondrial ribosomal RNAs (rRNAs) in excess over the rest of mitochondrial genes, has revealed that this termination signal occurs in equivalent positions in a wide variety of organisms from protozoa to mammals. Due to the compact organization of the mtDNA, the tridecamer motif usually appears as part of the 3′ adjacent gene sequence. Because in phylogenetically widely separated organisms the mitochondrial genome has experienced many rearrangements, it is interesting that its occurrence near the 3′ end of the large rRNA is independent of the adjacent gene. The tridecamer sequence has diverged in phylogenetically widely separated organisms. Nevertheless, a well-conserved heptamer—TGGCAGA, the mitochondrial rRNA termination box—can be defined. Although extending the experimental evidence of its role as a transcription termination signal in humans will be of great interest, its evolutionary conservation strongly suggests that mitochondrial rRNA transcription termination could be a widely conserved mechanism in animals. Furthermore, the conservation of a homologous tridecamer motif in one of the last 3′ secondary loops of nonmitochondrial 23S-like rRNAs suggests that the role of the sequence has changed during mitochondrial evolution.**

Animal mtDNA transcription has been investigated thus far mainly in mammalian cells. Promoters for transcription of the heavy and light strands are located in the displacement loop (D-loop) near the origin of replication (1–3). The ribosomal RNAs (rRNAs) located adjacent to the promoter region are transcribed 15–100 times in excess of the rest of the mitochondrial genes. This difference in the abundance of mitochondrial transcripts is the consequence of transcription termination of the rRNAs. Precise mapping with nuclease S1 in humans shows that the ends of the transcript map inside the $tRNA^{Leu(UUR)}$ gene (4, 5). A tridecamer sequence present in this tRNA, 5′-TGGCAGAGCCCGG-3′, can function bidirectionally, directing human mitochondrial 16S rRNA formation *in vitro* (5, 6), and a protein factor that specifically binds to this sequence has been purified (7, 8). The relevance of the termination signal has been supported by recent findings that point mutations in this sequence are associated in humans to a neurological syndrome characterized by myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS), which may be a consequence of the impairment of mitochondrial transcription (9). Here we show that this termination signal is a generally conserved motif in mtDNA across the animal kingdom and that it has evolved probably from a similar tridecamer motif present in a conserved secondary stem-loop at the 3′ end of the 23S-like rRNA present in most prokaryotic and eukaryotic genomes.

The Termination Signal Is Conserved in Animal mtDNA

In all vertebrate mtDNA sequences now available in the European Molecular Biology Laboratory (EMBL) and GenBank data bases, the tridecamer sequence is conserved in the $tRNA^{Leu(UUR)}$ gene located 3′ downstream adjacent to the 16S rRNA (Fig. 1). The similar mitochondrial genomic organization in vertebrates (10, 11) and the fact that the sequence conservation in the $tRNA^{Leu(UUR)}$ is very high (12) make it difficult to differentiate whether the conservation is due to evolutionary restrictions on this termination signal or to the tRNA structure itself. Therefore, we looked for the tridecamer sequence near the 3′ end of phylogenetically more distant organisms. For instance, in arthropod mtDNA, although the actual gene content is maintained, important rearrangements in the gene order have occurred. In this case, the $tRNA^{Leu(UUR)}$ gene maps between the cytochrome oxidase subunit I (COI) and III genes (13, 14). The $tRNA^{Leu(CUN)}$ maps in the 3′ region downstream of the 16S rRNA (14, 15). Fig. 1 shows that a sequence with clear homology to the vertebrate termination signal is present in the $tRNA^{Leu(CUN)}$ of all arthropod mtDNAs sequenced to date. Conservation is higher in the first 7 nt and the final guanidine in the tridecamer and is strictly conserved in *Drosophila*, *Locusta*, and *Artemia*. This heptamer sequence is located in the equivalent positions (8 to 14) of vertebrate $tRNA^{Leu(UUR)}$ and invertebrate $tRNA^{Leu(CUN)}$ genes. However, this region is not conserved between the different mitochondrial tRNAs. When we compare the sequence of the equivalent regions of both mammalian $tRNA^{Leu}$ genes, TGGCAGA in $tRNA^{Leu(UUR)}$ and AAGGATA in $tRNA^{Leu(CUN)}$, only three out of the seven positions are maintained. Thus, the presence of the heptamer in the arthropod $tRNA^{Leu(CUN)}$ cannot be related to tRNA structure conservation but instead is related to conservation of the transcription termination signal.

This interpretation is strengthened by data from echinoderm mtDNA analyses. When compared with other phyla, even more important rearrangements have occurred in the genome organization (16, 17). In particular, 15 tRNA genes cluster in a region that contains the putative replication origin. No tRNA is located 3′ downstream from the 16S rRNA. Instead, the gene encoding the COI maps in this region of the mtDNA. By using S1 nuclease-mapping analysis (18), the 16S rRNA transcription has been reported to terminate 3–5 nt inside the COI gene. Fig. 1 shows that a heptamer sequence homologous to the mammalian tridecamer is also found at the 5′ end of the COI gene. Interestingly, the termination motif is oriented in the opposite direction in

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

Abbreviations: rRNA, ribosomal RNA; COI, cytochrome oxidase subunit I.

*To whom reprint requests should be addressed.

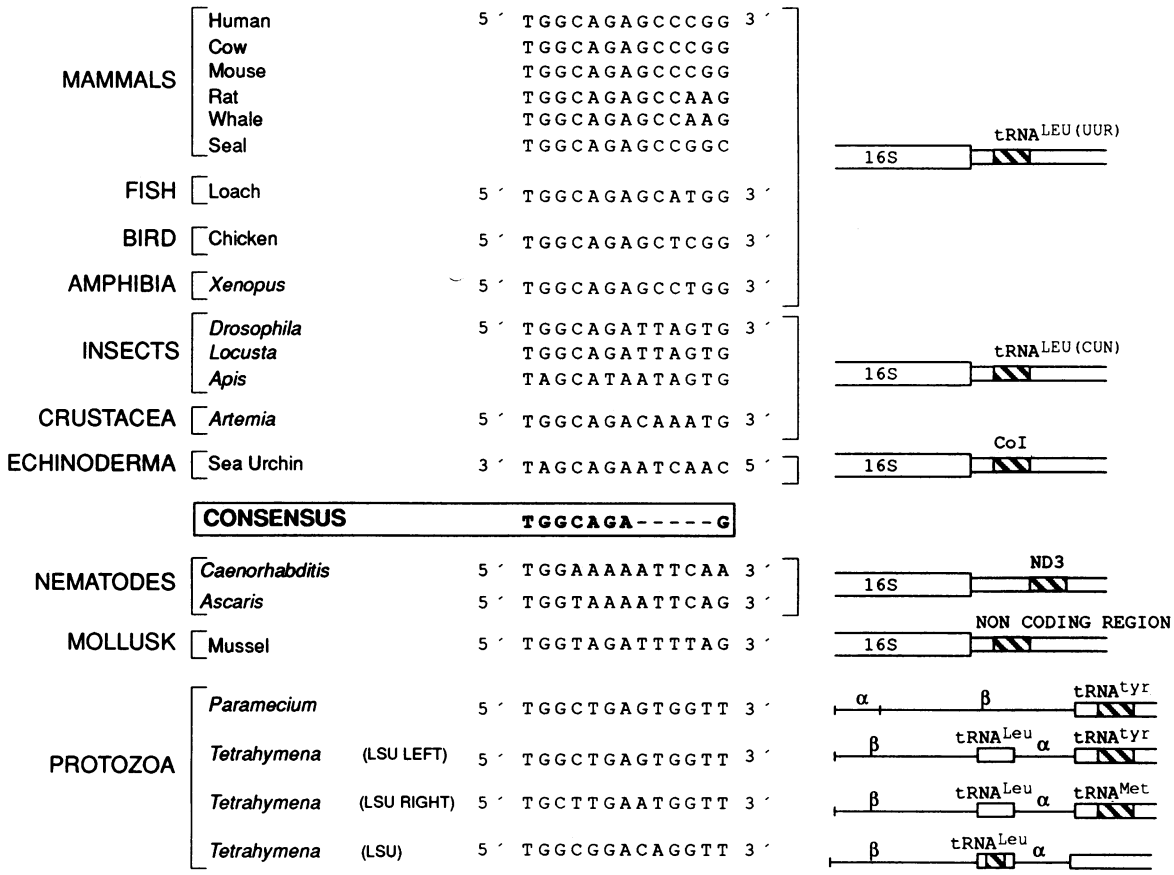

FIG. 1. Conservation of the termination signal in mtDNA from different phyla. A heptanucleotide sequence with high similarity to the first 7 nt of the human transcription termination signal can be found 3′ downstream of 16S rRNA in a variety of mtDNAs representing different phyla. The gene or region of the mitochondrial genome 3′ adjacent to the large rRNA (16SrRNA or LSU rRNA) is schematically shown at right; the position of the conserved sequence is boxed hatched. A consensus sequence has been deduced for vertebrates, arthropods, and echinoderma; the first 7 nt are maintained in other phyla with one to three mismatches. When the conserved region lies within a tRNA gene, it is always located from position 8 to 14 of the tRNA sequence. The equivalent conserved region starts in position 5 of the COI gene (sea urchin), position 135 of the ND3 gene (nematodes), and position 46 of the noncoding region (mussel). These sequences were obtained from EMBL and GenBank under the following accession numbers: human (*Homo sapiens*, V00662); cow (*Bos taurus*, V00654); mouse (*Mus domesticus*, V00711); rat (*Rattus norvegicus*, X61145); whale (*Balaenoptera physalus*, X61145); seal (*Phoca vitulina*, S37044); loach (*Crossostoma lacustre*, M91245); chicken (*Gallus gallus*, X52392); *Xenopus* (*Xenopus laevis*, X02890); *Drosophila* (*Drosophila yakuba*, X032340; *Drosophila melanogaster*, X400610); Locusta (*Locusta migratoria*, X05287); *Apis* (*Apis mellifera*, L06178); *Artemia* (*Artemia franciscana*, X69067); sea urchin (*Paracentrotus lividus*, J04815; *Strongylocentrotus purpuratus*, X12631); *Caenorhabditis* (*Caenorhabditis elegans*, X54252); *Ascaris* (*Ascaris suum*, X54253); mussel (*Mytilus edulis*, M83756); *paramecium* (*Paramecium tetraurelia*, X15917); *Tetrahymena* (*Tetrahymena pyriformis* M58010–M58011).

echinoderms, in agreement with the finding that in mammalian mtDNA this motif can function bidirectionally (6). Six out of the seven positions are identical to the conserved-box TGGCAGA, and beginning 5 nt inside the COI gene, this box appears in the exact position where the 16S rRNA transcript terminates *in vivo*.

Heptamer sequences potentially related to the termination signal were also found 3′ downstream of the 16S rRNA genes of invertebrates listed in the EMBL and GenBank data bases (Fig. 1), in spite of the significant rearrangements in their mtDNAs. Five positions of the heptamer were conserved in nematodes, 135 nt 3′ downstream of the 16S rRNA inside the ND3 gene; six positions out of seven were conserved in mollusks, 46 nt 3′ downstream of the 16S rRNA inside a large noncoding region. The only case in animal mtDNA where we have not found the heptamer sequence is in starfish mtDNA. Although the starfish belongs to the echinodermata, its mtDNA has major rearrangements compared with sea urchin mtDNA, and the 3′ region of the 16S rRNA is not adjacent to the COI gene but is adjacent to a relatively long noncoding region. In any case, our search has been done with partial information because the complete sequence of *Asterina pectinifera*, although determined, is not yet available (19).

The Termination Signal Is Also Conserved in Protozoan mtDNA

Extension of the analysis to additional nonanimal mtDNA sequences available in current data bases was difficult given the complexity and size of the mtDNAs of protoctista (see ref. 20 for the definition of this group), plantae, and fungi. Protozoan mtDNA, although larger than animal mtDNA, is not as complex as those of plantae and fungi. The mitochondrial large rRNA gene (LSU rRNA) is discontinuous in mtDNA of these organisms, consisting of two different sequences, the 5′ region α, ≈200 nt long, and the 3′ region β, ≈2300 nt long. In *Paramecium tetraaurelia* the $tRNA^{Tyr}$ is located 3′ downstream of the LSU rRNA (21). In *Tetrahymena pyriformis* the situation is more complex. (*i*) The β sequence is upstream from the α sequence, instead of being in the conventional 5′ (α) → 3′ (β) orientation (22, 23). There are two copies of the gene existing as inverted repeats close to the two ends of the linear mitochondrial genome of *Tetrahymena*. A $tRNA^{Leu}$ gene is located in the spacer region separating the α and β subunits in both copies. The $tRNA^{Met}$ gene is located 3′ downstream of the α sequence in the right repeat, and the $tRNA^{Tyr}$ gene is located 3′ downstream of the α sequence in the left repeat. As shown in Fig. 1, tridecamer

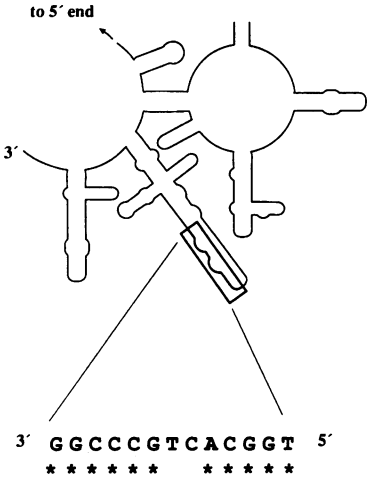

FIG. 2. Conserved 23S rRNA structure element containing a tridecamer with high identity to the mammalian termination signal. The consensus secondary structure (23, 24) of the 3′ end of 23S rRNA is shown schematically. The region containing the sequence with identity to the mammalian mitochondrial termination signal is boxed, and the *Escherichia coli* sequence is shown. Stars indicate positions conserved compared with the termination motif defined in human mtDNA.

sequences are present in all of these tRNA genes. The initial heptamer is conserved in all of them (six out of seven positions), with the one exception being $tRNA^{Met}$ of *Tetrahymena*, where only four of seven positions are maintained. Nevertheless, the remaining positions in the tridecamer are conserved.

The 23S-Like rRNA Contains a Sequence Homologous to the Termination Signal

In the large 23S-like rRNA subunit present in most nonanimal mitochondrial genomes, a tridecamer motif similar to the termination signal appears in a conserved secondary-structure stem-loop located at the 3′ end of the molecule (Fig. 2). This tridecamer exists in all the 23S-like rRNAs (except in animal 16S rRNAs) included in the secondary-structure rRNA data base (release 2.0; refs. 24 and 25). This data base includes a wide representation of large-subunit rRNAs from eubacteria (18), archaebacteria (13), plastids (13), mitochondria (3), and eukarya (4). The conservation is quite noticeable, ranging from 35% to 85% (Fig. 3). For example, the relatively high conservation of the mammalian tridecamer termination sequence compared with the tridecamer present in the 3′ secondary structure region of the *E. coli* 23S rRNA is shown in Fig. 2.

We also note that although animal 16S rRNA and other 23S-like rRNAs, including the prokaryotic 23S rRNA, can be folded in similar secondary-structure models (24, 25), the animal mitochondrial 16S rRNAs are smaller and have lost several secondary-structure elements. Among them, the stem-loop that contains the tridecamer sequence similar to the termination signal does not exist. In line with the endo-

Organism	Sequence	Matches
Human mtDNA (termination signal)	TGGCAG-AGCCCGG	13
Eubacteria (38% - 84%)		
Thermotoga maritima	CGGCATACGCCGGG	9
Thermus thermophilus	GGGCATAAGCTGGG	9
Rhodobacter sphaeroides	CGGCAG-TGCTGGG	9
Pseudomonas cepacia	TGGCAT-CGCCGGG	10
Escherichia coli	TGGCAC-TGCCCGG	11
Frankia sp. #1	GGGCAT-GGCTGGC	7
Micrococcus luteus	GTGCAT-GGCTGAT	5
Bacillus stearothermophilus	GGGCAC-CGCTGGG	8
Archaebacteria (61% - 77%)		
Archaeoglobus fulgidus	GGGCAT-CGCCGGG	9
Halobacterium halobium	GGGCATGTGCCGAG	8
Halococcus morrhuae	GGGCACGTGCCGGG	9
Methanococcus vannielii	AGGCAC-TGCCGAG	8
Desulfurococcus mobilis	GGGCAA-CGCCGGG	9
Termofilum pendens	GGGCAC-TGCCGGG	9
Sulfolobus solfataricus	GGGCAA-TGCCGGG	9
Chloroplast (46% - 61%)		
Chlamydomonas reinhardtii	TGGGGA-ACGCTGG	8
Chlorella ellipsoidea	GGGGATACGCTGGG	8
Astasia longa	CAGAATATGCTGGG	7
Nicotiana tabacum	CGGTAAACGCTGGG	8
Zea mays	CGGTAAACGCTGGG	8
Pisum sativum	CGGTAAACGCTGGG	8
Marchantia polymorpha	CGGTAAACGCTGGG	8
Eukarya (61%)		
Sacharomyces cerevisiae	AGGCAT-TGCCGCG	8
Giardia ardeae	GGGCAG-CCGCCGC	8
Plant mitochondria (77%)		
Zea mays	TAGCAG-CGCCGGG	10

FIG. 3. Comparison of the 23S-like rRNA sequence motif similar to the mitochondrial transcription termination signal. Aligned sequences of the equivalent region of 23S-like rRNAs (boxed in Fig. 2) from different organisms were compared, and the number of matches (excluding a single insertion/deletion introduced to maximize identity), compared with the human termination signal, were scored. The data have been obtained from the 23S and 23S-like rRNA data bank (release 2.0, 1993; supplied by Michael Gray, ref. 23). Only some representative examples of each group are shown.

symbiotic hypothesis of the origin of mitochondriae (26), the absence of this element raises the interesting possibility that while present in the original 23S-like rRNA, it was lost during evolution as the animal mtDNA became more compact. Nevertheless, the tridecamer motif was maintained 3′ downstream of the large rRNA subunit, where it has acquired another role as a transcription termination signal. In fact, the protozoa may represent an intermediate stage in this evolutionary trend, because for *Tetrahymena*, the 3′ stem-loop secondary structure containing the tridecamer homologue in 23S-like rRNAs is still present in the LSU rRNA, but the conserved sequence has been lost, changing into the sequence AAAGTAGGTTT in the equivalent position of the stem-loop.

Conclusion

In summary, our search for the tridecamer transcription termination signal originally found in mammalian mtDNA has shown that a homologous sequence is present in equivalent positions in a wide variety of organisms, from protozoa to mammals, for which extensive mtDNA-sequence information exists. Although the tridecamer sequence has diverged in phylogenetically widely separated organisms, a well-conserved heptamer TGGCAGA, the mitochondrial rRNA termination box, can be defined. The probability of finding such a heptamer is 1 in every 16,000 nt—i.e., once in the standard animal mitochondrial genome. Thus, the strict conservation of the sequence and its position relative to the mitochondrial rRNA genes across such a wide phylogenetic variety of species strongly supports the significance of its role as a transcription termination signal in human mtDNA.

This work is dedicated to the joint memory of the two founding fathers of Spanish Biochemistry, Alberto Sols (1990) and Severo Ochoa (1993). We thank Drs. Howard Jacobs and Julio Montoya for critically reading the first version of the manuscript and for suggestions and encouragement. This work has been supported by grants of the Fondo de Investigaciones Sanitarias de la Seguridad Social (93/0339) and Plan Nacional del Espacio (ESP91-627).

1. Montoya, J., Gaines, G. L. & Attardi, G. (1983) *Cell* **34,** 151–159.
2. Chang, D. D. & Clayton, D. A. (1984) *Cell* **36,** 635–643.
3. Yoza, B. K. & Bogenhagen, D. F. (1984) *J. Biol. Chem.* **259,** 3909–3915.
4. Van Etten, R. A., Bird, J. W. & Clayton, D. A. (1983) *J. Biol. Chem.* **258,** 10104–10110.
5. Christianson, T. W. & Clayton, D. A. (1988) *Mol. Cell. Biol.* **8,** 4502–4509.
6. Christianson, T. W. & Clayton, D. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6277–6281.
7. Kruse, B., Narasimhan, N. & Attardi, G. (1989) *Cell* **58,** 391–397.
8. Daga, A., Micol, V., Hess, D., Aebersold, R. & Attardi, G. (1993) *J. Biol. Chem.* **268,** 8123–8130.
9. Hess, J. F., Parisi, M. A., Bennett, J. L. & Clayton, D. A. (1991) *Nature (London)* **351,** 236–239.
10. Attardi, G. & Schatz, G. (1988) *Annu. Rev. Cell Biol.* **4,** 289–337.
11. Desjardins, P. & Morais, R. (1990) *J. Mol. Biol.* **212,** 599–634.
12. Gadaleta, G., Pepe, G., De Candia, G., Quagliariello, C., Sbisà, E. & Saccone, C. (1989) *J. Mol. Evol.* **28,** 497–516.
13. de Bruijn, M. H. L. (1983) *Nature (London)* **304,** 234–241.
14. Clary, D. O. & Wolstenholme, D. R. (1985) *J. Mol. Evol.* **22,** 252–271.
15. Garesse, R. (1988) *Genetics* **118,** 649–663.
16. Jacobs, H. T., Elliot, D. J., Math, V. B. & Fargharson, A. (1988) *J. Mol. Biol.* **202,** 1069–1082.
17. Cantatore, P., Roberti, M., Rinaldi, M., Gadaleta, N. & Saccone, C. (1989) *J. Biol. Chem.* **264,** 10965–10975.
18. Elliot, D. J. & Jacobs, H. T. (1989) *Mol. Cell. Biol.* **9,** 1069–1082.
19. Asakawa, S., Kumazawa, Y., Araki, T., Himeno, H., Miura, K. & Watanabe, K. (1991) *J. Mol. Evol.* **32,** 511–520.
20. Margulis, L. & Schwartz, K. V. (1988) *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth* (Freeman, New York).
21. Cummings, D. J. (1992) *Int. Rev. Cytol.* **141,** 1–64.
22. Heinomen, T. Y. K., Schnare, M. N., Young, P. G. & Gray, M. W. (1987) *J. Biol. Chem.* **262,** 2879–2887.
23. Heinomen, T. Y. K., Schnare, M. N. & Gray, M. W. (1990) *J. Biol. Chem.* **265,** 22336–22341.
24. Gutell, R. R., Schnare, M. N. & Gray, M. W. (1992) *Nucleic Acids Res.* **20** (Suppl.), 2095–2109.
25. Gutell, R. R., Gray, M. G. & Schnare, M. N. (1993) *Nucleic Acids Res.* **21,** 3055–3074.
26. Gray, M. W. (1989) *Trends Genet.* **5,** 294–299.