J Mol Evol (1987) 26:205–211

Journal of Molecular Evolution

Structural Elements Highly Preserved During the Evolution of the D-Loop-Containing Region in Vertebrate Mitochondrial DNA*

C. Saccone,[1] M. Attimonelli,[1] and E. Sbisà[2]

[1] Dipartimento di Biochimica e Biologia Molecolare, Università, 70126 Bari, Italy
[2] Centro Studi sui Mitocondri e Metabolismo Energetico, % Dipartimento di Biochimica e Biologia Molecolare, Università, 70126 Bari, Italy

Summary. A detailed comparative study of the regions surrounding the origin of replication in vertebrate mitochondrial DNA (mtDNA) has revealed a number of interesting properties. This region, called the D-loop-containing region, can be divided into three domains. The left (L) and right (R) domains, which have a low G content and contain the 5′ and the 3′ D-loop ends, respectively, are highly variable for both base sequence and length. They, however, contain thermodynamically stable secondary structures which include the conserved sequence blocks called CSB-1 and TAS which are associated with the start and stop sites, respectively, for D-loop strand synthesis. We have found that a "mirror symmetry" exists between the CSB-1 and TAS elements, which suggests that they can act as specific recognition sites for regulatory, probably dimeric, proteins. Long, statistically significant repeats are found in the L and R domains.

Between the L and R domains we observed in all mtDNA sequences a region with a higher G content which was apparently free of complex secondary structure. This central domain, well preserved in mammals, contains an open reading frame of variable length in the organisms considered.

The identification of common features well preserved in evolution despite the high primary structural divergence of the D-loop-containing region of vertebrate mtDNA suggests that these properties are of prime importance for the mitochondrial processes that occur in this region and may be useful for singling out the sites on which one should operate experimentally in order to discover functionally important elements.

Key words: Animal mitochondrial DNA — Origin of replication — Mirror symmetry — Mitochondrial regulatory region

Introduction

The animal mitochondrial (mt) genome is a closed circular molecule of about 5 μm contour length (about 16 kb) containing genes for ribosomal (r), transfer (t), and messenger (m) RNAs distributed between the two [heavy (H) and light (L)] DNA strands. The information for these mt products saturates the mt genome almost completely, leaving only a short and probably unique regulatory region whose length is different in various animal species. In vertebrates this region, which spans between the Phe-tRNA and the Pro-tRNA genes, is called the "D-loop-containing region," because, at the level of the heavy strand replication origin (O_H) contained within the D-loop, the nascent H-strand displaces the parental H-strand, creating a three-stranded structure called a displacement (D) loop. The promoters for heavy (HSP) and light (LSP) DNA strand transcription are also contained in this region, upstream in relation to the O_H (Clayton 1982, 1984; Attardi 1985).

The D-loop-containing region is considered to be the most rapidly evolving part of mtDNA which, on the whole, accumulates base substitutions, in-

Offprint requests to: C. Saccone
* Presented at the FEBS Symposium on Genome Organization and Evolution, held in Crete, Greece, September 1–5, 1986

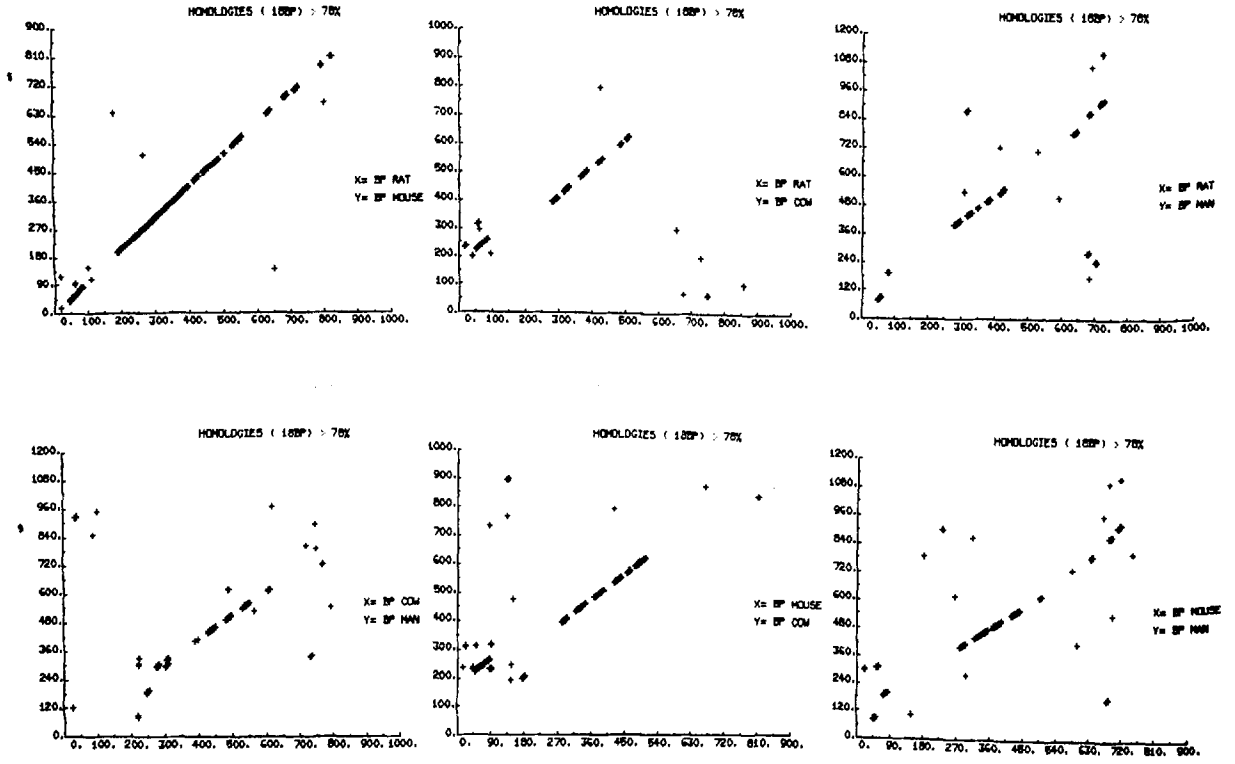


Fig. 1. Graphic representation of the sequence homology of the D-loop-containing region of mammalian mitochondrial DNA. The symbol + indicates the middle nucleotide position of two strings, 78% homologous, 18 nucleotides long. The 0 of the axes corresponds to the first position of the D-loop-containing region following the Pro-tRNA gene in all the organisms considered.

sertions, or deletions at a rate considerably faster than that of single-copy nuclear DNA (Dawid 1972; Brown et al. 1979; Ferris et al. 1981; Lanave et al. 1984, 1985).

We have recently sequenced the D-loop-containing regions of three rat mtDNAs, two from the species *Rattus norvegicus* and one from *Rattus rattus*. Comparisons made among the closely related sequences and with the sequences of mouse and other mammals have revealed a number of interesting features of both short- and long-term evolutionary change (Saccone et al. 1985; Brown et al. 1986).

We found that, although in comparison with more distantly related mtDNAs the D-loop region is the most divergent in the molecule, it does not diverge more than typical protein genes between the two species of rat, *R. norvegicus* and *R. rattus*. Another feature that has emerged from our comparisons of closely related mtDNAs is the fact that deletions and insertions occur approximately equally as frequently as base substitutions. Our data are in agreement with the observations of Greenberg et al. (1983) on the intraspecific sequence variability of the noncoding region of human mtDNA and indicate that the evolution of this mtDNA region does not proceed with continuity among species. When distantly related sequences are compared, however, we found that the primary sequence homology is confined to a central domain and to some conserved sequence blocks, which occur near the 5′ terminus [CSBs (Walberg and Clayton 1981)] or near the 3′ terminus [TAS (Doda et al. 1981)] of the D-loop strand and which are associated with the start and stop sites for DNA synthesis, respectively. At the level of these sites we found that, in spite of high primary structural divergence, the sequences can be folded in similar cloverleaf secondary structures whose evolutionary preservation suggests that they are of prime importance for processes occurring in the D-loop-containing region (Brown et al. 1986).

Thus, this region of the mitochondrial genome appears to be of particular interest both for its peculiar evolutionary dynamics and for its functional importance. Moreover, since the enzymes involved in mitochondrial transcription and replication are nuclear-coded, the study of this region can provide a model for the nuclear–mitochondrial concerted evolution of protein–nucleic acid interactions. This fact prompted us to perform a further, deeper analysis of this mtDNA segment in four mammalian species—rat, mouse, cow, and human—and in *Xenopus laevis*.

In this paper we describe a number of properties that characterize this region and try to define the sites on which one should operate experimentally to reveal functionally important elements.

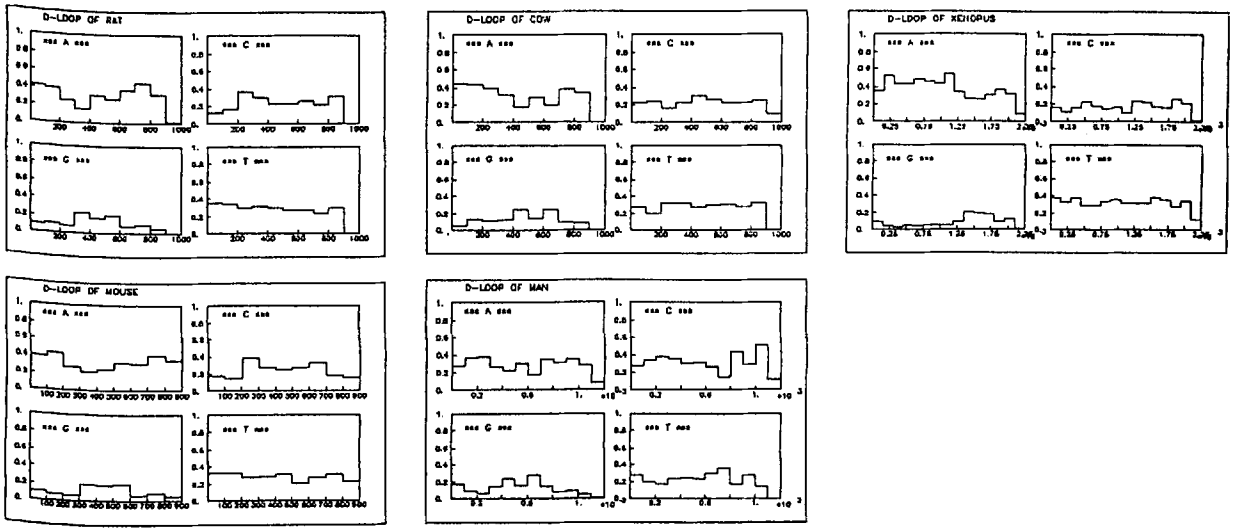


Fig. 2. Base distribution along the mitochondrial D-loop-containing region of vertebrates. The analysis is performed by counting the base occurrences over each 100 nucleotides for the mammals and over each 150 nucleotides for *X. laevis*. See legend to Fig. 1 for further details.

Table 1. Nucleotide frequencies

	Length (base pairs)	A %	C %	G %	T %
D-loop-containing region					
Rat	902	33	24	13	30
Mouse	879	34	25	12	29
Cow	910	33	24	14	29
Human	1122	30	33	14	23
Xenopus	2135	39	18	10	33
Three domains of the D-loop-containing region					
Left domain					
Rat	442	36	24	12	28
Mouse	418	36	24	12	28
Cow	342	29	26	14	31
Human	555	30	33	13	24
Xenopus	1350	45	16	6	33
Central domain					
Rat	206	21	32	19	28
Mouse	206	25	31	17	27
Cow	211	25	29	20	26
Human	201	27	31	18	24
Xenopus	450	27	19	19	35
Right domain					
Rat	254	37	18	10	35
Mouse	255	38	20	8	34
Cow	357	41	20	10	29
Human	366	33	33	12	22
Xenopus	335	34	22	11	33

Methods

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 using several programs (HOMT, DIRECT, HPLORUN) realized in our laboratory (C. Lanave and M. Attimonelli, unpublished). Sequence analyses also included use of graphic facilities which displayed the results in an easily intelligible way. The computer programs are all included in the analysis software, GLORIA (CABIOS in press).

Results

The D-loop-containing regions of the four mammals—rat, mouse, cow, and human—were compared using the computer program HOMT.

Figure 1 reveals the presence of a well-conserved domain, about 200 bp long, in the center of the region flanked by two less-conserved domains of variable length.

Figure 2 shows the base composition of the region (L strand) in the various organisms. Like the whole genome of vertebrates, this region is characterized by a low G content (10%–14%). However, in the central domain, which corresponds to the well-conserved sequence of Fig. 1, the percentage of G increases up to 20%, a value slightly higher than those found in the region of the mtDNA coding for rRNA (≅ 18%), tRNA (≅ 15%), and mRNA (≅ 12%) (Table 1). Moreover, using the computer program HPLORUN, which searches for possible runs, we have identified repeats as $(GG)_n$ (n ranging from 3 to 6) in vertebrates. These repeats are almost exclusively localized in the G-rich domain of the D-loop-containing region.

The search for long repeats along the same region was done using the computer program DIRECT. The statistical significance of the repeat lengths was evaluated according to Smith et al. (1985). We found that the shortest significant repeats in mammals are 18 bases long with four mismatches; in *Xenopus* repeats are 20 bases long with four mismatches. The central G-rich domain of mammals appears to be devoid of repeats in general, which instead are pres-

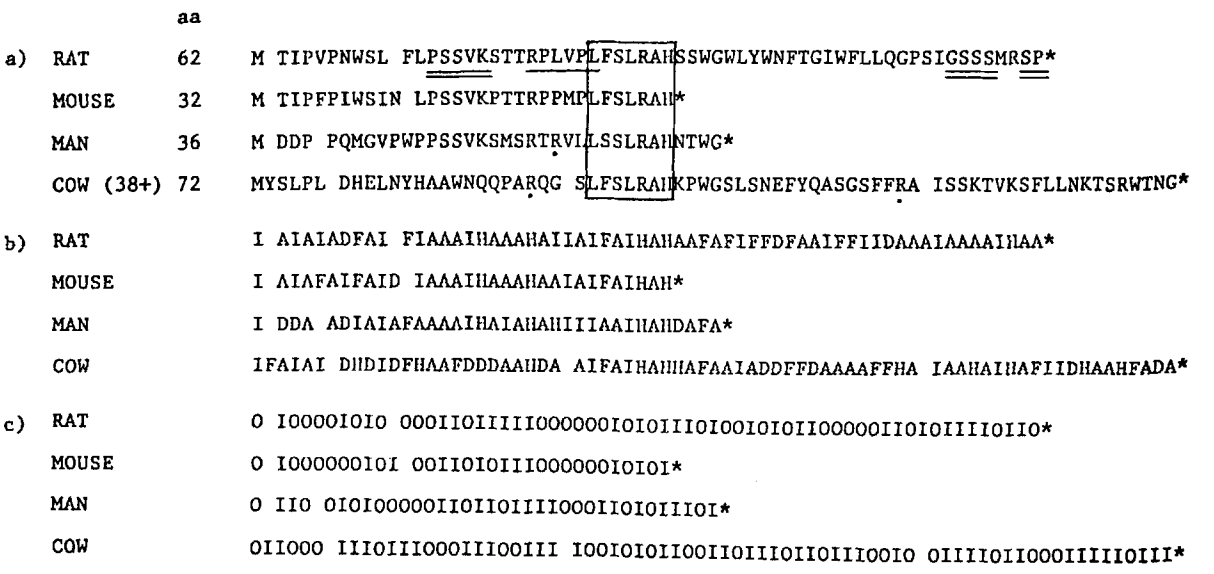
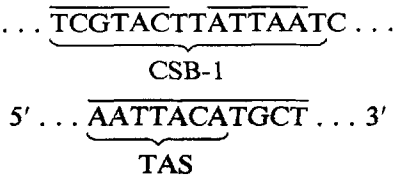


Fig. 3. Amino acid sequence comparison of the open reading frames present in the mitochondrial D-loop-containing region of mammals using different alphabets. **a** Best alignment among amino acids. The region of the ORF common to all the mammals has been boxed. The sequences doubly underlined and underlined in the rat ORF are homologous to the *E. coli* RNA polymerase β' chain and to human cytomegalovirus major immediate early gene, respectively. The dots indicate the AGA codon in human and the AGG codon in cow. **b** Best alignment according to functional alphabet. A = P, A, G, S, T (neutral weakly hydrophobic); D = Q, N, E, D, B, Z (hydrophilic acid amine); H = H, K, R (hydrophilic, basic); I = L, I, V, M (hydrophobic); F = F, Y, W (hydrophobic, aromatic); C = C (cross-link forming). **c** Best alignment according to hydropathic alphabet. I = hydrophilic; O = hydrophobic.

ent in the two flanking regions and are particularly evident in cow and human. In *Xenopus* repeats appear to be present in the first 1350 bases and also in the regions with low G content (data not shown).

We have previously reported (Saccone et al. 1985; Brown et al. 1986) that at the level of the 5′ terminus of the nascent H-strand DNA the sequences of the vertebrate mtDNAs, despite their high primary structural divergence, are capable of assuming similar cloverleaf secondary structural configurations (Brown et al. 1986). The evolutionary preservation of the potential to form such structures suggests that the structures are of prime importance for processes like mitochondrial transcription and regulation occurring in the D-loop-containing region. An additional interesting aspect of the sequences at the 5′ and 3′ D-loop strand ends is that a "mirror symmetry" exists between the TAS and CSB-1 elements. This symmetry, as it occurs in the mouse sequence, is as follows:



In an attempt to find a possible function for the G-rich conserved domain of the D-loop-containing region, which appears to be devoid of secondary structures, we have searched for the presence of possible reading frames in this DNA stretch. In all four mammals, proceeding counterclockwise (in the same direction as H-strand transcription), we found an open reading frame (ORF) of variable length. The ORF is 62 amino acids (aa) long in rat and 32 aa in mouse. In cow the ORF can start at several different sites, creating products ranging from 72 to 110 aa. These reading frames contain, however, two AGG codons, which, if used as stop codons, would reduce the product sizes by 51 or 22 aa. It is well known that the two arginine codons, AGG and AGA, have been suggested to act as stop codons in mammals (Attardi 1985). However, in the cow mtDNA, the AGG codon has never been used as a stop codon in mRNA. In humans the ORF is 36 aa long. It also contains an AGA codon which, if it acts as a terminator, should reduce the polypeptide size by 14 aa.

Figure 4 shows ORF comparisons among the four mammals using different alphabets. Great similarity is observed when the functional alphabet (see legend to Fig. 3) is used for the comparison. The codon strategy of this ORF is slightly different from the one adapted in the rest of the genome. In particular the C-ending codons exceed the T-ending codons, owing to the higher GC content of this region (see Table 1). The screening of protein data banks (NBRF) has revealed an interesting sequence homology of these ORFs with the RNA-polymerase β' chain of *Escherichia coli* (Ovchinnikov et al. 1982) and with the human cytomegalovirus major immediate early gene (Stenberg et al. 1984). In the *X. laevis* sequence

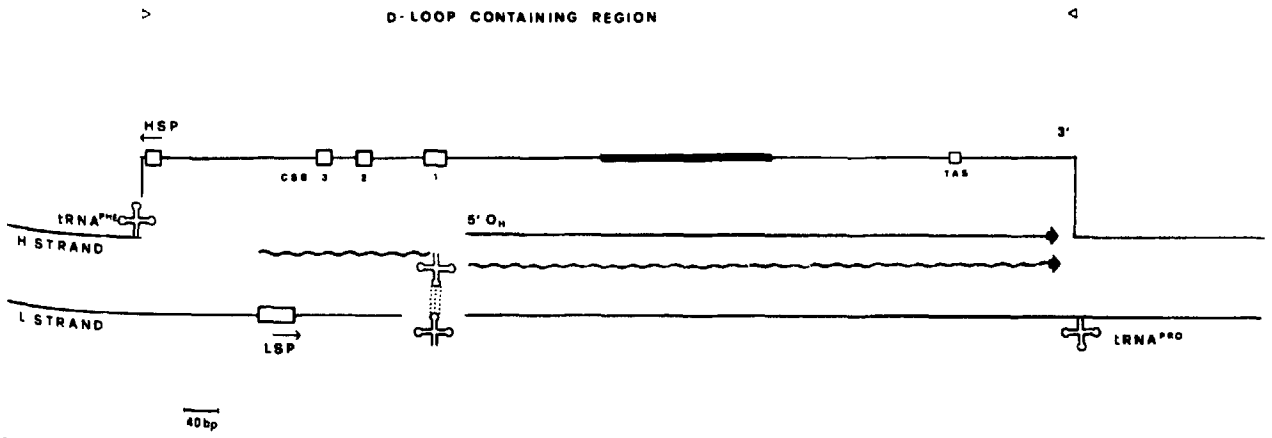


Fig. 4. General organization of the D-loop-containing region in vertebrates spanning from the Phe-tRNA to the ProtRNA genes. The central part (heavy line) is the most conserved region in mammals. We propose a model of regulation of the mt replication and transcription mechanisms that resembles that of Col E1 replication based on the interaction between the cloverleaf structures at the 5′ end of the new strand and other RNA molecules and/or protein factors. ——, DNA; ~~, RNA.

we found several ORFs. The longer ORF of 88 aa starts at position 924 and is at the right border of the L domain. Another ORF, 36 aa long, starting at position 1668, is localized in the G-rich region. This last ORF shows a slight degree of sequence similarity with the mammalian ORFs (data not shown).

Discussion

We can divide the D-loop-containing region into three domains on the basis of both base content (Table 1) and degree of conservation (Fig. 1). The left (L) domain, immediately adjacent to the Phe-tRNA gene, has different lengths in vertebrates and its base sequence greatly diverges except for the CSBs. It is probably the most important functional part of the region, containing O_H and the two promoters, HSP and LSP.

The central domain, approximately 200 bp long in mammals, is characterized by a lower A content and the highest G content of the whole L-strand DNA. The asymmetry in GC distribution between the two DNA strands (the L-strand is poor in G and the H-strand in C) has already been recognized as a peculiar feature of the mt genome in mammals. Since in mammalian mtDNA the majority of the products are transcribed from the H-strand, this means that they are very poor in G. This property is probably responsible for the codon strategies of mRNAs in which the use of G (Pepe et al. 1983) at the third codon position is extremely low. It also explains the behavior of the rate matrices obtained in the comparisons of silent codon sites of mammalian mRNAs, showing that during evolution the rate of replacement by G of bases other than G is not significantly different from zero (Lanave et al. 1984). The presence in the D-loop-containing region of a stretch relatively rich in G and well conserved during evolution is indicative of a functional constraint. It is noteworthy that in *X. laevis* mtDNA, whose base composition is on the whole different from that of mammals, this relatively G-rich stretch (Fig. 2) also exists (450 bp long).

The right (R) domain has the highest A content and the lowest G content; it spans the interval between the G-rich domain and the Pro-tRNA gene. It contains the site where the D-loop DNA strand stops and is of crucial importance because at this level, whether the synthesis of the H-strand should be arrested, as in resting molecules, or whether it should continue, thereby initiating the true replication cycle, is determined.

Another peculiar feature of the L and R domains is the presence of long repeats. Despite their high primary structural divergence, L and R domains contain well-preserved, thermodynamically stable, secondary structures (Saccone et al. 1985). Those present at the 5′ terminus of the nascent H-strand DNA have, in the most stable configuration, a cloverleaf structure (Brown et al. 1986) which resembles those of tRNAs and contains the CSB-1 element that has been suggested to act as the switching point between RNA and DNA synthesis (Chang et al. 1985). Also, at the 3′ terminus of the D-loop we find well-preserved secondary structures whose stable configuration may again be of cloverleaf type or may present multiple stems and loops (results not shown). The secondary structures at the 3′ end include the TAS sequences that have been found associated with the termination of D-loop synthesis (Doda et al. 1981). It is noteworthy that CSB-1 and TAS display the type of symmetry defined by Higgins and Ferro-Luzzi Ames (1982) as "mirror symmetry." Since a mirror symmetry does not allow secondary structures to form in either DNA or RNA,

the primary base sequence should be important for the nucleic acid–protein interaction. It must be noted, however, that the mirror symmetry is not a true symmetry because the sugar–phosphate backbone polarity is different in the two repeats. This implies that the protein–DNA interaction involving a mirror symmetry should be different from the protein–DNA interactions involving dyad symmetries (Ogata and Gilbert 1979; Anderson et al. 1981). Clearly, the functional significance of mirror symmetries remains to be demonstrated for bacterial and, in our case, for mitochondrial regulatory regions.

Finally, we want to stress the possible importance of an open reading frame in the central well-conserved domain of the D-loop-containing region. We identified several ORFs in the D-loop of mammals, but the only one common to all the organisms is the one reported in Fig. 3. It must be recalled that in other systems the translation of short RNA segments or the presence of small regulatory proteins plays a crucial role in the regulation of macromolecular processes like transcription and replication. For example, in the attenuation mechanism occurring in the transcription of bacterial operons, the translation of a short RNA segment acts as a modulator. In the ori C region of *E. coli* the replicative forks stop in a gene which codes for a protein involved in terminating the replication. The product has not been characterized, but a mutation, called dnaT, suggests the importance of this protein in the replication mechanism (Lewin 1983). The presence of ORFs in DNA segments containing the origin of replication has been reported also by other authors in mitochondria of different organisms. The ori sequence of *Paramecium* mtDNA contains four transcripts which encode three ORFs—209 aa, 113 aa, and 120 aa long, respectively (Pritchard et al. 1986). Short ORFs also exist in the ori sequences of yeast mtDNA. Those beginning with the AUG triplet and common to all the ori sequences are 27 aa long in the ori 3 and only 10 aa long in the ori 1, 2, and 5 for the presence of two ochre codons (De Zamaroczy et al. 1984).

As far as the replication mechanism is concerned, we recall that the initiation of replication of the Col E1 plasmid is under the control of a protein called Rop (or Rom) which is only 63 aa long (Tomizawa and Som 1984). In the replication of the plasmid pRBH1, a function similar to that of Rop is supported by a much bigger protein, Rep B, which is 231 aa long, but whose sequence is only slightly similar to that of Rop (Ano et al. 1986).

In Fig. 4 we propose a model of regulation for the D-loop-containing region that resembles that of Col E1 replication (Lacatena and Cesareni 1981; Tomizawa and Itoh 1981). The model is based on a possible interaction between the cloverleaf structures at the 5′ end of the new DNA strand and other RNA molecules and/or protein factors. This kind of interaction is essential for the regulation of DNA synthesis in plasmid systems since mutations that destroy the folding capacity lead to failure of replication (Masukata and Tomizawa 1984).

In our case we suggest that the interaction of the mt cloverleaf secondary structures with other RNA molecules and/or protein factors could be responsible for the modulation of both replication and transcription mechanisms (Fig. 4). Structural features that might support the present model include the sequence analysis of cow wherein we observed a possible base pairing between the secondary structure at the 5′ D-loop end and Pro-tRNA (data not shown).

Experiments performed in our laboratory (manuscript in preparation) clearly demonstrate the presence (in rat liver) of stable in vivo transcripts of the H-strand in the D-loop-containing region downstream from the Pro-tRNA gene. The size of the transcripts ranges between 200 and 300 and 800 and 1200 bases, and their precise mapping is now in progress. Characterization and isolation of the transcripts will allow us to study their ability to fold into secondary structures of the type described by us and to study their function.

To demonstrate the relative importance of the primary and secondary structural elements that we have described in this paper for processes involved in mitochondrial transcription and replication, it will ultimately be necessary to use in vitro systems—some already set up (Walberg and Clayton 1983; Shuey and Attardi 1985; Wong and Clayton 1985) and others to be developed—to perform site-directed mutagenesis experiments.

Acknowledgments. This work was supported by the Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie, by Progetto Strategico Biotecnologie of the Consiglio Nazionale delle Ricerche (Italy), and by Ministero della Pubblica Istruzione (40%).

References

Anderson S, Bankier AT, Barrell GB, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–464

Ano T, Imanaka T, Aiba S (1986) The copy number of *Bacillus* plasmid pRBH1 is negatively controlled by RepB protein. Mol Gen Genet 202:416–420

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

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

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76: 1967–1971

Chang DD, Hausweirth WW, Clayton DA (1985) Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. EMBO J 4:1559–1567

Clayton DA (1982) Replication of animal mitochondrial DNA. Cell 28:693–705

Clayton DA (1984) Transcription of the mammalian mitochondrial genome. Annu Rev Biochem 53:573–594

Dawid IB (1972) Evolution of mitochondrial DNA sequences in *Xenopus*. Dev Biol 29:139–151

De Zamaroczy M, Faugeron-Fonty G, Baldacci G, Goursot R, Bernardi G (1984) The ori sequences of the mitochondrial genome of a wild-type yeast strain: number, location, orientation and structure. Gene 32:439–457

Doda JN, Wright CT, Clayton DA (1981) Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc Natl Acad Sci USA 78:6116–6120

Ferris S, Brown WM, Davidson WS, Wilson AC (1981) Extensive polymorphism in the mitochondrial DNA of apes. Proc Natl Acad Sci USA 78:6319–6323

Gouy M, Gautier C, Attimonelli M, Lanave C, Di Paola G (1985) ACNUC—a portable retrieval system for nucleic acid sequence data-bases: logical and physical designs and usage. CABIOS 1:167–172

Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21:33–49

Higgins CF, Ferro-Luzzi Ames G (1982) Regulatory regions of two transport operons under nitrogen control: nucleotide sequences. Proc Natl Acad Sci USA 79:1083–1087

Lacatena RM, Cesareni G (1981) Base pairing of RNA I with its complementary sequence in the primer precursor inhibits Col E1 replication. Nature 294:623–626

Lanave C, Preparata G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rates. J Mol Evol 20:86–93

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

Lewin B (1983) Genes. Wiley J and Sons (eds) New York

Masukata H, Tomizawa J (1984) Effects of point mutations on formation and structure of the RNA primer for Col E1 DNA replication. Cell 36:513–522

Ogata RT, Gilbert W (1979) DNA-binding site of lac repressor probed by dimethylsulfate methylation of lac operator. J Mol Biol 132:709–728

Ovchinnikov Yu A, Monastyrskaya GS, Gubanov VV, Guryer SO, Salomatina IS, Shuvaeva TM, Lipkin VM, Sverdlov ED (1982) The primary structure of *E. coli* RNA polymerase. Nucleotide sequence of the rpoC gene and amino acid sequence of the B′ subunit. Nucleic Acids Res 10:4035–4044

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

Pritchard AE, Seilhamer JJ, Cummings DJ (1986) *Paramecium* mitochondrial DNA sequences and RNA transcripts for cytochrome oxidase subunit I, URF1, and three ORFs adjacent to the replication origin. Gene 44:243–253

Saccone C, Attimonelli M, Sbisà E (1985) Primary and higher order structural analysis of animal mitochondrial DNA. In: Quagliariello E, Slater EC, Palmieri F, Saccone C, Kroon AM (eds) Achievements and perspectives of mitochondrial research, vol II. Biogenesis. Elsevier BV (Biomedical Division), Amsterdam, p 37

Shuey DJ, Attardi G (1985) Characterization of an RNA polymerase activity from HeLa cell mitochondria, which initiates transcription at the heavy strand rRNA promoter and the light strand promoter in human mitochondrial DNA. J Biol Chem 260:1952–1958

Smith TF, Waterman MS, Burks C (1985) The statistical distribution of nucleic acid similarities. Nucleic Acids Res 13: 645–655

Stenberg RM, Thomsen RD, Stinski MF (1984) Structural analysis of the major immediate early gene of human cytomegalovirus. J Virology 49:190–198

Tomizawa J, Itoh T (1981) Plasmid Col E1 incompatibility determined by interaction of RNA I with primer transcript. Proc Natl Acad Sci USA 78:6096–6100

Tomizawa J, Som T (1984) Control of Col E1 plasmid replication: enhancement of binding of RNA I to the primer transcript by the Rom protein. Cell 38:871–878

Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421

Walberg MW, Clayton DA (1983) *In vitro* transcription of human mitochondrial DNA. Identification of specific light strand transcripts from the displacement loop region. J Biol Chem 258:1268–1275

Wong TW, Clayton DA (1985) *In vitro* replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. Cell 42:951–958