© 1990 Oxford University Press

Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans*

Ronald Okimoto, Jane L.Macfarlane and David R.Wolstenholme*
Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

Received May 21, 1990; Revised and Accepted August 31, 1990 EMBL accession nos X54252, X54254

ABSTRACT

Data obtained from alignments of nucleotide sequences of mitochondrial (mt) DNA molecules of the nematode worms *Ascaris suum* and *Caenorhabditis elegans* indicate that in six of the mt-protein genes of *A. suum* and three of the mt-protein genes of *C. elegans* TTG is used as the translation initiation codon. Also, GTT seems to be the translation initiation codon of the *A. suum* COIII gene. All of the five remaining *A. suum* mt-protein genes appear to begin with ATT and the remaining nine *C. elegans* mt-protein genes appear to begin with either ATT or ATA. Therefore, in contrast to all other metazoan mtDNAs sequenced so far, it is likely that none of the nematode mt-protein genes use the standard ATG translation initiation codon. Some *A. suum* and *C. elegans* mt-protein genes end in T or TA, suggesting that, as found in other metazoan mitochondria, 3'-terminal polyadenylation is occassionally necessary to generate complete translation termination codons in transcripts of nematode mt-protein genes.

INTRODUCTION

The mitochondrial (mt) genomes of metazoa are mainly circular DNA molecules with species-specific sizes in the range 14 to 39 kb. From complete and partial sequences of a number of metazoan mtDNA molecules it appears that all contain the same set of genes for two rRNAs, 22 tRNAs and 12 or 13 proteins (1–9). Metazoan mt-genomes have a number of distinguishing properties: gene arrangement is extremely compact, encoded tRNAs exhibit different degrees of structural change from the standard form, and there are various modifications to the genetic code (1–13).

Both translation initiation and translation termination of metazoan mt-protein genes have unusual features (1–10,14–17). Among many metazoan mtDNAs, triplets other than ATG (AUG) often appear to be used as translation initiation codons. All ATN codons are used in this way among mammals, *Drosophila yakuba* (except ATC) and sea urchin (except ATT). Also, ATAA has been suggested as the translation initiation codon of the COI gene of *Drosophila melanogaster* and *D. yakuba* mtDNAs, and GTG appears to be the translation initiation codon of a single protein gene in each of the mtDNAs of mouse, rat, *D. yakuba*, sea urchin and liver fluke. Some mt-protein genes in organisms from different metazoan phyla end in a T or TA rather than a complete termination codon. UAA codons in mature transcripts of these genes are generated by precise cleavage from multicistronic, primary transcripts followed by polyadenylation (15).

In this paper we present evidence for the occurrence of a novel initiation codon in metazoan mtDNA. We show that the triplet TTG, that normally specifies leucine, is the most plausible translation initiation codon of many of the mt-protein genes of two nematode worms, *Ascaris suum* and *Caenorhabditis elegans*. We further show that as in other metazoa, some of the mt-protein genes of these nematodes lack complete translation termination codons.

MATERIAL AND METHODS

Ascaris suum adults were obtained from the intestines of pigs at a local slaughterhouse. Mitochondria were isolated from body wall muscle, or from mature eggs, as described previously for *Drosophila* mitochondria (18). *Caenorhabditis elegans* (Bristol, N2 strain) were maintained, amplified and purified as given in references 19 and 20, except that *Klebsiella aerogenes* was used as the food source. Worms were ruptured by using a Dounce homogenizer (pestle A), and mitochondria were isolated as described for *A. suum*, except that mannitol was used in place of sucrose, and 0.1–0.2% bovine serum albumin was present in all solutions. *A. suum* and *C. elegans* mitochondria were lysed with 10% sarkosyl and mtDNAs were isolated by cesium chloride-ethidium bromide centrifugation (18).

DNA sequences were obtained (21) from sets of deletion clones (22,23) containing overlapping segments of the entire sequence of each complementary strand of the *A. suum* and *C. elegans* circular mtDNA molecules (R. Okimoto, J.L. Macfarlane, D.O. Clary and D.R. Wolstenholme, in preparation). Other details concerning sequencing, and computer assembly and analysis of sequences are given in reference 24.

* To whom correspondence should be addressed

RESULTS

Gene content, arrangement and structure

The nucleotide sequences of the entire circular mtDNA molecules of the parasitic nematode, *Ascaris suum* and the free-living soil nematode, *Caenorhabditis elegans* were obtained (Accession nos. X54252 and X54254). These molecules contain 14,284 ntp and 13,794 ntp, respectively, and each includes the genes for twelve proteins, two rRNAs and 22 tRNAs (Fig. 1). The protein genes are the same as twelve of the protein genes found in vertebrate, *Drosophila yakuba* and sea urchin mtDNAs (1,4,7). A gene for ATPase8, which in other metazoan mtDNAs precedes the ATPase6 gene, has not been located in either of the two nematode mtDNA molecules. The s-rRNA and l-rRNA genes are the shortest yet recorded among metazoan mtDNAs: 697 ntp and 953 ntp for *C. elegans*, and 701 ntp and 960 ntp for *A. suum*. All 22 tRNA genes are structurally unusual: in twenty, a simple loop of nucleotides replaces the TψC arm and variable loop, and in the remaining two (tRNA-ser(AGN) and tRNA-ser(UCN)), a simple loop of nucleotides replaces the dihydrouridine arm (6,13). Both the *A. suum* and *C. elegans* mtDNA molecules include a sequence that contains runs of AT dinucleotides, and in which genes have not been identified (the AT region: 886 ntp in *A. suum* and 466 ntp in *C. elegans*). Gene order is identical in the *A. suum* and *C. elegans* mtDNA molecules (except that the AT region is located between different genes (Fig. 1)), but differs from those of other sequenced metazoan mtDNAs. In both nematode mtDNA molecules all genes are transcribed in the same direction (Fig. 1). Data obtained from interspecific sequence comparisons indicate that in the *A. suum* and *C. elegans* genetic codes, TGA specifies tryptophan rather than termination, ATA specifies methionine rather than isoleucine, and AGA and AGG both specify serine rather than arginine (R. Okimoto, J.L. Macfarlane, D.O. Clary and D.R. Wolstenholme, in preparation).

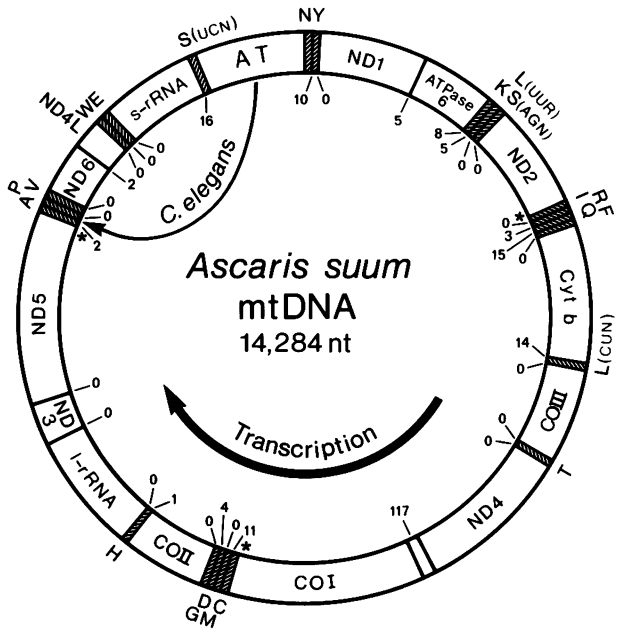

Figure 1. Gene map of the *Ascaris suum* mtDNA molecule (14,284 ntp). This molecule contains the genes for twelve proteins (Cyt b, cytochrome b; ATPase6, subunit 6 of the F_0 ATPase; COI-III, cytochrome *c* oxidase subunits I to III; ND1–6 and 4L, subunits 1–6 and 4L of the respiratory-chain NADH dehydrogenase), and the small and large subunit RNAs (s-rRNA and l-rRNA)). Ten of the protein genes were identified from similarities of amino acid sequences predicted from the various open reading frames to amino acid sequences of mouse and *D. yakuba* mt-protein genes (3,4). The remaining two nematode protein genes, ND4L and ND6, were tentatively identified from considerations of minimal amino acid sequence similarities, hydropathic profile comparisons with mouse and *D. yakuba* mt-protein genes, and size. Hatched areas indicate the locations of 22 tRNA genes identified by the single letter amino acid code. All genes would be transcribed in the direction shown by the arrow. In the *Caenorhabditis elegans* mtDNA molecule (13,794 ntp), gene order is identical to that of the *A. suum* molecule, except that the AT region (466 ntp compared to 886 ntp in the *A. suum* molecule) is located between the tRNA-ala and the tRNA-pro genes (as shown, arrow). The region between the ND4 and COI genes (117 ntp in *A. suum*, 109 ntp in *C. elegans*) includes a sequence with the potential to form a stable hairpin structure.

Translation initiation of nematode mt-protein genes

Of the two nematode mtDNA molecules, that of *A. suum* was sequenced first. The twelve open reading frames of this molecule were identified by amino acid sequence and hydropathic profile comparisons with the mt-protein genes of mouse and *D. yakuba* (Fig. 1). However, from these comparisons it was not possible to define with confidence the translation initiation codon of seven of the eight *A. suum* mt-protein genes that each follow either a tRNA gene or the l-rRNA gene (Fig. 2).

In vertebrate and *D. yakuba* mtDNAs many protein genes also follow tRNA genes (1,3–5). In each such case, the mt-protein gene putative translation initiation codon (ATN in mammals, ATR/T in *Drosophila* with rare exceptions; see Introduction) is located either immediately adjacent to or within a few nucleotides of the terminal nucleotide of the preceding tRNA gene. An ATG codon does not occur among the first 25 codons of the open reading frames of four of the eight *A. suum* mt-protein genes that follow a structural RNA gene; in the remaining four of these mt-protein genes the first ATG is 4, 6, 9, and 23 codons from the beginning of the open reading frame (Fig. 2). The *A. suum* Cyt b gene sequence begins with an ATT codon that immediately follows the preceding tRNA-phe gene. In the remaining seven *A. suum* mt-protein genes that follow a structural RNA gene, 3, 9, 15, 18, 18, 24 and 60 nt separate the first ATN codon from the beginning of the open reading frame. However, the open reading frame of each of five of these seven protein genes begins with a TTG (leucine) that is immediately adjacent to the 3′ nucleotide of a preceding tRNA gene. The most likely first codon in the ND3 open reading frame that follows the 3′ end of the l-rRNA gene is also TTG. In the remaining case, the COIII gene, a GTT codon (valine) immediately follows the preceding tRNA (leu(CUN)) gene. These observations suggested to us that TTG, and in one case GTT, are used as translation initiation codons in *A. suum* mtDNA. Support for this hypothesis was gained from observations made on the aligned sequences of the gene junctions of *A. suum* and *C. elegans* mtDNA molecules (Fig. 2).

In *C. elegans* mtDNA, of the six mt-protein gene-encoding open reading frames that are homologous to *A. suum* sequences that begin with TTG, two also begin with TTG (ND2 and ND4), but four begin with ATT (ND6, ND1, COII, ND3; Fig. 2). Also, the Cyt b gene open reading frame that in *A. suum* begins with ATT, begins with TTG in *C. elegans*. In two of these seven interspecific mt-protein gene comparisons, corresponding ATN codons do not occur in the first 25 codons (Fig. 2); in the remaining five, corresponding ATN codons are found at codon 2 (ND1), 6 (ND3), 9/14 (Cyt b), 15 (ND6), and 21 (COII).

An ATA codon is found at the beginning of the *C. elegans* COIII open reading frame consistent with the hypothesis that the translation initiation codon of the *A. suum* COIII gene is GTT.

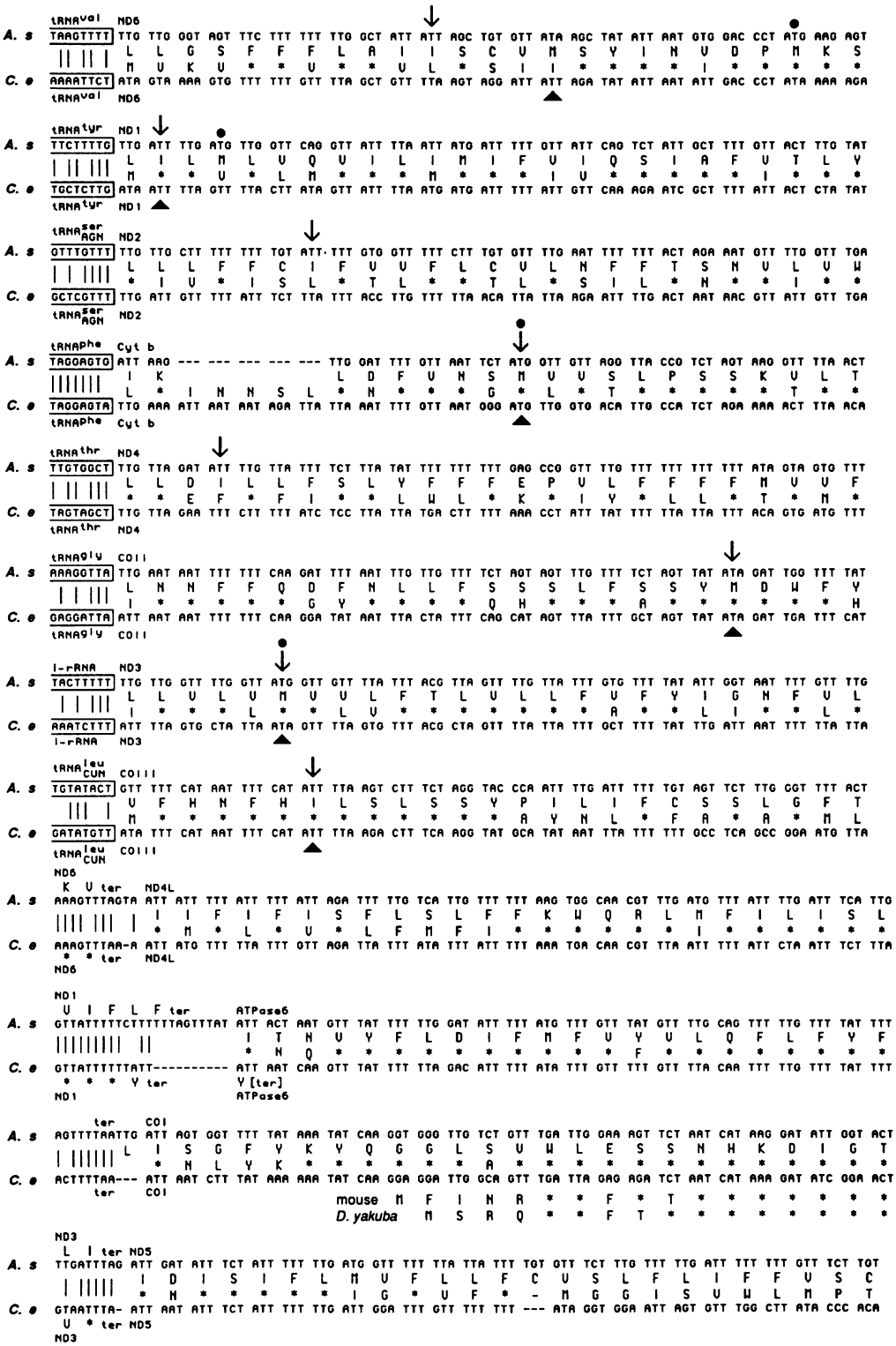

Figure 2. Comparisons of corresponding nucleotide sequences of *Ascaris suum* (A.s) and *Caenorhabditis elegans* (C.e) mtDNAs that contain the 5′ end-regions of each of the twelve mt-protein genes and the 3′ end-regions of the preceding genes. Gene abbreviation are: COI-COIII, cytochrome *c* oxidase subunit I-III; Cyt b, cytochrome b; ATPase6, subunit 6 of the F_0 ATPase complex; ND1-ND6 and ND4L subunits 1−6 and 4L, of the respiratory-chain NADH dehydrogenase complex. Predicted amino acid sequences are shown in the one letter code. An asterisk in a *C. elegans* sequence indicates the same amino acid as occurs in the *A. suum* sequence. Vertical lines indicate nucleotide sequence similarities in nematode gene 3′ end-regions. A dash indicates the absence in one nematode sequence of a nucleotide that occurs in the corresponding, other nematode sequence. A translation termination codon is indicated by ter. The [ter] in the ND1-ATPase6 gene junction region indicates a second possible termination codon for the ND1 gene that would overlap the 5′ end of the ATPase6 gene. The 3′ end-regions of tRNA genes and the large rRNA gene (l-rRNA) are boxed. All genes are transcribed in the same direction, left to right. The solid arrowheads beneath the aligned *A. suum* and *C. elegans* mt-protein sequences that follow a structural RNA gene indicate the most 5′, similarly located ATN codons. The dots and arrows above the *A. suum* mt-protein sequences indicate the most 5′ located ATG and ATN codons, respectively, in these sequences.

This interpretation is further supported by the finding that the following twelve codons in the *A. suum* and *C. elegans* sequences predict identical amino acids. A less likely candidate for the initiation codon in each of the nematode COIII genes is a TTG codon that occurs three and four codons upstream, respectively, within the preceding *A. suum* and *C. elegans* tRNA-leu(CUN) gene sequences.

In the *A. suum* and *C. elegans* mtDNA molecules there is little ambiguity regarding the initiation codon of the ND4L, ATPase6 and ND3 genes, each of which follows another protein gene (Fig. 2). Alignment of the *A. suum* and *C. elegans* sequences indicates corresponding ATT codons in the ND4L open reading frames beginning three and two nt, respectively, downstream from the termination codon of the preceding ND6 genes. Corresponding ATT codons occur in the *A. suum* and *C. elegans* ND5 genes immediately following the termination codon (TAG) and partial termination codon (TA; see below), respectively, of the preceding *A. suum* and *C. elegans* ND3 genes. The ATPase6 genes of *A. suum* and *C. elegans* seem most likely to initiate at corresponding ATT codons. In the *A. suum* sequence this codon is separated by five nt from the upstream termination codon (TAG) of the ND1 gene; in *C. elegans* it immediately follows a T nucleotide that appears to be an incomplete termination codon (see below). However, it is also possible that in *C. elegans*, the ND1 gene terminates with the TAA codon that overlaps the putative ATT initiation codon in the sequence 5′ TATTAA (Fig. 2, Table 1).

Upstream from the COI genes in *A. suum* and *C. elegans* mtDNAs are sequences of 117 ntp and 109 ntp that lack genes but include a segment with the potential to fold into a stable stem and loop structure (Fig. 1). Amino acid sequences predicted from the 5′ end-region of the *A. suum* and *C. elegans* COI open reading frames have high similarity to each other and also to the amino acid sequences predicted from the corresponding regions of the mouse and *D. yakuba* COI genes (Fig. 2). The codons most likely to act in translation initiation of the *A. suum* and *C. elegans* COI genes are corresponding ATTs located eight and nine codons upstream from the aligned positions of the putative translation initiation codons of the mouse and *D. yakuba* COI genes (Fig. 2). An in-frame translation termination codon is found two codons 5′ to the *A. suum* ATT and immediately 5′ to the *C. elegans* ATT. However, the codon that separates the termination codon from the ATT codon in the *A. suum* sequence is TTG, which, in view of the above discussion, also must be considered as a possible translation initiation codon for the *A. suum* COI gene. Corresponding ATNs (ATT and ATC) in the *A. suum* and *C. elegans* COI sequences, located 15 and 14 codons downstream from the beginnings of the aligned mouse and *D. yakuba* COI genes, seem less likely candidates for translation initiation codons as sequence similarity 5′ to these codons has been highly conserved in all four species considered. More likely, alternative possible initiation codons are corresponding TTGs located three and two codons downstream from the beginnings of the mouse and *D. yakuba* COI genes in the aligned sequences. However, similarity of amino acid sequences predicted from the *A. suum* and *C. elegans* nucleotide sequences that lie 5′ to these corresponding TTGs again favor the upstream corresponding ATT codons as the COI gene translation initiation codons.

Interestingly, of the six putative *A. suum* TTG translation initiation codons, three are followed by a second TTG and one is followed by ATT (Fig. 2). Also, one of the three *C. elegans* putative TTG initiation codons is followed by ATT. Further, of the putative ATN translation initiation codons, one in *A. suum* is followed by ATT and two in *C. elegans* are followed by ATT or ATG. These observations leave open the possibility for alternative translation initiation sites in some of the nematode mt-protein genes.

Table 1. Comparisons of putative initiation codons, and termination nucleotides and codons in the corresponding twelve mt-protein genes of *Ascaris suum* and *Caenorhabditis elegans*.

Gene	Initiation codons		Termination codons	
	A. suum	*C. elegans*	*A. suum*	*C. elegans*
Cyt b	ATT	TTG	TAG	TAA[f]
COI	ATT	ATT	TA(G)[c]	TA(G)[c]
COII	TTG	ATT	TAG	TAA
COIII	GTT	ATA	TAA	TAA
ATPase6	ATT	ATT[b]	TAG	TAA[f]
ND1	TTG	ATA	TAG[b]	T(TAA)[b]
ND2	TTG	TTG	T [d]	TAA
ND3	TTG[a]	ATT[a]	TAG	TA(A)[g]
ND4	TTG	TTG	TAA	TAA[f]
ND4L	ATT	ATT	TAG	TAG
ND5	ATT	ATT	T [e]	TAA
ND6	TTG	ATA	TAG	TAA

[a] The 3′ ends of the *A. suum* and *C. elegans* mt-l-rRNAs have not been directly mapped on the respective mtDNA molecules. The ND3 gene initiation codon assignments are therefore based entirely on similarities between the predicted amino acid sequences of the nematode ND3 gene-containing open reading frames, and similarities between these sequences and the amino acid sequences of the mouse and *D. yakuba* ND3 genes (3,4).

[b] In *A. suum* the ND1 gene termination codon (TAG) is separated by five nt from the ATT putative initiation codon of the ATPase6 gene, but in *C. elegans* the ND1 gene may terminate at a single T nucleotide that precedes the ATPase6 gene ATT putative initiation codon. Alternatively, the *C. elegans* ND1 gene may terminate at a TAA codon that overlaps the putative initiation codon of the ND1 gene in the sequence 5′ TATTAA (Fig. 2).

[c] (G) is the first nucleotide of the tRNA-^{cys} gene.

[d] This T is followed by TAT, the first three nucleotides of the tRNA-^{ile} gene.

[e] This T is followed by GGG, the first three nucleotides of the tRNA-^{ala} gene.

[f] Followed by a second TAA codon.

[g] (A) is the first nucleotide of the ATT initiation codon of the ND5 gene.

Translation termination of nematode mt-protein genes

In vertebrate, sea urchin, and *D. yakuba* mtDNAs, one or more of the protein genes end in a T rather than a termination codon, and this T is immediately adjacent to the 5′ terminal nucleotide of the sense strand of a tRNA gene (1–7,9). The primary transcription products of mammalian mtDNAs are multicistronic RNA molecules. Individual gene transcripts are produced by precise cleavage, and those protein gene transcripts ending in U acquire a complete termination codon by poladenylation (1,15).

An unambiguous, complete TAA or TAG codon is found at the end of each of nine *A. suum* and nine *C. elegans* mt-protein genes (Table 1). The *A. suum* ND2 and ND5 genes end in a T and the COI gene ends in TA and each of these mt-protein genes is followed by a tRNA gene. In the *C. elegans* mtDNA sequence, the COI gene also ends in TA, followed by a tRNA gene. However, a TA at the end of the *C. elegans* ND3 gene is adjacent to the putative ATT initiation codon of the ND5 gene in the sequence 5′ TAATT (Fig. 2). Also, the *C. elegans* ND1 gene appears to be terminated with a single T that is adjacent to the putative ATT initiation codon of the following ATPase6 gene. However, as discussed above, it remains possible that the termination codon of the *C. elegans* ND1 gene is a TAA codon that lies within a sequence overlap with the 5′ end of the ATPase6 gene. A small number of mt-protein genes ending in a TA that is adjacent to a putative ATN initiation codon are found in

vertebrate, sea urchin, and *D. yakuba* mtDNAs (1–7,9). It has been shown that in HeLa cells there is precise cleavage between the terminal UA of the ATPase6 gene transcript and the immediately adjacent AUG initiation codon of the COIII transcript (15). The occurrence in *A. suum* and *C. elegans* mtDNAs of exceptionally few non-coding nucleotides between genes (Fig. 1) is consistent with a transcription mechanism that involves production of primary multicistronic transcripts, and the finding that some mt-protein genes of these organisms end in T or TA suggests that their mitochondria also contain a cleavage-polyadenylation mechanism.

DISCUSSION

The data presented strongly support the interpretation that TTG (UUG) is used as the translation initiation codon of six of the twelve mt-protein genes of *A. suum* and three of the twelve mt-protein genes of *C. elegans* (Table 1). This is the first report of TTG as a putative translation initiation codon among mt-protein genes, and is the most extensive occurrence of putative TTG initiation codons among any particular set of genes, that we are aware of. Among metazoan mtDNA molecules that have been completely sequenced those of *A. suum* and *C. elegans* are the only ones that seem to totally lack protein genes with an ATG initiation codon. Five of the remaining six *A. suum* mt-protein genes have a putative ATT initiation codon and one, the COIII gene, has a putative GTT codon. Six *C. elegans* mt-protein genes appear to begin with ATT and three with ATA (Table 1).

Of the seven completely sequenced metazoan mtDNAs, only that of *Xenopus laevis* contains a full set of protein genes that all appear to begin with ATG (5). Among the four sequenced mammalian mtDNAs (man, mouse, cow and rat) it seems likely that between 9 and 10 of the mt-protein genes begin with ATG and the remainder begin with different ATN codons (1–3,9). The only exceptions are the ND1 genes of mouse and rat that have a putative GTT initiation codon (3,9). In the two other invertebrate mtDNAs sequenced, sea urchin and *D. yakuba*, eight and six protein genes, respectively, appear to begin with ATG. The remaining mt-protein genes of sea urchin appear to begin with ATA (three), ATC (one), and GTG (one), and of *D. yakuba* with ATA (one), ATT (four), GTG (one), and ATAA (one) (4,7).

The mt-protein genes of ciliated protozoa also seem to use all ATN codons, GTG and possibly GTA as translation initiation codons (25). A number of protein genes encoded in the maxicircular kinetoplast DNA (mtDNA) of the protozoan flagellates, *Trypanosoma brucei*, *Leishmania tarentolae* and *Crithidia fasciculata* were originally reported to start with non-ATG codons (26,27). However, it has recently been shown for many of these cases that a putative AUG initiation codon occurs in the mature protein gene transcript as the result of U addition (RNA editing; 28–30).

Translation initiation codons other than ATG are rare among fungal mt-protein genes. One exception is that the COIII gene of *Aspergillus nidulans* appears to begin with GTG (31). Also, it is likely that ATG is the initiation codon of all plant mt-protein genes so far sequenced.

Only ATG is found as the naturally occurring translation initiation codon of eukaryotic, nuclear-encoded genes, although ACG is the initiation codon of two eukaryotic viral genes (32,33). In contrast, rare alternative translation initiation codons are found among prokaryotic protein genes that include GTG, ATT and TTG (34–36).

From the results of experiments involving mutation of the *Escherichia coli* adenylate cyclase TTG initiation codon to GTG or ATG, Reddy et al. (37) concluded that this TTG initiation codon limited synthesis of adenylate cyclase in *E. coli*. It seems very unlikely that TTG plays any such role in regulating nematode mt-protein gene expression, because TTG is sometimes found as the putative initiation codon in different genes in *A. suum* and *C. elegans* mtDNAs. Also, in both species, some of the respiratory-chain NADH dehydrogenase (ND) subunit genes have a putative TTG initiation codon but others do not, and it might be expected that synthesis of different ND subunits (with the possible exception of ND2) is equimolar (38). In prokaryotic mRNAs, sequences located upstream from the translation initiation codon are essential for protein synthesis to start and, together with sequences downstream from the initiation codon, influence the efficiency of protein synthesis (34,35,39). Clearly different rules must apply for nematode and other metazoan mt-protein synthesis. The putative initiation codon is always at, or a few nucleotides downstream from the 5′ end of the mRNA and there is no indication of constant nucleotides downstream from the initiation codon (see Fig. 2). The absence of nucleotides upstream from the putative translation initiation codon in many nematode and other metazoan mt-mRNAs also excludes the possibility that initiation of metazoan mt-protein synthesis involves an mRNA/s-rRNA pairing analogous to the Shine-Dalgarno/ anti-Shine-Dalgarno sequence pairing found in prokaryotes. As all known *in vivo* protein synthesis begins with methionine or N-formyl-methionine (40), it seems more likely that in *A. suum* and *C. elegans* mitochondria the putative UUG initiation codons are recognized by the tRNA-f-met (anticodon CAU), than by the tRNA (anticodon CAA) that would be expected to recognize internal UUG codons.

ACKNOWLEDGEMENT

This work was supported by National Institutes of Health Grant No. GM 18375.

REFERENCES

1. Anderson,S., Bankier,A.T., Barrell,B.G., de Bruijn,M.H.L., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., Schreier,P.H., Smith,A.J.H., Staden,R. and Young,I.G. (1981) *Nature* **290**, 457–465.
2. Anderson,S., de Bruijn,M.H.L., Coulson,A.R., Eperon,I.C., Sanger,F. and Young,I.G. (1982) *J. Mol. Biol.* **156**, 683–717.
3. Bibb,M.J., van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) *Cell* **26**, 167–180.
4. Clary,D.O. and Wolstenholme,D.R. (1985) *J. Mol. Evol.* **22**, 252–271.
5. Roe,B.A., Ma,D.P., Wilson,R.K. and Wong,J.F.-H. (1985) *J. Biol. Chem.* **260**, 9759–9774.
6. Wolstenholme,D.R., Macfarlane,J.L., Okimoto,R., Clary,D.O. and Wahleithner,J.A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1324–1328.
7. Jacobs,H.T., Elliot,D.J., Math,V.B. and Farquharson,A. (1988) *J. Mol. Biol.* **202**, 185–217.
8. Garey,J.R. and Wolstenholme,D.R. (1989) *J. Mol. Evol.* **28**, 374–387.
9. Gadaleta,G., Pepe,G., De Candia,G., Quagliariello,C., Sbisa,E. and Saccone,C. (1989) *J. Mol. Evol.* **28**, 497–516.
10. Chomyn,A. and Attardi,G. (1987) *Current Topics in Bioenergetics* **15**, 295–329.
11. Barrell,B.G., Bankier,A.T. and Drouin,J. (1979) *Nature* **282**, 189–194.
12. Barrell,B.G., Anderson,S., Bankier,A.T., de Bruijn,M.H.L., Chen,E., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., Schreier,P.H., Smith,A.J.H., Staden,R. and Young,I.G. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3164–3166.
13. Okimoto,R. and Wolstenholme,D.R. (1990) EMBO J. **10**, in press.
14. Montoya,J., Ojala,D. and Attardi,G. (1981) *Nature* **290**, 465–470.
15. Ojala,D., Montoya,J. and Attardi,G. (1981) *Nature* **290**, 470–474.

16. de Bruijn,M.H.L. (1983) *Nature* **304**, 234–241.
17. Clary,D.O. and Wolstenholme,D.R. (1983) *Nucl. Acids Res.* **11**, 6859–6872.
18. Wolstenholme,D.R. and Fauron,C.M.R. (1976) *J. Cell Biol.* **71**, 434–448.
19. Brenner,S. (1974) *Genetics* **77**, 71–94.
20. Sulston,J.E. and Brenner,S. (1974) *Genetics* **77**, 95–104.
21. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
22. Dale,R.M.K., McClure,B. and Houchins,J.P. (1985) *Plasmid* **13**, 31–40.
23. Hong,G.F. (1982) *J. Mol. Biol.* **158**, 539–549.
24. Wahleithner,J.A. and Wolstenholme,D.R. (1987) *Curr. Genet.* **12**, 55–67.
25. Pritchard,A.E., Seilhamer,J.J., Mahalingam,R., Sable,C.L., Venuti,S.E. and Cummings,D.J. (1990) *Nucl. Acids Res.* **18**, 173–180.
26. Benne,R., van den Burg,J., Brakenhoff,J., de Fries,B., Nederlof,P., Sloof,P. and Voogd,A. (1985) In Achievements and Perspectives of Mitochondrial Research. Vol. II, Biogenesis, (eds.) Quagliariello,E., Slater,E.C., Palmieri,F., Saccone,C., Kroon,A.M. pp. 325–336. Elsevier Sciences Publishers, Amsterdam.
27. dela Cruz,V.F., Neckelmann,N. and Simpson,L. (1984) *J. Biol. Chem.* **259**, 15136–15147.
28. Feagin,J.E., Shaw,J.M., Simpson,L. and Stuart,K. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 539–543.
29. Shaw,J.M., Feagin,J.E., Stuart,K. and Simpson,L. (1988) *Cell* **53**, 401–411.
30. Simpson,L. and Shaw,J. (1989) *Cell* **57**, 355–366.
31. Netzker,R., Köchel,H.G., Basak,N. and Küntzel,H. (1982) *Nucl. Acids Res.* **10**, 4783–4794.
32. Becarra,S.P., Rose,J.A., Hardy,M., Baroudy,B.M. and Anderson,C.W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7919–7923.
33. Curran,J. and Kolakofsky,D. (1988) *EMBO J.* **7**, 245–252.
34. Stormo,G.D., Schneider,T.D. and Gold,L.M. (1982) *Nucl. Acids Res.* **10**, 2971–2996.
35. Kozak,M. (1983) *Microbiological Reviews* **47**, 1–45.
36. Eriani,G., Dirheimer,G. and Gaugloff,J. (1989) *Nucl. Acids Res.* **17**, 5725–5736.
37. Reddy,P., Peterkofsky,A. and McKenney,K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5656–5660.
38. Attardi,G., Chomyn,A., Doolittle,R.F., Mariottini,P. and Ragan,C.I. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 103–114.
39. Sprengart,M.L., Fatscher,H.P. and Fuchs,E. (1990) *Nucl. Acids Res.* **18**, 1719–1723.
40. Varshney,U. and RajBhandary,U.L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1586–1590.