The EMBO Journal vol.5 no.1 pp.21-28, 1986

The URF 5 gene of *Chlamydomonas reinhardtii* mitochondria: DNA sequence and mode of transcription

Poppo H.Boer[1] and Michael W.Gray

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

[1]Present address: Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N7, Canada

Communicated by C.Leaver

A gene homologous to unassigned reading frame (URF) 5 of the mammalian mitochondrial genome has been identified in the mitochondrial DNA of the unicellular green alga, *Chlamydomonas reinhardtii*. The algal URF 5 gene is closely flanked by the gene for subunit I of cytochrome oxidase (COI) and by an unidentified gene (ORF x). The URF 5 and ORF x genes are transcribed in the same direction, but opposite to that of the COI gene. Transcript analysis reveals a 1.9-kb mRNA whose major 5′ terminus maps to the putative URF 5 initiation codon and whose 3′ end abuts the 5′ end of the ORF x transcript. Characterization of other *C. reinhardtii* mitochondrial RNAs suggests a general pattern of abutting transcripts and mature mRNAs having little or no 5′ leader sequence. While this is reminiscent of post-transcriptional processing in animal mitochondria, different mechanisms must be employed in the two systems, since tRNA sequences (which appear to function as transcript processing signals in animal mitochondria) do not generally flank protein coding sequences in the *C. reinhardtii* mitochondrial genome. Nevertheless, characteristic secondary structure motifs do occur within the 3′-terminal regions of *C. reinhardtii* mitochondrial mRNAs, and their location close to mRNA termini suggests that such motifs may play a role in directing the precise endonucleolytic cleavage of long primary transcripts

Key words: Chlamydomonas reinhardtii/mitochondria/URF 5 gene/mRNA

Introduction

In unicellular green algae of the genus *Chlamydomonas*, the mitochondrial (mt) genome is among the smallest known, having a sequence complexity of only ~20 kbp (Boer *et al.*, 1985a). This is in marked contrast to the much larger and structurally more complex mitochondrial genomes of vascular land plants (reviewed by Leaver and Gray, 1982; Sederoff, 1984). *Chlamydomoans reinhardtii* contains a linear 15.8-kbp mtDNA species (Ryan *et al.*, 1978; Grant and Chiang, 1980) which we have termed '16-kbp DNA', and on which we have localized the genes for subunit I of cytochrome oxidase (COI), apocytochrome b (CYB), and the large subunit and small subunit rRNAs (Boer *et al.*, 1985a). These initial mapping studies indicated that although *C, reinhardtii* mtDNA is similar in size to animal mtDNA, the two genomes obviously differ in their organization; moreover, 16-kbp coding sequences appear more closely related to their counterparts in plant mtDNA than to the same genes in animal or fungal mtDNA (Boer *et al.*, 1985a; unpublished observations).

In mammals, unassigned reading frames (URFs) comprise over half of the coding capacity of the mtDNA (Bibb *et al.*, 1981; Anderson *et al.*, 1981, 1982). Open reading frames with amino acid sequence homology to some of the human mitochondrial URF proteins have also been identified in the mtDNA of *Drosophila* (Clary and Wolstenholme, 1983; de Bruijn, 1983; Clary *et al.*, 1984) and of the filamentous fungi *Aspergillus* and *Neurospora* (reviewed by Grivell, 1983), as well as in the maxicircle DNA of trypanosomatid protozoa (Hensgens *et al.*, 1984; de la Cruz *et al.*, 1984). Expression of all eight URF genes of mouse mtDNA has been demonstrated (Michael *et al.*, 1984). The evolutionary conservation of URF genes, coupled with the fact that they are indeed expressed (Mariottini *et al.*, 1983; Chomyn *et al.*, 1983; Michael *et al.*, 1984), suggests that the corresponding proteins fulfill important roles in mitochondrial function and biogenesis; however, what these roles may be has remained a question since the discovery of mitochondrial URFs. Recently, Chomyn *et al.* (1985) concluded that the gene products of six human mitochondrial URFs (1, 2, 3, 4, 4L, 5) are components of NADH:ubiquinone reductase of the inner mitochondrial membrane. Ise *et al.* (1985) also suggest that six mitochondrially synthesized subunits of the same enzyme complex are encoded by URFs in *Neurospora crassa*.

The presence in *C. reinhardtii* 16-kbp DNA of a long open reading frame, the 3′ region of which shows a low but significant amino acid identity with URF 2 proteins, has been reported by Pratje *et al.* (1984), and their identification of this URF 2 gene is supported by our sequence analysis and RNA mapping studies (unpublished). Here we report the DNA sequence of a region of *C. reinhardtii* mtDNA that encodes an URF 5 protein. We have also analysed transcript patterns to gain further insight into the mode of expression of the *C. reinhardtii* mitochondrial genome.

Results

Identification and DNA sequence of the URF 5 gene in C. reinhardtii mtDNA

A restriction map of the 15.8-kbp mitochondrial genome of *C. reinhardtii* (Figure 1) indicates the approximate positions of mitochondrial genes and their directions of transcription. An expansion of the URF 5 gene region (contained within *Hind*III fragment H1) illustrates the sequencing strategy. Northern blot hybridization of algal mtRNA (see below) suggested that three mRNAs are transcribed from this region, one of which encodes the previously identified COI protein (Boer *et al.*, 1985a). Translation of the DNA sequence into amino acid sequence (using the standard genetic code) revealed a long open reading frame of 1701 nucleotides, starting at position −63 (Figure 2). As discussed below, the likely initiation codon of the URF 5 protein is actually 63 nucleotides further downstream (position 1 in Figure 2). Upstream of this open reading frame, the COI gene is encoded on the other DNA strand, starting at position −82. Downstream of the URF 5 gene is a third open reading frame, termed ORF x, that runs out of the cloned region. Transcript

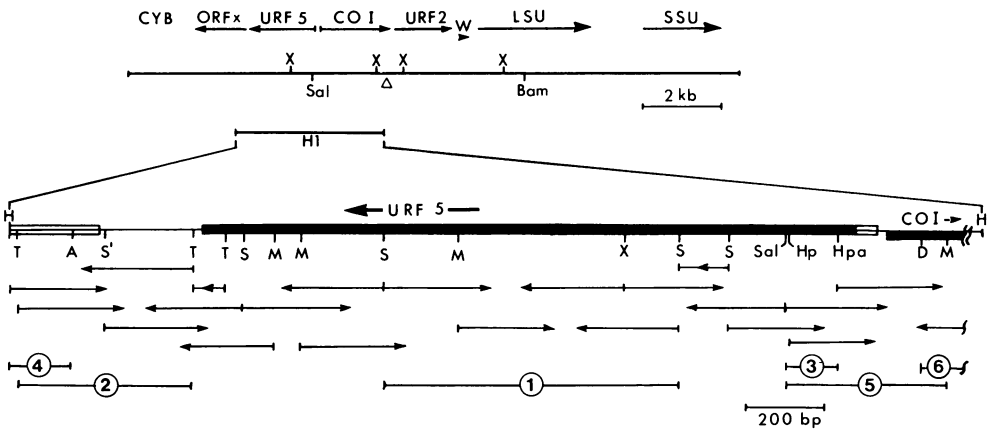

Fig. 1. (Upper) Restriction map of the *C. reinhardtii* mitochondrial genome, oriented as in Boer *et al.* (1985a). Arrows indicate known genes and their direction of transcription: W, $tRNA^{Trp}$; LSU, large subunit rRNA; SSU, small subunit rRNA. Open triangle, a cluster of three *Hind*III sites; Bam, *Bam*HI; Sal, *Sal*I; X, *Xba*I. (Lower) Expanded version of the URF 5 gene region contained within cloned *Hind*III restriction fragment H1. Here arrows show direction and extent of sequences determined, black bars indicate known genes, and open bars denote additional open reading frames. Only restriction sites used in the analysis are shown: A, *Ava*II; D, *Dde*I; H, *Hind*III; Hp, *Hinp*I; Hpa, *Hpa*I; M, *Msp*I; S, *Sau*3A; S', 'pseudo' *Sau*3A (i.e. a GCATC/GATGC sequence cloned into a *Bam*HI site); T, *Taq*I. The DNA probes used for transcript mapping are indicated by circled numbers.

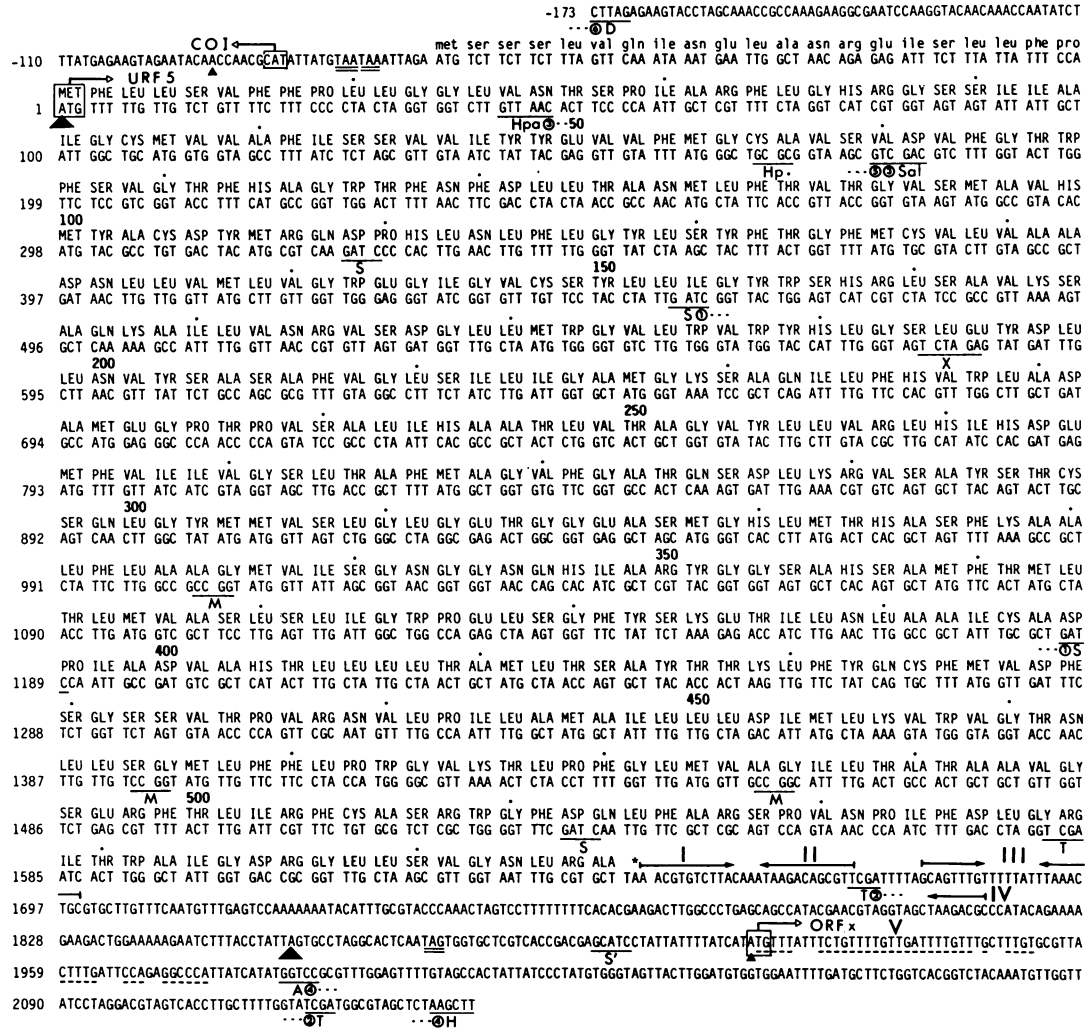

Fig. 2. Nucleotide sequence of the *C. reinhardtii* URF 5 gene region (note that the orientation is opposite to that shown in Figure 1). Nucleotide positions are numbered from the putative initiation codon of the URF 5 coding sequence. The derived amino acid sequence of the URF 5 protein (upper case, boldface numbering) and additional in-frame amino acids (lower case) are shown. Solid triangles indicate the most probable positions of transcript termini (see text), with triangle size reflecting RNA abundancy. Restriction sites used in the analysis (one-letter designation as in Figure 1) are underlined and the boundaries of the probes used for transcript mapping (see Figure 1) are indicated by the circled numbers. Putative initiation codons are boxed and in-frame stop codons are doubly underlined. Open arrows mark the start and relative orientation of coding sequences. Dashed underlining indicates nucleotides that are repeated within the 5′ region of the URF 2 gene. Arrows overline structural motifs I–IV (see text and Figure 8).

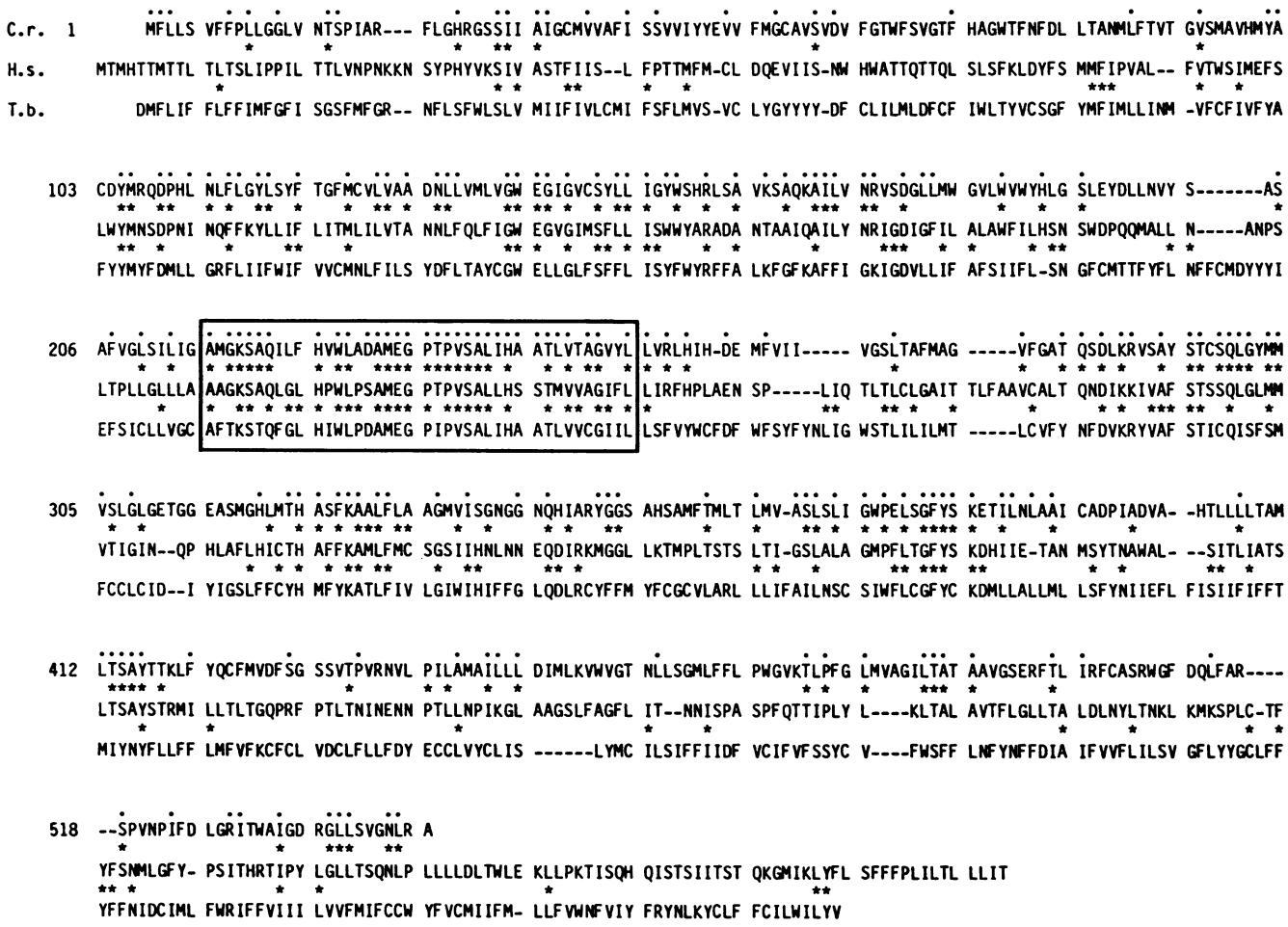

Fig. 3. Alignment of URF 5 protein sequences. The *C. reinhardtii* (C.r.)-derived amino acid sequence is compared with those of human (*Homo sapiens*, H.s.) and *Trypanosoma brucei* (T.b.) in the standard one-letter code. Black dots indicate amino acid identity in a three-way comparison among green algal, human and trypanosome sequences. Asterisks denote amino acid identity between the algal/human and human/trypanosome pairs. The box delineates a region of high amino acid sequence similarity that was used to align the three sequences.

mapping (see below) indicates that the ORF x mRNA begins at position 1857, with the first ATG codon appearing at position 1916 in the DNA sequence. The G+C content of the region shown in Figure 2 is 43.8%, close to the experimentally determined value of 46.9% for total 16-kbp DNA (Ryan *et al.*, 1978).

When the derived amino acid sequence of the long open reading frame was compared with those of proteins encoded by human mtDNA (Anderson *et al.*, 1981), its identity with URF 5 became apparent (Figure 3). Notably, the highly conserved region encompassing amino acid residues 216–255 in the algal URF 5 protein served as a focus for aligning its amino acid sequence with those of other URF 5 proteins. In a three-way comparison that includes the URF 5 protein of *Trypanosoma brucei* (Hensgens *et al.*, 1984), 195 out of 546 residues (35%) are identical between the *C. reinhardtii* and either the human or protozoan URF 5 proteins. In two-way comparisons, the algal URF 5 protein shares amino acid identities of 26% and 22% with its human and *T. brucei* analogues, respectively. The alignment used (Figure 3) requires the introduction of several gaps in the sequences to maximize amino acid homology; however, most of the gaps are shared between the *C. reinhardtii* and either human or *T. brucei* sequences. The algal URF 5 protein is very hydrophobic, with a 'polarity' (% polar amino acids: Capaldi and Vanderkooi, 1972) of only 27%, a value similar to that of other URF 5 proteins. Moreover, the hydrophobicity profiles of all these URF 5 proteins are much alike (cf. de la Cruz *et al.*, 1984; data for the algal protein not shown) and suggest that each could span a membrane a number of times.

The codon usage pattern of the *C. reinhardtii* URF 5 gene (Table I) closely parallels that of other *C. reinhardtii* mitochondrial protein coding sequences, including those specifying URF 2 and the URF 2 upstream reading frame (Pratje *et al.*, 1984), ORF x and COI (our unpublished observations). Codon usage is highly non-random and very distinctive. For example, in the single codon families ACN/Thr, GCN/Ala and GGN/Gly, a pyrimidine predominates in the third position, as it does in the mixed codon families for arginine and serine; in contrast, the GTA/Val and CCA/Pro codons are used frequently. Fifteen codons are not used at all in the algal URF 5 gene, and 11 of these do not appear in any of the *C. reinhardtii* mitochondrial protein genes we have sequenced to date. In contrast, three of these 'absent' codons (ATA, GAA, AGA) do appear among the codons that specify the 21 amino acid positions upstream and in frame with the URF 5 coding sequence. The amino acid sequence of this small reading frame does not align with other URF 5 proteins, and in fact is not encoded by the major URF 5 transcript (see below).

The indications of a specific evolutionary relationship between *C. reinhardtii* and plant mitochondrial genes (Boer *et al.*, 1985a) prompted us to use *C. reinhardtii* URF gene probes in hybridiza-

Table I. Codon usage in the *C. reinhardtii* URF 5 gene

	T			C			A			G			
	TTT	Phe	19	TCT	Ser	9	TAT	Tyr	7	TGT	Cys	3	T
T	TTC	Phe	16	TCC	Ser	8	TAC	Tyr	11	TGC	Cys	6	C
	TTA	Leu	0	TCA	Ser	0	TAA	*	1	TGA	*	0	A
	TTG	Leu	41	TCG	Ser	0	TAG	*	0	TGG	Trp	13	T
	CTT	Leu	9	CCT	Pro	1	CAT	His	6	CGT	Arg	10	T
C	CTC	Leu	0	CCC	Pro	2	CAC	His	9	CGC	Arg	5	C
	CTA	Leu	25	CCA	Pro	10	CAA	Gln	5	CGA	Arg	1	A
	CTG	Leu	2	CCG	Pro	0	CAG	Gln	3	CGG	Arg	0	G
	ATT	Ile	18	ACT	Thr	19	AAT	Asn	2	AGT	Ser	20	T
A	ATC	Ile	12	ACC	Thr	12	AAC	Asn	12	AGC	Ser	8	C
	ATA	Ile	0	ACA	Thr	0	AAA	Lys	8	AGA	Arg	0	A
	ATG	Met	28	ACG	Thr	0	AAG	Lys	1	AGG	Arg	0	G
	GTT	Val	23	GCT	Ala	31	GAT	Asp	11	GGT	Gly	42	T
G	GTC	Val	8	GCC	Ala	20	GAC	Asp	6	GGC	Gly	11	C
	GTA	Val	18	GCA	Ala	0	GAA	Glu	0	GGA	Gly	0	A
	GTG	Val	2	GCG	Ala	3	GAG	Glu	10	GGG	Gly	0	G

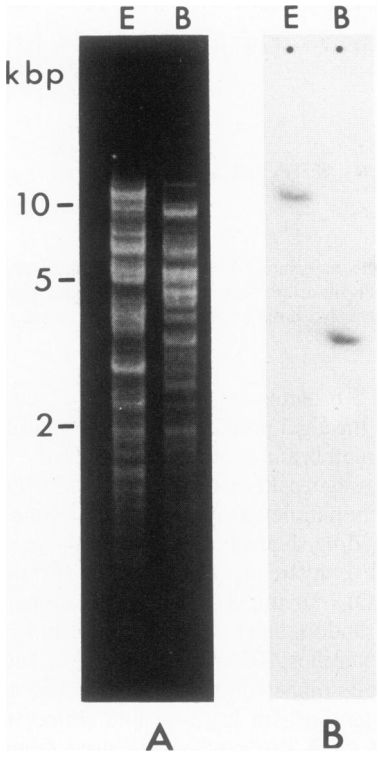

Fig. 4. Hybridization of a *C. reinhardtii* URF 5 gene probe to wheat mtDNA. (**A**) U.v. fluorescence patterns of wheat mtDNA digested with *Eco*RI (E) and *Bam*HI (B) and fractionated in 1% agarose gels. (**B**) Southern blot analysis using probe 1 (see Figure 1), a fragment internal to the algal URF 5 gene. DNA sizes of 10, 5 and 2 kbp are indicated. Conditions for heterologous hybridization were as described in Boer *et al.* (1985b).

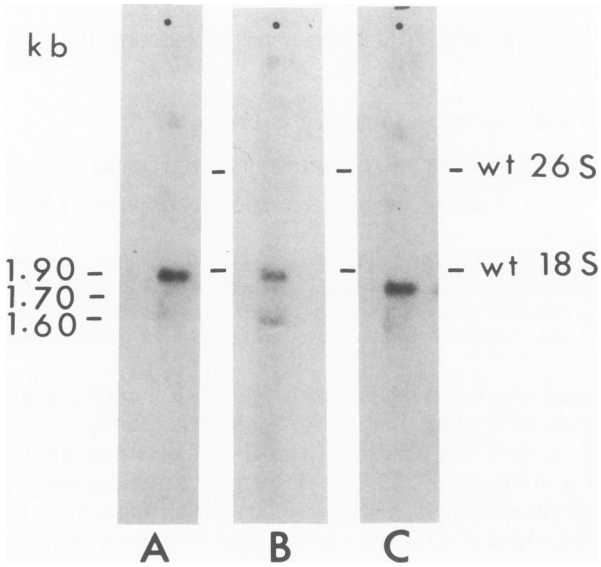

Fig. 5. Transcript analysis of the URF 5 gene region. Northern blots of *C. reinhardtii* mtRNA, fractionated in formaldehyde/agarose gels, were hybridized with (**A**) an internal URF 5 gene fragment (probe 1), (**B**) a fragment from the 3′ region of the URF 5 gene (probe 2) and (**C**) a probe internal to the COI gene (probe 6). From migration of size markers (wheat mitochondrial 18S and 26S rRNAs, right), transcript sizes in nucleotides (kb) were determined (left).

tion experiments with wheat mtDNA under conditions of reduced stringency. Selective hybridization was observed with the *C. reinhardtii* URF 5 probe (Figure 4), as well as with URF 2 gene probes (not shown). The results suggest the presence in wheat mtDNA of a single copy of an URF 5 gene that shares regions of high sequence similarity with the algal gene.

Transcript analysis of the C. reinhardtii URF 5 gene

In a Northern blot analysis using the gene-internal DNA probe 1 (see Figure 1 for map positions of probes), a 1.9-kb RNA species was detected (Figure 5A), showing that the URF 5 gene is transcribed. A DNA fragment from the 3′ region of the URF 5 gene (probe 2) identified, in addition, a 1.6-kb transcript that contains the ORF x coding sequence (Figure 5B), while a probe to sequences upstream of the URF 5 gene (probe 6) hybridized

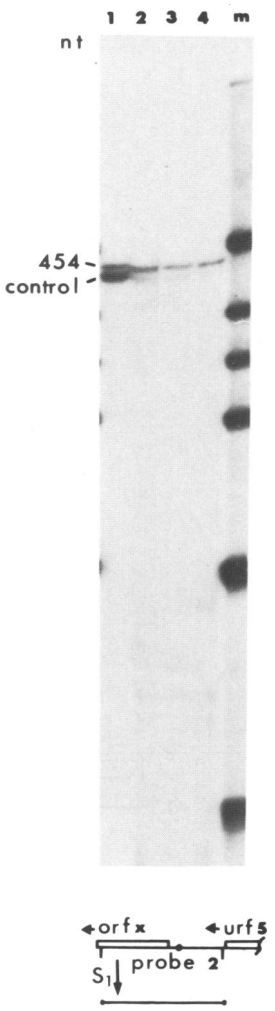

Fig. 6. S1 nuclease analysis of URF 5/ORF x transcripts. **Lane 1**: input DNA (5′ end-labeled 454 nucleotide *TaqI* fragment (probe 2) + control probe; see text); **lanes 2, 3, 4**: S1 nuclease at 50, 250 and 750 U/ml, respectively; m, a size marker of end-labeled pBR322 digested with *HinfI*. A scheme of the experiment is shown at the bottom of the figure.

only to a third transcript, the COI mRNA, ~1.7 kb in size (Figure 5C). Considering the map positions of the probes, the Northern data indicate that these three mRNAs must have short 5′ non-coding stretches and must map close together in this region of 16-kbp DNA. The direction of transcription was determined with strand-specific probes; in all three cases, transcripts were detected from only one of the two DNA strands.

Because probes derived from the region upstream of the URF 5 gene do not hybridize with the 1.9-kb mRNA, both the 3′ end of URF 5 mRNA and the 5′ end of ORF x mRNA must map within a 454-bp *TaqI* fragment located in this region of 16-kbp DNA. In S1 nuclease protection experiments to determine the ends of the transcripts, this *TaqI* fragment (probe 2) was isolated and labeled at its 5′ ends, together with a *TaqI* fragment of slightly smaller size from the pUC9 plasmid vector, taken as an internal control. The 454 nucleotide *C. reinhardtii* probe remained fully protected by *C. reinhardtii* mtRNA, while the unrelated pUC9 sequences were progressively digested by increasing amounts of S1 nuclease (Figure 6). This result implies that the 3′ end of the 1.9-kb URF 5 mRNA and the 5′ end of the 1.6-kb URF x mRNA map so close together on the mtDNA that S1 nuclease, under the conditions used, does not detect any single-strand-equivalent DNA, i.e. the transcripts must have ends that are abutting. Under similar experimental conditions, we have readily mapped other transcript termini by the S1 nuclease protection assay (data not shown). These results also argue that ORF x and URF 5 mRNAs are present in about equal amounts in total mtRNA, as suggested as well by Northern blot analysis (Figure 4B).

Determination of 5′ termini

Primer extension experiments were performed to determine the 5′ termini of the three algal mRNAs. As shown in Figure 7, a number of discrete primer extension products were obtained; all appeared homogeneous, implying that reverse transcription terminated at a unique position in each case. Based on the resolution of the sequencing gels used as a marker in these particular experiments, these unique positions could be mapped with an accuracy of $\pm 1-2$ nucleotides. The most probable position of each 5′ terminus is indicated by a solid triangle in Figures 2 and 7.

Two transcripts differing in length by 90 nucleotides at the 5′ end were detected using the *SalI-HpaI* fragment (primer 3) to extend URF 5 mRNA (Figure 7A, lane 2). The 5′ end of the major URF 5 transcript maps to the position of the initiation codon for translation of URF 5; the minor transcript has a leader of 90 nucleotides, which overlaps the COI coding region by 10 nucleotides. S1 nuclease protection experiments using probe 5 (Figure 1) gave results that were consistent with these primer extension experiments (data not shown).

For reverse transcription of ORF x mRNA we used primer fragment 4 (a *HindIII-AvaII* fragment) and obtained a major cDNA product of 306 nucleotides (#2), while a minor extension product of 247 nucleotides (#1) was also visible (Figure 7B, lane 2). Densitometric quantitation of the signals suggested that the minor band represents <10% of the total; interestingly, it maps to the first ATG of the ORF x gene. Most of the ORF x transcripts apparently have a discrete 5′ end, a result which agrees with Northern and S1 mapping experiments (Figures 5 and 6); thus, since we had concluded on the basis of S1 analysis (see above) that URF 5 and ORF x transcripts have abutting termini, the primer extension experiments localize the 3′ end of URF 5 mRNA to position 1856 in the DNA sequence.

The 5′ terminus of the COI mRNA also maps to the ATG start codon. Primer 6, a 208-nucleotide restriction fragment that terminates at the *DdeI* site at position −173 (Figure 2), directed reverse transcription of COI mRNA to give a major cDNA product of 299 nucleotides (Figure 7C, lane 2). The COI mRNA therefore starts at position −82 in the DNA sequence (−173 + 91, the number of nucleotides by which primer 6 was extended).

Conserved primary and secondary structural motifs in C. reinhardtii mitochondrial transcripts

Analysis of DNA sequences in the regions downstream of the *C. reinhardtii* mitochondrial genes for URF 5 (this paper) and COI (unpublished observations) has revealed a number of conserved sequence elements with potential secondary structure; these are summarized in Figure 8. Sequence block I (at positions 1641 − 1651, immediately downstream from the URF 5 stop codon) can form alternative stable hairpin structures either with the immediately adjacent block II (1656 − 1667) or with block IV, located further downstream at positions 1808 − 1815. Blocks II and IV are separated by a second inverted repeat structure, III (1676 − 1699), that is also present (with one mismatch in the loop) in the COI 3′ non-translated region. Downstream of the COI gene, the analogues of URF 5 blocks I, II and IV are sequences 8 − 11 nucleotides long that contain three closely linked *HindIII* sites (see Figure 1). Not only do the 3′ non-translated

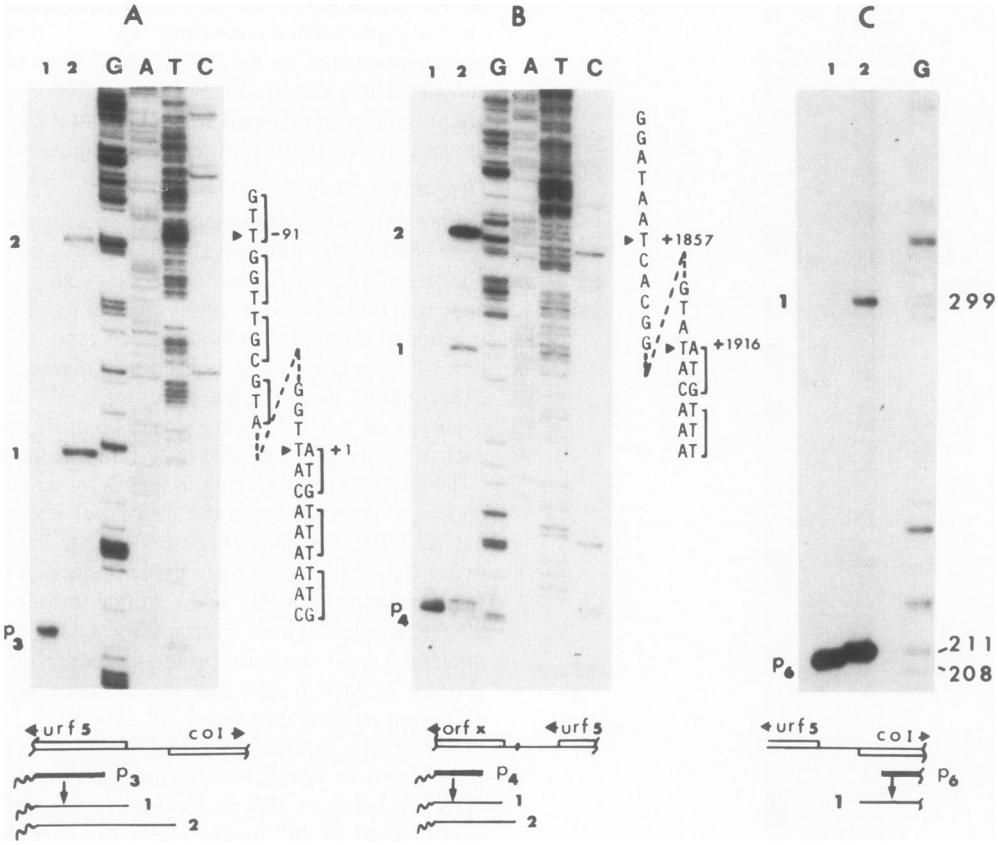

Fig. 7. Primer extension experiments. (**A**) A *SalI-HpaI* fragment (p_3) was used as primer for reverse transcription of URF 5 transcripts. (**B**) A *HindIII-AvaII* fragment (p_4) was used as primer for reverse transcription of ORF x mRNA. (**C**) A fragment terminating at the *DdeI* site (p_6) was used to map the 5′ end of CO1 mRNA. **Lane 1**: input DNA; **lane 2**: extension product(s). G, A, T, C: tracks of sequencing reactions of the same templates used to generate primers. Schematics of the experiments are shown at the bottom of the figure, with numbers indicating the major cDNAs. Wavy lines indicate M13 'tail' sequences that derive from probe constructions.

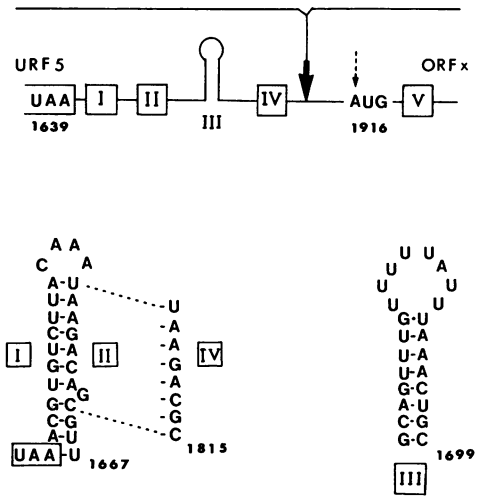

Fig. 8. Schematic representation of possible secondary structures in the 3′ non-translated region of the URF 5 mRNA. Nucleotide sequence block I is able to base pair either with sequence block II or with IV, which are themselves separated by hairpin III. Numbers refer to the DNA sequence in Figure 1. Arrows indicate major and minor transcript termini and V represents a region of sequence identity between the ORF x and URF 2 genes.

regions of the URF 5 and COI mRNAs have features in common, but the leaders of the transcripts of downstream genes are also very similar. The region designated V in Figure 8 (positions 1917 – 1977) is 70% identical with the corresponding 5′ region of the URF 2 gene, and contains a stretch of 20 nucleotides (starting at position 1925) that is identical to a region starting at position 26 in the URF 2 sequence (Pratje *et al.*, 1984). The conservation of these sequence blocks, their polarity and their locations close to mRNA termini strongly suggests a role in algal mtRNA processing.

Discussion

The data presented here show that *C. reinhardtii* mtDNA encodes an open reading frame of 546 amino acids that has obvious homology with human and trypanosome mitochondrial URF 5 proteins, and that is transcribed as part of a 1.9kb stable RNA species. We conclude that this open reading frame corresponds to the URF 5 gene of *C. reinhardtii* mtDNA. The same gene has been identified in mtDNA from other mammals, *Drosophila*, and a number of filamentous fungi, and the specific hybridization observed here between a *C. reinhardtii* URF 5 gene probe and wheat mtDNA strongly suggests that plants can be added to the list of organisms that encode URF 5 genes in mtDNA. Recently, Chomyn *et al.* (1985) concluded that the human mitochondrial URF 5 gene product is the largest subunit of complex I, the respiratory NADH dehydrogenase of the inner mitochondrial membrane. The human, *C. reinhardtii* and trypanosome URF 5 proteins have the characteristics of integral membrane components, in that all are very hydrophobic and have the potential (based on their similar hydrophobicity profiles) to span a membrane repeatedly.

In contrast to what has been observed in other mitochondrial systems, we have no indications so far that the mitochondrial genetic code in *C. reinhardtii* deviates in any way from the universal code. Of the 13 TGG-encoded tryptophan residues in the algal URF 5 protein, four align with tryptophan residues encoded by TGA in the human URF 5 gene. In concert with all plant mitochondrial protein genes sequenced to date (Fox and Leaver, 1979; Hiesel and Brennicke, 1983; Bonen *et al.*, 1984; Kao *et al.*, 1984; Dawson *et al.*, 1984; Boer *et al.*, 1985b; Schuster and Brennicke, 1985; Moon *et al.*, 1985; Dewey *et al.*, 1985; Isaac *et al.*, 1985), *C. reinhardtii* mitochondrial protein genes show a bias toward the use of codons with T in the third position (these constitute 35 – 40% of the total in both the algal and plant mitochondrial protein genes). However, codon usage is much more restricted in mitochondria of *C. reinhardtii* than in plant mitochondria. In the URF 5 gene, for example, 15 codons are not used, including CGG, which encodes tryptophan rather than arginine in plant mtDNA (Fox and Leaver, 1981; Hiesel and Brennicke, 1983; Bonen *et al.*, 1984), and AGA/G, which in human and bovine mtDNA (Anderson *et al.*, 1981, 1982) [but not mouse mtDNA (Bibb *et al.*, 1981)] appear to be termination rather than arginine codons. Although codon usage is also restricted in *C. reinhardtii* chloroplast DNA (Rochaix *et al.*, 1984), the patterns in the two organelles do not parallel one another. Rare synonyms in one organelle may be major ones in the other; for example, TTG/Leu is markedly preferred to TTA/Leu in *C. reinhardtii* mitochondria (TTA/Leu is used only once in the COI gene), while the reverse is true for *C. reinhardtii* chloroplast genes. Codon usage in *C. reinhardtii* mitochondrial genes is also quite distinct from the highly biased pattern seen in the nuclear β-tubulin genes of this alga (Youngblood *et al.*, 1984).

Transcripts from 16-kbp DNA derive from both DNA strands, as determined by DNA sequence and transcript analysis. In the present study we have characterized three mRNAs, which originate from the genes for URF 5, COI and the putative ORF x gene. The URF 5 mRNA is 1856 nucleotides long, begins with the AUG translation initiation codon, and contains a 3′ non-translated region of 218 nucleotides. The 3′ terminus of URF 5 mRNA appears to be juxtaposed with the 5′ end of URF x mRNA as they map to the genome; this is a strong indication that transcript termini result from RNA processing events, rather than via transcription termination and re-initiation of RNA synthesis. Likewise, the COI mRNA, encoded on the other DNA strains of 16-kbp DNA, begins with the AUG initiation codon and also has a short 3′ non-translated region. Again, RNA processing is implicated: preliminary transcript mapping results show that the 3′ end of COI mRNA and the 5′ end of URF 2 mRNA map very close together, while the URF 2 mRNA and $tRNA^{Trp}$ (see Figure 1) also map as abutting transcripts (unpublished observations). These observations recall transcript mapping data obtained for other organisms with small (16 – 19 kbp) mitochondrial genomes, such as mammals (Clayton, 1984) and the yeasts *Schizosaccharomyces pombe* (Lang *et al.*, 1983) and *Torulopsis glabrata* (Clark-Walker *et al.*, 1985). It should be noted, however, that *S. pombe* and *T. glabrata* encode mitochondrial genes on only one DNA strand, and they lack the homologues of human URF 2 and URF5 proteins that are encoded by *C. reinhardtii* mtDNA. Algal mitochondrial transcript data differ from those obtained with plant mitochondria, where mitochondrial mRNAs are not transcriptionally linked and possess long 5′ and 3′ non-translated regions (cf. Boer *et al.*, 1985b).

Major control of mitochondrial gene expression is thought to be exerted at the level of mitochondrial RNA processing (reviewed by Tabak *et al.*, 1983; Grivell, 1983; Clayton, 1984) and secondary structures or specific nucleotide sequence motifs within precursor transcripts have been implicated in processing. Moreover, such RNA processing elements have been conserved in mitochondrial genome evolution. For example, in mitochondria of humans (Ojala *et al.*, 1981), *S. pombe* (Lang *et al.*, 1983) and *N. crassa* (Burger *et al.*, 1985; Breitenberger *et al.*, 1985), tRNA sequences appear to signal precise endonucleolytic processing of long primary transcripts, while a conserved nucleotide sequence maps to the 3′ termini of most yeast mitochondrial mRNAs (Osinga *et al.*, 1984; Clark-Walker *et al.*, 1985). Our transcript mapping results with *C. reinhardtii* mitochondrial RNA suggest the presence of long precursor transcripts from which mature transcripts are produced by specific endonucleolytic cleavages. In this context, the minor URF 5 transcript that has a 5′ extension of 90 nucleotides may be a relatively stable precursor of URF 5 mRNA that has escaped RNA processing. This model can be tested critically by *in vitro* capping analysis of algal mitochondrial RNA to identify primary transcripts.

Although processing of *C. reinhardtii* mitochondrial mRNAs shows some similarities to mRNA processing in animal mitochondria, precise endonucleolytic cleavage of primary transcripts must be signalled by different mechanisms in the two systems, because tRNA sequences do not generally flank protein coding sequences in *C. reinhardtii* mtDNA. However, alternative elements of secondary structure, shown schematically in Figure 8, may be involved in RNA processing events in *C. reinhardtii* mitochondria, as suggested by their location close to mRNA termini and their presence in the 'leftward' and 'rightward' transcriptional units. No common sequence motif has been observed in the DNA regions where the 5′ termini of COI and URF 5 transcripts map; their promoters may be located far upstream.

Unicellular green algae such as *C. reinhardtii* are thought to share a direct common ancestor with vascular land plants (Margulis and Schwartz, 1981), as evidenced by sequence comparisons of, for example, nuclear-encoded 5S rRNA genes (Kumazaki *et al.*, 1983; Vandenberghe *et al.*, 1984). In view of this specific evolutionary relationship, it is striking that the organization and mode of expression of the *C. reinhardtii* mitochondrial genome should differ so radically from that of plants. Our results provide another dramatic example of the pronounced mitochondrial genome diversity (Gray, 1982; Sederoff, 1984) that complicates reconstruction of the evolutionary history of mitochondria (Gray *et al.*, 1984).

Materials and methods

The *C. reinhardtii* cell wall-less strain CW-15 was grown with continuous illumination in minimal medium supplemented with 0.2% sodium acetate, and mitochondria were isolated by differential centrifugation of cells disrupted in a French pressure cell, as described by Ryan *et al.* (1978). A 3.8-kbp *Hind*III fragment of the algal mtDNA cloned into the pUC9 plasmid vector (clone pH1, Boer *et al.*, 1985a) was the source of DNA for sequencing. Subfragments were isolated from low melting point agarose, further digested and cloned into compatible M13 vectors (Messing, 1983), and sequenced by the chain termination method of Sanger *et al.* (1977). Computer analysis of sequence data was conduced using the program of Queen and Korn (1984) (Microgenie, Beckman). For RNA preparation, the algal mitochondrial fraction was resuspended in an equal volume of 1 × SSC, lysed with a mixture of 2% SDS, 1% sodium 2-aminosalicylate, 0.5% sodium tri-isopropylnaphthalene disulphonate and 3% 2-butanol (van Ommen *et al.*, 1979), and repeatedly extracted with phenol, phenol/chloroform (1:1) and chloroform/isoamyl alcohol (24:1). Such mtRNA preparations contained variable amounts of contaminating cytoplasmic RNA. Southern and Northern hybridization conditions have been described (Bonen *et al.*, 1984). DNA probes for transcript mapping experiments were 5′ end-labeled with ^{32}P using T4 polynucleotide kinase (Maniatis *et al.*, 1982), or were uniformly labeled by second strand synthesis

on M13 templates (Messing, 1983), followed by restriction endonuclease digestion. Such probes were fractionated in and isolated from sequencing gels, electrophoresed into low melting point agarose, recovered, and then hybridized to mtRNA in 80% formamide, in the presence of 20 mM Pipes, pH 6.4, 0.4 M sodium chloride and 1 mM EDTA, at 50°C for >4 h. For S1 nuclease analysis (Weaver and Weissmann, 1979), the hybridization mixture was diluted 10-fold with 30 mM sodium acetate, pH 4.5, 4.5 mM zinc acetate, 0.2 M sodium chloride and 10 μg/ml sheared denatured salmon sperm DNA and treated with S1 nuclease (Boehringer) for 30 min at 37°C, usually as a titration series with increasing amounts of S1 nuclease, as specified in the figure legends. For primer extension experiments, annealed primer/template complexes, prepared as above, were recovered by ethanol precipitation and incubated [in 50 mM Tris-HCl (pH 8.3), 6 mM $MgCl_2$, 40 mM KCl] with reverse transcriptase (Life Sciences) at 200 U/ml for 20 min at 37°C. The products of these RNA mapping reactions were resolved on DNA sequencing gels, along with M13 sequencing reactions of the same templates so that signals obtained could be directly correlated with the DNA sequence.

Acknowledgements

We are grateful to L.Bonen for advice and encouragement throughout this study, and to L.Bonen, D.F.Spencer, and M.N.Schnare for a critical review of the manuscript. We thank F.Sasinek for expert assistance during production of the figures. This work was supported by grant MT-4124 from the Medical Research Council of Canada to M.W.G.

References

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**, 457–464.

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.

Bibb,M.J., Van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) *Cell*, **26**, 167–180.

Boer,P.H., Bonen,L., Lee,R.W. and Gray,M.W. (1985a) *Proc. Natl. Acad. Sci. USA*, **82**, 3340–3344.

Boer,P.H., McIntosh,J.E., Gray,M.W. and Bonen,L. (1985b) *Nucleic Acids Res.*, **13**, 2281–2292.

Bonen,L., Boer,P.H. and Gray,M.W. (1984) *EMBO J.*, **3**, 2531–2536.

Breitenberger,C.A., Browning,K.S., Alzner-DeWeerd,B. and RajBhandary,U.L. (1985) *EMBO J.*, **4**, 185–195.

Burger,G., Helmer Citterich,M., Nelson,M.A., Werner,S. and Macino,G. (1985) *EMBO J.*, **4**, 197–204.

Capaldi,R.A. and Vanderkooi,G. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 930–932.

Chomyn,A., Mariottini,P., Gonzalez-Cadavid,N., Attardi,G., Strong,D.D., Trovato,D., Riley,M. and Doolittle,R.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5535–5539.

Chomyn,A., Mariottini,P., Cleeter,M.W.J., Ragan,C.I., Matsuno-Yagi,A., Hatefi,Y., Doolittle,R.F. and Attardi,G. (1985) *Nature*, **314**, 529–597.

Clark-Walker,G.D., McArthur,C.R. and Sriprakash,K.S. (1985) *EMBO J.*, **4**, 465–473.

Clary,D.O. and Wolstenholme,D.R. (1983) *Nucleic Acids Res.*, **11**, 6859–6872.

Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R. (1984) *Nucleic Acids Res.*, **12**, 3747–3762.

Clayton,D.A. (1984) *Annu. Rev. Biochem.*, **53**, 573–594.

Dawson,A.J,. Jones,V.P. and Leaver,C.J. (1984) *EMBO J.*, **3**, 2107–2113.

de la Cruz,V.F., Neckelmann,N. and Simpson,L. (1984) *J. Biol. Chem.*, **259**, 15136–15147.

de Bruijn,M.H.L. (1983) *Nature*, **304**, 234–241.

Dewey,R.E., Schuster,A.M., Levings,C.S.,III and Timothy,D.H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1015–1019.

Fox,T.D. and Leaver,C.J. (1981) *Cell*, **26**, 315–323.

Grant,D. and Chiang,K.-S. (1980) *Plasmid*, **4**, 82–96.

Gray,M.W. (1982) *Can. J. Biochem.*, **60**, 157–171.

Gray,M.W., Sankoff,D. and Cedergren,R.J. (1984) *Nucleic Acids Res.*, **12**, 5837–5852.

Grivell,L.A. (1983) in Schweyen,R.J., Wolf,K. and Kaudewitz,F. (eds), *Mitochondria 1983*, de Gruyter, Berlin, pp. 25–45.

Hensgens,L.A.M., Brakenhoff,J., De Vries,B.F., Sloof,P., Tromp,M.C., Van Boom,J.H. and Benne,R. (1984) *Nucleic Acids Res.*, **12**, 7327–7344.

Hiesel,R. and Brennicke,A. (1983) *EMBO J.*, **2**, 2173–2178.

Isaac,P.G., Jones,V.P. and Leaver,C.J. (1985) *EMBO J.*, **4**, 1617–1623.

Ise,W., Haiker,H. and Weiss,H. (1985) *EMBO J.*, **4**, 2075–2080.

Kao,T.-h., Moon,E. and Wu,R. (1984) *Nucleic Acids Res.*, **12**, 7305–7315.

Kumazaki,T., Hori,H. and Osawa,S. (1983) *J. Mol. Evol.*, **19**, 411–419.

Lang,B.F., Ahne,F., Distler,S., Trinkl,H. and Kaudewitz,F. (1983) in Schweyen,R.J., Wolf,K. and Kaudewitz,F. (eds), *Mitochondria 1983*, de Gruyter, Berlin, pp. 313–329.

Leaver,C.J. and Gray,M.W. (1982) *Annu. Rev. Plant Physiol.*, **33**, 373–402.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.

Margulis,L. and Schwartz,K.V. (1982) *Five Kingdoms*, published by Freeman, San Francisco.

Mariottini,P., Chomyn,A., Attardi,G., Trovato,D., Strong,D.D. and Doolittle,R.F. (1983) *Cell*, **32**, 1269–1277.

Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.

Michael,N.L., Rothbard,J.B., Shiurba,R.A., Linke,H.K., Schoolnik,G.K. and Clayton,D.A. (1984) *EMBO J.*, **3**, 3165–3175.

Moon,E., Kao,T.-h. and Wu,R. (1985) *Nucleic Acids Res.*, **13**, 3195–3212.

Ojala,D., Montoya,J. and Attardi,G. (1981) *Nature*, **290**, 470–474.

Osinga,K.A., De Vries,E., Van der Horst,G. and Tabak,H.F. (1984) *EMBO J.*, **3**, 829–834.

Pratje,E., Schnierer,S. and Dujon,B. (1984) *Curr. Genet.*, **9**, 75–82.

Queen,C. and Korn,L.J. (1984) *Nucleic Acids Res.*, **12**, 581–599.

Rochaix,J.-D., Dron,M., Rahire,M. and Malnoe,P. (1984) *Plant Mol. Biol.*, **3**, 363–370.

Ryan,R., Grant,D., Chiang,K.-S. and Swift,H. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 3268–3272.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schuster,W. and Brennicke,A. (1985) *Curr. Genet.*, **9**, 157–163.

Sederoff,R.R. (1984) *Adv. Genet.*, **22**, 1–108.

Tabak,H.F., Grivell,L.A. and Borst,P. (1983) *CRC Crit. Rev. Biochem.*, **14**, 297–317.

Vandenberghe,A., Chen,M.-W., Dams,E., de Baere,R., de Roeck,E., Huysmans,E. and de Wachter,R. (1984) *FEBS Lett.*, **171**, 17–23.

Van Ommen,G.-J.B., Groot,G.S.P. and Grivell,L.A. (1979) *Cell*, **18**, 511–523.

Weaver,R.F. and Weissmann,C. (1979) *Nucleic Acids Res.*, **7**, 1175–1193.

Youngbloom,J., Schloss,J.A. and Silflow,C.D. (1984) *Mol. Cell. Biol.*, **4**, 2686–2696.

Received on 22 October 1984; revised on 25 November 1985

Note added in proof

After this paper was accepted, a sequence encompassing that shown in Figure 2 was published by another group (Vahrenholz,C., Pratje,E., Michaelis,G. and Dujon,B. (1985) *Mol. Gen. Genet.*, **201**, 213–224). Compared to our sequence, the one presented by Vahrenholz *et al.* shows a number of differences, some of which result in changes in the predicted amino acid sequence of the URF 5 protein. In particular, deletion of T_{1525} (Figure 2) produces a frame shift that would result in an URF 5 protein having a different (and truncated) C-terminal sequence than the one proposed here. The truncated C-terminal sequence (11 codons) proposed by Vahrenholz *et al.* contains 5 codons that are not otherwise used in the URF 5 gene.