Mol Gen Genet (1994) 242:65–72

MGG

RNA editing of a conserved reading frame in plant mitochondria increases its similarity to two overlapping reading frames in *Escherichia coli*

Sabine Sünkel, Axel Brennicke, Volker Knoop

Institut für Genbiologische Forschung, Ihnestr. 63, D-14195 Berlin, Germany

Received: 31 March 1993 / Accepted: 28 June 1993

Abstract. An open reading frame (*orfx*) in mitochondria of the higher plants *Oenothera berteriana* and *Arabidopsis thaliana* is homologous to *orf244* in the mitochondrial genome of *Marchantia polymorpha.* Homologous sequences are also present in carrot, potato and sugar beet. Profile analysis revealed similarity to two overlapping reading frames in the *Escherichia coli* genome. Potential translation initiation at conserved ATA (isoleucine) and TTG (leucine) codons is discussed. Transcripts of the open reading frame are altered by RNA editing in *Arabidopsis* and *Oenothera* downstream of these codons, suggesting this to be the functionally important region.

Key words: *Arabidopsis* – *Oenothera* – Profile analysis

Introduction

Several factors contribute to the large size of plant mitochondrial genomes (200–2000 kb) in comparison with those of animals (16 kb) and fungi (18–90 kb). Among these are the presence of introns, pseudogenes, additional gene copies and genes that are not mitochondrially encoded in other eukaryotes. In other eukaryotes, these latter genes have presumably been transferred to the nucleus during evolution, with the encoded proteins now being posttranslationally imported into mitochondria. Examples of such genes are those coding for ribosomal proteins or subunit α of the ATP synthase (reviewed in Bonen 1991). Alternatively, these additional genes in plant mitochondria might encode proteins for plant-specific requirements in mitochondrial biogenesis and function.

Analysis of the mitochondrial genes encoding subunit 5 of the NADH dehydrogenase *(nad5)* in two dicotyledonous plants has shown that *trans*-splicing of independent precursor mRNAs is necessary for assembly of the *nad5* reading frame (Knoop et al. 1991). The complex arrangement of the five *nad5* exons is conserved in monocotyledonous species (Pereira de Souza et al. 1991).

An important argument for *trans*-splicing ligation is the presence of other genes between the distant exons. An open reading frame *(orfx)* was identified upstream of the small central *nad5* exon c in *Oenothera* (Knoop et al. 1991). A corresponding cDNA clone covering the C-terminal part of this reading frame showed several RNA editing events. Since the initial reports on RNA editing in plant mitochondria (Gualberto et al. 1989; Covello and Gray 1989; Hiesel et al. 1989), C to U changes at the RNA level have been recognized as a general feature of plant mitochondrial protein gene expression (reviewed in Wissinger et al. 1992). RNA editing events in plant mitochondria are concentrated in open reading frames. The observation of RNA editing in an otherwise uncharacterized open reading frame can thus be taken as an indication of its functional expression.

Plant mitochondrial genomes vary considerably in size and structure, but functional genes are extremely well conserved at the nucleotide sequence level. We therefore tried to identify homologous counterparts in other species in order to substantiate this assumption. In this communication we have analysed the novel open reading frame *orfx* in mitochondria of *Arabidopsis* and *Oenothera* and report the presence of homologous sequences in *Marchantia,* potato, carrot, sugar beet and the bacterium *Escherichia coli.*

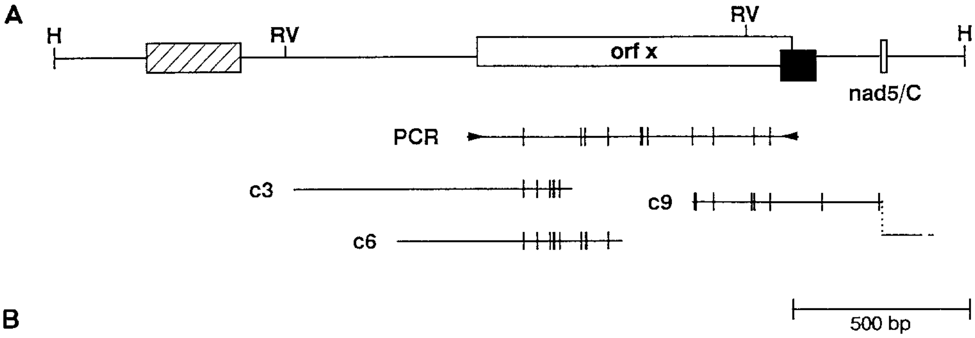
Materials and methods

Mitochondrial nucleic acids were extracted from plant cell cultures following published procedures (Schuster et al. 1988). RNA was extracted from isolated mitochondria in the presence of guanidinium thiocyanate with minor modifications from published procedures (Chomczynski and Sacchi 1987). A cosmid library of *Arabidopsis* total DNA was kindly provided by U. Halfter and

---

Communicated by R. Hagemann

Correspondence to: A. Brennicke



B

```
1191   ATG ATC AAG GTG CGG GGA GAC ATT TTA TTG ACC GGT CGC TAT TAC GGG GGT GTT ATA GAA TTA AAA CTC TCT TTT CTT ACC
   1    M   I   K   V   R   G   D   I   L   L   T   G   R   Y   Y   G   G   V   I   E   L   K   L   S   F   L   T

1272   TTC TGT CCT TTG AAT TAC TCC TAT ATA TCC TCA AAA TTT GAT TTT GCA|TcG|GAA ACT ATT CTA GAA GAA GTT CGA ATC CGT
  28    F   C   P   L   N   Y   S   Y   I   S   S   K   F   D   F   A |S→L| E   T   I   L   E   E   V   R   I   R

1353   TCC GTT|cGG|ATA TTG ATC GGT CTT GGT TTG ACA TGG TTT ACG|cGT|TAC TGG|TTc|cCG|GAA GAG TTC ATT|TcT|CTA TTA GCT
  55    S   V |R→W| I   L   I   G   L   G   L   T   W   F   T |R→C| Y   W |F→F|P→S| E   E   F   I |S→F| L   L   A

1434   AAA CCC TTT CTT AGC TTG CCT TTG GAC TCG TAT TTT GTT TGT ACA CAA|TcA|ACG GAG GCC|TcC|cCG|ACA TAT GTT GCA ACG
  82    K   P   F   L   S   L   P   L   D   S   Y   F   V   C   T   Q |S→L| T   E   A |S→F|P→S| T   Y   V   A   T

1515   TCT TCA ATA GCA TGC TCT TAC TTC GTC TTC CCT TTT ATA AGT|cAT|CAA ATT TGG TGC TTT TTG ATC CCC AGT TGC TAT GGG
 109    S   S   I   A   C   S   Y   F   V   F   P   F   I   S |H→Y| Q   I   W   C   F   L   I   P   S   C   Y   G

1596   GAA CAA AGG ACG AAA TAC AAT CGA TTC TTC CAT TTA AGT GGT|TcT|cGC|TTT TCC TTG TTC CTG TTC CTA ACT|CcT|CCC CGG
 136    E   Q   R   T   K   Y   N   R   F   F   H   L   S   G |S→F|R→C| F   S   L   F   L   F   L   T |P→L| P   R

1677   GTA GTT CCC AAT GTT TGG CAC TTT CCA TAC TTC GTG GGT GCA ACA TCA ACA AAT TCG CTC ATG ATC AAG TTA CAA CCT AAG
 163    V   V   P   N   V   W   H   F   P   Y   F   V   G   A   T   S   T   N   S   L   M   I   K   L   Q   P   K

1758   ATC TAT GAC CAT ATT ATG TTA ACT GTT CGT ATT|TcG|TTc|ATT|cCA|TCG GTA TGC TCC CTG GTA CCT GTA ATT TTG ATC TGT
 190    I   Y   D   H   I   M   L   T   V   R   I |S→L|F→F| I |P→S| S   V   C   S   L   V   P   V   I   L   I   C

1839   TTG CCA GAA|CcT|AGG GGT CTT TCT GTG GAA ACC TTC ACG AAC AAT CGT CGT TTT TTG ATG GTT TTT TCG CTT CTC ACA GCT
 217    L   P   E |P→L| R   G   L   S   V   E   T   F   T   N   N   R   R   F   L   M   V   F   S   L   L   T   A

1920   GCT CTT TCC ACA CCT CCG GAT ATC TGG TGC CAA|ATc|GTC GCC|cGT|TTC CTT ATT TCT TTG ATA ATA GAG TTG GCT ATC TTT
 244    A   L   S   T   P   P   D   I   W   C   Q |I→I| V   A |R→C| F   L   I   S   L   I   I   E   L   A   I   F

2001   GTG GCA|TcG|ATT GTA CAA GTT CGT GAG GAG GGC TGG ACG 'AGT GGA ATG AGG GAG AGC GGA TCG ACT CGA CTG ATA TGA TAG
 271    V   A |S→L| I   V   Q   V   R   E   E   G   W   T   S   G   M   R   E   S   G   S   T   R   L   I   *   *
```

Fig. 1A, B. Genomic arragement of *orfx* in *Oenothera*. **A** *Orfx* is encoded on a 2.65 kb *Hin*dIII fragment in the *Oenothera* mitochondrial genome, upstream of the central *nad5* exon c. Independent cDNA clones (c3, c6 and c9) were isolated from an *Oenothera* mitochondrial cDNA library, with clone c9 representing a *trans*-splicing event between *nad5* exons c and d. *Vertical lines* indicate RNA editing events identified in the independent cDNA clones. Polymerase chain reaction (PCR) amplification of *Oenothera* mitochondrial cDNA between specific oligonucleotide primers (*horizontal* arrowheads) yielded the complete cDNA sequence. Two editing sites are located between *orfx* and *nad5* exon c. The editing site proximal to *nad5* exon c had erroneously been reported as a genomically encoded T residue (Knoop et al. 1991). The two boxed regions indicate repeated sequences of the *Oenothera* mitochondrial genome as detailed in the text. Restriction sites are indicated for *Hin*dIII (H) and *Eco*RV (RV). **B** Sequence of the region encoding *orfx* beginning with the first in-frame methionine codon (numbering of nucleotides as in GenBank X60046). The reading frame is shown in triplets with the encoded amino acids given in the single letter code. Edited nucleotides are shown in *lower case letters* and are *underlined* with the resulting amino acid exchange indicated

L. Willmitzer. A set of cosmids with inserts of mitochondrial origin was selected from this library (W. Schuster, unpublished). Cloning procedures were performed as suggested by enzyme suppliers and by standard protocols (Sambrook et al. 1989). For agarose gel (1%) electrophoresis, RNA was denatured in the presence of glyoxal (Sambrook et al. 1989). Nucleic acids were transferred to Biodyne B membranes (Pall, UK) after agarose gel electrophoresis, as recommended by the manufacturer. Oligonucleotides were obtained commercially from TIB Molbiol, Berlin. Nucleic acids were labelled with [α-^{32}P]dCTP (Amersham, UK) by the multiprime method (Hodgson and Fisk 1987). Sequences were determined by the chain termination method (Sanger et al. 1977) using the T7 polymerase kit supplied by Pharmacia. Computer analysis was performed using version 7.1 of the UWGCG program package (Devereux et al. 1984) on a VAX/VMS system. Oligonucleotides used for PCR (polymerase chain reaction) amplification and/or sequencing of cDNAs were 5′-GGGATTACAGGCCGAAGGGGCG-3′ (upstream) and 5′-CATCATAACACAAGCTACCCATTC-3′ (downstream) for *Oenothera* and 5′-GCAAAATGGTGATGAATT(C)CGATTGG-3′ (P1), 5′-GAAGTTCG(G)ATCCGTTCCG-3′ (P2), 5′-GCTATTGAAGACGTTGC-3′ (P3) and 5′-NNNNGAATTCATTGATAGTTACTTTGCC-3′ (P4) for *Arabidopsis* as depicted in Figs. 1A and 3A, respectively. Single nucleotide exchanges (shown in parentheses) were introduced into P1 and P2 to create recognition sites for *Eco*RI and *Bam*HI, respectively.

Results

Genomic environment of orfx *in* Oenothera *mitochondria*

During investigation of the *nad5* gene arrangement an open reading frame was identified upstream of the central *nad5* exon c in *Oenothera* and was tentatively termed *orfx* (Knoop et al. 1991).

The entire nucleotide sequence of the 2.65 kb *Hin*dIII fragment encompassing *orfx* was determined (Fig. 1; GenBank X60046). Two regions in the investigated *Hin*dIII fragment were apparently duplicated by recombination events since they show high similarity (more than 95% identical residues, alignments not shown) to other sequences in the *Oenothera* mitochondrial genome (Fig. 1A). A sequence of 80 nucleotides (solid box in Fig. 1A) at the carboxy-terminal end of *orfx*, including the six terminal *orfx* codons, is nearly identical to a sequence located 50 nucleotides downstream of the gene for tRNA-Ser(UGA) (Binder et al. 1991). Another sequence of 260 nucleotides (hatched box in Fig. 1A) located 670 nucleotides upstream of *orfx* has high similarity to a region upstream of the *rps19* pseudogene (Schuster and Brennicke 1991) and is repeated several times elsewhere in the *Oenothera* mitochondrial genome (W. Schuster, personal communication).

Orfx *is conserved in higher plant mitochondria*

Since mitochondrial gene sequences are highly conserved between different plant species, the presence of *orfx* in plants other than *Oenothera* would support the assumption that *orfx* encodes a protein of functional importance. Southern blot analysis with an *orfx*-specific probe indeed detects homologous sequences in *Hin*dIII-digested mitochondrial DNA from *Arabidopsis thaliana* and carrot *(Daucus carota)* and total DNA from potato *(Solanum tuberosum)* on species-specific restriction fragments (Fig. 2A).

In *Oenothera orfx* is cotranscribed with the nearby exon c of the *nad5* gene (Knoop et al. 1991), supporting the functional relevance of this reading frame. A conserved function implies expression of *orfx* in other plants also. Northern blot hybridization experiments with *Arabidopsis* mitochondrial RNA indeed show the *Arabidopsis* homologue of *orfx* to be transcribed in this species (Fig. 2B). A major mRNA species of 2.4 kb and two minor RNAs of 1.4 and 1.1 kb are detected.

The homologue of *orfx* was analysed in detail in *Arabidopsis*. Screening of selected cosmids carrying *Arabidopsis* mitochondrial DNA with the *orfx*-specific probe from *Oenothera* identified cosmid clone 11A5. An internal 5.4 kb *Hin*dIII fragment of 11A5 corresponds in size to the signal obtained in the Southern blot analysis (Fig. 2A) and was subcloned. Upon further restriction analysis a 1.3 kb *Hin*dIII-*Eco*RI fragment was selected for sequencing (Fig. 3, GenBank X72616).

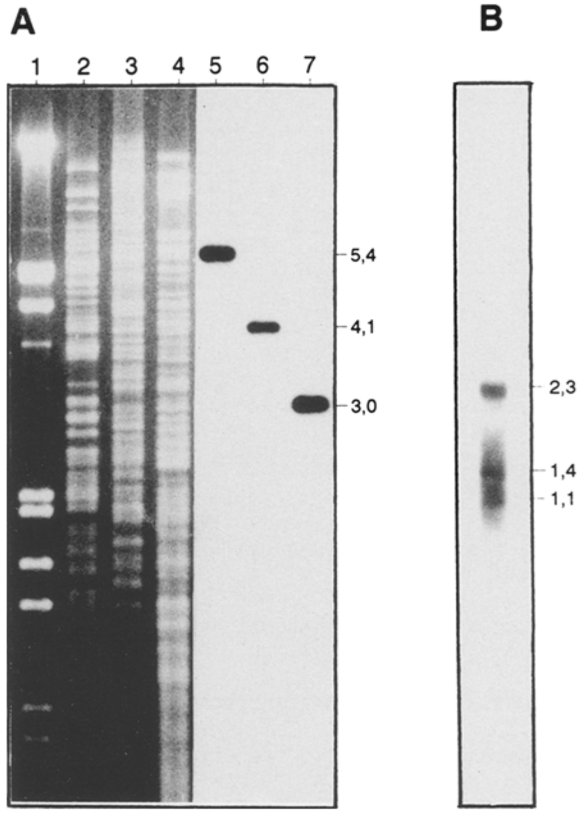
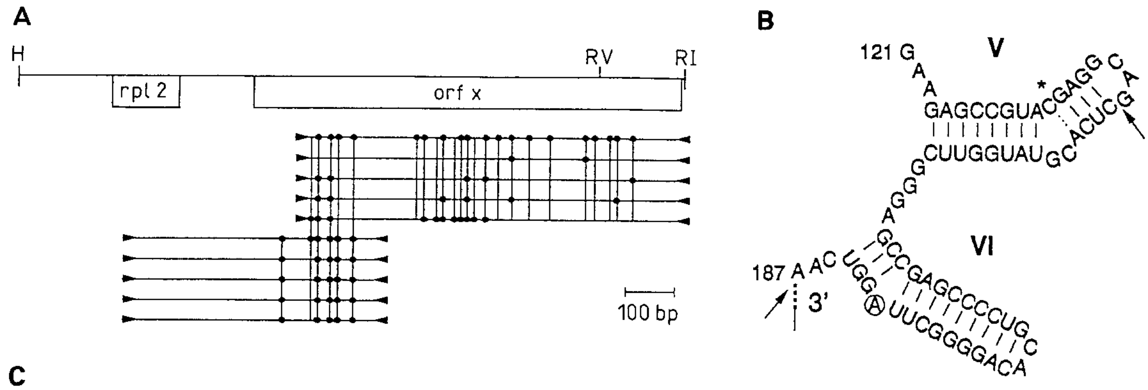


Fig. 2A, B. Conservation of *orfx* in other higher plant species. **A** Mitochondrial DNA from *Arabidopsis* (lanes 2, 5) and carrot (lanes 3, 6) and total DNA from potato (lanes 4, 7) was digested with *Hin*dIII and separated alongside a size marker (lane 1) by agarose gel electrophoresis (left panel, lanes 1–4). The *Eco*RV fragment derived from the *Oenothera orfx* clone (Fig. 1A) was used as a radiolabeled hybridization probe in a Southern blot (right panel, lanes 5–7) of the agarose gel. The different quantities of mitochondrial DNA in the three lanes are most probably responsible for the different hybridization signal intensities. **B** Northern blot with the same radiolabelled fragment as in **A** hybridized to mitochondrial RNA from *Arabidopsis* after electrophoresis and membrane transfer

The region of similarity, with 94% nucleotides identical, between *Arabidopsis* and *Oenothera* (alignment not shown) is confined to the region encoding *orfx* and does not include *nad5* exon c sequences. The 24 N-terminal amino acids of *orfx* beginning with a possible ATG start codon in *Oenothera* are, however, not conserved in *Arabidopsis*. Sequence similarity terminates where, in *Oenothera*, the similarity with the tRNA(Ser)-associated sequence begins, which includes the six C-terminal codons of the reading frame.

Using the FASTA algorithm (Pearson and Lipman 1988) an *orfx* homologous region was identified near the sugar beet *(Beta vulgaris) atp6* gene, which is published with the flanking regions as a database entry only (GenBank X55076). The homology of 850 nucleotides is located 400 nucleotides downstream of the *atp6* reading frame and starts at approximately the position where the *Ara-*



C

```
AAGCTTGACTGAGCAAAGAAGTCAAGGAACGAAGCTGCTTCTCTAATAGCCCCGTTGAATAGGAGGGCGAAGGCTTTAAAAAAAAGTTTGATTTAGGGAG  100

AGGGGGCTTCAAGTTCTTAGGAAGAGCCGTACGAGGCAGCTCACGTACGGTTCGGGAGCCGAGCCCCTGCACAGGGGCTTAGGTCAA C ACT TAT ATA 197
                    -------------------------------------------------------------------
                                                                                           T   Y   I

                                P1                          (C)
 ATA GCC AGT CAT CAA TTA GAA GCA GGC AAA ATG GTG ATG AAT TGC GAT TGG TCC AAA CCT TCG ACC AGC TCC TTC  272
  I   A   S   H   Q   L   E   A   G   K   M   V   M   N   C   D   W   S   K   P   S   T   S   S   F

 TTG CAA TCC GCC CAG AAT GAC CAT CCT AAG CCC TTA TTC ACT GTG TGA AGAACGGGGCAGTCCGCTGGAGTGTAGGCTTCTTG  355
  L   Q   S   A   Q   N   D   H   P   K   P   L   F   T   V   *

GCCACGCCCCCTGCTTATAGATACGAAATACTTGAGATAAATTATCAAATAGGAAATTGCATACCATTAGCCGATATACGTTTAGGAACATGGGTACATG  455

AACCCTA TAA AAT CCT TCA CTT TTA GCT TTG AAT TAC TTA TAT ATA TCC TAT GAA TTT AAT TTC GCA|TcG|GAA ACG   531
         *   N   P   S   L   L   A   L   N   Y   L   Y   I   S   Y   E   F   N   F   A |S→L| E   T

            P2      (G)
ATT CTA GGA GAA GTT CGA ATC CGT TCC GTT CGG ATA TTG ATC GGT CTT GGT TTG ACA TGG TTT ACG TGT TAC TGG   606
 I   L   G   E   V   R   I   R   S   V   R   I   L   I   G   L   G   L   T   W   F   T   C   Y   W

|TTc|cCG|GAA GAG TTA ATA|TcT|CcA|TTA GCG|TcA|CCC TTT CTT ACC CTG CCT TTT GAC|TcG|TAT TTT GTT TGT ACA   681
|F→F|P→S| E   E   L   I |S→F|P→L| L   A |S→L| P   F   L   T   L   P   F   D |S→L| Y   F   V   C   T

                                        P3
CAA TTA ACG GAG GCC TTT TCG ACA TTT GTT GCA ACG TCT TCA ATA GCA TGC TCT TAT TTC GTC TTT CCT TTA ATA   756
 Q   L   T   E   A   F   S   T   F   V   A   T   S   S   I   A   C   S   Y   F   V   F   P   L   I

AGT TAT CAA ATT TGG TGC TTT TTG ATC CCC AGT TGC TAT GGA GAA CAA AGG ACG AAA TAC AAT CGA TTC|cTC|cAT|  831
 S   Y   Q   I   W   C   F   L   I   P   S   C   Y   G   E   Q   R   T   K   Y   N   R   F |L→F|H→Y|

TTA AGT GGT|TcT|cGC|TTC TTC TTG TTC CTG TTC CTA ACT|ccT|cCC|cGG|GTC GTT CCC AAT GTT TGG CAC TTT|CcA|  906
 L   S   G |S→F|R→C| F   F   L   F   L   F   L   T |P→F|P→S|R→C| V   V   P   N   V   W   H   F |P→L|

TAC TTC GTG GGT GCA ACA TCA ACA AAT TCG|CTc|ATG ATC AAG TTA CAA CCT AAG ATC TAT GAC|cAT|ATT ATG TTA   981
 Y   F   V   G   A   T   S   T   N   S |L→L| M   I   K   L   Q   P   K   I   Y   D |H→Y| I   M   L

ACT GTT CGT ATT TCG TTC ATT|cCA|TCG GTA TGC TCC CAG GTA CCT GTA ATT GTG ATC TGT TTG|CcA|GAA CCA AGG  1056
 T   V   R   I   S   F   I |P→S| S   V   C   S   Q   V   P   V   I   V   I   C   L |P→L| E   P   R

GGT CTT TCT TTG GAA ACC TTC ACG AAC AAT CGT CGT TTT TTG ATG GTT TTT|cCG|CTT CTC ACA GCT GCT CTT|TcC|  1131
 G   L   S   L   E   T   F   T   N   N   R   R   F   L   M   V   F |P→S| L   L   T   A   A   L |S→F|

ACA CCT CCG GAT ATC TGG TGC CAA|ATc|GTC GCC|cGT|TTC CTT ATT TCT TTG ATA ATA GAG TTG GCT ATT TTT GTG  1206
 T   P   P   D   I   W   C   Q |I→I| V   A |R→C| F   L   I   S   L   I   I   E   L   A   I   F   V

GCA|TcG|ATT GTA CAA GTT CGT GAA GAG GGC TGG ACG AGT GGA ATG AGG GAG AGC GGC TCG ATC GAG AAA AAA AAT  1281
 A |S→L| I   V   Q   V   R   E   E   G   W   T   S   G   M   R   E   S   G   S   I   E   K   K   N

                                P4
AAG AGT CGC CCC CCC CCT AGA ACC TTG GCA AAG TAA CTATCAATGAATTCNNNN
 K   S   R   P   P   P   R   T   L   A   K   *
```

bidopsis and *Oenothera* homology begins (not shown). The *orfx* reading frame in sugar beet is conserved with the exception of a single G insertion. Whether this homologous region in the sugar beet mitochondrial genome indeed encodes the functional *orfx* gene remains to be investigated.

An rpl2 *gene is possibly encoded upstream of* orfx *in* Arabidopsis

Sequences upstream of *orfx* in *Arabidopsis* reveal high similarity to the gene encoding ribosomal protein L2 *(rpl2)* in the *Marchantia* mitochondrial genome (Fig. 3). In *Arabidopsis,* this region of homology is preceded by a nucleotide sequence that can be folded into the highly conserved secondary structure of domains V and VI of organellar group II introns (Michel et al. 1989) as shown in Fig. 3B. The intron structure conforms to all structural elements characteristic of group II introns. The only exception is the distal domain V stem, which is 2 bp shorter than the typical length. The single group II intron in the *Marchantia rpl2* gene is located in a different position, while the intron structure and position identified in *Arabidopsis* is conserved in *Oenothera* (W. Schuster, personal communication). PCR amplification of *Arabidopsis* mitochondrial cDNA using an oligonucleotide homologous to part of the 5′ exon of *Oenothera rpl2* as upstream and *Arabidopsis orfx* oligonucleotide P3 as downstream primers yielded products spliced at the predicted position, supporting the idea that a functional *rpl2* gene is present at this location.

The orfx *mRNA is edited*

The initially identified cDNA clone c9 from *Oenothera* showed several sites of RNA editing by C to U exchanges in *orfx*. The *Oenothera* cDNA library (Wissinger et al. 1991) was rescreened with an *orfx*-specific probe and two additional cDNA clones (c3, c6) were identified that cover the N-terminal region of the reading frame and show additional RNA editing events (Fig. 1).

Specific primers were synthesized as indicated in Figs. 1 and 3 to amplify *Arabidopsis* and *Oenothera* mitochondrial cDNA. The cloned and sequenced PCR products revealed 20 editing sites in the *Oenothera* and 25 in the *Arabidopsis orfx* coding regions. The two species have 10 editing sites in common, while the others are either genomically encoded as T or were not found to be edited in the other plant species. Two of the silent editing sites in third codon positions are identical in the two species. Six more editing sites are expected in each species in fully edited mRNAs to increase the number of amino acid identities to the *Marchantia* sequence (Fig. 4), but were not observed in the cDNA clones analysed. Editing of the ACG codon at positon 530 of the *Arabidopsis orfx* sequence (Fig. 3C) to an otherwise absent ATG startodon was not observed in 10 independent cDNA derived PCR clones. Five of these PCR clones with almost complete editing within *orfx* were completely sequenced (Fig. 3A). No editing events were detected in the part of the *rpl2* exon investigated nor in the intergenic region between *rpl2* and *orfx*.

Orfx *is homologous to* orf244 *in the* Marchantia polymorpha *mitochondrial genome and to two overlapping open reading frames in* E. coli

Database analysis with the *Oenothera* and *Arabidopsis orfx* sequences revealed high similarity of *orfx* to *orf244* in the *M. polymorpha* mitochondrial genome (Oda et al. 1992) (Fig. 4). The RNA editing events observed in the higher plants generally increase similarity to the *Marchantia* amino acid sequence. In *Oenothera* 15 of 17 and in *Arabidopsis* 16 of the 20 observed non-silent editing events establish amino acid identity with the *Marchantia* polypeptide. The investigated cDNA clones suggest that RNA editing in the higher plants is confined to the region of sequence homology to *Marchantia.* The homologous *orf244* in *Marchantia* is located between a tRNA gene and the gene encoding subunit 2 of the cytochrome oxidase *(cox2).*

From the four plant polypeptide sequences available at this time the profile of the hypothetical protein was calculated and used for database searches with the profile analysis method (Gribskov et al. 1987) of the UWGCG program package (Devereux et al. 1984). This strategy revealed homology between *orfx* and two overlapping reading frames (*orf154* and *orf131*) in the *E. coli* genome (Daniels et al. 1992) with 22% identical and 54% similar amino acids (Fig. 5A). In the *E. coli* genome these reading frames are located between the genes *udp* and *rfaH.* Interestingly, translation of the second orf *(orf131)* was suggested to be initiated at a GTG codon, a situation frequently found in *E. coli.* Considering the homology of the two reading frames to *orfx,* cotranslation of *orf154* and *orf131* by a (−1) frameshift may be possible, producing a single protein of homology to *orfx* over its entire length. The homology of *orfx* to *orf154* and *orf131*

Fig. 3A–C. Genomic arrangement of *orfx* in *Arabidopsis.* **A** In *Arabidopsis orfx* is encoded in a genomic environment different from that in *Oenothera.* For cloning in expression vectors cDNA was amplified using primers P2 and P4. For the complete analysis of RNA editing additional cDNA derived PCR-clones were sequenced betweed primers P1 and P3. *Dots* indicate edited sites in the individual cDNA clones. Restriction sites are indicated for *Hin*dIII (H), *Eco*RV (RV) and *Eco*RI (RI). **B** The region with similarity to *rpl2* is preceded by a nucleotide sequence that can assume a secondary structure typical of domains V and VI of group II introns. The unpaired A residue for branch site formation is *encircled,* and the 3′ splice site and a 2 bp deletion in domain V are indicated by *arrows.* Editing of the C residue marked by an *asterisk* would improve the secondary structure. **C** Nucleotide sequence of the *Hin*dIII-*Eco*RI fragment indicated in **A**. Sequence features are as in Fig. 1B with the sequence shown in **B** marked with a *broken line.* PCR primer binding sites are marked by *lines above* (for sense primers) and *below* (for antisense primers) the sequence and nucleotides altered in the primers for cloning are shown in *parentheses*

```
Oenothera (gen.)                                               S
Oenothera (cDNA)    MIKVRGDILLTGRYYGGVIELKLSFLTFCPLNYSYISSKFDFALETILEE  50
Marchantia                                                 MNFVLKTILEE
Arabidopsis (cDNA)                          NPSLLALNYLYISYEFNFALETILGE
Arabidopsis (gen.)                                             S

      R           R  FP    S                    S   SP
VRIRSVWILIGLGLTWFTCYWFSEEFIFLLAKPFLSLP-LDSYFVCTQLTEAFSTYVATSSIACSYFVFP 120
VRIRVFWILICFSFTWFTCYWFSEEFIFLLAKPFLTLPYLDSSFICTQLTEALSTYVTTSLISCFYFLFP
VRIRSVRILIGLGLTWFTCYWFSEELIFLLALPFLTLP-FDSYFVCTQLTEAFSTFVATSSIACSYFVFP
                     FP    SP  S

   H                          SR        P
FISYQIWCFLIPSCYGEQRTKYNRFFHLSGFCFSLFLFLTLPRVVPNVWHFPYFVGATSTNSLMIKLQPK 190
FLSYQIWCFLMPSCYEEQRKKYNKLFYLSGFCFFLFFFVTFVWIVPNVWHFLYKLSTTSTNLLIIKLQPK
LISYQIWCFLIPSCYGEQRTKYNRFFYLSGFCFFLFLFLTFSCVVPNVWHFLYFVGATSTNSLMIKLQPK
                         LH   SR        PPR        P          L

           SF P               P                                  I  R
IYDHIMLTVRILFISSVCSLVPVILICLPELRGLSVETFTNNRRFLMVFSLLTAALSTPPDIWCQIVACF 260
IFDYIMLTVRILFISSICSQVPVLVICLLESKG--VKTFIKNRRFFYSFFAFHSSFFYTPDIWCQIIACL
IYDYIMLTVRISFISSVCSQVPVIVICLLEPRGLSLETFTNNRRFLMVFSLLTAALFTPPDIWCQIVACF
   H          P             P                    P      S        I  R

            S
LISLIIELAIFVALIVQVREEGWTSGMRESGSTRLI
PIYFIIELTIFYALIIQVYKKQLAL
LISLIIELAIFVALIVQVREEGWTSGMRESGSIEKKNKSRPPPRTLAK
            S
```

Fig. 4. Comparison between the protein sequences deduced from *orfx* from *Arabidopsis* and *Oenothera* and the *Marchantia orf244* reading frames. Amino acid changes introduced by RNA editing are indicated. *White letters* on *black ground* highlight amino acids identical in at least two of the species compared

A

```
                                                            S            R
Oenothera orfx   MIKVRGDILLTGRYYGGVIELKLSFLTFCPLNYSYISSKFDFALETILEEVRIRSVWILIGLG   63
                                                    ¦| : : ¦| | | ¦¦| | ¦ :¦|¦¦
E.coli orf154                                       MSVEDTQPLITHLIELRKRLLNCIIAVI   28

     R  FP    S                    S   SP                   H
LTWFTCYWFSEEFIFLLAKPFLS-LPLDSYFVCTQLTEAFSTYVATSSIACSYFVFPFISYQIWCFLIPSCYGEQRTKYN  142
¦:¦¦:  ¦|:¦¦¦: |¦: |¦¦: || ¦| ¦¦  ¦¦::¦| | ¦  : ¦:¦ :¦  |:| ||¦|:|¦ |: | :¦|
VIFLCLVYFANDIYHLVSAPLIKQLPQGSTMIXXDVASPFFTPIKLTFMVSLILSAPVILYQVWAFIAPALYKHERRLVV  108

       SR        P                                         SF P               P
RFFHLSGFCFSLFLFLTLPRVVPNVWHFPY-FVGATSTNSLMIKLQPKIYDHIMLTVRILFISSVCSLVPVILICLPELR  221
:¦¦  |    ||¦¦¦¦:¦::  |:¦| |:¦  ¦¦     || |: ::| ¦ :   :: |
PLLVSS----SLLFYIGMHS--PTLWSFRW-HLA-----SLPIP-RRKGCRYPPTSPAI*                      154

                 P                                         SF P               P
Oenothera orfx   LPRVVPNVWHFPY-FVGATSTNSLMIKLQPKIYDHIMLTVRILFISSVCSLVPVILICLPELR  221
                        |¦:¦¦¦ |¦¦:|::¦¦¦ ¦: ¦   |  ¦¦¦:¦ ¦ |  |:   |¦¦¦|¦|¦ ¦¦
E.coli orf131          MVFPLAFGFLANTAPEGVQVSTDIASYLSFVMALFMAFGVSFEVPVAIVLLCWMGI-   56

                                 I  R              S
GLSVETFTNNRRFLMVFSLLTAALSTPPDIWCQIVACFLISLIIELAIFVALIVQVREEGW---TSGMRESGSTRLI*    295
  |:|:¦ ::|:¦¦¦| :¦¦:¦ | ||||¦¦¦|:¦ :¦ ¦  ¦¦|¦¦¦|:: ¦  :¦¦ :¦   ::¦  ||¦:|
-TSPEDLRKKRPYVLVGAFVVGMLLTPPDVFSQTLLAIPMYCLFEIGVFFSRFYVGKGRNREEENDAEAESEKTEE*     131
```

B

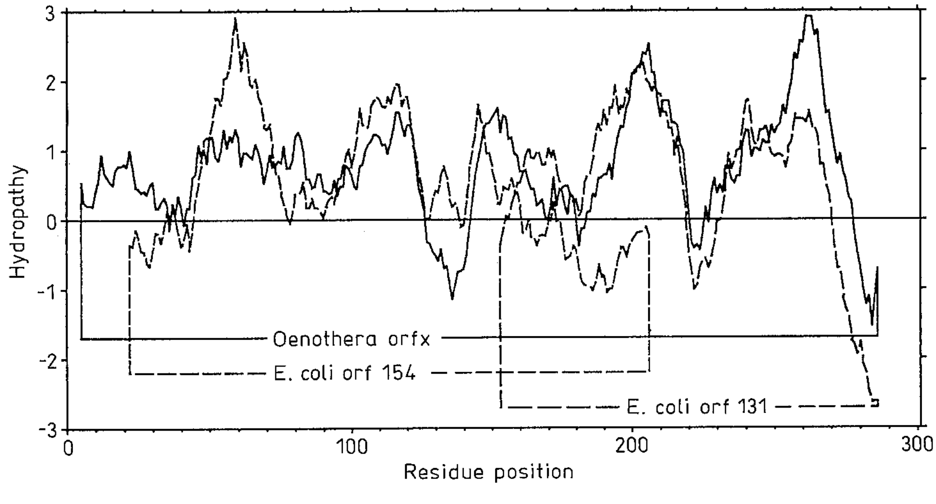


Fig. 5A, B. *Orfx* has homology to the two overlapping reading frames *orf154* and *orf131* in *Escherichia coli*. **A** *Vertical lines* in the alignment indicate identical residues, *broken lines* and *colons* indicate amino acid exchanges with high and low degrees of similarity, respectively, as introduced by the default settings of the program GAP of the UWGCG package. The ATG start codon of *orf154* is juxtaposed to an ATA codon in the higher plant species. The amino acid sequence derived from the *Beta vulgaris* entry is more similar to the *E. coli* sequence in the carboxy-terminal region, where a recombination event has altered the *Oenothera* reading frame (not shown). **B** The homology of both *orf154* and *orf131* to *orfx* is identified by comparison of the hydropathy plots of the three polypeptides. Hydropathy profiles were obtained using the PEPWINDOWS program of the UWGCG package with the window size set to 19 is strongly supported by hydropathy analysis of the polypeptide sequences (Fig. 5B). The hydropathy profiles are very similar over the entire length of the polypeptides except the amino-terminal extension of the *Oenothera* reading frame. The hydropathy analysis also suggests a translational switch between *orf154* and *orf131* in their overlapping region to produce a protein with colinearity to *orfx*. Alternatively, a functional homologue to *orfx* could be reconstituted as a heterodimer in *E. coli*.

Discussion

Is orfx *functionally expressed?*

Several arguments support *orfx* as a candidate for a functionally expressed gene in plant mitochondria. Firstly, sequences with homology to *orfx* are detected in at least five higher plant species and a homologue of *orfx* is also present in the *Marchantia* mitochondrial genome indicating a wide taxonomic distribution. Secondly, the reading frame is highly conserved between *Arabidopsis* and *Oenothera*. Thirdly, the more than 34% thymidine nucleotides observed for *orfx* at third codon positions are a common feature of functional genes in plant mitochondria. Fourthly, the RNA editing observed in *orfx* mRNAs is a feature indicative of functionally expressed proteins in these organelles. RNA editing in *orfx* increases its similarity to the *Marchantia* homologue *orf244*, while editing in mRNAs of an *rps19* pseudogene has been shown to decrease its similarity to functional copies of this gene (Schuster and Brennicke 1991).

The different genomic environments of *orfx* in the different plant species most likely do not interfere with its functional importance, since genes are generally highly conserved at the sequence level, but located in different genomic vicinities in mitochondrial genomes of different plant species (Palmer and Herbon 1988). In *Oenothera orfx* can be cotranscribed with the central exon c of the *nad5* gene and even stay linked to this exon after addition of the terminal *nad5* exons by *trans*-splicing (Knoop et al. 1991).

How is orfx *translation initiated in higher plants?*

No in-frame ATG start codons for the *orfx* reading frames are present in *Arabidopsis* and sugar beet. These incomplete reading frames could thus be interpreted as pseudogenes. This explanation, however, appears unlikely since all other features of the *orfx* reading frames, as outlined above, are suggestive of functional genes.

In this context it is intriguing to find an ATA (isoleucine) codon in all three plant species juxtaposed to the ATG start codon of *orf154* in *E. coli* in the alignment shown in Fig. 5A. A potential ATG start codon 35 codons upstream of the conserved ATA codon in the *Oenothera* reading frame is outside of the conserved sequence region, which extends only approximately 15 nucleotides upstream of this ATA codon in all three higher plants. The *Marchantia* reading frame begins four codons downstream of the conserved ATA codon. Could this ATA (isoleucine) codon be used for initiation of translation? In mitochondria of animals, ATA codons are frequently used for translation initiation instead of ATG (methionine) codons, but no evidence for such a divergent translation initiation mechanism has as yet been found in plant mitochondria.

The first codon that is conserved in the *orfx* reading frames of all three higher plant species for which sequence data are available is a TTG (leucine) codon. Evidence for translation initiation at TTG codons has been reported for mitochondria in nematodes (Okimoto et al. 1990) and appropriate TTG codons might likewise be used to initiate translation in plant mitochondria.

What is the function of the orfx *polypeptide?*

No function can as yet be ascribed to the protein product encoded by *orfx*. Attempts to express this protein in *E. coli* have failed up to now (data not shown). The *E. coli* transformants cease to grow immediately after induction of *orfx* expression constructs and no protein product is detected on electrophoresis of protein extracts. This phenotype may somehow be related to the observed homology between the two reading frames *orf154* and *orf131* and *orfx*. Again, no function has as yet been ascribed to these bacterial reading frames.

Complementation experiments with *orf154/orf131* deletion mutants of *E. coli* could confirm homologous functions of these open reading frames in the bacterium and *orfx* in higher plant mitochondria. These experiments will also allow investigation of differentially edited *orfx* cDNA constructs. Although RNA editing in plant mitochondria has indeed been shown to alter the encoded protein (Graves et al. 1990), it is as yet unclear whether unedited versions of proteins might fulfil their functions.

Acknowledgements. We thank Wolfgang Schuster for continuous interest and unpublished data, and Bernd Wissinger, Rainer Ternes, Stefan Binder and Wolfgang Schuster for gifts of DNAs. Work described in this manuscript was made possible by grants from the Deutsche Forschungsgemeinschaft and the Bundesministerium fur Forschung und Technologie.

References

Binder S, Knoop V, Brennicke A (1991) Nucleotide sequences of the mitochondrial genes *trnS*(TGA) encoding tRNA-Ser(TGA) in *Oenothera berteriana* and *Arabidopsis thaliana*. Gene 102:245–247

Bonen L (1991) The mitochondrial genome: so simple yet so complex. Curr Opin Genet Dev 1:515–522

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Covello PS, Gray MW (1989) RNA editing in plant mitochondria. Nature 341:662–666

Daniels DL, Plunkett IIIrd G, Burland V, Blattner FR (1992) Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. Science 257:771–778

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Graves PV, Bégu D, Velours J, Neau E, Belloc F, Litvak S, Araya A (1990) Direct protein sequencing of wheat mitochondrial ATP synthase subunit 9 confirms RNA editing in plants. J Mol Biol 214:1–6

Gribskov M, McLachlan AD, Eisenberg D (1987) Profile analysis: Detection of distantly related proteins. Proc Natl Acad Sci USA 84:4355–4358

Gualberto JM, Lamattina L, Bonnard G, Weil J-H, Grienenberger J-M (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. Nature 341:660–662

Hiesel R, Wissinger B, Schuster W, Brennicke A (1989) RNA editing in plant mitochondria. Science 246:1632–1634

Hodgson CP, Fisk ZZ (1987) Hybridization probe size control: optimized "oligolabeling". Nucleic Acids Res 15:6295

Knoop V, Schuster W, Wissinger B, Brennicke A (1991) *Trans* splicing integrates an exon of 22 nucleotides into the *nad5* mRNA in higher plant mitochondria. EMBO J 10:3483–3493

Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II introns – a review. Gene 82:5–30

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. J Mol Biol 223:1–7

Okimoto R, Macfarlane JL, Wolstenholme DR (1990) Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans*. Nucleic Acids Res 18:6113–6118

Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol 28:87–97

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Pereira de Souza A, Jubier M-F, Delcher E, Lancelin D, Lejeune B (1991) A trans-splicing model for the expression of the tripartite *nad5* gene in wheat and maize mitochondria. Plant Cell 3:1363–1378

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schuster W, Brennicke A (1991) RNA editing makes mistakes in plant mitochondria: editing loses sense in transcripts of a rpsl9 pseudogene and in creating stop codons in coxI and rps3 mRNAs of *Oenothera*. Nucleic Acids Res 19:6923–6928

Schuster W, Hiesel R, Wissinger B, Schobel W, Brennicke A (1988) lsolation and analysis of plant mitochondria and their genomes. In: Shaw CH (ed) Plant molecular biology. IRL Pres, Oxford, England, pp 79–102

Wissinger B, Schuster W, Brennicke A (1991) *Trans* splicing in *Oenothera* mitochondria: *nad1* mRNAs are edited in exon and *trans*-splicing group II intron sequences. Cell 65:473–482

Wissinger B, Brennicke A, Schuster W (1992) Regenerating good sense: RNA editing and trans-splicing in plant mitochondria. Trends Genet 8:322–328