Pea cytochrome oxidase subunit II gene has no intron and generates two mRNA transcripts with different 5'-termini

Eunpyo Moon, Teh-hui Kao and Ray Wu

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received 12 February 1985; Accepted 5 April 1985

ABSTRACT

We have isolated and sequenced the cytochrome oxidase subunit II (COII) gene from pea mitochondria. The coding sequence (777 bp) shows over 90% homology to the COII genes from three monocotyledonous plants (rice, maize and wheat) and one dicotyledonous plant (*Oenothera berteriana*). Several codons are deleted, however, in the pea COII gene. Of interest is the deletion in pea of the last three codons, including the stop codon, found at the 3' end of the other four COII genes. Instead, a new stop codon has been created due to a single-base substitution at the 13th bp downstream from the position of the original stop codon. This pea gene does not contain an intron which is found in all three monocots. Two distinct 5' termini of the pea COII transcripts have been identified by S1 nuclease mapping, one at 285 bp (site I) and the other at 302 bp (site II) upstream from the ATG codon. They are located at two identical sites within nearly perfect direct repeats. Transcripts with the 5' end corresponding to site I occur five time more frequently than those with the 5' end corresponding to site II. Both transcripts have the same 3' terminus which has been mapped to be at 193 to 195 bp downstream from the stop codon.

INTRODUCTION

There are approximately 18 to 20 polypeptides synthesized in plant mitochondria, a number representing only a small fraction of the proteins found within these organelles¹. Most of the mitochondrial proteins are encoded in the nuclear genome, synthesized on cytosolic ribosomes and then imported into the organelles². To our knowledge, none of the nuclear genes encoding plant mitochondrial proteins have been isolated yet. Only a few of the mitochondrial genes that code for this organelle's proteins have been identified, one of which is the gene for subunit II of cytochrome oxidase. This COII gene has been isolated and sequenced from three monocots: maize³, rice⁴ and wheat⁵; and one dicot: *Oenothera berteriana*⁶.

Comparison of the nucleotide sequences of the four COII genes reveals several interesting features:

(1) The COII genes of the three monocots contain an intron at exactly the same position, while the dicot COII gene, like the animal, fungal and

Drosophila genes, does not contain an intron.

(2) The coding sequence of the COII gene is highly conserved among the three monocots (99% homology). There is also strong homology between each of the three monocots and the evolutionarily distant dicot (88% homology). This phenomenon is in sharp contrast to what is observed in animal mitochondrial genes which have evolved much more rapidly⁷.

(3) In the 5'-noncoding region, no sequence homology can be detected between rice and maize, between rice and Oenothera, or between maize and Oenothera. The 5'-noncoding sequences of rice and wheat COII genes, however, show 97.4% homology up to -154 bp (the published sequence data for wheat only include 154 bp in the 5'-noncoding region).

(4) The 3'-noncoding sequence is highly conserved in COII genes of the three monocots (over 90% homology), but not in that of Oenothera.

(5) Codon usage is different between animal/fungal and plant mitochondrial genes in several cases. Two cases of particular interest are: CGG codes for Arg in animal/fungi but probably codes for Trp in plants³, and TGA codes for Trp in animal/fungi but is not found in the COII gene of these four plants. Perhaps TGA is used as a stop codon in plant mitochondria as in the universal genetic code³.

We extended the study of the COII gene to another dicot, pea, with two objectives. First, we wished to compare the sequence of the pea COII gene with those of the other four plants, examining in particular the features described above. Secondly, we hoped to determine the sequences (promoter and/or any other regulatory elements) which regulate the expression of plant mitochondrial genes. These sequences have already been determined for yeast⁸ and human⁹ mitochondrial genes, but not yet for plants.

Here, we report the isolation and sequence analysis of the pea COII gene. In addition, as the first step toward the study of the regulatory sequences, we have determined the location of the 5' and 3' termini of the pea COII transcripts.

MATERIALS AND METHODS

Isolation of pea mitochondrial DNA and RNA

Captan treated pea seeds (Pisum sativum L. Progress No. 9 Agway) were soaked in water overnight with aeration and then grown under dark conditions in vermiculite at room temperature for 2 weeks. Mitochondrial DNA was isolated as described previously⁴. To obtain mitochondrial RNA, the purified mitochondria in the homogenization buffer (50 mM Tris-HCl, pH

8, 3 mM EDTA, 0.44 M sucrose) were lysed with $\frac{1}{2}$ vol. of the lysis buffer (25 mM Tris-HCl, pH 7.6, 20 mM EDTA, 10% sodium sarkosyl) at 4°C for 15 min, extracted with phenol three times, and then precipitated with ethanol. Total RNA was dissolved in a buffer containing 25 mM Tris-HCl, pH 8, and 0.1 mM EDTA. All solutions used for RNA preparations were pretreated with 0.1% diethylpyrocarbonate overnight and the glassware was baked at 200°C overnight.

Construction of mitochondrial genomic library and isolation of COII clones

HindIII and EcoRI libraries were separately constructed essentially as described previously⁴. A 2.4 kb EcoRI fragment from pZmE1³ containing the maize COII gene was used as a probe to screen the libraries for the pea COII gene. Hybridization conditions were the same as described previously⁴.

Restriction enzyme digestion and DNA sequence analysis

Restriction enzymes were purchased from either New England Biolabs or Bethesda Research Laboratories. DNA samples were digested with restriction enzymes in a buffer containing 50 mM Tris-HCl, pH 7.6, 50 mM KCl, 10 mM MgCl₂, 10 mM dithiothreitol, and fractionated on 1% agarose gels. DNA fragments were transferred to nitrocellulose filters by the method of Southern¹⁰.

DNA fragments containing the pea COII gene were subcloned into M13 cloning vectors, mp10 and mp11, and DNA sequencing was carried out by using the dideoxynucleotide chain-termination method¹¹.

Northern blot analysis of RNA transcript

Total mitochondrial RNA was electrophoresed under denaturing conditions on 1% agarose-formaldehyde gels¹², and the RNA molecules were transferred to nitrocellulose filters¹³. Hybridization and washing conditions were the same as described previously⁴.

S1 nuclease mapping of 5' and 3' ends of COII transcript

Single-stranded DNA was prepared from recombinant M13 phage containing a 0.98 kb EcoRI-BglII fragment (-366 bp to 617 bp of pea COII gene) cloned into the EcoRI and BamHI sites of the M13 mp10 vector. It was then used as a template and a 152 bp HinfI-NcoI fragment (-99 bp to 53 bp) was used as the primer in synthesizing the radioactively labelled probe for mapping the 5' end of the transcript. Template and primer, in a molar ratio of 1 to 2, were mixed in a 14 μ l solution containing 60 mM NaCl, 7 mM Tris-HCl, pH 7.4, 7 mM MgCl₂, 100 μ g/ml bovine serum albumin, and heated at 100°C for 10 min. Annealing occurred during slow cooling to room temperature. Repair

synthesis was carried out at 30°C for 15 min with the addition of 8 μ Ci [α -³²P]dATP (410 Ci/mmol), 250 μ M each of unlabelled dGTP, dCTP and dTTP, and 4 units of large fragment of DNA polymerase I. The reaction was then chased with 250 μ M unlabelled dATP for another 15 min. After heating at 70°C for 10 min to inactivate the enzyme, 10 units of EcoRI were added and digestion proceeded for 1 hr. The single-stranded radioactive EcoRI-NcoI fragment of 419 bp was isolated on an 8% polyacrylamide-7 M urea gel. The subsequent annealing with total mitochondrial RNA and S1 nuclease digestions were performed essentially as described by Favaloro et al.¹⁴. Briefly, the radioactive DNA probe (about 10,000 cpm) and 20 μ g of total mitochondrial RNA were dissolved in 30 μ l of hybridization buffer containing 80% formamide, 0.04 M Pipes, pH 6.4, 0.4 M NaCl, 1 mM EDTA. The mixture was heated at 80°C for 15 min and then quickly transferred to a 50°C water bath. Annealing of DNA-RNA was allowed to proceed for 9 hrs. 300 μ l of S1 buffer (0.28 M NaCl, 0.05 M NaOAc, pH 4.65, 4.5 mM ZnSO₄, 20 μ g/ml denatured calf thymus DNA) was added and, after mixing, digestion of single-stranded DNA was carried out with the addition of 70 units S1 nuclease (BRL) at room temperature for 30 min. The reaction was stopped by the addition of 25 μ l 7.5 M NH₄OAc and 10 μ l 0.5 M EDTA. The S1-resistant DNA fragment was analyzed on an 8 % polyacrylamide-7 M urea gel.

The DNA probe for mapping the 3' end of the transcript was obtained as follows: A 0.37 kb BglIII-XbaI fragment (617 bp to 987 bp) was cloned into the BamHI and XbaI sites of the M13 mp10 vector. Single-stranded DNA prepared from this recombinant M13 phage was used as a template and a synthetic 17 bp sequencing primer, 5'-GTTTCCCGTCACGAC-3' (New England Biolabs), was used as the primer in synthesizing the radioactive probe essentially as described above. After repair synthesis, the mixture was digested with 10 units HindIII and the single-stranded radioactive HindIII-HindIII fragment of 206 bp was isolated on an 8% polyacrylamide-7 M urea gel. Annealing to mitochondrial RNA and S1 nuclease digestions were performed as described above except that the annealing temperature was 53°C.

RESULTS

Identification and isolation of pea COII gene

Two samples of total pea mitochondrial (mt) DNA were separately digested with EcoRI and HindIII, fractionated on a 1% agarose gel (Figure 1A, Lane 1 and 4), and then transferred to a nitrocellulose filter. A 2.4

Fig. 1 (A) Restriction digestions of pea mtDNA and pea COII gene - containing plasmids, pPSox2-E and pPSox2-H. Pea mtDNA was digested with EcoRI (Lane 1) and HindIII (Lane 4); pPSox2-E and pPSox2-H were digested with EcoRI (Lane 2) and HindIII (Lane 5), respectively. The digests were fractionated on a 0.7% agarose gel and stained with ethidium bromide. EcoRI and HindIII digested λ DNA was used as a size marker (Lane 3). The 1.8 kb EcoRI fragment (a) and the 1.4 kb Hind III fragment (b) containing the pea COII gene are indicated. (B) Southern blot analysis of the agarose gel shown in (A). DNA fractionated on the 0.7% agarose gel was transferred to nitrocellulose and hybridized with the nick-translated maize COII probe. Lanes 1-5 correspond to Lanes 1-5 shown in (A).

kb EcoRI fragment containing the maize COII gene³ was labelled with ³²P by nick-translation and then used as a hybridization probe to detect the corresponding pea gene. The results of the Southern analysis, shown in Figure 1B, reveal that a 1.8 kb EcoRI fragment and a 1.4 kb HindIII fragment hybridized to the maize probe (Lane 1 and 4).

Restriction Mapping and Sequencing Strategy of Pea COII Gene

Fig. 2 Restriction map of pea COII gene and DNA sequencing strategy. The restriction maps of the pea COII inserts in pPSox-2E and pPSox-2H were determined separately by gel electrophoretic analyses of several single or double enzyme digestions. The composite map is shown. The thick line indicates the coding region. Arrows below the map show the direction and extent of sequence determined from each restriction site. Abbreviations: B, BglII; E, EcoRI; H, HindIII; Hf, HinfI; N, NcoI; S, SstI; X, XbaI; Xh, XhoI.

Two separate mitochondrial genomic libraries were constructed by cloning EcoRI and HindIII digests of pea mtDNA into the plasmid vector pBR322. Approximately 3,000 clones from each library were screened with the radioactive maize probe. 4 and 3 positive clones were obtained from EcoRI and HindIII libraries, respectively. Restriction analysis showed that all the positive EcoRI clones contained the same 1.8 kb insert as had been identified by the Southern analysis (Lane 2 of Figures 1A and 1B shows the results of one of these clones). Likewise, all 3 positive HindIII clones contained the same 1.4 kb insert seen on the Southern analysis (Lane 5 of the same Figures shows one of these results).

The EcoRI clone was designated pPSox2-E. The HindIII clone was designated pPSox2-H. Plasmid DNA from these two clones was single or double digested with various restriction enzymes, separately, to determine the restriction maps. A composite map is shown in Figure 2.

DNA sequence analysis of pea COII gene

Restriction fragments of the entire 2.1 kb DNA fragment (a composite of the 1.8 kb EcoRI fragment and the 1.4 kb HindIII fragment) were subcloned into M13 vectors, mp10 and mp11. The DNA sequence was determined, according to the strategy indicated in Figure 2, by the dideoxynucleotide chain-termination method¹¹ using a synthetic 17-mer as a primer (New England Biolabs). Examination of the DNA sequence reveals that the 2.1 kb fragment contains a 777 bp coding sequence, a 628 bp 5'-noncoding sequence, and a 630 bp 3'-noncoding sequence (Figure 3).

Pea	-62R HindIII	-800	5'-AAGCTT TCTATGCGATT GCACCTTTAG AAAGGAAATC AGCATTTAGG ACAATTTTAC GAACCAAGCA /AAATGGGAC
Pea	-550	-500	ATTTCGATGT CTGCTAGGAA AGCAAAATCGG ATCTATGAAA AAATTCCTCA AAGAAAAGTG TCGATTGACG TTTTAAAGAG
Pea	-450	-400	TAAATAAAAA AGGTCCTCTG TAGCTTTTCC GGAGGAAGTT ACTTATTATA TATAATAATT GATACCTGAT
Pen	<u>EcoRI</u>	-350	GTANGAGAGG AAGTCTCTGG TTGCTTTCAA TTCAATCAAA AAAAGTACCA TTTTCAATTG AATCCTGAAT CCCTTATTCT
Pea	-300	-250	ATAAATTAC TAAGAGAACA AATCACGTAA GTGATAGAAA GAATCGTTCA GTAGCGTAAT CTTGAAAGT TTGCTTTTC
Rice	5'-GTGACGAT		TCGATA CTT C CGA TGGT GATTTCAG T TAAGGG G GAA G AT AAAAGGAAGG
Pea	<u>SacI</u>	-200	GATTCAGAAA GGGGACGCC CAAGACGCTC TCCAATAAGT GCACCGAAAG GGGGGCGGAG AGACGGTCAA CTTGAGATCG
Oenothera		5'-CCTA	C TTC T A C C T TC
Rice	A CGAGATT G AACC A A	CA AG A	ATG TGC C T G A C C TATAGT T TEATTG G C CTTGCTC
Pea	-150	-100 HinfI	GCTAAGATCG AAAAGTACT TGCTTTACAG CCAAGCAACG ACAAAGAAAA AAGAGTCCCA TGCTTTCTCT GATTGCTAAA
Oenothera	TTTGTTGAC T GC CG C	TT C TT T	TTTGAATTAA TTGGGC CTT TCA ACAG CCCTCGC A CTCGAG: T
Rice	A GC CAA CGCT CCAG	TA CGCTT A	GC AC TAGC TTTTG G TTAGAAGT CGTCAAGG ATG TG C T
Maize			5'-CA C AA TC CGA CTGAATTATG
Pea	-50	-1	CCCAACCAGC GATTTACAAC AAACAAGT-T CTTCCT-CT TGTGGGGGG AGC-AG--T TAAAAATCA ACCAAGCAA
Oenothera	AAA GGA A A A GG CG	G TTG C	AAT TC GG AG C -
Rice	TTTTCCTT G A AAAC	GCC G CGTC	AAGATCACTC CC TTCTA GAGCAGACG A G G GGA TT C
Maize	GTG GTACTA TG C CATT	C A TC A C TG CTA	GC ACTCTA TCTCGGTTTG CCT CT CT GGTGCTG C
Pea	MET LYS	LEU GLU TRP LEU PHE LEU THR ILE ALA PRO CYS ASP ALA ALA GLU PHE TRP	50 HcoI
Oenothera	ATG AAA --- --- TTA GAA TGG CTA TTC CTC ACG ATT GCT CCT TGT GAT GCA GCG GAA CCA TGG		
Rice	TT GTT --- --- AAC	C T T A	
Maize	TT CTT CGT TCA	T G A C T	
Pea	GLN LEU GLY PHE GLN ASP ALA ALA THR PRO MET MET GLN GLY ILE ILE ASP LEU HIS HIS ASP ILE		100
Oenothera	CAA TTA GGA TTT CAA GAC GCA ACA CCA CTT ATG ATG CAA GGA ATA ATA GAC TTA CAT CAC GAT ATC		
Rice	C		C T
Maize	C	T	C T
Pea	PHE PHE PHE LEU ILE LEU ILE LEU VAL PHE VAL SER TRP ILE LEU VAL ARG ALA LEU TRP HIS PHE		150
Oenothera	TTTATTC TTC CTC ATT CTT ATT TTG GTT TTC GTA TCA CGG ATC TTG GTT CCG GCT TTA TGG CAT TTC		
Rice	G	G	C T
Maize	C	G	
Pea	HIS TYR LYS LYS ASN PRO ILE PRO GLN ARG ILE VAL HIS GLY THR THR ILE GLU ILE LEU TRP THR		200 HinfI 250
Oenothera	CAC TAT AAA'AAA AAT CCA ATC CCG CAA AGG ATT GTT CAT GGA ACT ACT ATC GAG ATT CTT CGG ACC	A	
Rice	A G G C CT		A A
Maize	A G G C CT		A A
Pea	ILE PHE PRO SER ILE ILE PRO MET PHE ILE ALA ILE PRO SER PHE ALA LEU LEU TYR SER MET ASP		300
Oenothera	ATA TTT CCT ACT ATC ATT CCG ATG TTC ATT GCT ATA CCA TCA TTT GCT CTG TTA TAC TCA ATG GAC	C	
Rice	G T T	G	
Maize	T T A T	G	
Pea	GLY VAL LEU VAL ASP PRO ALA MET THR ILE LYS ALA ILE GLY HIS GLN TRP TYR TRP THR		350
Oenothera	GGG GTA TTA GTA GAT CCA GCC ATG ACT ATC AAA GCT ATT GGA CAT CAA TGG TAT CGG ACT	C T	
Rice	A G	T	G [intron 1265 bp]
Maize		T	G [intron 794 bp]

300

400

500

600

700

800

900

1000

1100

1200

1300

1400

1500

1600

1700

1800

1900

2000

2100

2200

2300

2400

2500

2600

2700

2800

2900

3000

3100

3200

3300

3400

3500

3600

3700

3800

3900

4000

4100

4200

4300

4400

4500

4600

4700

4800

4900

5000

5100

5200

5300

5400

5500

5600

5700

5800

5900

6000

6100

6200

6300

6400

6500

6600

6700

6800

6900

7000

7100

7200

7300

7400

7500

7600

7700

7800

7900

8000

8100

8200

8300

8400

8500

8600

8700

8800

8900

9000

9100

9200

9300

9400

9500

9600

9700

9800

9900

10000

10100

10200

10300

10400

10500

10600

10700

10800

10900

11000

11100

11200

11300

11400

11500

11600

11700

11800

11900

12000

12100

12200

12300

12400

12500

12600

12700

12800

12900

13000

13100

13200

13300

13400

13500

13600

13700

13800

13900

14000

14100

14200

14300

14400

14500

14600

14700

14800

14900

15000

15100

15200

15300

15400

15500

15600

15700

15800

15900

16000

16100

16200

16300

16400

16500

16600

16700

16800

16900

17000

17100

17200

17300

17400

17500

17600

17700

17800

17900

18000

18100

18200

18300

18400

18500

18600

18700

18800

18900

19000

19100

19200

19300

19400

19500

19600

19700

19800

19900

20000

20100

20200

20300

20400

20500

20600

20700

20800

20900

21000

21100

21200

21300

21400

21500

21600

21700

21800

21900

22000

22100

22200

22300

22400

22500

22600

22700

22800

22900

23000

23100

23200

23300

23400

23500

23600

23700

23800

23900

24000

24100

24200

24300

24400

24500

24600

24700

24800

24900

25000

25100

25200

25300

25400

25500

25600

25700

25800

25900

26000

26100

26200

26300

26400

26500

26600

26700

26800

26900

27000

27100

27200

27300

27400

27500

27600

27700

27800

27900

28000

28100

28200

28300

28400

28500

28600

28700

28800

28900

29000

29100

29200

29300

29400

29500

29600

29700

29800

29900

30000

30100

30200

30300

30400

30500

30600

30700

30800

30900

31000

31100

31200

31300

31400

31500

31600

31700

31800

31900

32000

32100

32200

32300

32400

32500

32600

32700

32800

32900

33000

33100

33200

33300

33400

33500

33600

33700

33800

33900

34000

34100

34200

34300

34400

34500

34600

34700

34800

34900

35000

35100

35200

35300

35400

35500

35600

35700

35800

35900

36000

36100

36200

36300

36400

36500

36600

36700

36800

36900

37000

37100

37200

37300

37400

37500

37600

37700

37800

37900

38000

38100

38200

38300

38400

38500

38600

38700

38800

38900

39000

39100

39200

39300

39400

39500

39600

39700

39800

39900

40000

40100

40200

40300

40400

40500

40600

40700

40800

40900

41000

41100

41200

41300

41400

41500

41600

41700

41800

41900

42000

42100

42200

<

Similar to the COII gene of the other dicot, Oenothera, the pea COII gene is not interrupted by an intron. Since all three monocots (rice, maize and wheat) so far analyzed contain an intron, it seems likely that in COII genes introns were either deleted from dicots or introduced into monocots after their evolutionary divergence.

Sequence comparison of pea COII gene with other species

Also shown in Figure 3 are the nucleotide sequences of Oenothera, rice and maize COII genes. We have compared the coding, 5'-noncoding and 3'-noncoding regions of each gene with pea. We have also compared these sequences of wheat with pea. The wheat sequence is not shown.

(1) Coding sequence: The coding sequence of the pea COII gene is 777 bp and lacks an intron. The nucleotide sequence shows 90.1% homology to rice, maize and wheat, and 90.7% homology to Oenothera. The deduced amino acid sequence shows 87.6% homology to the three monocots, and 89.5% homology to the dicot. The pea COII protein, however, is 2 amino acids shorter than that of rice, maize and wheat, due to the deletion of 7 codons and addition of only 5 codons in pea. Except for one codon deletion, these changes occur at the 5' and 3' ends of the pea gene. Three consecutive codons are deleted at the 5' end of the pea gene, and two out of these three are also deleted from Oenothera. Another three consecutive codons, including the TAA stop codon, found at the 3' end of all the other species are deleted in pea. Instead, in the sequence immediately following the position of the original stop codon, which is otherwise very homologous between pea and the other monocots, a new TAA stop codon has been created. This new TAA codon results from a single base change from G to T in the trinucleotide sequence GAA located at 13-15 bp downstream from that original stop codon. Thus, the 12 bp sequence preceding the new stop codon, which begins the 3'-noncoding region in the other species and ends the coding sequence of pea, codes for 4 additional amino acids. The remaining one codon deletion also occurs near the 3' end of the other genes, at the 26th codon upstream from the stop codon.

(2) 5'-noncoding sequence: There is limited sequence data available

Fig. 3 DNA sequence of pea COII gene aligned with the COII sequences of Oenothera, rice and maize. The DNA sequences are aligned to maximize homology. Only those nucleotides which differ from the pea gene are shown for Oenothera, rice and maize. Dashes indicate gaps that are necessary to align the sequences. The deduced amino acid sequence and the restriction sites for the pea gene are shown above the nucleotide sequence. The TAA termination codons are boxed. The CGG codons are indicated by arrows. The direct repeat sequences are underlined.

for the 5'-noncoding region of rice, wheat, maize and Oenothera COII genes (300, 150, 100, and 180 bp, respectively). In view of the considerable diversity observed, we believe that the transcriptional regulatory elements of the gene reside farther upstream. Accordingly, we sequenced 628 bp in the 5'-noncoding region in pea. Again, we found no homology to any of these four genes.

(3) 3'-noncoding sequence: We sequenced this region of the pea COII gene up to 630 bp downstream from the stop codon. We found that the sequence homology between pea and all three monocots is limited not only to the region immediately following the stop codon of the other monocots, but also extends to a fairly high degree farther downstream for as far as the sequence data is available for the other monocots. On the other hand, it is surprising that pea, a dicot, shows no homology at all to the other dicot, Oenothera, in the 70 bp published sequence in this region. Previously, when we compared the sequence of the rice COII gene with those of maize and wheat, we found that several short stretches of sequences were deleted from or inserted into the intron and 3'-noncoding region of the rice gene⁴. The deletion or insertion of these sequences, which are flanked by direct repeats, was proposed to have occurred through slipped-mispairing of direct repeats during DNA replication. These events are also seen in the 3'-noncoding region of the pea COII gene (nucleotide numbers 802-827, 840-851) when compared to the other COII genes. This observation suggests that deletion or insertion through the above-mentioned mechanism is very common to the plant mitochondrial genomes.

Codon usage of pea COII gene

The genetic codes used in the mitochondria of animals and fungi are different from the universal genetic code and are also different from each other¹⁵⁻¹⁷. The plant mitochondrial genetic code, surprisingly, more closely resembles the universal code than the mitochondrial codes of animals and fungi. Two codons of particular interest are CGG and TGA. CGG, which normally codes for Arg, very likely codes for Trp in plant mitochondria as pointed out by Fox and Leaver based on the sequence data of maize COII gene³. TGA, which codes for Trp in the animal and fungal mitochondria, has not been found in the plant mitochondrial genes sequenced to date^{3-6,18}. Thus, it may code for a stop codon as in the universal genetic code.

The pea COII gene contains four CGG codons, one more than maize, rice or wheat, and the same number as Oenothera. Three of them appear in the

Pea COII - Northern Blot

Fig. 4 Northern blot analysis of pea COII transcript. Total pea mRNA was fractionated on a 1% agarose-formaldehyde gel, transferred to nitrocellulose, and hybridized with the nick-translated pea COII probe (1.4 kb HindIII fragment). 26S (3.6 kb) 18S (1.9 kb) rRNA size markers were visualized by ethidium bromide staining.

same position as in the other four species, and the fourth in both pea and *Oenothera* occurs in a position where the corresponding codon in maize, rice and wheat is TGG. Therefore, these findings provide further evidence that CGG codes for Trp in the same way as TGG. The TGA codon is also not present in the pea COII gene, which supports the suggestion that TGA codes for a stop codon, although there are as yet no known examples of its use in this way. Moreover, we have not found the TGA codon in the COI gene of pea or rice (unpublished results).

Northern blot analysis of pea COII transcript

In order to show that the pea COII gene we have identified is actually transcribed, we looked for the presence of COII transcript by hybridization analysis (Northern blot) of total mitochondrial RNA. The RNA blot was hybridized with the nick-translated 1.4 kb HindIII fragment containing the pea COII gene. As shown in Figure 4, a single discrete transcript of about 1.2 kb was obtained. This finding is in sharp contrast to the complex transcription pattern of the maize COII gene³, but similar to that observed for *Oenothera* both in size and pattern⁶.

Determination of the 5' end of the pea COII transcript

We have determined the 5' end of the pea COII transcript, by S1 nuclease mapping, as a first step toward identifying the regulatory sequence(s) required for the transcription of mitochondrial genes. The DNA probe used was a uniformly labelled single-stranded DNA (419 bp) located

Fig. 5 Determination of 5' end of pea COII mRNA by S1 nuclease mapping. (A) Strategy of S1 mapping. The thick line represents the uniformly labelled single-stranded DNA fragment corresponding to -363 bp to +56 bp of the pea COII gene. Annealing of this DNA to the total mtRNA and subsequent digestion with S1 nuclease were described in Materials and Methods. (B) Gel electrophoretic pattern of the S1-resistant DNA fragments. The S1 nuclease digested sample (+S1) and the control sample without S1 treatment (-S1) were electrophoresed on an 8% polyacrylamide-7M urea gel at 3,000 V for 8 hours. Arrows indicate the S1-resistant DNA fragments. (C) Same as (B) above but run for 20 hours and matched against the nucleotide sequence ladder.

between -363 bp and +56 bp of the coding strand of the pea COII gene (Figure 5A). This DNA probe was annealed with total pea mitochondrial RNA and then digested with S1 nuclease as described in Materials and Methods. The S1-resistant DNA fragment was electrophoresed on an 8% polyacrylamide-7 M urea gel. At the same time, the same primer and template used in synthesizing the probe were used to generate in parallel a sequence ladder by the dideoxynucleotide chain-termination method.

The results show that two DNA fragments were protected by the pea COII mRNA from S1 nuclease digestion (Figure 5B). Comparing with the adjacent sequence ladder, we mapped the 5' ends of these two fragments to be at 285 bp and 302 bp upstream from the ATG start codon (Figure 5C). The former position was designated site I, and the latter site II. On close examination of these two sites, we found that they are located at two identical sites within nearly perfect direct repeat sequences, as shown in

Fig. 6 (A) Nucleotide sequence surrounding the 5' termini of the pea COII transcripts. Arrows point to sites I and II. The direct repeats are underlined. The line above the sequence shows a TATA box. (B) Direct repeat sequences flanking site I and site II. Sequences flanking sites I and II are aligned to show homology. Location for Sites I and II are indicated by arrows.

Figure 6B. The relative abundance of the pea COII transcripts with 5' ends corresponding to site I and site II was determined by measuring the relative intensity of the two bands shown in Figure 5B with a densitometer (Helena Laboratories). The resulting ratio was 84% and 16% for site I and site II, respectively. If these two sites are true transcription initiation sites, then these results would indicate that the transcription of pea COII gene initiates at site 1 approximately 5 times more frequently than at site 2.

Determination of the 3' end of the pea COII transcript

We have also determined the 3' end of the pea COII transcript by S1 nuclease mapping according to the strategy described in Materials and Methods. A uniformly labelled single-stranded HindIII-HindIII fragment (206 bp), located between 770 bp and 976 bp of the pea COII gene, was annealed to total mitochondrial RNA and digested with S1 nuclease (Figure 7A). The S1-resistant DNA fragment was electrophoresed on an 8% polyacrylamide-7 M urea gel together with a sequence ladder of a known sample as a size marker.

Figure 7B shows that the S1-resistant DNA fragment is about 196 to 198 bp in size, thus the 3' end of the pea COII mRNA is located between 193-195 bp downstream from the TAA stop codon. Combining S1 mapping results of the 5' and 3' ends of the pea COII transcript with the known length of the coding region, the size of pea COII mRNA is 1252 bp or 1268 bp, depending

Fig. 7 Determination of 3' end of pea COII mRNA. (A) Strategy of S1 mapping. The thick line represents the uniformly labelled single-stranded HindIII-HindIII fragment corresponding to 770 bp to 976 bp of the pea COII gene. (B) Gel pattern of the S1-resistant DNA fragment. The S1 nuclease digested sample (+S1) and the control sample without S1 treatment (-S1) were electrophoresed on an 8% polyacrylamide-7M urea gel together with a sequence ladder as a size marker.

on whether the calculation is based on site I or site II. This calculation is consistent with the results we obtained from Northern blot analysis as described above.

DISCUSSION

Study of the organization and structure of mitochondrial genomes has been carried out primarily for animal and fungal systems because of the relatively small size of their organelle genomes, and in the case of fungi, ease of genetic manipulations. Plant mitochondrial genomes, however, are much larger in size and more heterogeneous in physical form so that much less research has been done in this area¹⁹.

It was only a few years ago that the first protein-coding gene was isolated from plant mitochondria, i.e. the maize COII gene³. The nucleotide sequence of this gene reveals several interesting features, one of which is the presence of an intron which is not found in the

corresponding genes from animal, fungal or Drosophila mitochondria. Subsequently, another COII gene was isolated and sequenced from a dicot, Oenothera⁶. Comparison of this gene with that of the monocot maize shows a very high degree of coding sequence homology and reveals that Oenothera COII gene does not contain an intron.

Because of this observation, we are interested in determining whether or not it is generally true that the COII gene of dicots does not contain an intron whereas the COII gene of monocots does. We have previously reported the sequence of the COII gene from rice, a monocot⁴. Bonen *et al.* have sequenced this gene from wheat, another monocot⁵. Both COII genes contain an intron having very high sequence homology with the maize intron.

In this paper, we presented the data from a second dicot, pea. Similar to Oenothera, it does not contain an intron. This finding lends further support to the observation that in general the COII gene of dicots does not contain an intron. It is not known whether the intron was present in the progenitors of monocots and dicots before their divergence and subsequently deleted from dicots, or whether it was absent in the progenitors and inserted into monocots after divergence.

Study of the regulation of mitochondrial gene expression has also been mainly focused on animal and fungal systems, particularly man and yeast. The human mitochondrial genome is very compact in size (16 kb) and in gene organization. There are only two major transcripts: the heavy strand transcript and the light strand transcript²⁰. The promoter sequences for both transcripts have been identified, and a consensus sequence, 5'-CANACCCCAAGAPyA-3' has been derived⁹. An interesting feature of these regulatory elements is that the transcription initiation site is an integral part of each promoter, that is, the initiation site is required for transcription. This is reminiscent of the eukaryotic polymerase I promoter²¹.

The yeast mitochondrial genome is larger than that of man (78 kb for Saccharomyces cerevisiae) and a number of primary transcripts have been identified^{22,23}. A 9 bp consensus sequence, 5'ATATAAGTA-3', has been found immediately upstream from the transcription start sites of all the identified primary transcripts^{8,23}. This sequence is likely to be involved in the promoter functions.

Although the yeast promoter sequence shows no homology to its human counterpart, both are located very close to the transcription start sites (less than 10 bp upstream).

We have extended the study of the regulation of mitochondrial gene expression to the plant system, which has not previously been done. Our objective is to determine the transcription start sites of a number of different mitochondrial genes from a variety of plant species and then compare the upstream sequences in order to identify a consensus regulatory sequence.

In this report we have analyzed the 5' and 3' termini of the pea COII transcript by the S1 nuclease mapping method. Transcripts have been detected with 5' termini mapped at two different positions; one at 285 bp (site I) and the other at 302 bp (site II) upstream from the ATG start codon. These transcripts, however, have the same 3' terminus mapped at 193 to 195 bp downstream from the stop codon. After examining the sequences surrounding the two sites, we found that they are located at identical positions in nearly perfect direct repeat sequences (Fig. 6).

If sites I and II are indeed true transcription start sites (see below for further discussion), then the sequences immediately upstream would likely be promoter sequences. This assumes they follow the pattern of human and yeast promoters, where the promoter sequence is less than 10 bp away from the transcription start site. The similarity of these two putative promoter sequences would explain why it is at these two sites that transcription begins. The greater frequency of using site I as a start site might reflect the discrepancies between the two sequences, the sequence surrounding site I being closer to a consensus promoter sequence. It is interesting to note that there is a 7 bp sequence, 5'-TATAAAT-3', located at 23 bp to 29 bp and 6 bp to 12 bp upstream from site I and site II, respectively, which is identical to the eukaryotic consensus promoter sequence²⁴.

Since the S1 nuclease mapping method does not distinguish 5'-termini resulting from primary transcription from those resulting from subsequent processing events, there are alternative interpretations to our S1 mapping data as follows:

- (1) Only site II is the transcription start site. The primary transcript of which 84% is processed at site I initiates at this site.

- (2) Neither site I nor site II is the transcription start site. The COII transcript is contained in a multigene primary transcript with transcription initiating upstream from these two sites and this primary transcript is then processed at both sites to give rise to two COII transcripts. Therefore, the sequences flanking sites I and II serve as

signals for processing enzyme(s) rather than as promoters.

These two possibilities are equally likely as the interpretation mentioned above, and therefore we cannot rule them out yet. We are currently examining them.

To definitively prove that the sequences immediately upstream from site I and site II do function as promoters, we have to await the development of an in vitro transcription system for plant mitochondria, which, as far as we know, is not yet available. We plan to develop such an in vitro system in order to study the regulation of plant mitochondrial gene expression more fully.

ACKNOWLEDGEMENTS

We thank Tom Fox for the maize cytochrome oxidase subunit II gene, Richard McCarty and Konrad Howitz for providing the facility for growing peas, Tom Fox, Deborah Hanson and Peggy McCann for comments on the manuscript, and Debra Sepe for typing the manuscript. This work was supported by research grant GM29179 from the National Institutes of Health, U.S. Public Health Service.

REFERENCES

1. Leaver, C.J. and Gray, M.W. (1982) *Annu. Rev. Plant Physiol.* 33, 373-402.
2. Schatz, G. and Butow, R.A. (1983) *Cell* 32, 316-318.
3. Fox, T.D. and Leaver, C.J. (1981) *Cell* 26, 315-323.
4. Kao, T.H., Moon, E. and Wu, R. (1984) *Nucleic Acids Res.* 12, 7305-7315.
5. Bonen, L. Boer, P.H. and Gray, M.W. (1984) *EMBO J.* 3, 2531-2536.
6. Hiesel, R. and Brennicke, A. (1983) *EMBO J.* 2, 2173-2178.
7. Brown, W.M. (1983) in *Evolution of Genes and Proteins*, Nei, M. and Koehn, R.K., Eds., pp. 62-88, Sinauer Associates, Massachusetts.
8. Osinga, K.A., Vries, E. De, Van der Horst, G.T.J. and Tabak, H.F. (1984) *Nucleic Acids Res.* 12, 1889-1900.
9. Chang, D.D. and Clayton, D.A. (1984) *Cell* 36, 635-643.
10. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
11. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
12. Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) *Biochemistry* 26, 4743-4751.
13. Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
14. Favaloro, J., Treisman, R. and Kamen, R. (1980) in *Methods in Enzymology* 65, Grossman, L. and Moldave, K., Eds., pp. 718-749, Academic Press, New York.
15. Bibb, M., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) *Cell* 26, 167-180.
16. Barrell, B.G. *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77, 3164-3166.
17. Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E., Tzagoloff, A. and Macino, G. (1980) *J. Biol. Chem.* 255, 11927-11941.

18. Dawson, A.J., Jones, V.P. and Leaver, C.J. (1984) *EMBO J.* 3, 2107-2113.
19. Scandalios, J.G. (1984) *Advances In Genetics. Volume 22. Molecular Genetics of Plants.* pp. 109-134. Academic Press, Orlando.
20. Clayton, D.A. (1984) *Ann. Rev. Biochem.* 53, 573-594.
21. Sollner-Webb, B., Wilkinson, J.A.K., Roan, J. and Reeder, R.H. (1983) *Cell* 35, 199-206.
22. Levens, D., Ticho, B., Ackerman, E. and Rabinowitz, M. (1981) *J. Biol. Chem.* 256, 5226-5232.
23. Christianson, T. and Rabinowitz, M. (1983) *J. Biol. Chem.* 258, 14025-14033.
24. Breathnach, R. and Chambon, M. (1983) *Annu. Rev. Biochem.* 50, 349-384.