The EMBO Journal vol.4 no.9 pp.2159–2165, 1985

A nuclear gene encoding the beta subunit of the mitochondrial ATP synthase in *Nicotiana plumbaginifolia*

Marc Boutry and Nam-Hai Chua

Laboratory of Plant Molecular Biology, The Rockefeller University, New York, NY 10021-6399, USA

Communicated by C.J.Leaver

The beta subunit of the mitochondrial ATP synthase in *Nicotiana plumbaginifolia* is encoded by two nuclear genes, *atp2-1* and *atp2-2*, which are both expressed. The complete nucleotide sequence of *atp2-1* has been determined. It contains eight introns ranging from 88 to 1453 bp. The last intron contains a putative insertion element (*Inp*), of 812 bp bordered by 35-bp inverted repeats which share an 11-bp homology with *Agrobacterium tumefaciens* T-DNA borders. Sequences homologous to *Inp* are present in multiple copies in the *N. plumbaginifolia* and the *N. tabacum* genome but not in more distant species. The *atp2-1* encoded polypeptide is highly homologous to beta subunits from other ATP synthases but it contains an extension at the N terminus which is probably involved in mitochondrial targeting. A sequence homology between exon 4 of *atp2-1* and exon 1 of the human *ras* genes suggests a common ancestral origin for these exons.

Key words: ATPase/insertion element/intron/mitochondria/plant nuclear gene

Introduction

The terminal step in the main energy production of living organisms is the synthesis of ATP from ADP. This process is catalyzed by the ATP synthase complex located in bacterial, mitochondrial or chloroplast membranes. In all three cases, the ATP synthase complexes perform the same catalytic function and are strikingly similar in their structure (Amzel and Pedersen, 1983; Senior and Wise, 1983). The complex is made up of a soluble component linked to a membrane sector. The soluble part (F_1), which contains the catalytic activity for ATP synthesis, is composed of five non-identical subunits. The two largest subunits, alpha and beta, are present in three copies per ATP synthase complex. There is a high degree of sequence conservation among the beta subunits of the bacterial, chloroplast and mitochondrial complexes (Saltzgaber *et al.*, 1983; Runswick and Walker, 1983). This subunit plays a central role in ATP synthesis since it contains amino acids involved in substrate binding at the catalytic site (Amzel and Pedersen, 1983; Senior and Wise, 1983).

The genes encoding the polypeptide subunits of the complex show diverse organization. The best characterized example is the *Escherichia coli* ATP synthase complex which is encoded by an operon of nine genes (Futai and Kanazawa, 1983). In eukaryotes, where the enzyme can be located both in mitochondria and chloroplasts, genes for the subunits are dispersed between nuclear and organelle genomes but in a different fashion depending on the organism or the organelle. For example, the alpha and beta subunits of the mitochondrial enzyme are encoded by the nucleus of mammals and yeast (Tzagoloff *et al.*, 1979). In contrast, the corresponding subunits of the chloroplast enzyme are products of the chloroplast genome (Whitfeld and Bottomley, 1983). The plant mitochondrial F_1 enzyme is unique since only the alpha subunit is synthesized by mitochondria, implying that the beta subunit is encoded in the nucleus (Boutry *et al.*, 1983; Hack and Leaver, 1983).

Here, we show that the beta subunit of the mitochondrial ATP synthase in *Nicotiana plumbaginifolia* is encoded by two nuclear genes. We have isolated cDNA and genomic clones for both genes (*atp2-1* and *atp2-2*) and characterized one of them (*atp2-1*) in detail. Unlike the mitochondrial beta subunit gene of yeast (Saltzgaber *et al.*, 1983), the coding sequence of the plant *atp2-1* gene is interrupted by eight introns, some of which could delimit protein domains. In particular, one *atp2-1* exon has a counterpart exon in mammalian *ras* genes. The eighth intron contains a 950-bp structure that has features characteristic of an insertion element. Sequences homologous to this element are present in multiple copies in the genome of *N. plumbaginifolia* and *N. tabacum* but not in more distant species.

Results

Isolation of genes encoding the mitochondrial ATP synthase beta subunit

In order to characterize control sequences involved in plant gene transcription, we were interested in isolating a gene that is expressed in all plant tissues. Mitochondria are present in all plant cells and the ATP synthase is one of the most abundant proteins in mitochondria. Since sequence data have shown very good homology between the yeast and bovine mitochondrial ATP synthase beta subunits (Saltzgaber *et al.*, 1983; Runswick and Walker, 1983), we screened a plant genomic library for this gene using the yeast gene as a probe. We chose *N. plumbaginifolia* because this species has a small genome size and can be genetically transformed with the T-DNA of *Agrobacterium tumefaciens*.

Under low stringency conditions, several clones were identified using the yeast beta subunit gene as a probe. Restriction mapping and genomic Southern blots analysis indicated that the *N. plumbaginifolia* genome contains two different genes for the mitochondrial ATP synthase beta subunit (see below). One clone (Np9) which contains one (*atp2-1*) of the two genes was characterized in detail (Figure 1) and its complete nucleotide sequence determined (Figure 2). Another clone (Np5) containing the second gene (*atp2-2*) was partially sequenced (data not shown). To localize the exon-intron junctions more precisely, clones were also selected from a leaf cDNA library of *N. plumbaginifolia*. Restriction analysis and partial sequence of several of these clones indicated that both *atp2-1* and *atp2-2* are expressed and equally represented in the cDNA library.

Structure and organization of atp2-1

Overlapping fragments of Np9 containing *atp2-1* were subcloned in pEMBL8 vectors (Dente *et al.*, 1983) and their nucleotide sequences determined (Biggin *et al.*, 1983). A corresponding cDNA clone (cNp10) containing 5′ and 3′ non-translated sequences and a poly(A) tail was also sequenced. Figure 1 presents

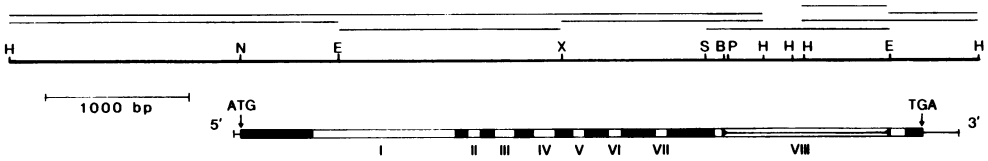

Fig. 1. Structure and organization of *atp2-1*. The upper panel shows the restriction map of a portion of the genomic clone Npg9 for the enzymes *Hind*III (H), *Nco*I (N), *Eco*RI (E), *Xho*I (X), *Sal*I (S), *Bam*HI (B) and *Pst*I (P). The lines on top represent the fragments subcloned in pEMBL8 vectors and sequenced. The lower panel depicts the organization of the *atp2-1* gene as deduced from the sequence data presented in Figure 2. Exons are represented in closed boxes and introns (numbered I–VIII) in open boxes. The line inserted in intron VIII represents the putative insertion element. The lines at the left and right ends indicate 5′ and 3′ non-translated regions, respectively, of the transcript.

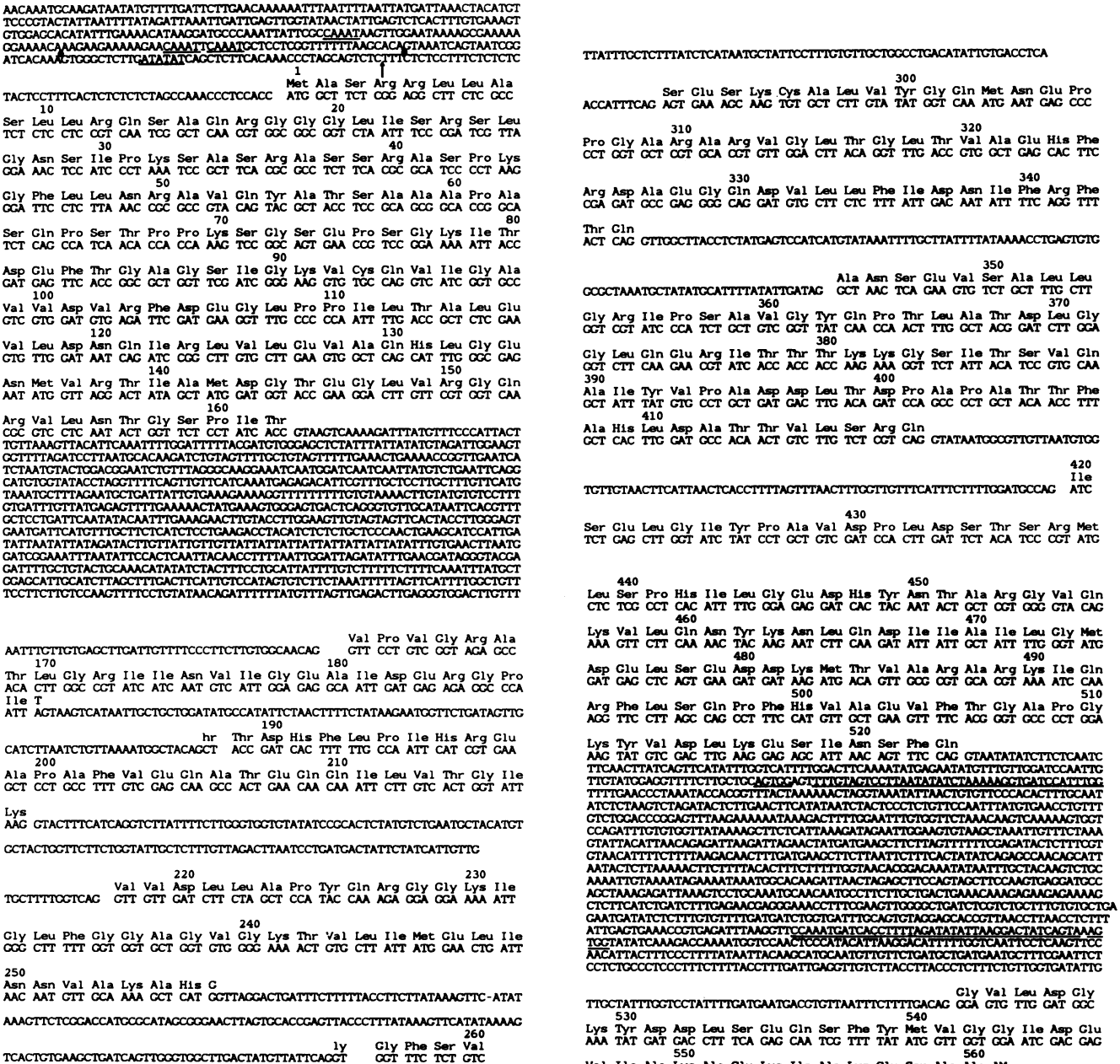

Fig. 2. Nucleotide sequence of *atp2-1*. Npg9 fragments shown in Figure 1 (upper panel) and the cDNA clone cNp10 were sequenced as indicated in Materials and methods. The 5′ boundary and the polyadenylation site of the cDNA clone are indicated by arrows. The deduced amino acid residues are shown above the nucleotide triplets. Dots denote the 5′ and 3′ termini of the transcript mapped by S1 nuclease protection assays. Putative transcription and polyadenylation controls and direct and inverted repeats of *Inp* are underlined.

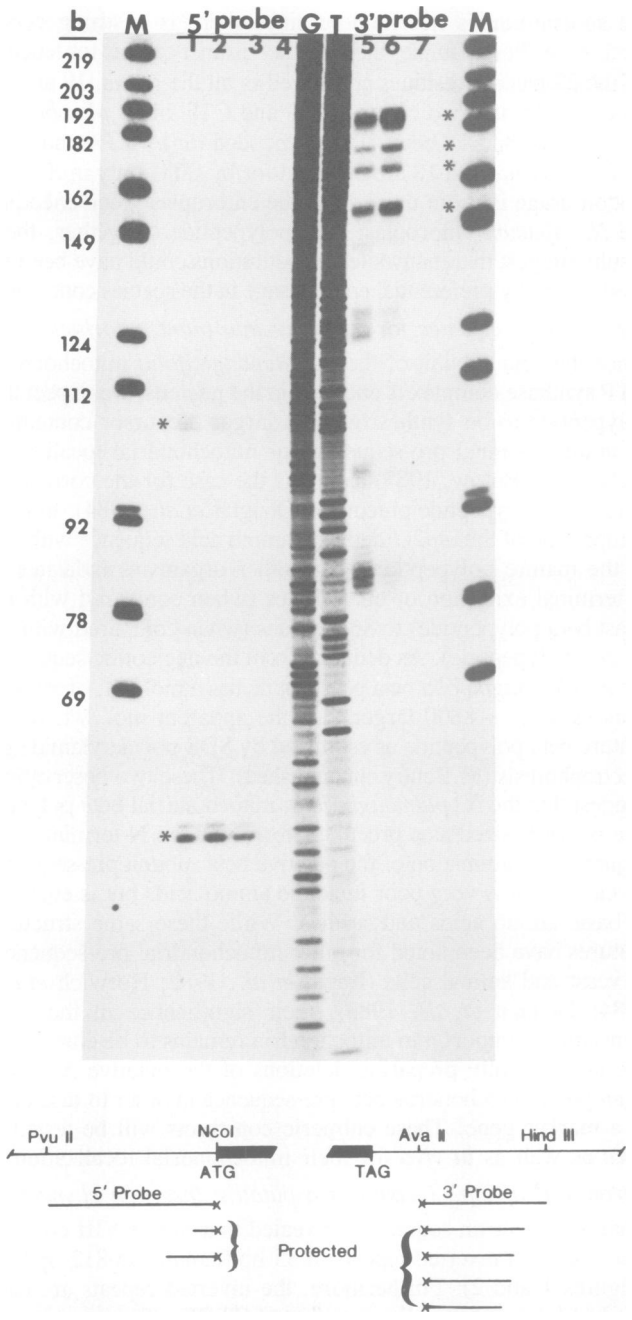

Fig. 3. Mapping of the 5′ and 3′ termini of the *atp2-1* transcript by S1 nuclease protection. In **lanes 1–4**, the probe was a 440-bp *Pvu*I-*Nco*I fragment end-labeled at the 5′ end (*Nco*I site) with kinase. In **lanes 5–7**, the probe was a 239-bp *Ava*II-*Hind*III fragment end-labeled at the *Ava*II site with the Klenow polymerase. Probes and protected fragments are displayed in the lower panel. A deletion clone of *atp2-1* in pEMBL8+ starting 20 bp downstream the *Nco*I site was sequenced in the strand complementary to the mRNA (only G and T reactions) and subsequently cut with *Nco*I to generate the sequence ladder. **Lanes 1,2,5,6**: *N. plumbaginifolia* leaf total RNA; **lane 3**: *N. plumbaginifolia* root total RNA; **lanes 4** and **7**: yeast tRNA. S1 nuclease digestion was performed at 37°C (**lanes 1,3,4,5,7**) or 32°C (**lanes 2** and **6**). On both sides are *Hpa*II-pBR322 markers (M).

the structure and organization of *atp2-1* as deduced from the sequence shown in Figure 2. Unlike the yeast *atp2* which has an uninterrupted coding sequence (Saltzgaber *et al.*, 1983), the plant gene *atp2-1* contains eight introns ranging in size from 88 to 1453 bp. Comparison of exon-intron junctions gives the following consensus sequence:

$$\frac{A}{T}\,\frac{A}{G}\ .\ G\ T\ \frac{A}{T}\,\frac{T}{A}\ \text{at the 5′ splice site}$$

$$C\ A\ G\ .\ \frac{G}{A}\ \text{at the 3′ splice site}$$

We have defined the 5′ and 3′ termini of the transcript by S1 nuclease protection assays (Figure 3). A major transcription start site is located 53 bp 5′ to the first ATG, giving a non-translated region that is 85% pyridimine rich. This S1 start site is three nucleotides downstream of the 5′ terminus of the corresponding cDNA clone. This discrepancy is probably an S1 artifact caused by 'breathing' of the RNA/DNA hybrids at the terminal three TA base pairs. Thirty seven nucleotides upstream of the transcription start site is a putative 'TATA' box (Breathnach and Chambon, 1982): ATATAT, and at position −98 there is a repeated 'CAAT' box-like sequence: CAAATTCAAAT. A second and minor initiation site occurs 102 nucleotides upstream of the first ATG. There is no obvious 'TATA' box immediately 5′ to this second start site. However, the sequence 'CAAAT' is found 103 nucleotides upstream of this minor initiation site.

At the 3′ end, four major polyadenylation sites were detected by S1 mapping within a 35-nucleotide region located 243–277 bp downstream of the stop codon (Figure 3). One of these sites corresponds to the polyadenylation site of cDNA clone cNp10. Multiple polyadenylation sites have also been reported for other plant genes (Messing *et al.*, 1983). Fifty four nucleotides upstream of the first polyadenylation site is the consensus sequence 'AATAAA' (Montell *et al.*, 1983). This hexanucleotide sequence is part of a putative stem structure:

ACCCCCUUUUCUUUUAUGUCACAUAAUAAAGGGGGU

Three additional S1 signals are located further upstream. They probably are 'breathing' artifacts since they correspond to AT-rich regions and decrease in intensity when the S1 reaction is performed at a lower temperature.

The beta subunit of the ATP synthase is encoded by two nuclear genes

Restriction analysis and partial sequence determinations of genomic and cDNA clones indicated the presence of at least two different genes encoding the beta subunit. Two heterologous cDNA clones were used as probes for hybridization to genomic Southern blots (Figure 4A). The hybridization patterns, together with the restriction maps of *atp2-1* (Figure 1) and *atp2-2* (see legend to Figure 4), clearly indicate that the *N. plumbaginifolia* genome contains only two *atp-2* genes. The sizes of the restriction fragments which hybridized to the probes are in agreement with those predicted from the restriction patterns of the genomic clones. Since a stronger signal is obtained when the cDNA probe is hybridized to its cognate gene, we conclude that *atp2-1* and *atp2-2* have sequence divergence in their coding regions. Partial sequence of *atp2-2* indicates that the homology within exonic sequences of both genes is close to 90% (data not shown).

Northern blot analysis shows that the *atp-2* transcript is ~2.1 kb (Figure 4B), a size consistent with our S1 mapping data and with the length of the largest cDNA clone recovered in our experiments. The *atp-2* transcript is present in leaves and roots (Figure 4B) and calli (not shown).

Comparison of the mitochondrial, chloroplast and bacterial ATP synthase beta subunits

Since the ATP synthase is present in both prokaryotes and

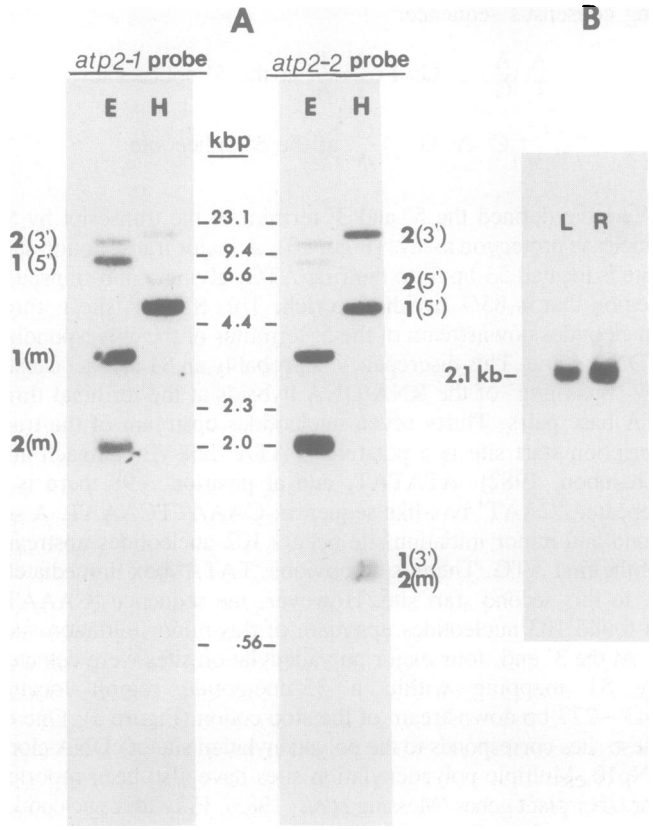

Fig. 4. Southern and Northern blot analysis of *atp2*. (**A**) 10 μg of *N. plumbaginifolia* nuclear DNA was cut with *Eco*RI (E) or *Hind*III (H), electrophoresed and transferred to nitrocellulose filters as indicated in Materials and methods. The probe was a nick-translated fragment of a cDNA clone from *atp2-1* (cNp10, *Nco*I-*Sal*I fragment, see Figure 1) or *atp2-2* (cNp4, an *Eco*RI fragment covering 80% of the coding region of *atp2-2*, not shown). Both *atp2* genes lie on several restriction fragments which are denoted as follows: 1 (for *atp2-1*) or 2 (for *atp2-2*) followed by 5′, 3′ or m corresponding to the 5′, 3′ or middle fragment, respectively, of the gene. In the Np5 clone, *atp2-2* lies on three *Eco*RI fragments: a 0.4-kb upstream fragment (not detected in the genomic blot), a 2.0-kb middle fragment and a ≥2.4-kb downstream fragment whose 3′ end corresponds to the insert border. The same gene lies on three *Hind*III fragments: a ≥5.5-kb upstream fragment whose 5′ end corresponds to the insert border, a 1.2-kb middle fragment and a ≥3-kb downstream fragment whose 3′ end corresponds to the insert border. (**B**) Leaf (L) and root (R) *N. plumbaginifolia* RNA was electrophoresed and transferred to nitrocellulose filters. Hybridization was performed with the *atp2-1* probe.

eukaryotes, it serves as a useful paradigm to investigate evolution of genes from a common ancestor which presumably originated before the eukaryote-prokaryote divergence. The beta subunit, in particular, is probably the most conserved subunit of the entire complex (Runswick and Walker, 1983). A comparison of the ATP synthase beta subunits from the *E. coli* enzyme (Saraste *et al.*, 1981; Kanazawa *et al.*, 1982), the *S. cerevisiae* mitochondrial enzyme (Saltzgaber *et al.*, 1983; M.Douglas, personal communication), the *N. plumbaginifolia* mitochondrial enzyme, the bovine mitochondrial enzyme for which only the amino acid sequence is available (Runswick and Walker, 1983) and the *N. tabacum* chloroplast enzyme (Shinozaki *et al.*, 1983) indicates a high homology all along the gene (59−67% at the nucleotide level and 67−80% at the amino acid level); the highest homology is found between the three mitochondrial polypeptides (data not shown). At the nucleotide level, the divergence is mostly found in silent substitutions suggesting a constraint to maintain the same amino acid sequence. However, the substitution is not random since there is a strong codon preference. For example, there are six distinct codons for leucine. Of the 22 leucine residues conserved in all the genes, 19 are encoded for by the two codons TTG and CTT in *N. plumbaginifolia* (mitochondrial beta), 19 are encoded for by TTA and TTG in *S. cerevisiae* and 18 are encoded for by CTG only in *E. coli*. Codon usage is more uniform in the chloroplast gene encoding the *N. tabacum* chloroplast beta polypeptide. Together, these results suggest that many silent substitutions could have been indeed driven by preferential codon usage in the species concerned.

A precursor sequence for targeting into plant mitochondria?

Since the beta subunit of the *N. plumbaginifolia* mitochondrial ATP synthase complex is encoded in the nucleus, we expect this polypeptide to be synthesized as a larger precursor containing an amino-terminal pre-sequence for mitochondrial localization (Schatz and Butow, 1983) as this is the case for the corresponding yeast polypeptide precursor (Douglas *et al.*, 1984). Indeed, comparison of the *atp2-1* deduced amino acid sequence with that of the mature polypeptides from other organisms indicates an N-terminal extension of 60 residues (when compared with the yeast beta polypeptide) to 86 residues (when compared with the *E. coli* polypeptide). As deduced from the nucleotide sequence, the *N. plumbaginifolia* beta polypeptide has a mol. wt. of 59 866. This value is ~8800 larger than the apparent mol. wt. of the mature beta polypeptide as estimated by SDS-polyacrylamide gel electrophoresis (M.Boutry, unpublished). These two observations suggest that the *N. plumbaginifolia* mitochondrial beta polypeptide is synthesized as a precursor form with an N-terminal pre-sequence. Examination of the putative beta subunit pre-sequence reveals that it is very poor in acidic amino acids but is enriched in basic amino acids and serines. While these same structural features have been noted for other mitochondrial pre-sequences in yeast and animal cells (Kaput *et al.*, 1982; Horwich *et al.*, 1984; Dettaan *et al.*, 1983), their significance in the post-translational import into mitochondria remains to be elucidated. We are currently preparing deletions of the putative *N. plumbaginifolia* mitochondrial beta pre-sequence in order to fuse them to a marker gene. These chimeric constructs will be tested *in vitro* as well as *in vivo* for their mitochondrial localization.

Intron VIII of atp2-1 contains a putative insertion element

Analysis of intron sequences revealed that intron VIII contains nearly perfect inverted repeats of 35 bp flanking an 812-bp loop (Figures 1 and 2). Furthermore, the inverted repeats are each bordered by a 5-bp direct repeat AGTGG. These structural features are reminiscent of those reported for insertion elements of maize, snapdragon and soybean (for a review, see Freeling, 1984). Accordingly, we have designated this putative insertion element as *Inp* for *Insertion Nicotiana plumbaginifolia*. *Inp* is probably not autonomous since the loop sequence is rich in AT (65%) and does not contain any long open reading frame.

To see if *Inp* is also present elsewhere in the *N. plumbaginifolia* genome, we performed Southern blot analysis using the *Inp* insert as a hybridization probe (Figure 5A). An *atp2-1* cDNA clone was used as a control for non-repetitive DNA sequence (Figure 5B). Figure 5A shows that the *N. plumbaginifolia* DNA contains a large number of sequences that hybridized to the *Inp* probe. A major hybridizing band (which can be seen on an under-exposed film, not shown) corresponds, in size, to the one that hybridized to *atp2-1* but the genomic bands which contain *atp2-2* did not give a strong signal. In a separate experiment, we found that *Inp* does not hybridize to a genomic clone containing the *atp2-2* gene (data not shown). These results indicate that *Inp* or

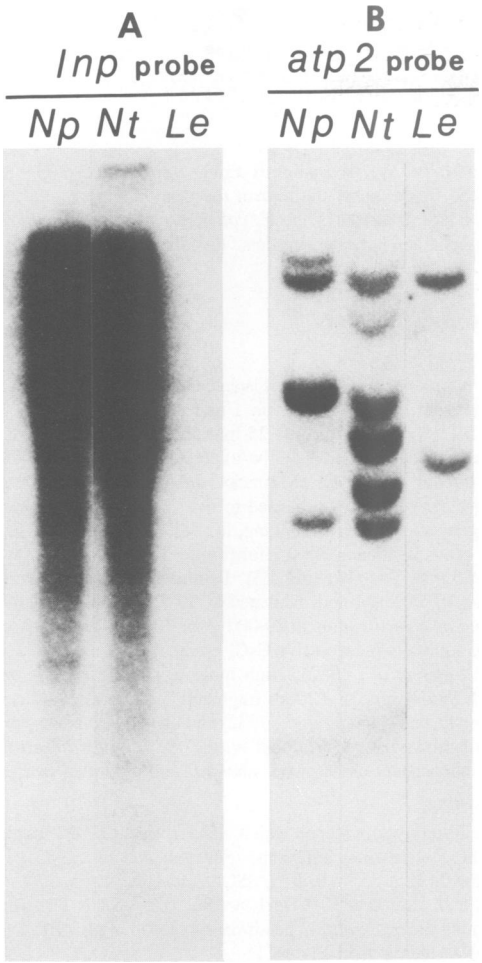

Fig. 5. Southern blot analysis of *Inp*. 10 μg of *N. plumbaginifolia* (Np), *N. tabacum* (Nt), *Lycopersicon esculentum* (Le) [tomato] nuclear DNA were restricted with *Eco*RI, electrophoresed, transferred to nitrocellulose filter and hybridized as indicated in Materials and methods. In **A**, the *Inp* probe was a *Pst*I-*Bcl*I fragment (nucleotides 3398−4267 of *atp2-1*). In **B**, the *atp2-1* cDNA probe (Figure 4) was used.

homologous sequences are present in multiple copies in the *N. plumbaginifolia* genome and that these sequences are not associated with the *atp2-2* gene. Similar Southern blot hybridization experiments show that *Inp*-related sequences are present in *N. tabacum* but not detectable in tomato (Figure 5), petunia, pea or wheat (not shown).

Discussion

We have shown that the *N. plumbaginifolia* genome contains two nuclear genes (*atp2-1* and *atp2-2*) for the beta subunit of the mitochondrial ATP synthase. Southern blot analysis indicated that other plant species also contain two or more genes for the same subunit (not shown). In contrast, yeast contains only one beta subunit gene (Saltzgaber *et al.*, 1983; Vassarotti *et al.*, 1984). Although cDNA clones for both *atp2-1* and *atp2-2* have been found at the same frequency in a *N. plumbaginifolia* leaf cDNA library, we do not know whether both genes have similar expression levels in other tissues. The sequence of >50% of the *atp2-2* coding region has been determined and there is >98% homology with *atp2-1* at the amino acid level (not shown). We do not believe that these two beta subunit polypeptides have distinct physiological functions. The structure and function of the ATPase complex among various organisms are well conserved and it is unlikely that plants would have developed a more complex enzyme. As a comparison, the small subunit of the ribulose 1,5-bisphosphate carboxylase, also a multimeric enzyme, is encoded by a gene family. Although there is a polymorphism of the small subunit polypeptides in tobacco (Stobaek *et al.*, 1976) and in wheat (Broglie *et al.*, 1983), there is no evidence so far that the different polypeptides perform distinct functions.

Among the various beta subunit genes sequenced thus far, the *N. plumbaginifolia atp2-1* is the only one which contains introns. Two observations derived from our sequence analysis of *atp2-1* support the hypothesis proposed by Gilbert (1978) that genes encoding proteins arose by combinations of exon modules.

First, the *atp2-1* introns are located at regions where the amino acid sequence is less constrained. Comparison among the beta subunits from different ATP synthases reveals that with the exception of introns III and VI, all the other introns fall in positions of sequence divergence. In particular, introns II and V are located at two of the three internal regions of the beta polypeptides where the amino acid divergence extends to a few residues. Interestingly, intron V lies in a position where deletions of 1−8 amino acids are found in other beta polypeptides when compared with the plant mitochondrial beta polypeptide. One possible explanation is that an elimination of an intron at this position was associated with a loss of some coding sequences. These observations are in agreement with the hypothesis that introns are located at junctions of domains, where amino acid conservation is less essential (Craik *et al.*, 1982). Recently, Lonberg and Gilbert (1985) have reported that the exons of the chicken pyruvate kinase gene show fairly uniform size distribution. This is not the case with the plant *atp2-1* which, for example, contains exon 1 of 486 bp and exon 8 of 222 bp. A heterogeneous exon size distribution is also found in the maize alcohol dehydrogenase genes (Dennis *et al.*, 1984, 1985), the introns of which have been shown to delimit structural domains (Branden, *et al.*, 1984). It is possible, however, that the largest exons from the two plant genes could have been interrupted with introns which were subsequently lost during evolution. There is some evidence that differential loss of introns had occurred for actin and globin genes in different organisms (Shah *et al.*, 1983; Antoine and Niessing, 1984).

A second type of observation supports the coincidence of exons with protein domains. Previous reports noted that regions of the beta subunit of different ATP synthases have homologous counterparts in the amino acid sequence of other nucleotide-binding enzymes (Walker *et al.*, 1982; Futai and Kanazawa, 1983). Gilbert's hypothesis would predict that the homologous sequences in different genes were derived from a common ancestral exon. If this were the case, it should be possible to find instances where a similar exon is found in genes encoding different proteins. Figure 6 depicts the most convincing homology region which corresponds to residues 223−251 of the *atp2-1* product and which is homologous to a sequence in the *Rec* A protein (Walker *et al.*, 1982) and *ras* proteins (Gay and Walker, 1983). Clearly, the homology region lies in exon 4 and does not extend beyond its borders. The homology with the human *ras* genes is particularly interesting since these also contain introns (Capon *et al.*, 1983; McGrath *et al.*, 1983; Shimizu *et al.*, 1983; Taparowsky *et al.*, 1983). Figure 6 shows that the homologous sequences in *atp2-1* and the human *ras* are both encoded by short exons. Between residues Lys (230) and Val (252), the overall homology is 58% at the nucleotide level and 43% at the amino acid level. Moreover, most of the amino acid substitutions are conservative. There is no sequence homology at the intron/exon

Fig. 6. Alignment of the amino acid sequence of *atp2-1* with the homology regions of *Rec*A and *ras*. We have aligned amino acid residues 193–278 of the *N. plumbaginifolia* mitochondrial beta polypeptide, the amino acid sequence corresponding to the first exon of the human *ras* gene (Capon *et al.*, 1983), amino acid residues 1–45 of a yeast RAS gene (Defev-Jones *et al.*, 1983) and amino acid residues 24–109 of the *E. coli Rec*A protein (Sancar *et al.*, 1980). Intron positions of *atp2-1* are indicated by triangles. Intron positions of the human *ras* gene are indicated by dots. Residues homologous to the *atp2* residues are boxed.

borders which do not coincide exactly. However, as mentioned earlier, exon borders may have much less structural or functional roles (Craik *et al.*, 1982) and the actual borders may have resulted from minor deletions or alterations in the splicing recognition site of the same exon. Consistent with this is the observation that two yeast *ras* genes which share a 90% amino acid sequence homology with the mammalian *ras* genes in the region depicted in Figure 6, contain seven extra amino acids at the N terminus (Defev-Jones *et al.*, 1983; Powers *et al.*, 1984). These additional amino acids are probably not essential since a mammalian *ras* gene can complement a defective *ras* yeast gene (Kataoka *et al.*, 1985). Moreover, as mentioned earlier, the comparison of beta subunits from different organisms indicates that some of them contain deletions at a position corresponding to an *atp2-1* intron. We believe that these observations are consistent with the hypothesis (Gilbert, 1978) that at least some of current proteins are the result of ancestral combinations of distinct exons.

The *atp2-1* intron VIII contains an insertion element-like structure (*Inp*) composed of 35-bp inverted repeats flanking a 812-bp loop. Although we do not know if the *atp2-1 Inp* element is mobile, its insertion origin is supported by the presence of multiple sequences in *N. plumbaginifolia* and *N. tabacum* which hybridize to *Inp*. Sequences homologous to *Inp* were not found in more distant species and we are currently investigating the presence of *Inp* sequences in other *Nicotiana* species.

Comparisons of the *Inp* inverted repeat sequence with those of other plant insertion elements do not reveal any significant homology. However, this sequence shares an 11-bp homology with the 25-bp left or right borders of the T-DNA of *A. tumefaciens* (Zambryski *et al.*, 1982):

right border
GACAGGATATATTGGCGGGTAAAC
...TTTTAGATATATTAAGGACTATCA...
Inp

Moreover, the 5′ border of *Inp* has the sequence GCTGCAGTGG which is similar to the Chi sequence (GCTGGTGG) found in the junction sequence of an octopine tumor line (Simpson *et al.*, 1982). Recent data show that during plant transformation, the T-DNA is present in *Agrobacterium* as a circular form resulting from recombination between the direct repeats (Koukolikova-Nicola *et al.*, 1985) within a 12-bp sequence (overlined above). The latter corresponds to the region highly homologous to *Inp*. *Inp* is probably not a vestige of a T-DNA since it has inverted repeats whereas the T-DNA has direct repeats. Moreover, there is no significant sequence homology between *Inp* and the T-DNA. A common feature between T-DNA and the putative insertion element *Inp* is their integration into the plant genome, which presumably involves excision and recombination. The conserved sequence may be a core recognition sequence for one or more of these molecular events. We are currently testing whether the *Inp* inverted repeat sequence may function as a T-DNA border.

Materials and methods

Genomic library

N. plumbaginifolia nuclear DNA was isolated as follows. 100 g of young leaves were homogenized with a Polytron in 1 liter of 0.44 M sucrose, 2.5% (w/v) Ficoll-400, 5% (w/v) Dextran 40, 25 mM Hepes (pH 7.5), 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and 0.1% (w/v) polyvinylpyrolidone. The homogenate was filtrated through three layers of Miracloth and centrifuged at 3000 *g* for 5 min. The crude nuclear pellet was subjected to five cycles of resuspension in 50 ml of the homogenization buffer containing, in addition, 0.5% (w/v) Triton X-100, followed by centrifugation at 3000 *g* for 5 min. The final pellet was solubilized in 20 ml of 50 mM Tris-HCl (pH 7.5), 10 mm EDTA, 0.4 mg/ml Proteinase K and 2% (w/v) Sarkosyl and incubated at 50°C for 1 h. Insoluble materials were removed by centrifugation at 20 000 *g* for 20 min. The DNA was purified by centrifugation on two successive CsCl gradients in the presence of ethidium bromide (Maniatis *et al.*, 1982). High molecular nuclear DNA was partially digested with *Sau*3A and 15–20 kb fragments were isolated by centrifugation on sucrose gradients (Maniatis *et al.*, 1982) and ligated to *Bam*HI-generated arms of EMBL 4 lambda vector (Frischauf *et al.*, 1983). After *in vitro* packaging, ~7 × 10^5 independent clones were obtained and amplified in *E. coli* K802.

Library screening

The genomic library was screened with a nick-translated *Eco*RI fragment of pJ14 containing the *S. cerevisiae atp2* gene (Saltzgaber *et al.*, 1983). Filters were hybridized for 24 h at 55°C in 6 × SSC, 2 × Denhardt's, 25 μg/ml salmon sperm DNA and washed at 50°C in 1 × SSC, 0.1% SDS. The genomic clone Npg9 containing the entire *atp2-1* gene was subcloned in pEMBL vectors (Dente *et al.*, 1983) as indicated in Figure 1.

A cDNA library of *N. plumbaginifolia* leaf polyadenylated RNA in lambda gt11 (Young and Davis, 1983) was provided by Drs. S.Tingey and G.Coruzzi and was screened with the 1432-bp *Xho*I-*Hind*III fragment of Npg9. Positive clones were subcloned in pEMBL8 vectors and mapped by restriction enzymes. One cDNA clone (cNp10) which encodes *atp2-1* was sequenced.

Sequencing

Restriction fragments to be sequenced were subcloned in pEMBL vectors (Dente *et al.*, 1983). Progressive deletions were made according to Barnes *et al.* (1983). Single-stranded clones were sequenced by the dideoxy method (Biggin *et al.*, 1983). More than 90% of the sequence was obtained for both strands. At least two overlapping clones were used when only one strand was sequenced.

S1 nuclease mapping

DNA fragments used as hybridization probes for 5′ and 3′ S1 nuclease mapping were isolated and labeled with ^{32}P at either 5′ or 3′ ends and the complementary strands separated by polyacylamide gel electrophoresis (Maniatis *et al.*, 1982). 20 μg of total RNA were hybridized to end-labeled probes at 42°C for 16 h in 10 μl of 80% formamide, 0.4 M NaCl, 40 mM Pipes pH 6.4, 1 mM EDTA. 100 μl of S1 buffer (0.25 M NaCl, 30 mM NaAc, 1 mM $ZnSO_4$, 20 μg/ml denatured salmon sperm DNA, pH 4.0) containing 100 U of S1 nuclease were added (Berk and Sharp, 1977). After incubating at 32° or 37°C for 45 min, nucleic acids were ethanol-precipitated and electrophoresed on a 6% polyacrylamide-7 M urea gel.

Southern analysis

10 μg of *N. plumbaginifolia* nuclear DNA was cut with restriction enzymes, electrophoresed on agarose gel and transferred to nitrocellulose filters (Southern, 1975). Hybridization was performed for 24 h at 42°C in 6 × SSC, 50% formamide, 2 × Denhardt's, 25 μg/ml denatured salmon sperm DNA. Filters were washed at 60°C in 0.4 × SSC, 0.1% SDS.

Northern analysis

N. plumbaginifolia RNA was prepared from leaves and roots (Glisen *et al.*, 1974) and polyadenylated RNA was isolated by chromatography on poly(U)-Sepharose (Broglie *et al.*, 1981). 1.5 μg of polyadenylated RNA was denatured by formaldehyde, electrophoresed on agarose gel and transferred to nitrocellulose filter (Thomas, 1980). Hybridization was performed as indicated above.

Acknowledgements

We thank I.Roberson for excellent technical help, Dr. S.Tingey and G.Coruzzi for providing the cDNA library, Dr. M.Douglas for providing the pJ14 plasmid, Dr. C.Kuhlemeier for synthesizing sequencing primers, Dr. M.Durand-Tardif for providing tomato DNA and Dr. J.Odell for helpful discussion and critical reading of the manuscript. M.B. was a recipient of an NIH Fogarty International Center fellowship TW03224. This work was supported by a NIH grant GM 30726 and by a grant from Monsanto.

References

Amzel,L.M. and Pedersen,P.L. (1983) *Annu. Rev. Biochem.*, **52**, 801-824.
Antoine,M. and Niessing,J. (1984) *Nature*, **310**, 795-798.
Barnes,W.M., Bevan,M. and Son,P.H. (1983) *Methods Enzymol.*, **101**, 98-122.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Biggin,M.,D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963-3965.
Boutry,M., Briquet,M. and Goffeau,A. (1983) *J. Biol. Chem.*, **258**, 8524-8526.
Branden,C.-I., Eklund,H., Cambillan,C. and Pryor,A.J. (1984) *EMBO J.*, **3**, 1307-1310.
Breathnach,R. and Chambon,P. (1982) *Annu. Rev. Biochem.*, **50**, 349-384.
Broglie,R., Bellemare,G., Bartlett,S.G., Chua,N.-H. and Cashmore,A. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7304-7308.
Broglie,R., Coruzzi,G., Lamppa,G., Keith,B. and Chua,N.-H. (1983) *Biotechnology*, **1**, 55-61.
Capon,D., Ellson,Y., Levinson,A., Seeburg,P. and Goeddel,D. (1983) *Nature*, **302**, 33-37.
Craik,C. Sprang,S., Fletterick,R. and Rutter,W.J. (1982) *Nature*, **299**, 180-182.
Defev-Jones,D., Scolnick,E., Koller,R. and Dhar,R. (1983) *Nature*, **306**, 707-709.
Dennis,E.S., Gerlach,W.L. Pryor,A.J., Bennetzen,J.L., Inglis,A., Llewellyn,D., Sachs,M.M., Ferl,R.J. and Peacock,W.J. (1984) *Nucleic Acids Res.*, **12**, 3983-4000.
Dennis,E.S., Sachs,M.M., Gerlach,W.L., Finnegan,E.J. and Peacock,W.J. (1985) *Nucleic Acids Res.*, **13**, 727-743.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645-1655.
Dettaan,M., Van Loon,A., Kreike,J., Vaessen,R. and Grivell,L. (1983) *Eur. J. Biochem.*, **138**, 169-177.
Douglas,M., Geller,B. and Emr,S. (1984) *Proc. Natl. Acad. Sci.USA*, **81**, 3983-3987.
Freeling,M. (1984) *Annu. Rev. Plant Physiol.*, **35**, 277-298.
Frischauf,A.M., Lehrach,H., Pouska,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827-842.
Futai,M. and Kanazawa,H. (1983) *Microb. Rev.*, **47**, 285-312.
Gay,N.J. and Walker,J.E. (1983) *Nature*, **301**, 262-264.
Gilbert,W. (1978) *Nature*, **271**, 501.
Glisen,V., Crkvenjalov,R. and Byns,C. (1974) *Biochemistry (Wash.)*, **13**, 2633-2637.
Hack,E. and Leaver,C. (1983) *EMBO J.*, **2**, 1783-1789.
Horwich,A.L., Fenton,W.A., Williams,K.R., Kalousek,F., Kraus,J.P., Doolittle,R.F., Konigsberg,W. and Rosenberg,L.E. (1984) *Science (Wash.)*, **224**, 1068-1074.
Kanazawa,H., Kayano,T., Kiyasu,T. and Futai,M. (1982) *Biochim. Biophys. Res. Commun.*, **82**, 1257-1264.
Kaput,J,. Goltz,S. and Blobel,G. (1982) *J. Biol. Chem.*, **257**, 15054-15058.
Kataoka,T., Powers,S., Cameron,S., Fasano,O., Golfarb,M., Broach,J. and Wigler,M. (1985) *Cell*, **40**, 19-26.
Koukolikova-Nicola,Z., Shillito,R.D., Hohn,B., Wang,K., Van Montagu,M. and Zambryski,P. (1985) *Nature*, **313**, 191-196.
Lonberg,N. and Gilbert,W. (1985) *Cell*, **40**, 81-90.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
McGrath,J.P., Capon,D., Smith,D.H., Chen,E.Y., Seeburg,P.H., Goedell,D. and Levinson,A. (1983) *Nature*, **304**, 501-506.
Messing,J., Geraghty,D., Heidecker,G., Hu,N.-T., Kridl,J. and Rubinstein,I. (1983) in Kosuge,T., Meredith,C.P. and Hollaender,A. (eds.), *Genetic Engineering of Plants*, Plenum Press, NY, pp. 211-227.
Montell,C., Fisher,E.F., Caruthers,M.H. and Berk,A.J. (1983) *Nature*, **305**, 600-605.
Powers,S., Kataoka,T., Fasano,O., Goldfarb,M., Strahern,J., Broach,J. and Wigler,M. (1984) *Cell*, **36**, 607-612.
Runswick,M.J. and Walker,J.E. (1983) *J. Biol. Chem.*, **258**, 3081-3089.
Saltzgaber,J., Kunapuli,S. and Douglas,M.G. (1983) *J. Biol. Chem.*, **258**, 11465-11470.
Sancar,A., Stachelek,C., Konigsberg,W. and Rupp,W.D. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 2611-2615.
Saraste,M., Gay,N.J., Eberle,A., Runswick,M.J. and Walker,J.E. (1981) *Nucleic Acids Res.*, **9**, 5287-5296.
Schatz,G. and Butow,R.A. (1983) *Cell*, **32**, 316-318.
Senior,A.E. and Wise,J.G. (1983) *J. Membr. Biol.*, **73**, 105-124.
Shah,D.M., Hightower,R.C. and Meagher,R.B. (1983) *J. Mol. Appl. Genet.*, **2**, 111-126.
Shimizu,K., Birnbaum,D., Ruley,M.A., Fasano,L., Suard,Y., Edlund,L., Taparowsly,K., Goldfarb,M. and Wigler,M. (1983) *Nature*, **304**, 497-500.
Shinozaki,K., Deno,H., Kato,A. and Suguria,M. (1983) *Gene*, **24**, 147-155.
Simpson,R.B., O'Hara,P.J., Kwok,W., Montoya,A.L., Lichtenstein,C., Gordon,M.P. and Nester,E.W. (1982) *Cell*, **29**, 1005-1014.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Strobaek,S., Gibbons,G.C., Haslett,B., Boulder,D. and Wildman,S.G. (1976) *Carlsberg Res. Commun.*, **41**, 335-343.
Taparowsky,E., Shimizu,K., Goldfarb,M. and Wigler,M. (1983) *Cell*, **34**, 581-586.
Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5205.
Tzagoloff,A., Macino,G. and Sebald,W. (1979) *Annu. Rev. Biochem.*, **48**, 419-441.
Vassarotti,A., Boutry,M., Colson,A.M. and Goffeau,A. (1984) *J. Biol. Chem.*, **259**, 2845-2849.
Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N. (1982) *EMBO J.*, **1**, 945-951.
Whitfeld,P.R. and Bottomley,W. (1983) *Annu. Rev. Plant Physiol.*, **34**, 279-310.
Young,R.A. and Davis,R.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1194-1198.
Zambryski,P., Depicker,A., Kruger,K. and Goodman,H. (1982) *J. Mol. Appl. Genet.*, **1**, 361-370.

Received on 24 May 1985; accepted on 24 June 1985