Mol Gen Genet (1986) 205:331–338

MGG

The ATP synthase subunit 9 gene of *Aspergillus nidulans*: sequence and transcription

Michael Ward and Geoffrey Turner
Department of Microbiology, The Medical School, University of Bristol, Bristol BS8 1TD, UK

Summary. We have determined the nucleotide sequence of the *Aspergillus nidulans* nuclear gene *oliC31*, which encodes subunit 9 of mitochondrial ATP synthase. The open reading frame contains no introns and specifies a predicted protein of 143 amino acids comprising a pre-sequence of 62 residues and a mature protein of 81 residues. The amino acid homology with the equivalent *Neurospora crassa* protein is 50% for the pre-sequence and 80% for the mature protein. A comparison with this and other imported mitochondrial proteins has revealed conserved regions which may be important for transport or subsequent processing. Multiple transcription initiation and polyadenylation sites have been identified. The promoter region of the *oliC31* gene is characterised by long pyrimidine-rich tracts preceding the transcription initiation sites.

Key words: *Aspergillus nidulans* – ATP synthase – Transcription – Sequence – Mitochondria

Introduction

Subunit 9 of the F_0 portion of mitochondrial ATP synthase is a small (8,000 mol. wt.), hydrophobic protein which is also named the proteolipid or dicyclohexylcarbodiimide (DCCD)-binding protein. It is present in the ATP synthase complex as an oligomer (possibly hexamer) and contains sites of interaction with DCCD and oligomycin, both inhibitors of ATP synthesis (Enns and Criddle 1977; Sebald et al. 1977). Amino acid sequences of this subunit from a variety of organisms are available and a pattern of highly conserved residues has been identified (for review see Sebald and Hoppe 1981). The other components of fungal F_0 are less well characterised, but there may be a further five subunits including subunit 6 (Macino and Tzagoloff 1980; Grisi et al. 1982) and a small "proteolipid"-like subunit (Grisi et al. 1982), both mitochondrially encoded.

In both *Aspergillus nidulans* and *Neurospora crassa* subunit 9 is encoded by a nuclear gene (Ward et al. 1986; Sebald et al. 1977), translated on cytoplasmic ribosomes (Turner et al. 1979; Jackl and Sebald 1975) and imported into mitochondria. However, in *Saccharomyces cerevisiae*, another Ascomycete, the gene is in the mitochondrial genome and translation occurs on mitochondrial ribosomes (Macino and Tzagoloff 1979; Tzagoloff and Meagher 1972). Interestingly, a mitochondrial DNA sequence which potentially encodes a protein similar to subunit 9 is also present in each of the filamentous fungi (van den Boogaart et al. 1982; Brown et al. 1984) though no evidence for its expression during vegetative growth has been found.

We have previously cloned a nuclear gene *oliC31*, which encodes subunit 9, from an oligomycin-resistant strain of *A. nidulans* and found it useful as a selectable marker for transformation of this organism (Ward et al. 1986). Here we present the primary structure of the *oliC31* gene. The equivalent gene from *N. crassa* has also been cloned and sequenced (Viebrock et al. 1982; Sebald and Kruse 1984) and it provides an interesting comparison.

In common with many other imported mitochondrial proteins (Hay et al. 1984) the *A. nidulans* and *N. crassa* subunit 9 proteins are initially translated as larger precursors (pre-subunit 9), from which an amino-terminal leader sequence is cleaved following entry into the organelle (Schmidt et al. 1984). Several other genes that encode mitochondrially imported proteins have been sequenced (Sebald and Kruse 1984; Upshall et al. 1986; Takiguchi et al. 1984; Maarse et al. 1984; Wright et al. 1984; Sachs et al. 1986) but no consensus sequence has been identified as a signal for import. The information within the leader sequence will presumably differ according to the site that the mature protein will eventually occupy. Several discrete pieces of information may be encoded in a leader sequence and some studies have shown that a smaller subsection of the leader is sufficient for mitochondrial uptake (Hurt et al. 1985; Keng et al. 1986; Horwich et al. 1986). In addition different proteins may interact with different receptors on the mitochondrial outer surface (Hay et al. 1984). Comparison of pre-subunit 9 leader sequences from different species may identify characteristics important for the uptake of this protein by mitochondria.

The *oliC31* gene is probably highly and perhaps constitutively expressed, and as such may possess a comparatively efficient promoter. It has been demonstrated that the genes from a variety of filamentous Ascomycetes are expressed when transferred to other filamentous fungi (Ballance et al. 1983; Mullaney et al. 1985; Weiss et al. 1985; Buxton et al. 1985; Kelly and Hynes 1985) suggesting that their promoter regions are conserved. Interestingly one of these (*A. nidulans argB*) codes for a mitochondrial protein and import occurs in *N. crassa*, suggesting that the signals directing up-

Offprint requests to: G. Turner

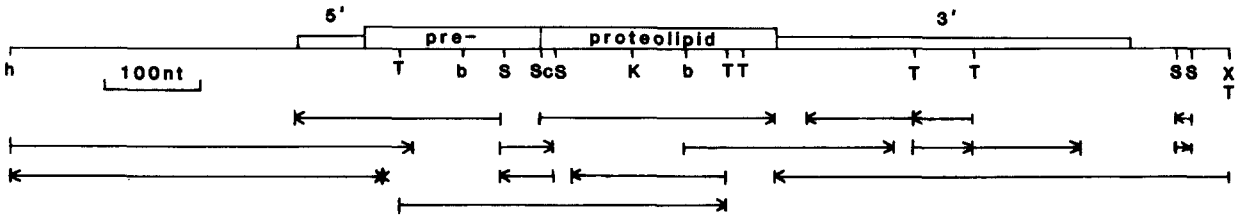

Fig. 1. Sequencing strategy and structure of the *oliC31* gene. The *large open box* denotes the open reading frame and the *smaller open boxes* indicate the 5′ and 3′ untranslated regions. *Arrows* indicate direction and extent of sequencing from various restriction sites or using a synthetic oligonucleotide (*). Restriction sites are; b, *Bal*I; h, *Hin*dII; K, *Kpn*I; S, *Sau*3A; Sc, *Sca*I; T, *Taq*I; X, *Xho*I

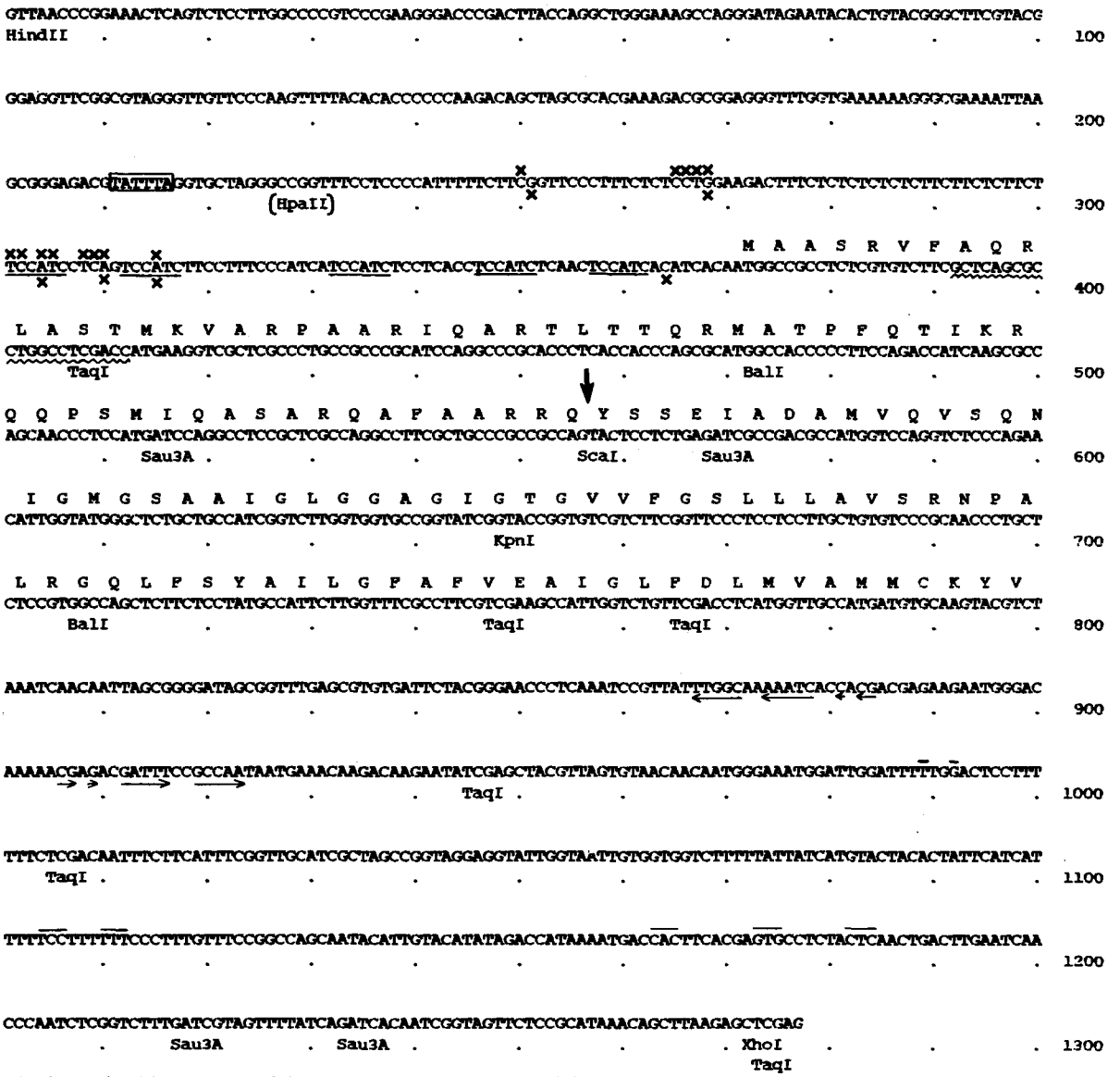

Fig. 2. Nucleotide sequence of the *A. nidulans oliC31* gene and derived amino acid sequence. A potential TATA box is *boxed*. Transcription initiation sites are indicated by *crosses*, those above the sequence determined by S1 mapping and those below determined by primer extension. A repeating hexamer at the 5′ end is *underlined* and the sequence underlined with a *wavy line* is complementary to that of an oligonucleotide synthesised for sequencing and primer extension. The *Hpa*II site (*bracketed*) used for S1 mapping studies is not a unique site. The regions at which polyadenylation occurs are *overlined* and a potential hairpin loop structure is indicated by *arrows* (→). A *bold arrow* (↓) identifies the proteolytic cleavage point which yields the mature protein

take are recognised in the heterologous host (Weiss et al. 1985). Several genes from filamentous fungi have recently been sequenced allowing comparison of promoter, terminator and translated regions.

Materials and methods

DNA sequencing. The dideoxynucleotide chain termination method of sequencing (Sanger et al. 1977) was used with

Table 1. Codon usage for selected amino acids in *oliC31*

Amino acid	Codon	Frequency (%)	Amino acid	Codon	Frequency (%)
Phe	TTT	0 (0)	Gly	GGT	10 (83)
	TTC	8 (100)		GGC	2 (17)
				GGA	0 (0)
Gln	CAA	1 (8)		GGG	0 (0)
	CAG	11 (92)			
			Leu	TTA	0 (0)
Ile	ATT	3 (33)		TTG	0 (0)
	ATC	6 (67)		CTT	3 (27)
	ATA	0 (0)		CTC	6 (55)
				CTA	0 (0)
Thr	ACT	0 (0)		CTG	2 (18)
	ACC	7 (100)			
	ACA	0 (0)	Ser	AGT	0 (0)
	ACG	0 (0)		AGC	0 (0)
				TCT	3 (27)
Val	GTT	1 (10)		TCC	7 (64)
	GTC	8 (80)		TCA	0 (0)
	GTA	0 (0)		TCG	1 (9)
	GTG	1 (10)			
			Arg	AGA	0 (0)
Ala	GCT	7 (28)		AGG	0 (0)
	GCC	18 (72)		CGT	2 (17)
	GCA	0 (0)		CGC	10 (83)
	GCG	0 (0)		CGA	0 (0)
				CGG	0 (0)

M13 single-stranded templates (Messing 1983) and a universal primer (Amersham). Templates were constructed by subcloning restriction fragments from lambda clone CR1 (Ward et al. 1986) which contains the *oliC31* gene. A synthetic oligonucleotide primer (Fig. 2), supplied by Genentech, Inc., was also used in one case.

RNA isolation and manipulation. Total *A. nidulans* RNA was extracted from ground mycelium (Timberlake and Barnard 1981) and poly(A)$^+$ RNA was supplied by W.E. Timberlake. Gel electrophoresis of RNA, Northern blotting and hybridization with nick-translated pMW11 plasmid DNA, that comprises the *oliC31* gene and pUC9 (Ward et al. 1986), were by standard methods (Maniatis et al. 1982).

Transcript analysis. For S1 mapping, single-stranded probes were synthesised on M13 templates (Burke 1984), hybridised with poly(A)$^+$ or total RNA and subjected to S1 nuclease digestion as described by Yelton et al. (1983). Primer extension, using the synthetic oligonucleotide provided by Genentech, Inc., hybridized with poly(A)$^+$ or total RNA was essentially as described by Mullaney et al. (1985). In both cases results were analysed on polyacrylamide sequencing gels using sequencing reactions as size markers.

Results

Primary sequence and open reading frame

The cloned *oliC31* gene can be reinserted into the nuclear genome of *A. nidulans* by transformation, conferring oligomycin-resistance on a drug-sensitive recipient (Ward et al. 1986). Integration either at the *oliC* locus or elsewhere in the genome can be distinguished according to the resistance phenotype. The expression of the *oliC31* gene following integration at a site other than the *oliC* locus would be good evidence that the entire gene was present on that plasmid. Using this method (results not shown) we have identified a 1.3 kb *Hin*dII/*Xho*I fragment of *A. nidulans* genomic DNA (subcloned from lambda CR1; Ward et al. 1986) that contains the intact *oliC31* allele. Figure 1 outlines the sequencing strategy employed for this fragment and Fig. 2 presents the primary sequence obtained. The position and extent of the open reading frame was deduced from homology with the *N. crassa* pre-subunit 9 gene (Viebrock et al. 1982; Sebald and Kruse 1984). The *Neurospora* gene contains two introns, one within the leader sequence and the other between the penultimate and final codons. A comparison of the sequence data indicates that the *A. nidulans* gene does not contain the first intron, but it is possible that the second intron is present. If no intron is postulated in this second position, the open reading frame would terminate to give a protein of the expected length. The sequence GTCTAA follows the penultimate codon, and may be sufficiently similar to the consensus sequence (GTA$^{A}_{C}$GT) for the 5′ end of introns from filamentous fungi (Legerton and Yanofsky 1985) to suggest that this could represent an intron splice point. However S1 nuclease mapping studies of the 3′ end of the gene have ruled out this latter possibility (see later).

```
        -60        -50        -40        -30        -20        -10          ▼
An MAASRVFAQRLASTM----KVARPAARIQARTLTTQRMATPFQTIKRQQPSMI-QASARQAFAARRQYSSEIADA
Nn MASTRVLASRLASQMAASAKVARPAVRVAQVSKRTIQTGSPLQTLKRTQMTSIVNATTRQAF-QKRAYSSEIAQA
   ++  ++ + ++++       ++++++ +        +      + ++ ++ +    +  +  ++++    + ++++++ +
                                                              Nd F-QRRGYAQ

   +10        +20        +30        +40        +50        +60        +70        +80
Am MLQSARIIGTGLATTGLIGAGVGIGVVFGALILGVARNPALRGQLFSYAILGFAFAEATGLFALMMAFLLLYVA
An MVQVSQNIGMGSAAIGLGGAGIGTGVVFGSLLLAVSRNPALRGQLFSYAILGFAFVEAIGLFDLMVAMMCKYV
Nn MVEVSKNLGMGSAAIGLTGAGIGIGLVFAALLNGVARNPALRGQLFSYAILGFAFVEAIGLFDLMVALMALFT
Nm MIQVAKIIGTGLATTGLIGAGIGIGVVFGSLIIGVSRNPSLKSQLFAYAILGFAFSEATGLFALMMAFLLLYVA
   *+-++ +-*+*+*++**-***+*-*-**- *+ -* ***+*++****+********+***+***+**-+-----
      - -  - - --                  -                          -   -  -  -
```

Fig. 3. Amino acid sequences of ATP synthase subunit 9. The numbering is derived from the *N. crassa* sequence; a *dash* indicates introduced spacing for maximum homology. An, *A. nidulans* nuclear gene product; Nn, *N. crassa* nuclear gene product; Am, *A. nidulans* potential mitochondrial gene product; Nm, *N. crassa* potential mitochondrial gene product; Nd, sequence around the cleavage point of the *N. crassa* ATP synthase delta subunit. An *arrow* indicates the point of proteolytic cleavage which yields the mature protein. *, identical residue in all four proteins; +, identical residue in the two nuclear gene products; −, identical residue in the two mitochondrial gene products

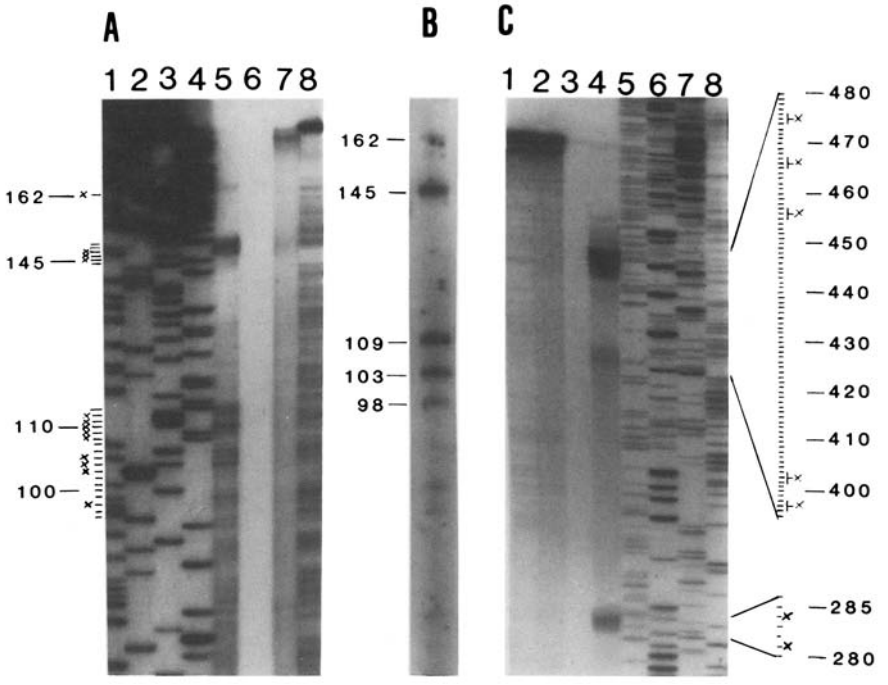

Fig. 4A–C. Mapping of 5′ and 3′ transcript end points. **A** Lanes 1–4: sequencing reactions as size markers; lane 5: S1 nuclease digestion products of a single-stranded DNA probe from the 5′ end of the gene hybridized with total RNA; lane 6: digestion of the DNA probe without RNA; lane 7: digestion of the DNA probe hybridized with the M13 clone on which it was synthesised; lane 8: undigested DNA probe. **B** cDNA products obtained by extension of a primer (Fig. 2) hybridized with poly(A)$^+$ RNA. **C** Lanes 5–8: sequencing reactions; lane 4: S1 nuclease digestion products of a single-stranded DNA probe from the 3′ end of the gene hybridized with total RNA; lane 3: digestion of the DNA probe without RNA; lane 2: digestion of DNA probe hybridized with the M13 clone on which it was synthesised; lane 1: undigested DNA probe

Codon usage

Table 1 gives codon usage values for the more abundant amino acids in the pre-subunit 9. While many amino acids are poorly represented in this small protein, there is obviously heavy bias towards the use of certain codons, as might be expected from a highly expressed gene (Bennetzen and Hall 1982). In particular, codons with A in the third position are seldom used, whereas C in this position is favoured (except for glycine, when T is favoured). The bias is similar to that noted for other highly expressed genes from filamentous fungi (Kinnaird and Fincham 1983; Clements and Roberts 1986).

Pre-protein

The first residue of the mature subunit 9 from *A. nidulans* was shown by Turner et al. (1979) to be tyrosine, as in *N. crassa* (Sebald et al. 1979); the total length of the mature protein was estimated to be 83 residues. Thus the only likely point for proteolytic cleavage of the *A. nidulans* pre-subunit 9 is at the position equivalent to that cleaved in *N. crassa* (Fig. 2). This gives a leader polypeptide of 62 residues (4 shorter than that of *N. crassa*) and a mature protein of 81 residues. The leader sequence contains 15 serine and threonine residues, 12 basic (arginine and lysine) and no acidic residues, and is highly hydrophilic [50% hydrophobicity as defined by Capaldi and Vanderkooi (1972)]. In all these respects it is very similar to the *N. crassa* leader sequence and there is 50% exact homology (though higher functional homology) between the two polypeptides if aligned as in Fig. 3.

Mature protein

The amino acid homology between the mature *A. nidulans* and *N. crassa* proteins is 80% (Fig. 3). Amino acid sequences derived from the *A. nidulans* and *N. crassa* mitochondrial genes (van den Boogaart et al. 1982; Brown et al. 1984) that have homology with subunit 9 are also shown aligned with the nuclear gene products. There is 84% homology between the two mitochondrial versions but only 60% homology between the two *Neurospora* proteins and 67% homology between the two *Aspergillus* proteins.

All four proteins share elements common to subunit 9 obtained from other species (Sebald and Hoppe 1981). However, the nuclear *A. nidulans* protein contains an alanine residue at position 42 as opposed to the invariant glycine previously observed in this position. Similarly methionine is found in position 76 where only leucine or phenylalanine were previously observed, and tyrosine occurs at position 80 where only phenylalanine or glycine have been found (although tyrosine is present in this position in the potential mitochondrially encoded subunit 9). The gene we have sequenced contains a mutation conferring oligomycin resistance but the alteration concerned is unknown. Changes to subunit 9 known to produce oligomycin resistance in *N. crassa* are substitution of the phenylalanine at position 61 with serine or threonine, or the phenylalanine at position 70 with tyrosine; neither of these alterations appears to have occurred in *oliC31*.

Nontranslated sequence

A synthetic oligonucleotide (indicated in Fig. 2) was used for both S1 mapping and primer extension studies to map the 5′ ends of the *oliC* transcripts. For S1 nuclease mapping, the oligonucleotide was used to prime the synthesis on an M13mp template of a ^{32}P-labelled, single-stranded DNA probe that terminates at the *Hpa*II site at nucleotide 227 (Fig. 2). The sizes of fragments obtained following hybridization with total RNA and S1 nuclease digestion are shown in Fig. 4A. These results were in good agreement with fragments identified by cDNA synthesis, using the oligonucleotide as a primer hybridised with either total or poly(A)$^+$

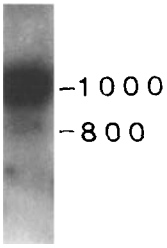

Fig. 5. Poly(A)$^+$ RNA (3 μg) was separated by electrophoresis, transferred to a filter and hybridized to nick translated pMW11. Sizes are given in nucleotides

```
       +     +          + +     +
1 MAA-SR-VFAQRLASTM----KVARP--AARI
2 MAS-TR-VLASRLASQMAASAKVARP--AVRV
3 MNS-LR-I-A-RAA--L----RV-RPT-AVRA
4 MLS--RAIF--RNPVINRTLLR-ARPG-AYHA
5 MLS-LR---------QSI---RFFKP--ATRT
6 --------------------RAAKPTMAVRA
7 MAS-LRSVL-----------------------
8 MLSNLR-IL-----------------------

  MAS-LR-I              RVARP--AVRA
   L     V
```

Fig. 6. Alignment of regions of the presequence of some imported mitochondrial proteins. 1, *A. nidulans* ATP synthase subunit 9; 2, *N. crassa* ATP synthase subunit 9 (Viebrock et al. 1982); 3, *N. crassa* ATP synthase delta subunit (Sebald and Kruse 1984); 4, yeast cytochrome c oxidase subunit VI (Wright et al. 1984); 5, yeast cytochrome c oxidase subunit IV (Maarse et al. 1984); 6, *N. crassa* cytochrome c oxidase subunit V (Sachs et al. 1986); 7, *A. nidulans* ornithine transcarbamylase (Upshall et al. 1986); 8, rat ornithine transcarbamylase (Takiguchi et al. 1984); +, basic residues used to align the sequences. A consensus sequence is given underneath the sequences

RNA (Fig. 4B). The major transcription start points appeared to be at positions 268, 304 and 310 (Fig. 2) with the middle one of these probably being the most frequently used. The multiple bands corresponding to these sites seen with S1 mapping were probably due to incomplete nuclease digestion. Initiation also appeared to occur to a lesser degree around position 251. An apparently major start point at position 364 was also indicated by some of the primer extension experiments, but this was not observed using S1 nuclease mapping and may be an artifact.

In order to identify the 3′ transcript end points a ^{32}P-labelled, single-stranded DNA fragment with one end at the *BalI* site at position 708 and extending beyond the termination codon by >550 nucleotides was synthesised on an M13mp template. S1 nuclease mapping of this fragment identified several regions in which transcripts appeared to terminate (Fig. 4C), the major sites being around nucleotides 1,174 and 1,183 (Fig. 2). It was also apparent from this mapping that there was no intron near the 3′ end of the gene.

There is a potential hairpin loop structure approximately 60 nucleotides beyond the translation stop codon which, if formed in the mRNA, would have a stem of 18 nucleotide pairs, a loop of 21 nucleotides and a $\Delta G(25°\,C)$ of −15.3 kcal calculated by the method of Tinoco et al. (1973).

Northern blotting demonstrated a major transcript band of approximately 1,000 nucleotides and a minor band of around 800 nucleotides (Fig. 5). This is in agreement with the transcript end-point mapping data, the major band representing initiation and polyadenylation at the major sites and the minor band possibly representing initiation at the major sites and polyadenylation at the minor site around nucleotide 984.

Discussion

Mitochondrial import

Although no consensus sequence for mitochondrial import has previously been identified, it has been noted that leader sequences tend to contain high proportions of threonine, serine and basic (arginine and lysine) residues, and low numbers of acidic residues. This generalisation holds true for the *A. nidulans* pre-subunit 9. Substitution of arginine residues in the leader peptide of human ornithine transcarbamylase prevented import and processing (Horwich et al. 1985). The pre-subunit 9 of *N. crassa* is cleaved as a two-stage process to yield the mature protein, possibly reflecting separate functional units within the pre-sequence (Schmidt et al. 1984). While for the most part there is a high degree of complete or functional homology between the *N. crassa* and *A. nidulans* pre-subunit 9 leader sequences, there is little homology in the region around the *N. crassa* intermediate cleavage point (between residues −32 and −31). It is possible that only a single cleavage occurs in *A. nidulans* or that an intermediate cleavage point exists elsewhere. There is interesting homology between the intermediate and final cleavage sites of pre-subunit 9 from *N. crassa*, and the final cleavage sites of *A. nidulans* and the *N. crassa* pre-subunit delta of ATP synthase (Sebald and Kruse 1984; Fig. 3). A somewhat similar sequence is present around the −22 to −19 position in both the *N. crassa* and *A. nidulans* pre-subunit 9. This could represent an intermediate cleavage site for the *Aspergillus* preprotein and cannot be ruled out as a possible third cleavage site in *N. crassa*.

All the information required for mitochondrial uptake may reside within approximately the first 30 residues (i.e. before the intermediate cleavage point of the *Neurospora* protein) of the leader sequence of pre-subunit 9. If the *Aspergillus* and *Neurospora* preproteins are aligned with the precursors of the *Neurospora* ATP synthase delta subunit (Sebald and Kruse 1984) and the yeast cytochrome c oxidase subunits IV and VI (Maarse et al. 1984; Wright et al. 1984) using the basic residues as reference points, some homology is evident (Fig. 6). The regions for which a consensus has been derived may have some functional significance, either for import or subsequent processing. The homology in the region before the first basic residue can also be seen in the ornithine transcarbamylase presequences from *A. nidulans* and the rat (Upshall et al. 1986; Takiguchi et al. 1984) although the second region of homology is not evident. Additionally, some proteins imported by plant chloroplasts have the sequences MAAS, MAST or MASS at their amino-terminal end (Karlin-Neuman and Tobin 1986) which are similar to that found in pre-subunit 9. The second region of homology can also be identified in the *N. crassa* precursor of cytochrome c oxidase subunit V (Sachs et al. 1986) but in this case it is very close to the cleavage site which yields the mature protein. Although this sequence is present in the precursor of yeast cytochrome c oxidase subunit IV (just prior to an intermediate cleavage

site) it is apparently not required for import since the first 12 residues of the presequence (ending within the second region of homolgy) are sufficient to target a cytosolic protein into mitochondria, although they do not also direct proteolytic cleavage (Hurt et al. 1985). Clearly, if these regions of homology are important for some aspect of import they are only required by a subset of imported proteins since they cannot be found in many other imported preproteins.

5′ untranslated region

Subunit 9 is an abundant protein (10% of the ATP synthase complex and 0.2%–0.4% of total cellular protein in *N. crassa*; Sebald et al. 1979) and this is reflected in a high amount of *N. crassa* pre-subunit 9 mRNA (Viebrock et al. 1982). An abundant message was confirmed for *A. nidulans* by the strength of signal observed with Northern blotting. It is reasonable to suppose that sequences 5′ to the *oliC* gene will represent a strong promoter for *A. nidulans*, although the possibility of a highly stable mRNA species leading to high levels of expression cannot be discounted.

Higher eukaryotic genes almost invariably possess a well defined TATA box at a constant distance (approximately 30 bp) from the, usually single, transcription initiation point (Corden et al. 1980). In comparison, the presumptive TATA box of yeast genes, while generally recognisable, can vary greatly from the consensus sequence and can be situated at variable distances (up to 200 bp) from the transcription start points (Dobson et al. 1982; Hahn et al. 1985). In addition multiple initiation points are a common feature of yeast genes (Hahn et al. 1985) as they are also of genes from filamentous fungi (Sebald and Kruse 1984; Upshall et al. 1986; Mullaney et al. 1985; Legerton and Yanofsky 1985; Kinnaird and Fincham 1983; Clements and Roberts 1986; Munger et al. 1985; Rutledge 1984). However, a TATA box is only recognisable in about half of the genes from filamentous fungi sequenced to date, where it is located at various distances from the likely transcription initiation sites (Kinnaird and Fincham 1983; Clements and Roberts 1986; Munger et al. 1985; Shoemaker et al. 1983; Boel et al. 1984; McKnight et al. 1985; Arends and Sebald 1984). A potential TATA box, situated a plausible distance from the presumed transcription initiation sites is present in the *oliC31* gene (Fig. 2) but cannot be identified in the *N. crassa* subunit 9 or delta subunit genes (Sebald and Kruse 1984) both of which also would be expected to be highly and constitutively expressed.

One striking feature of the 5′ region of the *oliC31* gene is the long pyrimidine-rich tracts between the putative TATA box and the suggested transcription start points. Transcription appears to initiate at the first purine following these tracts in each case. Similar regions have been noted in highly expressed yeast genes for which it has been suggested that the level of expression correlates with the extent of these CT-rich areas (Dobson et al. 1982; Kozak 1986). However, in these yeast genes transcription was initiated at a sequence (CAAG) approximately 10 nucleotides downstream. CT-rich regions can also be found within the promoter regions of other, possibly highly expressed, genes from filamentous fungi (Sebald and Kruse 1984; Kinnaird and Fincham 1983; Clements and Roberts 1986; Shoemaker et al. 1983; Arends and Sebald 1984), although these are not as extensive as in the *oliC31* gene even in the *N. crassa* subunit 9 gene. There is no clear correlation with expected level of expression and extent of pyrimidine-rich regions, since the *A. niger* glucoamylase (Boel et al. 1984) gene does not contain an obvious CT-rich region whereas the weakly expressed *trpC* gene of *A. nidulans* does, albeit a short one (Mullaney et al. 1985).

The repeating hexamer, TCCATC, in the 5′ end of the *oliC31* transcribed sequence is also present, but as a single copy, close to the 5′ end of the major transcripts of several other potentially highly expressed genes of filamentous fungi (Legerton and Yanofsky 1985; Kinnaird and Fincham 1983; Clements and Roberts 1986; Arends and Sebald 1984). However, this sequence is not present as an exact copy in the *N. crassa* subunit 9 or delta subunit genes (Sebald and Kruse 1984).

The sequence immediately surrounding the translation initiation codon may be important for ribosome assembly and modulation of translation (Kozak 1986). The sequence immediately preceding the ATG codon matches perfectly the consensus sequence of ATCAC$^{A}_{C}$*ATG* derived from 11 *N. crassa* genes by Legerton and Yanofsky (1985).

Multiple polyadenylation sites were evident at the 3′ end of the *oliC31* gene as have been observed for other genes from filamentous fungi (Upshall et al. 1986; Mullaney et al. 1985; Clements and Roberts 1986; Rutledge 1984). In higher eukaryotes, polyadenylation of transcripts occurs at a position 5′ to the transcription termination point and may be signalled by the sequences AAUAAA and CAYUG nearby (Berget 1984). Similar sequences have been found close to the polyadenylation sites in a few genes from filamentous fungi (Mullaney et al. 1985; Clements and Roberts 1986). However, in yeast transcription termination and polyadenylation may be linked, and other conserved sequences have been reported in this region (Zaret and Sherman 1982; Henikoff et al. 1983). In the *oliC31* gene there is the sequence AAUAA which lies next to a potential hairpin loop structure (Fig. 2) but this is a large distance from the major polyadenylation sites. No significant homology was found around these sites with sequences from genes of yeast or filamentous fungi.

We have identified several potentially important features within the promoter region of the *oliC31* gene and also in the presequence of pre-subunit 9. Determination of the functional significance of these with respect to gene expression or mitochondrial protein import will require in vitro mutagenesis and gene replacement, techniques which are now feasible for *A. nidulans*.

Acknowledgements. We thank K. Miller and W.E. Timberlake for assistance with S1 nuclease mapping and primer extension studies and P. Sterling for photographic assistance. This work was supported by grants from the SERC and from Genencor, Inc.

References

Arends H, Sebald W (1984) Nucleotide sequence of the cloned mRNA and gene of the ADP/ATP carrier from *Neurospora crassa*. EMBO J 3:377–382

Ballance DJ, Buxton FP, Turner G (1983) Transformation of *Aspergillus nidulans* by the orotidine-5′-phosphate decarboxylase gene of *Neurospora crassa*. Biochem Biophys Res Comm 112:284–289

Bennetzen JL, Hall BD (1982) Codon selection in yeast. J Biol Chem 257:3026–3031

Berget SM (1984) Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature 309:179–182

Boel E, Hansen MT, Hjort I, Hoegh I, Fiil NP (1984) Two different types of intervening sequences in the glucoamylase gene from *Aspergillus niger*. EMBO J 3:1581–1585

Brown TA, Ray JA, Waring RB, Scazzocchio C, Davies RW (1984) A mitochondrial reading frame which may code for a second form of ATPase 9 in *Aspergillus nidulans*. Curr Genet 8:489–492

Burke JF (1984) High-sensitivity S1 mapping with single-stranded ^{32}[P] DNA probes synthesized from bacteriophage M13mp templates. Gene 30:63–68

Buxton FP, Gwynne DI, Davies RW (1985) Transformation of *Aspergillus niger* using the *argB* gene of *Aspergillus nidulans*. Gene 37:207–214

Capaldi RA, Vanderkooi G (1972) The low polarity of many membrane proteins. Proc Natl Acad Sci USA 69:930–932

Clements JM, Roberts CF (1986) Gene, in press

Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C, Chambon P (1980) Promoter sequences of eukaryotic protein-coding genes. Science 209:1406–1414

Dobson MJ, Tuite MF, Roberts NA, Kingsman AJ, Kingsman SM (1982) Conservation of high efficiency promoter sequences in *Saccharomyces cerevisiae*. Nucleic Acids Res 10:2625–2637

Enns RK, Criddle RS (1977) Affinity labelling of yeast mitochondrial adenosine phosphatase by reduction with [^{3}H]borohydride. Arch Biochem Biophys 182:587–600

Grisi E, Brown TA, Waring RB, Scazzochio C, Davies RW (1982) Nucleotide sequence of a region of the mitochondrial genome of *Aspergillus nidulans* including the gene for ATPase subunit 6. Nucleic Acids Res 10:3531–3539

Hahn S, Hoar ET, Guarente L (1985) Each of the "TATA elements" specifies a subset of the transcription initiation sites at the *CYC-1* promoter of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 82:8562–8566

Hay R, Bohni P, Gasser S (1984) How mitochondria import proteins. Biochim Biophys Acta 779:65–87

Henikoff S, Kelly JD, Cohen EH (1983) Transcription terminates in yeast distal to a control sequence. Cell 33:607–614

Horwich AL, Kalousek F, Rosenberg LE (1985) Arginine in the leader peptide is required for both import and proteolytic cleavage of a mitochondrial precursor. Proc Natl Acad Sci USA 82:4930–4933

Horwich AL, Kalousek F, Fenton WA, Pollock RA, Rosenberg LE (1986) Targetting of pre-ornithine transcarbamylase to mitochondria: definition of critical regions and residues in the leader peptide. Cell 44:451–459

Hurt EC, Pesold-Hurt B, Suda K, Oppliger W, Schatz G (1985) The first twelve amino acids (less than half of the presequence) of an imported mitochondrial protein can direct mouse cytosolic dihydrofolate reductase into the yeast mitochondrial matrix. EMBO J 4:2061–2068

Jackl G, Sebald W (1975) Identification of two products of mitochondrial protein synthesis associated with mitochondrial adenosine triphosphatase from *Neurospora crassa*. Eur J Biochem 54:97–106

Karlin-Neuman GA, Tobin EM (1986) Transit peptides of nuclear-encoded chloroplast proteins share a common amino acid framework. EMBO J 5:9–13

Kelly JM, Hynes MJ (1985) Transformation of *Aspergillus niger* by the *amdS* gene of *Aspergillus nidulans*. EMBO J 4:475–479

Keng T, Alani E, Guarente L (1986) The nine amino terminal residues of α-aminolevulinate synthase direct β-galactosidase into the mitochondrial matrix. Mol Cell Biol 6:355–364

Kinnaird JH, Fincham JRS (1983) The complete nucleotide sequence of the *Neurospora crassa am* (NADP-specific glutamate dehydrogenase) gene. Gene 26:253–260

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292

Legerton TL, Yanofsky C (1985) Cloning and characterization of the multifunctional *his-3* gene of *Neurospora crassa*. Gene 39:129–140

Maarse AC, van Loon APGM, Riezman H, Gregor I, Schatz G, Grivell LA (1984) Subunit IV of yeast cytochrome c oxidase: cloning and nucleotide sequencing of the gene and partial amino acid sequencing of the mature protein. EMBO J 3:2831–2837

Macino G, Tzagoloff A (1979) Assembly of the mitochondrial membrane system. The DNA sequence of a mitochondrial ATPase gene in *Saccharomyces cerevisiae*. J Biol Chem 254:4617–4623

Macino G, Tzagoloff A (1980) Assembly of the mitochondrial membrane system: Sequence analysis of a yeast mitochondrial ATPase gene containing the *oli-2* and *oli-4* loci. Cell 20:507–517

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, New York

McKnight GL, Kato H, Upshall A, Parker MD, Saari G, O'Hara PJ (1985) Identification and molecular analysis of a third *Aspergillus nidulans* alcohol dehydrogenase gene. EMBO J 4:2093–2099

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101:20–78

Mullaney EJ, Hamer JE, Roberti KA, Yelton MM, Timberlake WE (1985) Primary structure of the *trpC* gene from *Aspergillus nidulans*. Mol Gen Genet 199:37–45

Munger K, Germann UA, Lerch K (1985) Isolation and structural organization of the *Neurospora crassa* copper metallothionein gene. EMBO J 4:2665–2668

Rutledge BJ (1984) Molecular characterization of the *qa-4* gene of *Neurospora crassa*. Gene 32:275–287

Sachs MS, David M, Werner S, Raj-Bhandary UL (1986) Nuclear genes for cytochrome c oxidase subunits of *Neurospora crassa*. Isolation and characterization of cDNA clones for subunits IV, V, VI and possibly VII. J Biol Chem 261:869–873

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

Schmidt B, Wachter E, Sebald W, Neupert W (1984) Processing peptidase of *Neurospora* mitochondria. Two-step cleavage of imported ATPase subunit 9. Eur J Biochem 144:581–588

Sebald W, Hoppe J (1981) On the structure and genetics of the proteolipid subunit of the ATP synthase complex. Curr Top Bioenerg 12:1–64

Sebald W, Kruse B (1984) Nucleotide sequences of the nuclear genes for the proteolipid and delta subunit of the mitochondrial ATP synthase from *Neurospora crassa*. In: Papa S, Altendorf K, Ernster L, Packer L (eds) H^+-ATPase (ATP synthase): Structure, function, biogenesis. The F_0F_1 complex of coupling membranes. Adritica Editrice, Bari, pp 67–75

Sebald W, Sebald-Althaus M, Wachter E (1977) Altered amino acid sequence of the DCCD-binding protein of the nuclear oligomycin-resistant mutant AP-2 from *Neurospora crassa*. In: Bandlow W, Schweyen RJ, Wolf K, Kaudewitz F (eds) Mitochondria 1977 – Genetics and biogenesis of mitochondria. W. de Gruyter, Berlin, pp 433–440

Sebald W, Graf T, Lukins HB (1979) The dicyclohexylcarbodiimide-binding protein of the mitochondrial ATPase complex from *Neurospora crassa* and *Saccharomyces cerevisiae*. Eur J Biochem 93:587–599

Shoemaker S, Schweickart V, Ladner M, Gelfand D, Kwok S, Myambo K, Innis M (1983) Molecular cloning of the exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27. Biotechnology 1:691–696

Takiguchi M, Miura S, Mori M, Tatibana M, Nagata S, Kaziro Y (1984) Molecular cloning and nucleotide sequence of cDNA for rat ornithine carbamoyltransferase precursor. Proc Natl Acad Sci USA 81:7412–7416

Timberlake WE, Barnard EC (1981) Organization of a gene cluster expressed specifically in the asexual spores of *A. nidulans*. Cell 26:29–37

Tinoco IJ, Borer PN, Dengler B, Levine MD, Uhlenbeck OC, Crothers DM, Gralla J (1973) Improved estimation of secondary structure in ribonucleic acids. Nature New Biol 246:40–41

Turner G, Imam G, Küntzel H (1979) Mitochondrial ATPase complex of *Aspergillus nidulans* and the dicyclohyxylcarbodiimide-binding protein. Eur J Biochem 97:565–571

Tzagoloff A, Meagher P (1972) Assembly of the mitochondrial membrane system. VI. Mitochondrial synthesis of subunit proteins of the rutamycin-sensitive adenosine triphosphatase. J Biol Chem 247:594–603

Upshall A, Gilbert T, Saari G, O'Hara PJ, Weglenski P, Miller K, Timberlake WE (1986) Mol Gen Genet: in press

Van den Boogaart P, Samallo J, Agsteribbe E (1982) Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes of *Neurospora crassa*. Nature 298:187–189

Viebrock A, Perz A, Sebald W (1982) The imported pre-protein of the proteolipid subunit of the mitochondrial ATP synthase from *Neurospora crassa*. Molecular cloning and sequencing of the mRNA. EMBO J 1:565–571

Ward M, Wilkinson B, Turner G (1986) Transformation of *Aspergillus nidulans* with a cloned, oligomycin-resistant ATP synthase subunit 9 gene. Mol Gen Genet 202:265–279

Weiss RL, Puetz D, Cybis J (1985) Expression of *Aspergillus* genes in *Neurospora*. In: Bennett JW, Lasure LL (eds) Gene manipulations in fungi. Academic Press, Orlando, pp 279–292

Wright RM, Ko C, Cumsky MG, Poyton RO (1984) Isolation and sequence of the structural gene for cytochrome c oxidase subunit VI from *Saccharomyces cerevisiae*. J Biol Chem 259:15401–15407

Yelton MM, Hamer JE, de Souza ER, Mullaney EJ, Timberlake WE (1983) Developmental regulation of the *Aspergillus nidulans trpC* gene. Proc Natl Acad Sci USA 80:7576–7580

Zaret KS, Sherman F (1982) DNA sequence required for efficient transcription termination in yeast. Cell 28:563–573

Communicated by C.P. Hollenberg

Received July 7, 1986