The EMBO Journal vol.3 no.9 pp.2121 – 2128, 1984

Three variant introns of the same general class in the mitochondrial gene for cytochrome oxidase subunit 1 in *Aspergillus nidulans*

R.B.Waring[1], T.A.Brown[1], J.A.Ray[1], C.Scazzocchio[2] and R.W.Davies[1,3]

[1]Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology, P.O.Box 88, Manchester M60 1QD, and [2]Department of Biology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

[3]Present address: Molecular Biology Department, Allelix Inc., 6850 Goreway Drive, Mississauga, Ontario L4V 1P1 Canada

Communicated by D.C.Burke

The *oxi*A gene of *Aspergillus nidulans*, coding for cytochrome oxidase subunit 1, is shown by DNA sequencing to contain three introns. An AUG start codon is not present at the beginning of the sequence, suggesting that either another codon, possibly the four base codon AUGA, is used for initiation or there is a further short intron between the true start codon and the beginning of the recognisable coding region. The second and third introns have long open reading frames, which could code for maturase proteins. The lack of conservation of amino acid sequence in the putative region of proteolytic cleavage for maturase formation suggests that the first conserved decapeptide may act as the recognition signal for protein processing. The third intron is remarkably (70%) homologous to the second intron of the cytochrome oxidase subunit 1 gene of *Schizosaccharomyces pombe* and both are located in exactly the same position. The third *Aspergillus* intron has an in-frame insertion of a 37-bp GC-rich DNA sequence which is now flanked by a 5-bp repeat, a well-known feature of transposable elements. All three introns in the *oxi*A gene have a 'core' RNA secondary structure found in a class of introns fitting the RNA splicing model of Davies *et al.* (1982). This core RNA structure may play a catalytic as well as a structural role in intron splicing. A sequence within the intron could act as a guide to align the splice sites of two of the introns in accordance with the model of Davies *et al.*

Key words: cytochrome oxidase subunit 1 nucleotide sequence/RNA splicing/ribozyme/*Schizosaccharomyces pombe*/initiation codons

Introduction

We have previously reported that all the introns present in the mitochondrial genome of *Aspergillus nidulans* have some striking regularities in their internal structure (Waring *et al.*, 1982). Four short sequences are strongly conserved and always occur in the same order. When base-pairing between these sequences is taken as a starting point for RNA secondary structure formation, other conserved base-pairings (usually eight) between similarly or identically placed but non-conserved sequences become evident, and all these introns form identical RNA secondary structures despite wide variations in base sequence. These regularities occur in some, but not all the introns of the mitochondrial genome of *Saccharomyces cerevisiae*. We have constructed a model of intron excision based on these regularities (Davies *et al.*, 1982). Secondary structure models of these introns have also been made by Michel *et al.* (1982). They are now referred to as class I introns (Michel and Dujon, 1983); class II introns also occur in fungal mitochondria and have their own characteristic secondary structure. Examples of introns with the characteristic RNA secondary structure of class I introns have since been found in nuclear rRNA genes of protozoa, and chloroplast rRNA and tRNA genes of higher plants, as well as in fungal mitochondrial genes (Waring *et al.*, 1983; Michel and Dujon, 1983). The detailed structure of the intron RNA has proved to be particularly interesting, because one of the introns which fits all aspects of the model of Davies *et al.* (1982), that in the nuclear rRNA gene of certain species of *Tetrahymena*, has been conclusively shown to be excised *in vitro* in the absence of any protein (Cech *et al.*, 1981; Kruger *et al.*, 1982; Zaug *et al.*, 1983). Therefore, this intron, and thus potentially all introns in this class, must possess intrinsic catalytic activity. Genetical and biochemical work aimed at understanding how RNA can work as a catalyst is underway in several laboratories; all these studies are based on the clear delineation of intron RNA secondary structure models.

Three of the four *Aspergillus* introns on which the RNA splicing model of Davies *et al.* (1982) was based occur within one gene, *oxi*A. Here we report the complete sequence of the *oxi*A gene of *A. nidulans* coding for subunit 1 of cytochrome oxidase, analyse all three introns in detail, and compare them with the RNA structures of other related introns. One of these introns is homologous to an intron in the same mitochondrial gene of *Schizosaccharomyces pombe*, and the significance of this is considered by Lang (accompanying paper). Comparison of these sequences reveals an insertion in the *Aspergillus* intron which provides the first suggestion of the existence of transposable elements in filamentous fungi.

Results

General structure of the oxiA gene of A. nidulans

Figure 1 shows the structure of the *oxi*A gene of *A. nidulans* and the DNA sequencing strategy while Figure 2 shows the complete sequence of the *oxi*A gene. A sequence coding for a histidine tRNA is 135 bp upstream from the start of the open reading frame of the *oxi*A gene. A comparison with the amino acid sequence of the corresponding genes of *S. cerevisiae, Homo sapiens* and *Neurospora crassa* establishes beyond doubt that the *A. nidulans* gene codes for subunit 1 of cytochrome oxidase. Three intervening sequences are present which we have called NOX1, NOX2 and NOX3. On the basis of maximising amino acid homology with the coding sequences of yeast and human there are only two likely splice point positions for each intron (Figure 2). The splice junctions which we consider to be most likely are those consistent with the model of RNA splicing proposed by Davies *et al.* (1982) in which the last exon base should be a pyrimidine, usually a U, while the last intron base is always a G. The application of this rule leads in each case to an unambiguous

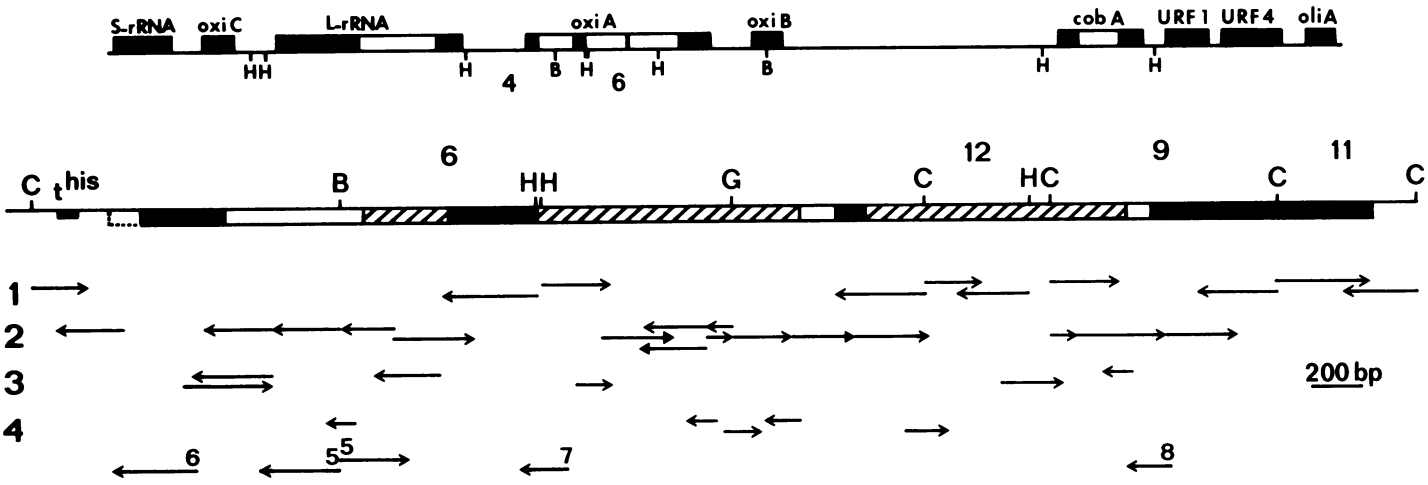

Fig. 1. The structure of the *oxi*A gene of the *A. nidulans* mitochondrial genome with the strategy used for DNA sequence analysis. The top line shows the location of *oxi*A relative to some other known genes in *A. nidulans*. The genome is circular and is shown broken at the junction of *Hind*III fragments 5 and 3. The two *Bam*HI sites are labelled B, the *Hpa*II sites H (fragments 4 and 6 are indicated). The lower section shows the *oxi*A gene preceded by the tRNA His gene. The exact start of the gene is not known (see text). Solid bar, exon region; hatched bar, open reading frame of intron; open bar, remainder of intron. Transcription is from left to right. *Bcl*I sites are labelled C (fragments 6, 12, 9 and 11 are indicated) and a *Bgl*II site is labelled G. All sequencing was done using the chain-terminator method of Sanger *et al.* (1977), using single-stranded DNA derived from M13 clones and either a universal primer (Anderson *et al.* 1980) or internal primers. Sequencing strategy: **Level 1**: M13mp7 clones of *Hpa*II fragments 4 and 6 and M13mp6 clones of *Bcl*I/*Bgl*II double-restricted mitochondrial fragments. **Level 2**: M13mp6 *Sau*3A clones either derived from larger mitochondrial fragments, previously cloned into M13 or pBR327 (see Materials and methods) or derived from *Sau*3A-digested total mitochondrial DNA (arrows with solid heads). **Level 3**: M13mp2 clones of *Eco*RI*-digested total mitochondrial DNA. **Level 4**: selected *Sau*3A fragments were prepared as in Waring *et al.* (1981) and used as internal primers to obtain overlaps of shorter sequences. These were pretreated with exonuclease III (Sanger *et al.*, 1980; Waring *et al.*, 1982) and then annealed to templates derived from clones described for **Level 1**. **5**: the M13mp7 clone of *Hpa*II fragment 4 was digested with *Bam*HI to give two *Bam*HI mitochondrial fragments (there are two *Bam*HI sites in the multi-restriction site region of M13mp7 split by an *Acc*I site into which fragment 4 was cloned and a single *Bam*HI site in fragment 4). These were recloned into M13 and sequenced as shown. Three internal primers were prepared to close the remaining gaps. **6**: a 39-bp *Alu*I fragment. **7**: an *Mbo*II/*Sau*3A fragment. **8**: a *Pvu*II/*Sau*3A fragment, pre-treated with exonuclease III.

positioning of the splice junctions.

Introns have been found in eight different positions in the homologous *oxi*3 gene in two different strains of *S. cerevisiae* (Bonitz *et al.*, 1980; Hensgens *et al.*, 1983). None of the *A. nidulans* introns in *oxi*A occur in the same position as any of these introns. The second intron of the homologous *cox*1 gene in *S. pombe* is however in precisely the same position as NOX3 (Lang, accompanying paper).

Where is the start codon of the A. nidulans oxiA gene?

The amino-terminal ends of mammalian, yeast, *N. crassa* and *A. nidulans* cytochrome oxidase subunit 1 proteins are strongly conserved. The *S. cerevisiae* protein starts with an AUG methionine codon. The homology between the *S. cerevisiae* and the *A. nidulans* protein starts at the fourth and fifth residues which are arginine and tryptophan, respectively. The *A. nidulans* protein has a tryptophan in the position where the yeast protein has the methionine start codon and moreover the frame is open for a further 40 residues upstream. However, there is no AUG start codon in this stretch. An unusual start codon or a splicing reaction are two possible ways of resolving this situation.

There are one AUA and three AUU isoleucine codons at the start of the open reading frame. AUA and AUU are probably used as start codons in the mammalian mitochondrial genomes (Anderson *et al.*, 1981; Bibb *et al.*, 1981). It is not known if such codons can be used in *A. nidulans* although at least one *A. nidulans* mitochondrial gene is known not to start with AUG; GUG is used as the start codon for the *oxi*C gene (Netzker *et al.*, 1982). The most downstream of the AUU codons follows immediately an A-rich sequence similar to the one found preceding the start codons of the *cob*A (Waring *et al.*, 1981) and *oxi*C *A. nidulans* genes (Netzker *et al.*, 1982). It is also likely that *N. crassa* (Burger *et al.*, 1982) and *Drosophila* (Clary and Wolstenholme, 1983) have a non-standard start codon at the beginning of the *cox*I gene. Thus while there are several indications that an unusual start codon may be used, it is not possible to identify a clear candidate, particularly in view of the following unusual codons; a leucine CUC and threonine ACC and ACG codons are present in the open reading frame upstream of the conserved arginine and tryptophan amino acids described above. These codons have never been found before in any of the structural genes of the *A. nidulans* mitochondrial genome (ACC and ACG codons are found in intron open reading frames). The codons between the recognisable coding sequence and the nearest threonine ACG are all poor candidates for start codons, which suggests that these codons could be in a short intron located just after an upstream standard AUG start codon for which there are a number of possibilities.

It has been postulated that the *cox*I gene in *Drosophila yakuba* starts with the four letter codon AUAA, coding for a formyl-methionine. In the *A. nidulans oxi*A sequence AUGA is found in the homologous position, that also corresponds to the position of the yeast and mammalian initiation codons (Bonitz *et al.*, 1980; Anderson *et al.*, 1982). If indeed a four letter initiation codon is used, it raises the problem of how it is identified in the correct frame. It might be relevant that the sequence upstream from the putative four letter codon in the *oxi*A gene is unusual in having a relatively high GC content and more potential secondary structure that regions upstream from regular initiation codons in any other mitochondrial gene in this organism.

Internal structure of the oxiA introns

Open reading frames. The second and third introns (NOX2

```
      tRNA his
1     AATCTATAGGTGGGTGTAGTTCAAAGGTAGAACAGCTGTATGTGGCATAGTATATCCTAG
61    TTCAATTCTAGGTATCCACCCTAACGTACTTAATGTAAAAGGTTATAACCTAGTTTGGTT
121   AAAGGTAACCACATTAACTAAATAATAAAAAAGATTATACACTATTTATTATTCATATAA
181   ATAAAAGTTCATGTATAGCATACGTAAAGTATGTTATTAATGTAAGGTAAAGTTATTTTT
241   ATAATTAATTATTGATCTCAATACAAATAATGTATTAGGTAAAAAATTTTCAACCTCAAC
       *** L  I  I  D  L  N  T  N  N  V  L  G  K  K  F  S  T  S  T
301   TAAAAAAGAAAATATTAAACAAATAGAAAGTTCATCGTTCTTAACGTTTAAACAACCTAC
       K  K  E  N  I  K  Q  I  E  S  S  S  F  L  T  F  K  Q  P  T
361   AGAATGACAAGAAAGATGATATTTATCATCAAATGCTAAAGATATTGGAACATTATATTT
       E  W  Q  E  R  W  Y  L  S  S  N  A  K  D  I  G  T  L  Y  L
421   AATGTTTGCATTATTTTCAGGTTTATTAGGTACAGCTTTTTCTGTACTTATAAGATTAGA
       M  F  A  L  F  S  G  L  L  G  T  A  F  S  V  L  I  R  L  E
481   ATTATCAGGGCCTGGTGTTCAATATATAGCAGATAACCAATTATATAATAGTATAATAAC
       L  S  G  P  G  V  Q  Y  I  A  D  N  Q  L  Y  N  S  I  I  T
541   AGCTCATGCCATAATGATGATTTTCTTTATGGTTATGCCAGCTTTAATAGGAGGTTTCGG
       A  H  A  I  M  M  I  F  F  M  V  M  P  A  L  I  G  G  F  G
601   AAATTTTTTATTACCATTATTAGTAGGAGGTCCAGATATGGCAAAATAATAAGTCTTAAA
       N  F  L  L  P  L  L  V  G  G  P  D  M  A  K ******
661   TAATAAAGACTTTTTGGAACTATTATAAATAATAATTATTTATTATAATGACTTAAAAAT
721   ATACTTAGCAGGTTTATTTTAAGGAAATATTTGAATCAATAATCTAGAAGAACATAATTC
781   AAGATTTTAGATAAAAAAATTAATGGATATAATAGGATCAGGATTTATTAGATTTAAATA
841   TAATTATATAATAATTATTATTATTAATATAATTCCAGAATTAATGCCGTTAAAAAGTGT
901   AAACTTTTTAATTATTACTTCGCTGTATGCATGGAAAATCCCGAAATAACAAATACATAG
961   ACACCTAATTGGAGTTTATATACTTATAAAGTTATATTTCCAACTGAATTAGTCGTAATA
1021  CTTATGTATTCATGGGAATAATATGCAGGAAACCAAAGTAATTTTGAATAAAATTATTAG
1081  TAGGATCCTCAGAGACTACACGCGAGGCTCCATTATAATATAATGGATGAAGACATAGTC
1141  CGTTGTCCAGTTCTAACTATTAACTGGACCTAAAAAAAAAAATTCTTACTCTTAAATTTT
                           *** L  D  L  K  K  K  I  L  T  L  K  F
1201  TTTGAAGTAACGAATTTAGGACTAAACAAAGAGAGATAAGGGGGAAATACGACCCAGATG
       F  W  S  N  E  F  R  T  K  Q  R  E  I  R  G  K  Y  D  P  D
1261  TAGGGTGATTTAGCTATTTTTTTAGATCATTAATAACGTGTATACTGACGTTAAACCAAA
       V  G  W  F  S  Y  F  F  R  S  L  I  T  C  I  L  T  L  N  Q
1321  ATTATGTGGTTACAATTTCACAAATTTTAGGTAATGATTTAGAAAGTAGATTAGCTAATA
       N  Y  V  V  T  I  S  Q  I  L  G  N  D  L  E  S  R  L  A  N
1381  AAGTATTACAAGATTATGTTTTTAATAGCTTAAGTTTACCAAATAAATATAATTTTATTA
       K  V  L  Q  D  Y  V  F  N  S  L  S  L  P  N  K  Y  N  F  I
1441  TGTCAATTGGTAATTATAATTCAATACCCCATAATCCGACTATCACAAAATTTTAGAATT
       M  S  I  G  N  Y  N  S  I  P  H  N  P  T  I  T  K  F ***
1501  AGGTTTCCTAGATTAAATAATATAAGTTTCTGATTATTAGTACCTAGTTTACTATTATTT
          F  P  R  L  N  N  I  S  F  W  L  L  V  P  S  L  L  L  F
1561  GTATTCTCTGCAACAATAGAAAATGGAGCAGGTACAGGTTGAACTCTTTACCCTCCTTTA
       V  F  S  A  T  I  E  N  G  A  G  T  G  W  T  L  Y  P  P  L
1621  TCAGGAATACAATCACACAGTGGGCCAAGTGTTGATTTAGCTATTTTTGGTTTACACTTA
       S  G  I  Q  S  H  S  G  P  S  V  D  L  A  I  F  G  L  H  L
1681  AGTGGAATTAGTAGTATGTTAGGAGCTATGAACTTCATAACAACAATATTAAATATGAGA
       S  G  I  S  S  M  L  G  A  M  N  F  I  T  T  I  L  N  M  R
1741  AGTCCAGGTATACGTTTACACAAATTAGCTTTATTTGGTTGAGCAGTAATTATAACAGCT
       S  P  G  I  R  L  H  K  L  A  L  F  G  W  A  V  I  I  T  A
1801  GTGTTATTATTATTATCATTACCTGTATTAGCCGGTAAACATAGAAAACTGCCGGCATTA
       V  L  L  L  L  S  L  P  V  L  A  G  K  H  R  K  L  P  A  L
1861  AATTTGGCTATATACTGGAAACTGTTATATATTTATAGATATATATCACAATCAGTAGGA
       N  L  A  I  Y  W  K  L  L  Y  I  Y  R  Y  I  S  Q  S  V  G
1921  AACCATCTAGGTTTAAACCTAAATGGGTTCTTCAGAGACTATACGCCAAAGTATTTTCAT
       N  H  L  G  L  N  L  N  G  F  F  R  D  Y  T  P  K  Y  F  H
1981  TGTAAATTCTTACATTCGACGAACAATAATAATTTACCTTTTGCTGCTTATTTAGCAGGT
       C  K  F  L  H  S  T  N  N  N  N  L  P  F  A  A  Y  L  A  G
2041  CTTATTGAAGGTGACGGTACTATAATAGTACCAAATACTTTAAGATCACCTAAAGGTAAA
       L  I  E  G  D  G  T  I  I  V  P  N  T  L  R  S  P  K  G  K
2101  TTGAATTATCCTTCAATTCAAATTGTTTTTCATTTAAAAGATTTACCTCTAGCTTTAATG
       L  N  Y  P  S  I  Q  I  V  F  H  L  K  D  L  P  L  A  L  M
2161  ATTCAAAAAGAATTAGGTTTTGGTTCTCTTTCAAGAAAAAAAGGTGTTAATGCTTATATT
       I  Q  K  E  L  G  F  G  S  L  S  R  K  K  G  V  N  A  Y  I
2221  TTGACTATAAATAATAAGGAGGGTATATTATTTGTTATATCTTTATTAAATGGTAATATG
       L  T  I  N  N  K  E  G  I  L  F  V  I  S  L  L  N  G  N  M
2281  AGAACACCTAAGATAAATAGTTTATATAAATTGATTGATTTTTATCAAGATAATATAAAT
       R  T  P  K  I  N  S  L  Y  K  L  I  D  F  Y  Q  D  N  I  N
2341  ATTGAAAAAAAGCCTTTAAATAATGAACCTTTAGAATCTAATGCCTGATTATCAGGATTT
       I  E  K  K  P  L  N  N  E  P  L  E  S  N  A  W  L  S  G  F
2401  ATAGAAGCAGATGGTTCTTTTCAAGTTAGAAGTACTCTTAGTGGTAAATATCCTAAATTC
       I  E  A  D  G  S  F  Q  V  R  S  T  L  S  G  K  Y  P  K  F
2461  GAATGTAAGTTGGAAATATCTCAAAGACAAACAGATCATAAAGGTTTTAGTAATCAGGAG
       E  C  K  L  E  I  S  Q  R  Q  T  D  H  K  G  F  S  N  Q  E
2521  TTTTTAAATAAAATTGCTGATTTCTTTCACACTGAAGTTAAAGAAACAAGATTAAATAGA
       F  L  N  K  I  A  D  F  F  H  T  E  V  K  E  T  R  L  N  R
2581  TCTACTCCAGAGTATAGAGTTAGAACTACTAATTTACAAGGTAATAATCAAGCAAAGAGT
       S  T  P  E  Y  R  V  R  T  T  N  L  Q  G  N  N  Q  A  K  S
2641  TATTTTATTAAATATCCATTGTTTGGAACTAAGTATTTAGATTCTATAGATTGAATGAAA
       Y  F  I  K  Y  P  L  F  G  T  K  Y  L  D  S  I  D  W  M  K
2701  GTTGTAGATTTATTTAATAATGGTGAACATAAAACTGAATTAGGTAAAGAAAAAATTTTA
       V  V  D  L  F  N  N  G  E  H  K  T  E  L  G  K  E  K  I  L
2761  AATATAAAATCTAATATGAATGATAAAAGAACTGTTTTTACTTGAGATCATTTACAAAAT
       N  I  K  S  N  M  N  D  K  R  T  V  F  T  W  D  H  L  Q  N
2821  TTTTACAAATTGAAAATATAAAATATAGTCCAAACAATTTCGAGAGAAATTGAGTGTCTA
       F  Y  K  L  K  I ***
2881  CCTTAAATAACTTACCCCTATAGGGTAAGTTATTTTTGAAACTAAGTTTTATACTTATAG
2941  TTATGGGATTTTATCCAGTAGACCAATAATAATAGGGTATTACTATGGTTTTAACAGATA
                            ********* G  I  T  M  V  L  T  D
3001  GAAACTTTAATACATCATTCTTCGAAGTAGCTGGTGGTGGTGATCCTATCTTATTCCAAC
       R  N  F  N  T  S  F  F  E  V  A  G  G  G  D  P  I  L  F  Q
3061  ATTTATTCTGATTCTTCGGGCACCCAGAGGTCAAAAATATAGGCTTCTTAACGTTGCTGT
       H  L  F  W  F  F  G  H  P  E  V  K  N  I  G  F  L  T  L  L
3121  ATGCTGGGACCACTTCAATATTAAGTTTTAAATACTCAATCTTAAATGACACAGTAAAAA
       Y  A  G  T  T  S  I  L  S  F  K  Y  S  I  L  N  D  T  V  K
3181  AGTTAAAACGATGAAGTATATCAGCAGGTAACATTTTAAATGGAACCTCAGAGACTTTAT
       K  L  K  R  W  S  I  S  A  G  N  I  L  N  G  T  S  E  T  L
3241  GCAACGAAATTGTAGTAAACACAGAAAATATAAAATCTATTTCTGTTCATGTACCTAAAC
       C  N  E  I  V  V  N  T  E  N  I  K  S  I  S  V  H  V  P  K
3301  ATTTAAAACCAGTTAGTGATGATCAATTTGGACATTATTTAGCTGGTTTAATAGATGGTG
       H  L  K  P  V  S  D  D  Q  F  G  H  Y  L  A  G  L  I  D  G
3361  ATGGTCATTTTAATAGTAAACAACAATTAGTTATTGCATTTCACTCATTAGATGTTTCTT
       D  G  H  F  N  S  K  Q  Q  L  V  I  A  F  H  S  L  D  V  S
3421  TAGCCTATTATATAAAAAAAAGATTAGGTTATGGAAGTGTTAAAAAAGTTAAAAATAAAA
       L  A  Y  Y  I  K  K  R  L  G  Y  G  S  V  K  K  V  K  N  K
3481  ATGCTTTTATTCTAGTTGTAACAGCTAGAGAAGGTATAGAAAAAGTAATTAATTTAATAA
       N  A  F  I  L  V  V  T  A  R  E  G  I  E  K  V  I  N  L  I
3541  ATGGCAAAATTAGAACAGAAAATAAATTTAATCAAATAACTAATAATATATTAAATCATG
       N  G  K  I  R  T  E  N  K  F  N  Q  I  T  N  N  I  L  N  H
3601  AGAATTATGCTGAATTTAGTAAAAAAATTAGTTTAAAATTAAATTTAAGTAATAATTTTA
       E  N  Y  A  E  F  S  K  K  I  S  L  K  L  N  L  S  N  N  F
3661  AAAATCATTGATTAGCTGGTTTTTCTGATGCTGATGCTAGCGGAGCTGGCCTTCACCCCG
       K  N  H  W  L  A  G  F  S  D  A  D  A  S  G  A  G  L  H  P
3721  GCTGCGTCAGCCCAAAGGCTAGTTTTCAAATAAAAAATACTTAATCGTAATAAAAAAATTG
       G  C  V  S  P  K  A  S  F  Q  I  K  I  L  N  R  N  K  K  I
3781  AAGTTAGATTAAATTTTCAAATTGATCAAAAAAAAGAATATATTTTACTATTAATTAAAG
       E  V  R  L  N  F  Q  I  D  Q  K  K  E  Y  I  L  L  L  I  K
3841  AATTTTTAGGTGGTAATATTGGCTATAGAAAAAGTCAAGATACATATTACTATGGATCTA
       E  F  L  G  G  N  I  G  Y  R  K  S  Q  D  T  Y  Y  Y  G  S
3901  CTAGTTTTGGTTCAGCTAAAAATGTTATTAATTATTTTGATTACTTTCATTTATTATCTA
       T  S  F  G  S  A  K  N  V  I  N  Y  F  D  Y  F  H  L  L  S
3961  GTAAACATGTTAACTATTTAAAATGAAGAAAAGCATATTTAATAATTCAAAATAAAGACC
       S  K  H  V  N  Y  L  K  W  R  K  A  Y  L  I  I  Q  N  K  D
4021  ATTTAAACAATGATGGACTTGAAAAAATTATAAAATTAAAAAGTACAATGAACAGAGTAA
       H  L  N  N  D  G  L  E  K  I  I  K  L  K  S  T  M  N  R  V
```

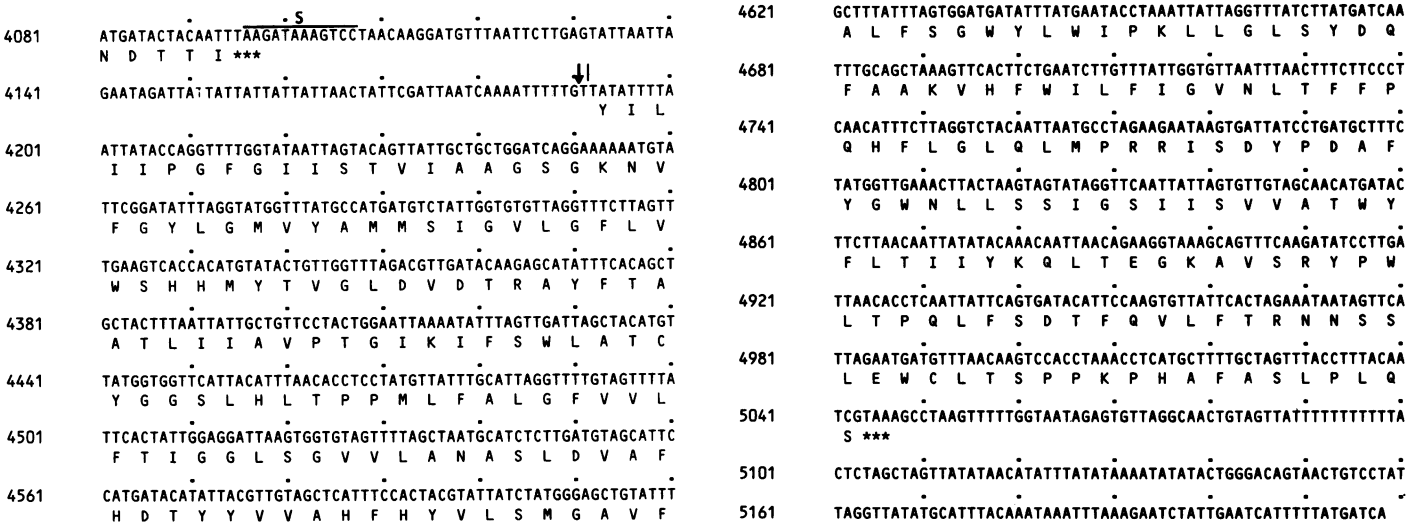

Fig. 2. DNA sequence of *oxi*A, the *A. nidulans* mitochondrial gene for subunit 1 of cytochrome oxidase. The likely splice sites are indicated by arrows. The lines adjacent to them indicate alternative splice sites on the basis of maximising amino acid homologies with yeast (Bonitz *et al.*, 1980) and bovine (Anderson *et al.*, 1982) sequences. The initiation codon cannot be identified – strong homology with the yeast and mammalian amino acid sequences begins at the arginine amino acid encoded by nucleotides 374-376. The two boxed sequences in intron 3 are 5-bp repeats, which may have arisen by duplication during insertion of the 37-bp sequence between them. The amino acids of the conserved decapeptide sequences are underlined. The conserved nucleotide sequences P, Q, R and S are overlined. Base sequences at positions 287–289 and 308 were ambiguous but have been established by Netzker *et al.* (1982). This reference contains the sequence from nucleotide 1 to the *Bam*HI site at nucleotide 1083.

and NOX3) contain long open reading frames continuous with the preceding exon. In contrast, the first intron, NOX1, has no open reading frame contiguous with the upstream exon, although there is a potential reading frame coding for 110 residues located after the S sequence, which ends in a UAG codon seven nucleotides upstream from the 3′ splicing point. The ninth residue of this open reading frame is a possible AUU initiation codon that follows a long stretch of A bases (see preceding section for the possible significance of this sequence). The codons of the NOX1 open reading frame are relatively GC-rich and the reading frame does not contain either of the two conserved decapeptides which are probably diagnostic of maturase proteins (see below). The location of this open reading frame in the secondary structure of the intron is shown in Figure 4, where the three *oxi*A introns are compared with the generalised structure proposed by Davies *et al.* (1982). By way of comparison, the 5_β intron of *oxi*3 in yeast has a large open reading frame with a maturase type amino acid sequence located in a similar position. The open reading frame starts just after the S sequence (see below) and has no AUG start codon (Hensgens *et al.*, 1983), just as in NOX1.

Conserved decapeptides in the open reading frames. Figure 3 shows some of the conserved decapeptides which occur twice in open reading frames of mitochondrial introns. It has been proposed that these decapeptides are diagnostic of a maturase protein (Waring *et al.*, 1982; Hensgens *et al.*, 1983). Hensgens *et al.* (1983) have divided the decapeptide-containing sequences into two classes based on the degree of homology – we only show one class. The first amino acid is always an aromatic residue, a tyrosine or tryptophan in the first decapeptide, a tryptophan in the second. The first decapeptide always ends with the sequence: glutamic or aspartic acid, glycine, aspartic acid, glycine, while the second ends with glutamic or aspartic acid, alanine, glutamic or aspartic acid, glycine (alanine in the case of NOX3). Both NOX2 and NOX3 have convincing decapeptide sequences (Figure 3) and thus probably code for maturase-like proteins. The 5′ end of NOX1 contains a short interrupted reading frame in which the sequence YLAGLFOchG occurs (Och represents an ochre terminator codon) (Figure 3). This sequence is very similar to the first decapeptides of NOX2 and NOX3 and the intron in the *cob*A gene of *Aspergillus* (Waring *et al.*, 1982) (Figure 3). It is possible that NOX1 originally contained a maturase-coding region which has since degenerated to such a degree that the abbreviated peptide sequence is the only recognisable feature left.

The two decapeptides are the most obviously conserved sequences of the intron open reading frames. However other less obvious regularities can be identified so that the amino acid sequences of NOX2 and NOX3 both show 20% homology to those of both the yeast *cob* fourth intron (YC4) and the *A. nidulans cob*A intron (NC) and 29% and 32% respectively to the yeast *oxi*3 third intron (YOX3). (Homologies are calculated from the first decapeptide to the end of the reading frame.) The homology of NOX3 to the *S. pombe* cytochrome oxidase subunit 1 second intron is 70% (88% over one stretch of 102 amino acids) which is remarkably high. This compares with a homology of 61% between the two *S. cerevisiae* maturases sequences YC4 and the fourth intron of *oxi*3 (YOX4). These were the most similar maturase sequences previously identified and the two maturases can act on their own and each other's intron (reviewed in Dujon, 1981; Dujardin *et al.*, 1982). The *S. cerevisiae cob* third intron and the single *A. nidulans cob*A intron which are located in equivalent positions show 48% homology.

Inserted sequence in the NOX3 intron. Comparison of the maturase sequences from NOX2 and NOX3 suggested that there is an extra short GC-rich DNA sequence located in the NOX3 intron just downstream from the region encoding the second decapeptide. This was verified when the sequence for the *S. pombe cox-1* second intron became available (Lang, accompanying paper) as 37 of 39 amino acids spanning this

```
YC4      DSNIRFNQ WLAGLIDGDG YFCITKNK- 80-
YOX4     SINKKFNQ WLAGLIDGDG YFGIVSKK- 80-
NC       HAGDLTYA YLVGLFEGDG YFSITKKG-106-
YOX3     LNYDKLGP YLAGLIEGDG SITVQNSS-106-
NOX2     NNNLPFAA YLAGLIEGDG TIIVPNTL- 93-
NOX3     VSDDQFGH YLAGLIDGDG HFNSKQQL- 85-
YOX5β    YTEEEKGY YLSGLFEGDG NIYTRCFS- 99-
Y RF1          fMISGFTDGDG SFYIKLND-110-
NOX1     IYYNDLKI YLAGLF*GNI WINNLEEH

YC4    -IKLTKDNA WFIGFFDADG TINYYYSG-
YOX4   -IKLTKDNS WFVGFFDADG TINYSFKN-
NC     -INTSYFSA WLVGFIEAEG CFSVYKLN-
YOX3   -TSDIGSNA WLAILTDADG NFSINLIN-
NOX2   -NEPLESNA WLSGFIEADG SFQVRSTL-
NOX3   -LSNNFKNH WLAGFSDADA SFQIKILN-
YOX5β  -EFDLTLNP WLTGFNDADG YFYTGFQK-
Y RF1  -PYDKINDY WILGFIEAEG SFDTSPKR-
```

Fig. 3. Conserved decapeptide sequences in proteins encoded by introns. The top group lists the first occurrences of the decapeptides, the bottom group the second. The neighbouring amino acids in the proteins are shown on either side. The single letter amino acid code is used (IUPAC-IUB Commission, 1969). YRF1 is a maturase-type protein sequence encoded by an open reading frame not located in an intron of a known gene (Corruzzi *et al.*, 1981). NOX1 shows amino acids from a closed section of the intron, which contains a region, similar to a decapeptide (see text) – the asterisk represents a stop codon. For the second decapeptide of NOX3, the amino acid sequence is given after removal of a proposed insertion sequence and one of the proposed five basepair duplicated sequences (see text). References are as follows: YC4, Nobrega and Tzagoloff (1980); YOX3 and YOX4, Bonitz *et al.* (1980) NC, Waring *et al.* (1982) and YOX5β, Hensgens *et al.* (1983).

region are perfectly homologous. The inserted region shows no homology and is flanked by a 5-bp direct repeat (Figure 2), indicative of an insertion event. This insert does not affect the frame of reading, but inserts 14 additional amino acids not present in the homologous open reading frame of *S. pombe*. If the insert and one of the direct repeats are removed, perfect homology at the amino acid level is restored between the open reading frames of these two introns. This coupled with the fact that the inserted DNA is exceptionally GC rich, more so than any other comparable stretch of the 95% of the *A. nidulans* mitochondrial genome sequenced to date, suggests that it did not originate from the mitochondrial DNA. Short GC-rich repetitive elements occur in the yeast mitochondrial genome, but they are not bordered by direct repeats (Sor and Fukuhara, 1982).

Secondary structure and splicing alignment of the introns in the oxiA gene

In this section and Figure 4 we describe the variations present in the secondary structure of the *oxi*A introns with reference to the generalised secondary structure presented by Davies *et al.* (1982). Four conserved sequences P, Q, R and S, with two other key sequences, E and E′, pair to produce three key pairings of a core RNA secondary structure. Besides these pairings three other stem loop structures (P6/L4, P8/L5 and P9/L6) are always present in the core secondary structure. The L5 loop usually contains the open reading frame coding for the putative maturase protein. The very short L3 loop present in NOX2 strengthens the argument for the occurrence of the P-Q (P4) specific pairing since in this case it would occur by short range RNA-RNA interaction. Similarly, the very short loop L5 in NOX1 (that does not contain an open reading frame in this position, see below) strengthens the arguments for the occurrence of the P8 and RS (P7) pairings. The stop codon of the open reading frame (in L5) in both NOX2 and NOX3 occurs immediately after the P8 pairing and overlaps the start of the S sequence.

In their generalised model for the splicing of class I introns, Davies *et al.* (1982) proposed that an internal guide sequence (IGS) aligns the splice sites by base pairing with the ends of the upstream (P1 pairing) and downstream exons (P10 pairing) to bring the exon ends within the distance of a phosphodiester bond (Figure 4). The P1 pairing includes a base pair between a pyrimidine, usually a U, at the end of the exon and a G base in the IGS, the base pair being located proximal to the loop formed by P1 (Figure 4). All three *oxi*A introns contain the conserved sequence 5′-GCC-3′ in the region of the IGS which pairs with the end of the upstream exon to make the P1 pairing (Figure 4). The IGS lies between the E sequence and the 5′ splicing site. Downstream from the P1 pairing there is usually a stem-loop structure, P2/L2 (Figure 4a). This structure is absent in both NOX2 and NOX3; the E sequence follows the IGS directly, separated by only six and three nucleotides respectively. NOX2 and NOX3 clearly have an RNA sequence which could act as an internal guide sequence to align the splice sites. In NOX1 a P1 pairing can be formed by bases in the correct position relative to the P2/L2 stem-loop, but no significant P10 pairing can be found.

A comparison between NOX1 and two introns of *S. cerevisiae* is illuminating. These are the third intron of *oxi*3 (YOX3) and the fifth intron of *cob* (YC5). YC5 was shown to be a class I intron by Michel *et al.* (1982) and was not included in the analysis of Davies *et al.* (1982). (i) The three introns have a typical core structure but the distance between the 5′ splice junction and the P2/L2 stem loop is >200 nucleotides which is much longer than usual. (ii) YOX3 and NOX1 are very similar in their sequences in regions where usually only the pairing is conserved, namely the P2/L2, P6/L4 and P8/L5 stem loops (Figure 4). (iii) All three introns have a GC-rich sequence immediately upstream of the P2/L2 stem-loop, which could pair with the end of the upstream exon to give a typical P1 pairing (i.e. with a pyrimidine-G base-pair proximal to the loop side of the pairing). The sequences are identical in NOX1 and YOX3 (UGCCG). However in YOX3 this particular P1 pairing may not form as the sequence at the end of the upstream exon can be sequestered into an alternative pairing with a sequence just downstream from it (see also Michel *et al.*, 1982). Of the three, only YC5 can form a P10 pairing. The situation for YC5 is described more fully in Waring and Davies (1984).

Thus of the three introns, it is possible that NOX1 and YOX3 still require part of an IGS alignment system to aid

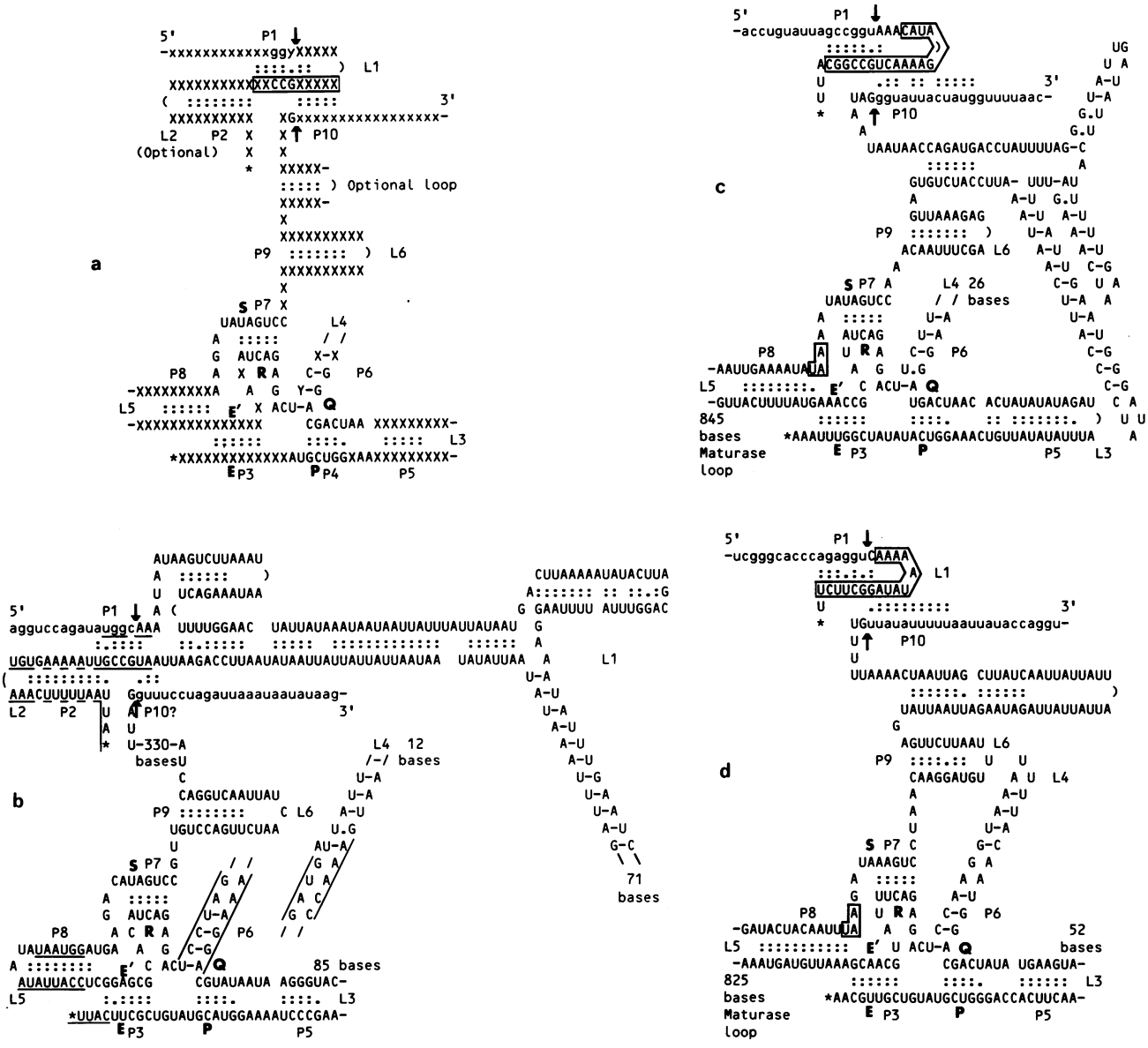

Fig. 4. Model secondary structure of the three *oxi*A class I introns. **(a)** Generalised secondary structure proposed by Davies *et al.* (1982), showing conserved bases and conserved pairings between non-conserved bases. RNA secondary structure of **(b)** NOX1, **(c)** NOX2 and **(d)** NOX3. For convenience of two-dimensional representation, the RNA structure has been shown as if cut at the point of the asterisks and so the sequence is continuous between these two points. In the generalised structure Y represents a pyrimidine base and X any base. Exon sequence is in lower case letters, intron sequence in capitals. The sequences P, Q, R, S, E and E′ are described in the text. Base-paired regions are labelled P1 to P10 and loops L1 to L6. In cases where all the sequence is not shown the number of bases omitted is shown at the ends of the pairings. The internal guide sequence (IGS) in NOX2 and NOX3 is boxed. Although the pairings of P1 and P10 with the IGS are shown as both occurring simultaneously prior to splicing, it is likely that P10 occurs after cleavage of the upstream splice site and release of the 5′ end of the intron. The P1 pairing would remain intact until splicing is complete (see text). The underlined sequences in NOX1 denote those sequences in P1, P2, P8 and P6 which are identical in the same region of the third intron of *oxi*3 in yeast (YOX3). In NOX1 the region between P9 and the 3′ splice site contains an open reading frame of 110 codons, which are not shown. The boxed UAA sequences in NOX2 and NOX3 are the stop codons of the proposed maturase reading frames. They are identically situated and overlap the start of the S sequence.

splicing while YC5 still actually uses an IGS for aligning the splice junctions.

Discussion

The *oxi*A gene of *A. nidulans* has three introns, which clearly belong to a large class (class I) of mitochondrial, nuclear ribosomal and chloroplast ribosomal and tRNA introns (Waring *et al.*, 1983; Michel and Dujon, 1983) on the basis of their having a particular conserved core RNA secondary structure as defined by Davies *et al.* (1982). In all class I introns there is a sequence immediately upstream of the L2/P2 stem loop which can pair with the end of the upstream exon (the alternative pairing for YOX3 has already been noted). In most cases this pairing (P1) enables the end of the downstream exon to be aligned adjacent to the end of the upstream exon (the P10 pairings). The work of Cech's group (Cech *et al.*, 1981; Zaug and Cech, 1982; Kruger *et al.*, 1982) suggests that the upstream splice site is cleaved first so that the P10 pairing probably occurs after the P1 pairing rather than simultaneously. In a subclass of introns, such as NOX1, there is no obvious P10 type pairing. In contrast the P1 pairing is always conserved. Thus even in those cases where the full IGS pairing with both upstream and downstream exons is unlikely

to form, the pairing of the 3′ end of the IGS-related sequences with the end of the upstream exon may be needed for cleavage of the 5′ splice site by providing the structure required for recognition and for breakage of this site. In all other aspects the detailed analysis of these *oxi*A introns precisely supports the model of Davies *et al.* (1982).

The second and third introns have an open reading frame which could code for maturase-type proteins. The YC4 maturase of yeast probably has a precursor protein encoded by the upstream exon codons fused to the intron open reading frame, the latter being contiguous with the upstream exon reading frame. The site of cleavage is probably downstream of the amino acids encoded by the R and E′ sequences (where *box*9 mutations map) but upstream of the first decapeptide (Jacq *et al.*, 1982; de la Salle *et al.*, 1982; Weiss-Brummer *et al.*, 1982; Anziano *et al.*, 1982; Waring *et al.*, 1982). As NOX2, NOX3 and NC are structurally similar to yeast YC4, maturases may well be produced from these introns in an equivalent manner. The very highly conserved 12 base R sequence is read in the three different possible phases of the reading frame in these three *Aspergillus* introns, thus generating different amino acids. This is consistent with the idea that the amino acid sequence in this region is not important and presumably lies in the portion of the precursor which is cleaved off. We have previously suggested that the first decapeptide was at or near the N terminus of the mature maturase (Waring *et al.*, 1982). We have compared the amino acids situated after those encoded by the R and E′ sequence and before the start of the decapeptide in the yeast YC4 and YOX4 and in NC, NOX2 and NOX3 to look for a conserved proteolytic site. There is no conserved pair of amino acids in all five introns and indeed in NC there are only four amino acids in this region. It seems that the putative proteolytic activity recognises more than just primary structure, or perhaps that the decapeptide itself is recognised.

There is a remarkable degree of homology between NOX3 and the second intron of the *oxi*1 gene of *S. pombe*. Since it is probable that these species have been sexually isolated for a very long time, this indicates that the details of intron sequence and structure are very important, and suggest that introns may be functionally important. However, the function, if any, of the introns in the *Aspergillus* species is not clear. In yeast the splicing of the fourth intron of *oxi*3 (YOX4) is dependent upon the YC4 maturase. It has been suggested that this arrangement co-ordinates the stoichiometric expression of the *cob* and *oxi*3 genes, which code for components of different complexes in the respiratory chain (Jacq *et al.*, 1982). As the *cob*A intron (Waring *et al.*, 1981) is not very homologous to any of the *oxi*A introns and as *A. quarilineatus* probably has no *cob*A intron (Earl *et al.*, 1981; Turner *et al.*, 1982), this concept is implausible for *Aspergillus*.

The major difference between NOX3 and its *S. pombe* equivalent is the presence in NOX3 of an extra GC-rich sequence, flanked by a 5-bp direct repeat. This is very reminiscent of the typical result of integration of transposable elements in both prokaryotes and eukaryotes. The small size and lack of inverted repeats at the ends of this element suggest that this is a remnant of the original element, produced possibly by a deletion event that restored the reading frame. Nevertheless, it is probable that this is the first indication that transposable elements occur in filamentous fungi.

Materials and methods

DNA sequencing

The preparation of *A. nidulans* mitochondrial DNA, plasmid DNA and restriction fragments as primers for DNA sequence analysis has been described (Waring *et al.*, 1981). The following restricted fragments of mitochondrial DNA from the *oxi*A region were cloned: *Bam*HI fragment 2 into pBR327 (Soberon *et al.*, 1981); *Hpa*II fragments 4 and 6 into M13mp7 (Messing *et al.*, 1981); *Bcl*I fragments 6, 9, 11 and 12 into M13mp6 (Messing *et al.*, 1981). *Hpa*II fragments 4 and 6 were purified from the clones by digestion with *Eco*RI and separation from M13mp7 DNA using a malachite green AT base affinity column (Boehringer, Mannheim) as described in Waring *et al.* (1982); *Bcl*I fragment 9 was purified by digestion with *Eco*RI, separation on an agarose gel and elution of the fragment as described in Waring *et al.* (1981). These purified fragments were digested with *Sau*3A and subcloned into M13mp6. Sequences were obtained using the chain-termination method of Sanger *et al.* (1977) with a 30-bp universal primer (Anderson *et al.*, 1980) as described by Davies (1982). As part of our overall project to sequence the mitochondrial genome total mitochondrial DNA had also been digested with *Sau*3A and *Eco*RI* (Hsu and Berg, 1978) and cloned into M13mp6 and M13mp2 respectively. Some of the sequence from these clones overlapped other sequences in *oxi*A. The *Sau*3A sites were sequenced across using internal primers pre-treated with exonuclease III on templates of M13 clones, containing larger fragments such as *Hpa*II fragment 4 and 6. Primers were prepared from mitochondrial DNA cloned into pBR327 and M13. Residual gaps in the sequence were obtained using other internal primers as described in Figure 1.

Acknowledgements

We thank B.F.Lang for providing us with sequence data prior to publication. This work was supported by project grant G7904290CB from the Medical Research Council to R.W.D. and C.S.

References

Anderson,S., Gait,M., Mayol,L. and Young,I.G. (1980) *Nucleic Acids Res.*, **8**, 1731-1743.

Anderson,S., Bankier,A.T., Barrell,B.G., de Bruijn,M.H.L., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., Schreier, P.H., Smith,A.J.H., Staden,R. and Young,I.G. (1981) *Nature,* **290**, 457-465.

Anderson,S., de Bruijn,M.H.L., Coulson,A.R., Eperon,I.C., Sanger,F. and Young,I.G. (1982) *J. Mol. Biol.,* **156**, 683-717.

Anziano,P.Q., Hanson,D.K., Mahler,H.R. and Perlman,P.S. (1982) *Cell,* **30**, 925-932.

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

Bonitz,S.G., Coruzzi,G., Thalenfeld,B.E., Tzagaloff,A. and Macino,G. (1980) *J. Biol. Chem.,* **255**, 11927-11941.

Burger,G., Scriven,C., Machleidt,W. and Werner,S. (1982) *EMBO J.,* **1**, 1385-1391.

Cech,R.T., Zaug,A.J. and Grabowski,P.J. (1981) *Cell,* **27**, 487-496.

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

Coruzzi,G., Bonitz,S.G., Thalenfeld,B.E. and Tzagoloff,A. (1981) *J. Biol. Chem.,* **256**, 12780-12787.

Davies,R.W. (1982) in Rickwood,D. and Hames,B.D. (eds.), *Gel Electrophoresis of Nucleic Acids – A Practical Approach*, IRL Press Ltd., Oxford pp. 117-172.

Davies,R.W., Waring,R.B., Ray,J.A., Brown,T.A. and Scazzocchio,C. (1982) *Nature,* **300**, 719-724.

de la Salle,H., Jacq,C. and Slonimski,P. (1982) *Cell,* **28**, 721-732.

Dujardin,G., Jacq,C. and Slonimski,P. (1982) *Nature,* **298**, 628-632.

Dujon,B. (1981) in Strathern,J.N., Jones,E.W. and Broach,J.R. (eds.), *Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance,* Cold Spring Harbor Laboratory Press, NY, pp. 505-635.

Earl,A.J., Turner,G., Croft,J.H., Dales,R.B.G., Lazarus,C.M., Lunsdorf, H. and Kuntzel,H. (1981) *Curr. Genet.,* **3**, 221-228.

Hensgens,L.A.M., Arnberg,A.C., Roosendaal,E., Van der Horst,G., Van der Veen,R., Van Ommen,G.J.B. and Grivell,L.A. (1983) *Cell,* **32**, 379-389.

Hsu,M. and Berg,P. (1978) *Biochemistry,* **17**, 131-138.

Jacq,C., Pajot,P., Lazowska,J., Dujardin,G., Claisse,M., Groudinsky,O., de la Salle,H., Grandchamp,C., Labouesse,M., Gargouri,A., Guiard,B., Spyridakis,A., Dreyfus,M. and Slonimski,P. (1982) in Slonimski,P., Borst,P. and Attardi,G. (eds.), *Mitochondrial Genes*, Cold Spring Harbor

Laboratory Press, NY, pp. 155-183.
Kruger,K., Grabowski,P.J., Zaug,A.J., Sands,J., Gottschling,D.E. and Cech,T.R. (1982) *Cell,* **31**, 147-157.
Messing,J., Crea,R. and Seeburg,P.H. (1981) *Nucleic Acids Res.,* **9**, 309-321.
Michel,F. and Dujon,B. (1983) *EMBO J.,* **2**, 33-38.
Michel,F., Jacquier,A. and Dujon,B. (1982) *Biochimie,* **64**, 867-881.
Netzker,R., Kochel,H.G., Basak,N. and Kuntzel,H. (182) *Nucleic Acids Res.,* **10**, 4783-4794.
Nobrega,F.G. and Tzagoloff,A. (1980) *J. Biol. Chem.,* **255**, 9828-9837.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5463-5467.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.,* **143**, 161-178.
Soberon,X., Covarrubias,L. and Bolivar,F. (1981) *Gene,* **9**, 287-305.
Sor,F. and Fukuhara,H. (1982) *Nucleic Acids Res.,* **10**, 1625-1633.
Turner,G., Earl,A.J. and Greaves,D.R. (1982) in Slonimski,P.P., Borst,P. and Attardi,G. (eds.), *Mitochondrial Genes,* Cold Spring Harbor Laboratory Press, NY, pp. 411-414.
Waring,R.B., Davies,R.W., Lee,S., Grisi,E., McPhail Berks,M. and Scazzocchio,C. (1981) *Cell,* **27**, 4-11.
Waring,R.B., Davies,R.W., Scazzocchio,C. and Brown,T.A. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 6332-6336.
Waring,R.B., Davies,R.W., Brown,T.A. and Scazzocchio,C. (1983) *J. Mol. Biol.,* **167**, 595-606.
Waring,R.B. and Davies,R.W. (1984) *Gene,* in press.
Weiss-Brummer,B., Rodel,G., Schweyen,R.J. and Kaudewitz,F. (1982) *Cell,* **29**, 527-536.
Zaug,A.J. and Cech,T.R. (1982) *Nucleic Acids Res.,* **10**, 2823-2838.
Zaug,A.J., Grabowski,P.J. and Cech,T.R. (1983) *Nature,* **301**, 578-583.

Received on 8 May 1984

Note added in proof

In Figure 4d there should be four not five consecutive U bases in the downstream exon in the P10 pairing region and therefore one less base pair in P10.