Making ends meet: a model for RNA splicing in fungal mitochondria

R. Wayne Davies*, Richard B. Waring*, John A. Ray*, Terence A. Brown* & Claudio Scazzocchio†

*Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK

†Department of Biology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, Essex, UK

On the basis of available nucleotide sequence and genetic data, we present a model for RNA splicing in fungal mitochondria. Seven intron RNAs of two fungal species can form identical secondary structures, involving four conserved sequences, which bring the ends of each intron together and allow an internal guide RNA sequence to pair with exon bases adjacent to the splice junctions. The splicing sites are thus aligned precisely within a conserved structure, which we suggest could present specific recognition signals to the proteins that catalyse the splicing reaction.

MANY eukaryotic genes are interrupted by DNA sequences which do not code for any part of the gene product. These DNA sequences are known as introns (or intervening sequences), whereas the coding regions of the gene are called exons. Introns have been found in eukaryotic mRNA, tRNA and rRNA nuclear genes, and also in mitochondrial mRNA and rRNA genes of fungi[1-3] and at least one plant mitochondrial mRNA gene[4]. In all cases introns are transcribed as part of the primary RNA transcript and are then excised by 'splicing'—a precise cleavage and ligation reaction which leaves the various exons joined together end to end in their order of synthesis. The mechanism of splicing has not been fully elucidated for any intron. We now present a model which shows how the RNA of one class of mitochondrial intron could take up a well defined secondary structure, allowing the precise and unambiguous alignment of the ends of exons flanking the intron ready for the cleavage and ligation reactions. We have used two lines of evidence in constructing the model. Comparison of the DNA sequences of four mitochondrial introns of *Aspergillus nidulans* which we recently determined (ref. 5 and R.B.W. *et al.*, in preparation) and five of *Saccharomyces cerevisiae*[6-8] highlights certain key conserved sections of intron sequence. When these are taken as the basis of the formation of base-paired regions of the RNA molecule, further conserved RNA–RNA pairing between non-conserved sequences is revealed, and quite remarkably all these introns can form very similar RNA secondary structures. Genetic evidence, provided largely by the work of Slonimski and co-workers[9-12], shows that particular short segments of the intron sequence distant from the splicing sites are important in *cis* for splicing; our model explains the phenotypes resulting from *cis*-acting mutations.

The introns studied are listed in Table 1 and the terminology used is described in Fig. 1 legend. The sequences that are highly conserved in introns of both species[5] are named P, Q, R and S; their sequences and locations are given in Table 1. Seven of the introns have the internal primary structure shown in Fig. 1[9,13]: a *cis*-acting region occupies the 5′ end of the intron (which in mRNA introns has an open reading frame), a central protein-coding region extends from just beyond conserved sequence R to the stop codon and there is a 3′ *cis*-acting region with no open reading frame. The protein-coding region of YC4 has been shown[9,13] to code for a protein essential for the excision of YC4 ('mRNA maturase')[11], and this is probably the case for several other introns. The general features of the model will

Table 1 Important sequences within mitochondrial introns

Intron	5′ Splice Site	IG	E	P	Q	R	E′	S	3′ Splice Site
		*		Sequence					
YC4	↑- 6-	AUGGCCU	-32- UGUGC	-3-AUGCUGGAAA-	192-AAUCAGCAGG-	65-UCAGAGACUACA-	1-GCACA	-1137-AAGAUAUAGUCC-	21-↑
YOX4	↑- 6-	GUGGCC	-34- UUUUUC	-3-AUGCUGGAAA-	78-AAUCAGCAGG-	50-UCAGAGACUACA-	1-GAAAAA	- 740-AAGAUAUAGUCC-	36-↑
Y_{ω}	↑- 3-UUACCCCCUUGUCCC	-64-AUUGGGUG	-2-UUGCUUAGAU-	25-AAUAAGCAGC-	49-UCAACGACUAGA-	34-CGCCCAAU-	848-AUGAUAUAGUCU-	61-↑	
YOX3	↑-1005- UAAUAAGAUGCCG	-28- CAACAC	-3-AUGCGGGAAA-	165-AAUCCGCAGG-	95-UCAGAGACUACA-	1-GUGUUG	- 28-AAGAUAUAGUCC-	29-↑	
NOX1	↑- 250-	AAUGCCG	-27- CUUCGC	-3-AUGCAUGGAA-	102-AAUAUGCAGG-	39-UCAGAGACUACA-	1-GCGAGG	22-AAGACAUAGUCC-	362-↑
NOX2	↑- 9-AUAGAAAACUGCCGGC-	6- UUUGGC	-3-AUACUGGAAA-	28-AAUCAGUAGG-	32-UCAGAGACUAUA-	1-GCCAAA	869-AAAAUAUAGUCC-	124-↑	
NOX3	↑- 1-AAAAAUAUAGGCUUCU-	3- CGUUGC	-3-AUGCUGGGAC-	69-UAUCAGCAGG-	19-UCAGAGACUUUA-	1-GCAACG	- 851-AAGAUAAAGUCC-	83-↑	
NC	↑- 4-	ACAGAUGAACC	-26- UCCUGC	-3-AUGCUGGAAA-	74-AAUCAGCAGG-	30-UCAGAGACUACA-	1-GCAGGA	- 764-AUGAUAUAGUCC-	76-↑
YC3	↑- 2-	AUAUAGAGGAUCC	-						↑
Consensus				7878667586 AUGCUGGAAA	7886687887 AAUCAGCAGG	888778888758 UCAGAGACUACA		867878788887 AAGAUAUAGUCC	

The nucleotide sequences and locations within nine fungal mitochondrial introns of regions that are important for RNA splicing, including the conserved sequences P, Q, R and S, are shown. The latter are named in their order of occurrence in the 5′ to 3′ direction within the introns. The E, E′ and IG sequences are described in the text. Numbers indicate distances in bases between particular sequences. The asterisk marks the conserved G base of the IGS. P and Q are the same as A and B, respectively, in ref. 5.

0028-0836/82/510719—06$01.00

be presented first, followed by a detailed consideration of the supporting evidence.

A model of mitochondrial RNA splicing

Figure 2 gives a diagrammatic representation of the steps in the splicing process, and Fig. 3 shows the actual secondary structure of the folded intron RNA. The model proposes three major stages in the excision of these introns.

(1) Intron RNA forms a specific secondary, and presumably tertiary, structure involving a number (usually nine) of conserved RNA–RNA interactions, with the result that the ends of the intron are brought close together. The RNA–RNA pairings involving the conserved sequences P, Q, R and S are among the most critical. The P and Q sequences are synthesized early, and their pairing P4 and the associated P5 (Fig. 3) are important in nucleating the secondary structure. The R–S pairing P7 and the associated P8 (Fig. 3) bring the 3′ end of the intron close to R, which is in turn brought close to the 5′ end of the intron by the pairing of the E and E′ sequences. Proteins may also be involved, binding at specific sites on the intron RNA either to stabilize secondary structure or to aid the formation of the precise alignment structure in stage 2. We regard the characteristic secondary structure formed by pairings P3 to P9 in Fig. 3 as the 'core structure' of these introns. The formation of the core structure is a prerequisite for the next stage.

(2) An internal guide sequence (IGS) located just 5′ to the E sequence (Figs 1, 3) pairs with short stretches of RNA sequence at the end of the upstream exon and the beginning of the downstream exon, acting as an adaptor molecule that brings the two splicing sites tightly together, separated only by the length of a phosphodiester bond. The IGS–upstream exon pairing is P1 in Fig. 3, the IGS–downstream exon pairing is P10. The P1 and P10 pairings involving IGS make up the precise alignment structure, details of which are given in Fig. 6 for nine introns. In all cases the splice sites are precisely aligned, and are in similar chemical environments at the ends of symmetrically placed helices. As shown in Fig. 2, this structure also ensures that exon–exon ligation is favoured, because when the RNA is cut at the exon–intron junction the exon ends are held in the double-stranded region, while the intron ends are free to diffuse away.

(3) The completed precise alignment structure is recognized by the splicing protein(s). The features recognized are the RNA structure *per se*, and certain chemical groups that only occur together near the splice sites in the complete structure, in particular the U · G base pair (bp) of P1 adjacent to the 5′ splice site and the G base at the 3′ end of the intron. In addition, the protein(s) recognizing this structure may also need to bind to components of the core structure or its associated proteins so as to bind efficiently. The specificity of the reaction is dictated by other non-conserved chemical signals and by the mRNA maturase[13]. Endonucleolytic scission at the 5′ and 3′ splice sites may be simultaneous or sequential. The free ends of intron RNA diffuse away immediately but the free ends of exon RNA are still held in the P1 and P10 pairings, and are ligated together by an RNA ligase which need not be specific. The splicing proteins are released from the IGS; the P1 and P10 pairings fall apart, releasing the linear intron (which is presumably subsequently destroyed) and the spliced mRNA.

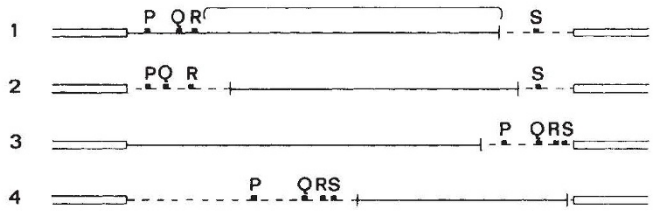

Fig. 1 A schematic representation of the internal primary structure of a class of fungal mitochondrial introns. Four different arrangements of open reading frames are shown; class 1 is the majority arrangement, found in six of the nine introns discussed here, while the other classes have one representative each. Internal features of each intron are shown to scale, but all introns have been scaled to the size of the class 1 intron. Symbols are as follows: open bars, exon sequences; continuous line, open reading frame within an intron; discontinuous line, intron sequence with no open reading frame. The bracketing line above the class 1 intron indicates the part of the yeast cob I4 intron in which *trans*-acting mutations of the *box*7 class have previously been located; this defines the maturase coding region. P, Q, R and S are the conserved sequences shown in Table 1. The results presented in this article are based on the comparison of nine intron DNA sequences, four from *A. nidulans* and five from yeast. We refer to these and other introns by species, gene and number only, symbolizing *S. cerevisiae* by Y and *A. nidulans* as N, and dropping 'intron'. Thus, yeast introns are written as follows: cob I3, YC3; cob I4, YC4; oxi3 I3, YOX3; oxi3 I4, YOX4; 21S rRNAω, Yω. *A. nidulans* introns are written as follows: cobA I1, NC; oxiA I1, NOX1; oxiA I2, NOX2; oxiA I3, NOX3. The data for Yω (ref. 8), YC4 (ref. 7), YOX3 and YOX4 (ref. 6) have been published, and those for NC, NOX1, NOX2 and NOX3 are from our laboratory and will be published in detail elsewhere (ref. 5 and R.B.W. *et al.*, in preparation). Nine introns NC, NOX1, NOX2, NOX3, YC3, YC4, YOX3, YOX4 and Yω, have been studied but we only show analyses based on eight of the introns plus those portions of the YC3 sequence that have already been published[21]. In this figure intron 1 = YC4 (YOX4, YC3, NC, NOX2 and NOX3 are similar to YC4), 2 = Yω, 3 = YOX3 and 4 = NOX1.

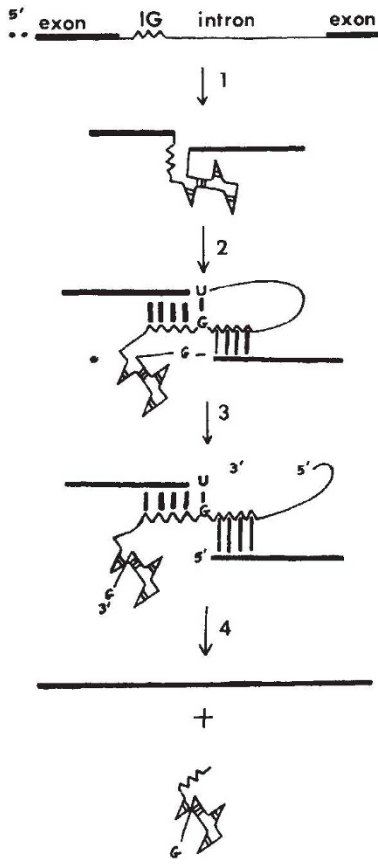

Fig. 2 Model of the role of the RNA substrate in mitochondrial RNA splicing. The model is described in the text. The steps shown here are: (1) formation of internal secondary structure; (2) formation of the precise alignment structure; (3) endonuclease binding followed by specific cuts to produce the ends shown and diffusion of free intron ends away from the splicing complex; (4) ligation of exon ends and release of linear intron RNA from the splicing complex. The precise alignment complex which is P1 + P10 of Fig. 3, is shown greatly enlarged between stages (2) and (4). The thick line represents exon RNA, the thin line intron RNA, the zigzag line the internal guide RNA sequence (IGS).

The core structure

The fact that all introns considered here possess the same core secondary structure argues for the functional importance of this structure. This is strongly supported by our ability to explain the phenotype of all the sequenced *cis*-acting mutations affecting splicing, and by the maintenance of intact base-pairing despite variations from the consensus sequences. We stress that if any one intron RNA sequence is considered alone it is possible to find alternative ways of pairing the conserved sequences, but the model secondary structure presented in Fig. 3 is applicable to all the introns of both species. We have only considered pairings to be potentially biologically meaningful if they occur between identical or identically placed sequences of all the introns studied.

The extreme regularity and conservation of detail of RNA secondary structure within introns can be seen in Fig. 3*a*; all the conserved bases shown in the core structure in this figure are present in at least six out of eight introns and the conserved base pairs in all or seven out of eight introns. One of the most remarkable features that these introns have in common is the existence of a pair of sequences named by us E and E′ (Table 1, Fig. 3). Two bases after the conserved sequence R, there is a conserved G base in all the introns followed by a C base in five out of seven introns. These two bases together with the next four (three in YC4, five in YOX4 and Yω) make up the sequence we call E′. The amazing property of E′ is that, although the last four bases are completely different from intron to intron, they and the GC preceding them are always capable of pairing perfectly with a sequence E that precedes sequence P by three bases (Figs 3, 6). This conservation of precise base-pairing ability of sequences in identical positions in all introns, independent of actual nucleotide sequence conservation, demands the interpretation that these sequences really do pair.

Further evidence for the functional importance of the RNA–RNA interactions between P and Q, R and S is provided by looking at the effects on the pairings of variations from the consensus sequence. Sequences P and Q are strongly conserved between introns and species, but there are several variations from the consensus sequences. In all cases where a different nucleotide sequence is found within the section of P that can pair with Q, a compensatory alteration in sequence Q is found so that the base-pairing between the sequences is maintained. There are five such compensatory exchanges, the most striking being the case of NOX1, where an insertion of a base in P is exactly compensated for by two changes in Q. There are no alterations in the proposed base-pairing region of P that are not compensated for by complementary alterations in Q. The pairing regions of the R and S sequences are very precisely conserved; only in NOX3 does a change occur in part of the R sequence that pairs with S, and here too a compensatory alteration in S is found. All these compensatory exchanges in the variant sequences are shown in Fig. 4.

This model also accounts for the phenotypes due to several *cis*-acting mutations affecting splicing which have been sequenced by Slonimski and co-workers[9,12]. The mutations fall into two clusters, which in YC4 have been called *box*9 and *box*2. All the *box*9 mutations that have been sequenced (including the equivalent class in YOX4) fall in R or E′, and reduce the strength of base-pairing either of R and S or of E and E′; as R and E′ are separated by only one base, they have previously been considered to be part of one functional sequence, whereas we propose that they have different roles, although both are involved in secondary structure formation. Details of the sequenced mutations are given in refs 9 and 12, and their effects on the R–S and E–E′ interactions are shown in Fig. 5. Two mutations of the *box*2 class which affect sequence S of YC4 have also been sequenced[9]. Both W300-3 and G2590 alter G (ref. 8) to A; this would destroy a central G · C base pair in the pairing of S with R (Fig. 5). Thus, our proposal also explains the observed phenotype of *box*2 mutations. Weakening of the R–S or E–E′ pairings by these mutations will reduce the stability of the core structure, thus preventing efficient

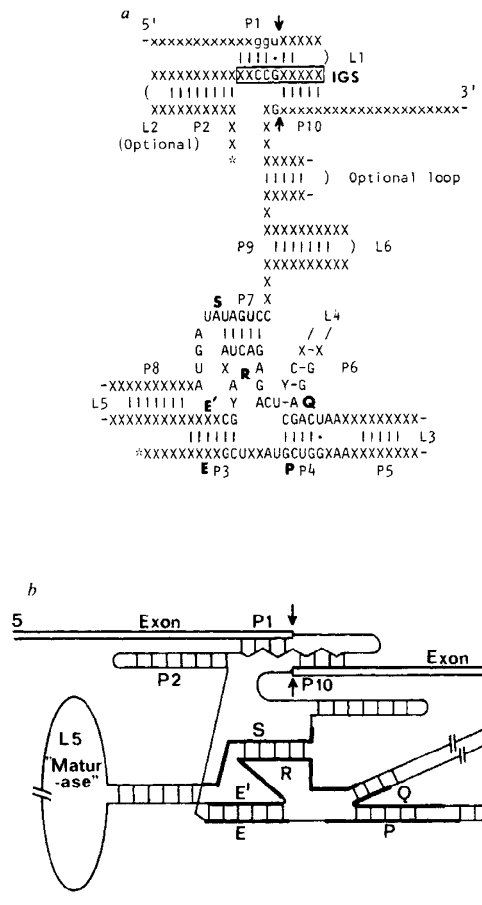

Fig. 3 Model RNA secondary structure for a class of fungal mitochondrial introns. *a*, A generalized secondary structure model derived from the RNA sequence of eight introns (YC3 data were not used), showing conserved bases and conserved pairings between non-conserved bases. *b*, A diagrammatic representation of the secondary structure showing only the key components. For convenience of two-dimensional representation, the RNA structure has been shown in *a* as if cut between P2 and P3 (or just 5′ to E)—the arbitrary ends that are left are marked with an asterisk and the sequence is continuous between these points. Intron bases are in capitals, exon bases in lower-case letters. A, U, C or G is written when these bases occur in a particular position in at least six out of eight introns, while Y indicates a pyrimidine and R a purine base in six out of eight introns; otherwise X or x signifies any base. Except for P1 and P10 (see Fig. 6) when a base pair is given, it occurs in at least seven out of eight introns in this position. When the end of a stem–loop structure is closed by a round bracket, the number of bases shown is that of the consensus intron; when the end is shown with a dash after each nucleotide, then a large variable number of bases is found in the loop (see Table 1). Base-paired regions are labelled P1 to P10, loops L1 to L6. L5 contains the maturase-coding region of the majority class intron; the putative maturase-coding region of YOX3 occurs in L1, and this loop is also large (though non-coding) in NOX1. YOX3 and NOX1 have small L5 loops, all others have small L1 loops of the average size shown in the consensus structure. P3 is formed by the pairing of P and Q, P7 by the pairing of R and S, P1 and P10 by the pairing of IGS with exon sequences. The central core of the structure is P3, P4 and P7 with the associated P5, P6 and P8. P2 and P9 are peripheral stems which bring the splice sites close to the folded central core; P1 and P10 form the precise alignment structure. In *b* exons are shown as open bars, conserved sequences as thick solid lines, IGS as a wavy line, and base pairs are shown where conserved. The line joining P2 to E is of necessity out of scale, and represents only four bases on average. The L5 loop is shown as coding for the maturase protein as in YC4, but it can code for non-maturase protein (Yω), be defective for protein coding (YOX4, NOX3), or very small (NOX1, NOX3).

formation of the precise alignment structure. This is why mutations within an intron cause splicing defects.

No mutations have so far been isolated in P, Q or E. This is probably because the genetic map is not saturated, that is, only a small proportion of all the possible mutations affecting splicing have been identified. Most of the mutations that have been isolated in R, S and E′ are leaky[13], so if point mutations in P and Q had slightly weaker effects they could be difficult to detect. Alternatively, the mutations that have been isolated may affect not only RNA–RNA pairing but also alter the specific binding sites of proteins involved in splicing, and the P–Q pairing might not bind any protein at all. The model predicts that the distribution of small deletion mutations would be much more widespread and include P and Q, that point mutations will be found in E, and that E mutations will be found as suppressors of E′ mutations and vice versa.

The precise-alignment structure and the internal guide sequence

At some stage of splicing the splice sites must be aligned precisely and held together long enough for the excision of the intron and the ligation of at least the exon ends to be achieved. Although this could conceivably be done by proteins alone, a very reliable and precise way of aligning two nucleotide sequences is by way of complementary base-pairing with an adaptor molecule[14,15]. We propose that this class of fungal mitochondrial intron RNAs possesses an internal guide sequence that functions as an adaptor, pairing with upstream and downstream exon sequences adjacent to the splice sites in such a way as to place the splice sites together exactly. The base-pairing in this precise-alignment structure need not be very extensive because we propose that the previous folding of the RNA core structure strongly favours the pairing of the internal guide sequence with the exon RNA sequences. In fact, most of the precise-alignment structures shown in Fig. 6 involve base-pairings (P1, 3–6 bp; P10, 3–8 bp) that are comparable with those proposed for the pairing of the U1 external guide RNA with intron sequences in the alignment of the ends of nuclear mRNA introns[14,15] without previous folding.

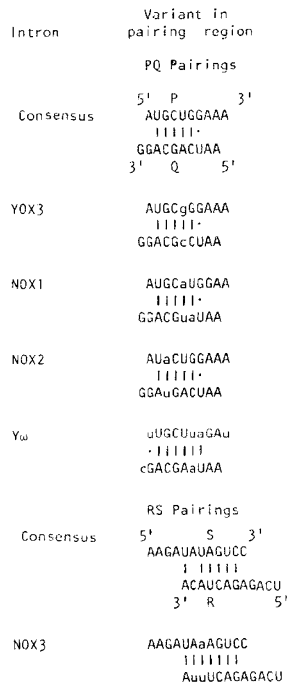

Fig. 4 The P–Q and R–S pairings with variants. Consensus pairings and all known exceptions are shown, demonstrating the conservation of base-pairing in the variant introns. Bases that differ from consensus are shown in lower-case, consensus bases in capitals. Intron terminology is given in Fig. 1 legend.

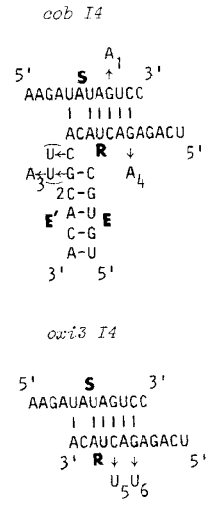

Fig. 5 Mutational evidence for the R–S and E–E′ pairings. The R–S pairing of YOX4 and the R–S and E–E′ pairings of YC4 are shown with all the sequenced *cis*-acting mutations affecting splicing[9,12]. Numbers next to bases in mutants indicate the identity of the mutant[9,12,13]: 1 = G2590 and W300-3; 2 = M4474 (a double base change); 3 = M7832; 4 = V353; 5 = G2190; 6 = G192.

The internal guide sequences are given in Table 1, and the IGS–exon pairings making up the precise-alignment structures are shown in Fig. 6; these correspond to P1 and P10 of Fig. 3. A similar though slightly shifted version of an alignment structure was proposed for Yω by Bos *et al.*[16]. If P1 is considered in isolation it involves an extra 1 or 2 bp beyond the U · G pair shown as the end of P1 in Fig. 6; these base-pairings are between the first bases of the intron and one or two of the IGS bases that can also take part in the P10 pairing.

The first argument in favour of this structure is the regularity with which it can be formed by a sequence (IGS) between the 5′ splice site and sequence E. There are several conserved features, which can be located in Fig. 3 and seen in detail in Fig. 6. The finding that the precise-alignment structures are variable apart from these features agrees well with the considerable specificity of splicing reactions; note that the alignment structures and IGS of YOX4 and YC4, which share the same splicing specificity[13], are very similar.

The phenotypes resulting from two known mutations can be explained if the mutations reduce the strength of the pairings in the precise-alignment structure. G2457 is a mit$^-$ mutation of yeast that is clearly defective in the excision of YC4, but has been shown by DNA sequencing to be a single G→A transition affecting the last but one nucleotide of the upstream exon[9]; this G→A change would significantly reduce the stability of the pairing, as well as affecting the potential protein-binding site. This G and its pairing with a C in IGS within the P1 pairing are strongly conserved. The other example is the class of mutations known as ω^n, which occur in the gene for the large mitochondrial rRNA of *S. cerevisiae*; ω^n mutations remove the strong polarity (biased transmission of parental genotypes to offspring) that is otherwise found in crosses between ω^+ and ω^- strains of *S. cerevisiae*. ω^+ and ω^- strains differ in the presence and absence of the Yω intron, and ω^n mutations only occur in ω^- strains, never in ω^+ strains. ω^n mutations have been shown by DNA sequencing[8] to be a G→A transition three bases before the end of the upstream exon. This would reduce the strength of the P1 pairing in our model. Clearly, if this prevents RNA splicing, ω^n mutations could never occur in ω^+ strains because the Yω intron would not be excised and this would be lethal.

The proposed structure thus fulfils all the criteria for a structure having a critical role in determining the precision of the splicing reaction. As for P, Q and E, we expect mutations to be found in the IGS and further mutations to be found in the

exon sequences near splice junctions as the genetic map becomes saturated.

Two introns which definitely have the core structure but only weak precise-alignment structures differ from the seven other introns in having extra DNA present between at least one splice junction and the core structure sequences (Fig. 1). The core structure of these introns is still necessary (hence its conservation) but not sufficient for bringing the ends of the intron together. There are two ways of interpreting the structure of NOX1 and YOX3. First, these introns are now spliced in a different way from the other seven, and either a completely new recognition mechanism has been superimposed or intron excision occurs in multiple steps, one of which may correspond to the standard model. Second, they are still spliced in the standard way, the extra DNA contributing additional secondary structure to correct for its own presence.

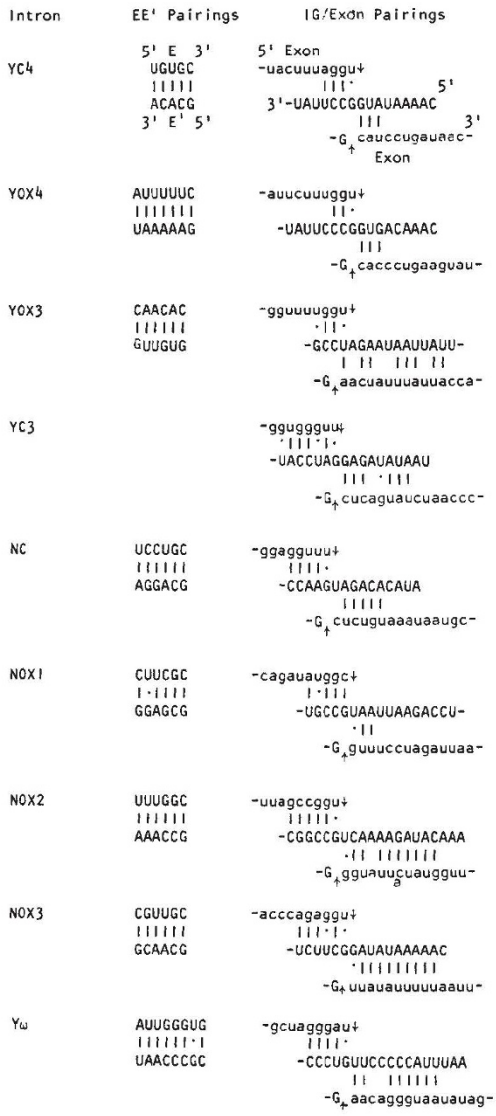

Fig. 6 Base pairing between the internal guide sequences and exon sequences adjacent to the 5′ and 3′ splice sites, and base pairing between E and E′ in nine fungal mitochondrial introns. Intron bases are in capitals, exon bases in lower-case lettering. The precise alignment structures are shown for convenience as if both the 5′ and 3′ cuts have been made. The intron bases that were in the L1 loop (Fig. 3*a*) are shown as a linear array in the middle line with the first (5′) base of the intron on the right, except where a dash indicates that a large number of bases are present 5′ to the sequence shown. Arrows indicate the cut sites. The last (3′) base of the intron is shown to the left of the 3′ cut site because this G is a conserved feature of the precise alignment structure. The one or two IGS bases that can pair either in P1 or P10 (Fig. 3*a*) are thus shown in their P10 pairing mode.

Proteins and the splicing reaction

The model deals with the role of RNA sequences in the splicing process, and not with the role of proteins or the mechanism of the reaction itself. It demonstrates, however, that proteins are probably involved in stabilizing RNA secondary structure as well as in the cutting and ligation steps. Splicing proteins may also need to recognize parts of the secondary structure instead of, or as well as, sequences near the splice junctions.

The form of the precise-alignment structure has consequences for the mechanism of the splicing reaction. IGS bases have two alternative pairings as shown in Fig. 3*a*. Once the core structure, perhaps aided by proteins, allows the P10 pairings to form, the two forms of the structure must exist in an equilibrium due to branch migration unless one is favoured by the local environment. If the fully formed P10 pairing is present for a reasonable fraction of the time, or is stabilized by proteins, then simultaneous cuts could be made at the 5′ and 3′ cut sites in a structure which has both the cut sites at the junction of a very short (half-turn) helical segment with single-stranded RNA, and separated only by the length of a phosphodiester bond.

It is important to make the two cuts spatially and temporally close together so as to reduce the error rate of the reaction, and this can be achieved if recognition and cutting of the 5′ and 3′ ends only occur efficiently once the whole structure has formed. Given this proviso, a sequential mechanism is possible with a nick first occurring at the 5′ scission site in the top strand of the full P1 pairing of Fig. 3*a*; the last base pair or two of P1 then dissociate or are displaced by P10 branch migration, and the 3′ cut is made. Even a fully sequential mechanism is possible if P1 is stabilized by proteins after the 5′ cut has been made. In certain nuclear mutant[17] and *cis*-acting mutant[18] strains of yeast, the 5′ cut can occasionally be made independently of the 3′ cut.

A strong body of evidence[13] supports the involvement of intron-coded mRNA maturases in the excision of some of these introns. These could be important at any stage in the splicing process, including the endonuclease reaction. One specific problem of mitochondrial RNA splicing is that RNA synthesis and splicing and protein synthesis occur in the same compartment in mitochondria, so that the RNA being spliced is simultaneously being translated. The hypothesis that ribosomes have a positive role in splicing[9,13] has received a little experimental support[19]. Ribosomes are likely to disrupt any secondary structure that does form, and in most of these introns ribosomes must cross the 5′ splice junctions to synthesize the maturase. A positive role for ribosomes within the framework of our model might be to inhibit premature 5′-end cutting by disrupting the P1 pairing, although there are alternative ways of achieving this. However, ribosomes travelling into the intron will tend to disrupt the core structure. We have assumed that ribosomes are sufficiently well spaced to allow the splicing complex to form (ribosomes within loops should have little effect). If ribosomes are closely spaced, one of the roles of proteins binding to the core structure may be to prevent the disruption of the structure by ribosomes. In this case it is possible that the maturase is involved in stabilizing the structure against disruption by ribosomes, as well as acting as at least a specificity factor in the splicing reaction.

Conclusions

The class of fungal mitochondrial introns which we have considered possess, according to our proposal, all the specific elements required for their own excision. The model presented here accounts for all the data at present pertaining to the precise excision of these introns, and makes clear predictions of classes of mutations that should be found as well as suggesting the

possible types of activity of proteins involved in splicing. In this model there can also be no ambiguity in the splicing of multiple introns within one RNA precursor molecule, as the dependence of the formation of the splicing complex on previous folding of the intron RNA ensures that the 5′ and 3′ ends of the same intron are always chosen for the excision reaction.

This mechanism of RNA splicing may be of more general importance, because at least one nuclear intron, that in the gene for the large nuclear rRNA of *Tetrahymena*, fits this model precisely (R.B.W. *et al.*, in preparation).

We thank Drs J. Lazowska and P. P. Slonimski for communicating some of their data on YC3 before publication, and Dr B. Dujon for pointing out that Yω had certain features of this class of intron. We acknowledge the help of many workers in the field, but particularly Professor Slonimski and his group whose precise genetic analysis of mitochondrial RNA splicing provided essential evidence. This work was supported by MRC Hproject grant G7904290CB to R.W.D. and C.S.

Note added in proof: One prediction of the model has already been fulfilled: Dr P. S. Perlman and co-workers (personal communication) have obtained a mutation in E that blocks splicing of YC4. Dr R. Schweyen and co-workers (personal communication) have obtained the same base change in E as a revertant of a *box*9 mutation in E′, and the double mutant has regained an intact pairing region. Computer-aided analysis of intron RNAs of *S. cerevisiae* and *Kluyveromyces thermotolerans*[20] has provided RNA structures in excellent agreement with the consensus core structure presented here.

Received 23 June; accepted 1 November 1982.

1. Borst, P. & Grivell, L. A. *Nature* **285,** 439–440 (1981).
2. Waring, R. B. *et al. Cell* **27,** 4–11 (1981).
3. Davies, R. W. *et al.* in *Mitochondrial Genes* (eds Borst, P., Attardi, G. & Slonimski, P. P.) 405–410 (Cold Spring Harbor Laboratory, New York, 1982).
4. Fox, T. & Leaver, C. J. *Cell* **26,** 315–32 (1981).
5. Waring, R. B., Davies, R. W., Scazzocchio, C. & Brown, T. A. *Proc. natn. Acad. Sci. U.S.A.* **79,** 6332–6336 (1982).
6. Bonitz, S. G. *et al. J. biol. Chem.* **225,** 11922–11926 (1980).
7. Nobrega, F. G. & Tzagoloff, A. *J. biol. Chem.* **255,** 9829–9837 (1980).
8. Dujon, B. *Cell* **20,** 185–197 (1980).
9. De La Salle, H., Jacq, C. & Slonimski, P. P. *Cell* **28,** 721–732 (1982).
10. Lamouroux, A. *et al.* in *The Organisation and Expression of the Mitochondrial Genome* (eds Kroon, A. M. & Saccone, C.) 153–155 (Elsevier, Amsterdam, 1980).
11. Lazowska, J., Jacq, C. & Slonimski, P. P. *Cell* **22,** 333–348 (1980).
12. Netter, P., Jacq, C., Carignani, G. & Slonimski, P. P. *Cell* **28,** 733–738 (1982).
13. Jacq, C. *et al.* in *Mitochondrial Genes* (eds Borst, P., Attardi, G. & Slominski, P. P.) 155–183 (Cold Spring Harbor Laboratory, New York, 1982).
14. Rogers, J. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **77,** 1877–1879 (1980).
15. Lerner, M. R. *et al. Nature* **283,** 220–224 (1980).
16. Bos, J. L. *et al. Cell* **20,** 207–214 (1980).
17. Dieckman, C. L., Pape, L. K. & Tzagoloff, A. *Proc. natn. Acad. Sci. U.S.A.* **79,** 1805–1809 (1982).
18. Lamb, M. R. *et al. J. biol. Chem.* (in the press).
19. Schmelzer, C. & Schweyen, R. J. *Nucleic Acids Res.* **10,** 513–524 (1982).
20. Michel, F., Jacquier, A. & Dujon, B. *Biochemie* (in the press).
21. Lazowska, J., Jacq, C. & Slonimski, P. P. *Cell* **27,** 12–14 (1981).

Characterization of a regulatory region upstream of the *ADR2* locus of *S. cerevisiae*

David R. Beier & Elton T. Young

Department of Biochemistry SJ-70, University of Washington, Seattle, Washington 98195, USA

We have used in vitro *mutagenesis to identify a DNA sequence that is required for glucose regulation of the* ADR2 *locus of* Saccharomyces cerevisiae. *This region, located between 200 and 1,000 base pairs upstream of the structural gene for ADH II, has been identified by genetic and biochemical analysis as the site of* ADR3, *a* cis*-acting regulatory locus. If this region is deleted,* ADR2 *is no longer repressed by glucose. Moreover, if the sequence is excised and ligated in front of the* ADC1 *gene, it puts that gene under glucose control.*

THE eukaryotic promoter has been identified by site-specific mutagenesis and deletion analysis of cloned genes[1–6]. In several cases it has been possible to identify regions which have regulatory functions[7–10], and occasionally, DNA sequences immediately 5′ to the transcription initiation site have been implicated in regulating the expression of the adjacent structural gene. In other cases regulatory DNA sequences have been identified at a considerable distance from the affected gene[11–14].

We have chosen to examine the regulatory regions which affect the expression of the alcohol dehydrogenase (ADH) isozymes of the yeast, *Saccharomyces cerevisiae*. The 'classical' ADH I is synthesized during fermentative growth and is repressed about fivefold during growth on ethanol-containing medium[15]. A second isozyme, ADH II, is repressed when yeast are grown on glucose and is derepressed 100-fold when yeast are grown on a nonfermentable carbon source[16]. The structural genes coding for both ADH I (*ADC1*) and ADH II (*ADR2*) have been identified genetically[18], cloned[19,20] and their DNA sequences determined[21,22].

We show here that deletion of a specific 5′-flanking sequence of *ADR2* results in the loss of glucose control of ADH II synthesis. This sequence overlaps *ADR3*, a genetically defined *cis*-acting locus which regulates *ADR2*[23]. In addition, we show that these sequences can be removed from their normal context in front of *ADR2*, placed upstream from the promoter of a different gene (*ADC1* in this case), and still retain their function.

Construction of an *ADR2* deletion

There is extensive nucleotide sequence homology in the region between the TATAA boxes and the transcription initiation sites of *ADC1* and *ADR2*[22]. The presence of an *Sph*I restriction site immediately 5′ to this homologous region in *ADR2* allowed us to test the hypothesis that regulatory sequences would be 5′ to the *ADR2* TATAA box. If this idea were correct, a deletion of sequences 5′ to the *Sph*I site would prevent glucose repression of *ADR2*. To minimize gene dosage effects, the *ADR2* gene constructs were inserted into pBC3T1, a stable, single-copy vector containing a yeast centromere from chromosome 3 (*CEN3*)[24] and the *TRP1* gene[17] (see Fig. 1). A 2.6-kilobase (kb) *Bam*HI/*Sph*I fragment containing *ADR2* and 1.2 kb of 5′-flanking sequence was cloned into pBC3T1 and designated pADR2. A second DNA fragment containing *ADR2* and only 176 base pairs (bp) of 5′ sequence was inserted into the *Sph*I site of pBC3T1. Clones with the fragment inserted in both orientations [pADR2(−176)R and pADR2(−176)L] were isolated.

These plasmids were used to transform the trp$^-$ yeast strain 302-21 which contains defective alleles of all three ADH structural genes: *adc1-11*, *adr2-43* and *adm* (the last gene codes for a minor ADH isozyme found in the mitochondria). Single TRP$^+$ colonies were isolated and grown on medium containing either glucose or ethanol. Cell extracts were prepared and assayed for ADH enzyme activity. As the results given in Table

0028–0836/82/510724—05$01.00