A Baroque Turn for Intron Processing

Once way out in left field in the world of intron splicing, research on mitochondrial genes is beginning to get more of the play

Ever since the discovery in 1977 that some eukaryotic genes are mosaics of coding and noncoding regions there has been intense interest in the mechanism by which intact mature RNA is formed from precursor molecules. In other words, how are the noncoding sequences, or introns, excised and the remaining coding sequences, or exons, ligated in an orderly fashion?

By the beginning of this year, it was already clear that no single mechanism was applicable to all types of RNA. Precursors of messenger RNA, transfer RNA, and ribosomal RNA might, it seemed, be processed by three different systems. Moreover, the split genes of mitochondria appeared to be a class distinct from those in the nucleus, insofar as removal of noncoding regions was concerned. A simple picture of RNA processing had therefore become relatively complex.

A flurry of results, some just published and others still in press, reveals the pattern to be yet more baroque. While most of the ramifications remain to be fully explored, three striking results emerge. First, higher order structure in the introns of certain genes is crucial to the processing of those gene transcripts. Second, there is a striking, and unexpected, similarity between mitochondrial and certain nuclear introns. Third, echoes of transposability are apparent among certain introns, perhaps suggesting an origin as true transposable elements.

Consider, for example, a paper in the current issue of *Cell* on the intron in the gene for the large ribosomal RNA of *Neurospora crassa* mitochondria. Uttam RajBhandary and John Burke of the Massachusetts Institute of Technology were interested in the intron because it promised to display certain intriguing properties. The most prominent of these was that, like the intron of the equivalent gene in yeast (*Saccharomyces cerevisiae*), which was sequenced by Bernard Dujon, there was a good possibility that part of it would code for a protein.

Now, the nomenclature applied to introns—or intervening sequences or noncoding sequences, depending on one's preference—has always been a matter of contention. So, when a couple of years ago a French group discovered that certain introns in yeast mitochondria appeared to code for proteins, the simple descriptive distinction based on coding and noncoding sequences broke down. Introns can code for proteins, but their putative function appeared to be limited to controlling the splicing of the intron from which they derived or even the splicing of introns in different genes. The proteins have been termed "maturases."

Since Piotr Slonimski and his colleagues in Gif sur Yvette, France, made their unexpected finding, another nine introns of yeast mitochondria have been added to the list of those with a potential to code for proteins, presumably maturases. Indeed, the possession of an open reading frame, which might code for a protein, has provided a clear demarcation between mitochondrial and nuclear introns. Although only some mitochondrial introns have open reading frames, no nuclear intron has yet been reported to have one.

> There is a striking, and unexpected, similarity between mitochondrial and certain nuclear introns.

The large ribosomal RNA mitochondrial intron of *N. crassa* did indeed turn out to have a long open reading frame, a 1278 nucleotide section out of a total of 2295. An important distinction between this one and the putative maturase coding regions is that in the *N. crassa* intron the open reading frame is centrally located rather than being contiguous with one of the flanking exons, a pattern that is also found in the yeast ribosomal intron. Moreover, there is good suggestive evidence that the protein encoded by the intron's open reading frame is not a maturase at all, but instead is a component, S5, of the small ribosomal unit.

RajBhandary and Burke note that, in common with most coding sequences in the *N. crassa* mitochondrial genome, the open reading frame in the large ribosomal RNA intron is flanked by short, highly conserved, pallindromic sequences. This, they suggest, is evidence that the open reading frame is indeed translated. Once the intron is excised from the precursor RNA it is unusually stable, another observation indicating that the intron RNA has some function. The MIT workers have sequenced the intron and find that the molecular weight and amino acid composition inferred for the product of the open reading frame matches closely that of protein S5. RajBhandary and Burke await the amino acid sequence data for S5, to be determined in collaboration with Alan Lambowitz of St. Louis University Medical School, for final confirmation that it is encoded by the intron.

One might have anticipated that the open reading frame in the intron of the equivalent gene in yeast would also code for a ribosomal protein. But, observe RajBhandary and Burke, this is not the case. The equivalent ribosomal protein in yeast, denoted var1, is coded for by a free-standing gene. Although there is substantial divergence between S5 and var1, they are of roughly similar size and there are short segments of strong homology in isolated regions of the DNA sequences. Moreover, some of the codon choice in the var1 gene is reminiscent of that in intron encoded genes. The significance of these echoes of similarity is not yet clear, but conceivably there may have been an ancient transposition.

Other indications of apparent mobility of mitochondrial introns come from a comparison of the large ribosomal RNA gene intron of *N. crassa* with its counterpart in yeast. When RajBhandary and Burke began their search for the *Neurospora* intron in collaboration with Dujon and Joyce Heckman, they used sequences of the yeast intron as a probe. To their surprise they found that the probe hybridized only weakly with the ribosomal gene in *Neurospora* and yet strongly with sequences elsewhere in the genome. The introns in the ribosomal genes of the two organisms, although located at precisely the same site within the coding sequences, were clearly different. And yet, sequences akin to the yeast intron were to be found elsewhere in the *Neurospora* mitochondrial DNA.

RajBhandary and Burke suggest that one possible interpretation of these data is that intron sequences in an ancestor of

0036-8075/82/1224-1293$01.00/0

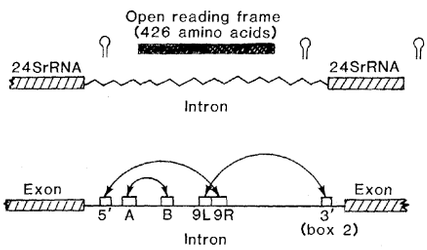
these two fungi have become rearranged over evolutionary time. They also note that in the slime mold *Physarum polycephalum*, one of the two introns in the large ribosomal RNA gene in the nucleus shares the same location as the introns in *N. crassa* and yeast.

Meanwhile, it is worth noting that in yeast there are two pairs of introns with reading frames that are closely related to each other in sequence: one pair has a 70 percent homology, the other has 50 percent. This, says Philip Perlman of Ohio State University, is suggestive evidence of an ancient transposition of the intron sequences. Further support for the mobility of introns comes from work by Slonimski's group on splicing-defective mutants of yeast. This group found that a revertant of a mutant of the cytochrome b gene that was previously blocked in splicing at intron 4 had lost the defective intron entirely. In addition, the next intron downstream had also disappeared.

When RajBhandary and Burke looked for the region in the *N. crassa* large ribosomal RNA gene that had hybridized weakly with the yeast probe they found a 57-nucleotide block of strong homology. And within this block was a 16-nucleotide segment that was closely conserved. The MIT workers searched through published sequences of other introns and found that of 14 from mitochondrial genes, 12 contained the sequence or something close to it. (Eight of these were from cytochrome oxidase and cytochrome b genes in *S. cerevisiae*, one from the cytochrome b gene of *Aspergillus nidulans*, one from the cytochrome oxidase gene of *Zea mays*, and two from ribosomal RNA genes of *N. crassa* and *S. cerevisiae*.) Surprisingly, four more examples came from the only nuclear ribosomal RNA introns so far sequenced (one each from *Tetrahymena thermophila* and *T. pigmentosa* and two from *Physarum polycephalum*).

Overall, RajBhandary and Burke established a consensus sequence 16 nucleotides long which, they infer, is probably important as a splicing signal in the processing of introns containing it. Their inference is based principally on a coincidence with work in a number of other laboratories on splicing-defective mutants. For instance, Slonimski's group has shown that point mutations in this region can block splicing. Perlman's group, in collaboration with that of Henry Mahler, at Indiana University, has similar results.

Indeed, analysis of splicing-defective point mutants in yeast mitochondrial introns has been especially useful of late in identifying apparently functional areas



Top shows the N. crassa *large ribosomal RNA gene, with hairpins denoting pallindromic regions. Bottom shows potential interactions between intron sequences.*

within introns. This approach has so far revealed four important regions in intron 4 of the mitochondrial cytochrome b gene: there is a short sequence in the 5′ region of the intron; a section in the middle, denoted box 9, which is in fact two adjacent functional elements together; and a sequence in the 3′ region of the intron, denoted box 2. Box 9, whose separate components are often referred to as box 9L and box 9R, is congruent with RajBhandary and Burke's consensus sequence.

One particularly striking fact to emerge from this work on mutants is that, in contrast with introns from nuclear protein genes, mitochondrial introns have sequences important to splicing in their internal regions, not at their boundaries. The boundaries of mitochondrial introns are not particularly conserved, again in contrast with the well-established consensus sequences at the splice junctions of nuclear protein genes.

The number of conserved segments that are probably important to splicing in mitochondrial introns was recently increased by two by the publication in the *Proceedings of the National Academy of Sciences U.S.A.* of new data on *Aspergillus nidulans*. Wayne Davies and his colleagues at the University of Manchester Institute of Science and Technology, England, found two separate sequences, each ten nucleotides long, conserved in those mitochondrial introns of *A. nidulans* and *S. cerevisiae* that have the other four elements. These are located between the 5′ sequence and box 9 and are denoted A and B. The Manchester group used DNA sequencing, not mutants, in their study.

As Perlman notes, the interesting thing about this group of six elements—the 5′ sequence, sequences A and B, box 9L, box 9R, and box 2—is not so much that they are conserved (either in sequence or position) but that each can pair specifically with one other. To wit: box 9L can form a 5 base pair stem with box 2, box 9R with 5′, and A with B. One obvious consequence of these interactions is that the intron would fold into an ordered configuration involving secondary and tertiary components, a fact remarked upon by several workers in the field but first put into print by Davies in his paper in the *Proceedings of the National Academy of Sciences U.S.A.* Perhaps higher order structures formed in this way are central to the proper alignment of intron boundaries prior to splicing?

In the case of the *Neurospora* large ribosomal RNA intron, and possibly in other split ribosomal genes too, it appears possible that assembly of the precursor RNA into a ribosomal subunit might normally precede splicing. Lambowitz and his colleagues described, at the RNA processing meeting at Cold Spring Harbor earlier this year, data on protein-binding experiments with this RNA. It seems that interactions between the RNA and proteins in the assembled unit serve to stabilize the alignment of the intron boundaries in a position favorable for splicing. The folding of the RNA in the intron, as well as that within the ribosomal unit, is probably important in setting up the proper splicing substrate.

Slonimski and his colleagues, in a paper earlier this year in *Cell*, touched on the possible role of higher order structure in the splicing of introns in yeast cytochrome b genes. They remarked on a complementarity of sequence between box 9 and a segment of the small ribosomal RNA sequence. As box 9 is frequently, but not always, in the middle of an open reading frame, it is possible, they say, that, as the ribosome passes along the intron, presumably translating a maturase sequence, it would stutter at box 9 because of the complementarity. This interruption in the mobility of the ribosome might effect a higher order folding of the intron, they postulate. This scheme cannot apply to introns that lack reading frames.

In any case, one other feature of box 9 and its fellow conserved regions is worth noting. Although each can pair faithfully with one other, the interaction usually involves only half of the conserved sequence. For what role is the rest of the sequence conserved? Perlman speculates that in addition to the folding of the naked RNA, there might be important binding with, say, proteins, which recognize the sequences adjacent to the stems.

Until recently mitochondrial introns appeared to be in a class, or rather classes, of their own. Their possession, in some cases at least, of open reading frames, of various conserved segments and the lack of splice junction consensus sequences set them apart from nuclear

introns. Then, just a few weeks ago, the distinction between nuclear and mitochondrial introns was suddenly breached.

Tom Cech and his colleagues at the University of Colorado published their latest information on a remarkable intron in the nuclear ribosomal RNA gene of *Tetrahymena thermophila*. The intron, just 413 nucleotides long, is excised and the remaining exons ligated in the absence of enzymes. The splicing reaction appeared to be mediated by catalytic activity inherent to the RNA. Perlman, while sitting in on a graduate seminar on the self-splicing intron, noticed that the intron contained the box 9L sequence. Intrigued with this discovery, he looked for box 2, which he quickly found. Further scrutiny revealed the 5′ and box 9R sequences too. And, over the next few days, careful study showed up the A and B sequences. Burke independently noticed the coincidence of box 9 between mitochondrial introns and the ribosomal intron of *T. thermophila*. That made six out of six signatures of mitochondrial introns in this rather short nuclear intron, a surprising finding indeed.

The coincidence in the conserved sequences between the long mitochondrial introns, many of which have open reading frames for maturases, and the short nuclear self-splicing intron, which has no substantial reading frame, has many possible ramifications. Could it be that, like the nuclear intron from *Tetrahymena*, the mitochondrial introns are self-splicing? This seems unlikely in view of the data on maturases, but splicing-enzyme involvement has yet to be conclusively demonstrated. Perlman wonders whether mitochondrial introns might once have been self-splicing but that, as they acquired more and more sequences through insertions, they came under selection pressure to evolve maturase activity. At the very least there is the implication that normal self-splicing of the *Tetrahymena* intron might require the binding of proteins to parts of the higher order structure induced by the interaction between the conserved sequences. This might explain why self-splicing is slower in the total absence of protein than when it occurs in the cell.

The *Physarum* nuclear ribosomal RNA introns have four of the six functionally important elements: they lack sequences A and B. Examination of nuclear introns from other lower eukaryotes promises to turn up other examples like *Physarum* and *Tetrahymena*. And the discovery of new conserved sequences involved in splicing can be expected too. In any case, as Perlman points out, there could have been relatively recent exchange between nuclear and mitochondrial introns. Whether this involved the wholesale movement of genes between the compartments, or just the intervening sequences acting as transposable elements, remains to be established.

The probability that higher order structure is important in the splicing of mitochondrial, and some nuclear, introns now seems to have moved close to certainty. As yeast mitochondria offer an extremely suitable experimental system in which to isolate splicing-defective mutants, this, plus the study of revertants, promises a fruitful avenue of inquiry in which to pin down further the structures involved.—**ROGER LEWIN**

Additional Reading

1. John M. Burke and Uttam RajBhandary, *Cell* **31**, 509 (1982).
2. Richard Waring, R. Wayne Davies, Claudio Scazzocchio, Terence A. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6332 (1982).
3. Henri De La Salle, Claude Jacq, Piotr P. Slonimski, *Cell* **28**, 721 (1982).
4. Paul Q. Anziano, Deborah K. Hanson, Henry R. Mahler, Philip S. Perlman, *ibid.* **30**, 925 (1982).

Fetal Hemoglobin Genes Turned On in Adults

Scientists used a cancer drug to turn on fetal hemoglobin genes and thereby corrected anemia in patients with thalassemia and sickle cell anemia

In an experiment that has been described as bringing molecular biology to the bedside, researchers have used a drug to unmask fetal hemoglobin genes that normally are suppressed before birth. As a consequence, they have been able to partially correct severe anemia in patients with β thalassemia and sickle cell anemia. Although these results are a natural culmination of a large body of recent research in molecular biology and genetics, the scientists themselves express amazement that the experiments worked so dramatically well in humans.

Over the past several years, it was learned that patients with the genetic disease β thalassemia, like those with sickle cell anemia, have mutations in β globin genes. Adult hemoglobin consists of two β globin subunits and two α globin subunits. But during fetal life a different but equally effective hemoglobin is made in which two subunits from a gene called γ replace the β globin subunits. Fetal hemoglobin, then, consists of two α globin subunits and two γ globin subunits. It was known that fetal hemoglobin is perfectly effective in adults since patients with a rare genetic disorder that causes them to produce only fetal hemoglobin throughout their lives appear quite normal and healthy. So, the investigators reasoned, if they could just turn on γ globin genes in patients with β thalassemia or sickle cell anemia they could substitute for the defective β globin genes.

The next thing that was learned is that the γ globin DNA seems to be chemically modified when these genes are turned off in adult life. Gary Felsenfeld of the National Institutes of Health and other molecular biologists discovered that active genes frequently have few methyl groups attached to them. In contrast, genes that are not being expressed often are covered with methyl groups. About 2 years ago Richard Flavell and L. H. T. van der Ploeg, then at the University of Amsterdam, reported that DNA in the region of the γ globin genes is undermethylated during fetal life and is methylated in adult life. The clue to turning on γ globin genes, then, might be to remove methyl groups from them.

At about the same time as these methylation discoveries were being made, several groups of researchers, including Mark Groudin and Harold Weintraub of the Fred Hutchinson Cancer Research Center in Seattle, found that 5-azacytidine, a drug used to treat leukemia patients, might do the trick. This drug is an analog of cytidine, one of the DNA bases, and is incorporated into newly synthesized DNA. When 5-azacytidine is supplied to cells in tissue culture, newly synthesized DNA is undermethylated and repressed genes sometimes become active.

A year ago Joseph DeSimone, Paul Heller, and their colleagues at the Uni-

0036-8075/82/1224-1295$01.00/0