Introns in chloroplast protein-coding genes of land plants

AINE L. PLANT & JOHN C. GRAY
Botany School, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

Received 14 September 1987; accepted 1 December 1987

Key words: intron, splice, exon, lariat

Abstract. Several protein-coding genes from land plant chloroplasts have been shown to contain introns. The majority of these introns resemble the fungal mitochondrial group II introns due to considerable nucleotide sequence homology at their 5' and 3' ends and they can readily be folded to form six hairpins characteristic of the predicted secondary structure of the mitochondrial group II introns. Recently it has been demonstrated that some mitochondrial group II introns are capable of self-splicing in vitro in the absence of protein co-factors. However evidence presented in this overview suggests that this is probably not the case for chloroplast introns and that trans-acting factors are almost certainly involved in their processing reactions.

Abbreviations: kbp—kilobase pairs, ORF—Open Reading Frame, pre-RNA—precursor ribonucleic acid

Introduction

Introns were first discovered in the ovalbumin and β-globin genes of chickens and mammals respectively (Breathnach et al. 1977, Jeffreys and Flavell 1977). They are DNA sequences within genes, that interrupt and thereby split the protein-coding regions (exons). Introns are transcribed together with the exons into precursor RNA molecules and are subsequently removed in a series of reactions leading to splicing of the exons to produce the mature transcript. Based on their nucleotide sequence, introns can be divided into 4 different classes. One class comprises the introns of nuclear mRNA precursors. These include the ovalbumin and β-globin gene introns mentioned above and are characterized by invariant GU and AG dinucleotides at their boundaries (Breathnach et al. 1978). Splicing of these introns requires assembly of pre-RNA into a complex structure involving ribonucleoprotein particles (Brody and Abelson 1985, Frendewey and Keller 1985, Grabowski et al. 1985). Self-splicing introns typified by the Tetrahymena rRNA gene (Kruger et al. 1982) form another class. Similar
introns are also found in fungal mitochondrial genes; these introns have been termed group I introns on the basis of conserved sequence elements (Davies et al. 1982, Michel and Dujon 1983). Another group of self-splicing introns occurring in yeast mitochondrial genes form another class. These introns were termed group II introns and possess a set of conserved sequences distinct from those of group I introns (Michel and Dujon 1983). Finally there are the introns characteristic of nuclear tRNA genes (Abelson 1979).

**Introns in chloroplast protein-coding genes**

The first protein-coding gene from higher plant chloroplasts reported to contain an intron was the *rpL2* gene for the ribosomal protein L2 from *Nicotiana debneyi* (Zurawski et al. 1984). This intron was discovered as a result of a direct comparison between the nucleotide sequences of the *rpL2* genes from *Spinacea oleracea* and *Nicotiana debneyi* and showed that the *Nicotiana debneyi* gene contained a 666 bp insertion which was absent from the *Spinacea oleracea* gene. Prior to this, the discovery of introns in chloroplast genes was confined to tRNA genes (Koch et al. 1981, Deno et al. 1982, Steinmetz et al. 1982, Bonnard et al. 1984, Deno and Sugiura 1984) and those present in the rRNA and protein-coding genes of *Chlamydomonas reinhardii* and *Euglena gracilis* (Rochaix and Malnoe 1978, Erickson et al. 1984, Karabin et al. 1984, Koller et al. 1984). It has been estimated that the *Euglena gracilis* chloroplast genome contains a minimum of 50 introns accounting for approximately 32 kbp of the 145 kbp genome (Koller and Delius 1984); the *rbcL* gene, coding for the large subunit of ribulose 1,5-bisphosphate carboxylase, alone has a total of 9 introns (Koller et al. 1984, Gingrich and Hallick 1985). The abundance of introns in protein-coding genes of *Euglena gracilis* chloroplasts contrasted sharply with their rare occurrence in genes from land plant chloroplasts. However, since the discovery of the intron in *rpL2*, a number of introns have been reported in chloroplast genes (see Table 1). Of these, however, only a few have been shown to have the intron correctly removed with ligation of the exons in vivo. For the *Triticum aestivum atpF* gene it was shown by SI nuclease analysis that the mature RNA has the intron removed at the predicted exon-intron boundaries (Bird et al. 1985). Sequencing of *Spinacea oleracea* chloroplast cDNA clones also revealed that the exons of *atp F* were correctly ligated (Hudson et al. 1987). Similarly, primer extension of *Marchantia polymorpha* chloroplast RNA using exon-specific primers has shown that the *petB*, *petD* and *trans*-spliced *rps12* introns are precisely removed with correct ligation of the exons (Fukuzawa et al. 1987, Zaita et al. 1987). The