The EMBO Journal vol.5 no.11 pp.2749–2758, 1986

Accurate *in vitro* splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract

J.W.S.Brown, G.Feix and D.Frendewey[1]

Institute for Biology III, Albert Ludwigs University, D-7800 Freiburg, and [1]Division of Molecular Biology, Institute of Cell and Tumour Biology, German Cancer Research Centre, D-6900 Heidelberg, FRG

Communicated by W.Keller

Two plant introns along with flanking exon sequences have been isolated from an amylase gene of wheat and a legumin gene of pea and cloned behind the phage SP6 promoter. Pre-mRNAs produced by *in vitro* transcription with SP6 RNA polymerase were tested for their ability to be spliced in a HeLa cell nuclear extract. The plant introns were accurately spliced and the predicted splice junctions were used. Lariat RNAs were observed as both intermediates and final products during the splicing reaction. The branch points were mapped to adenosine residues lying within sequences that showed good homology to the animal branch point consensus. Consensus sequences for the 5′ and 3′ splice junctions and for putative branch point sequences of plants were derived from an analysis of 168 plant intron sequences.

Key words: messenger RNA splicing/splice junction sequences/branch point/lariat/primer extension

Introduction

Interruption of protein coding genes by intervening sequences (IVS, intron) is a feature that is common to all known eukaryotic genomes. Introns are especially prevalent in the genes of higher eukaryotes. The expression of a significant portion of a typical eukaryotic genome therefore requires removal of introns from messenger RNA precursors (pre-mRNAs) by the process of splicing.

The analysis of the biochemical mechanism of pre-mRNA splicing has been greatly facilitated by the development of *in vitro* systems that accurately and efficiently splice exogenously added pre-mRNA substrates. The two systems that have yielded the most information on the mechanism of splicing are based on nuclear extracts from HeLa cells (Hernandez and Keller, 1983; Hardy *et al.*, 1984; Krainer *et al.*, 1984) or whole cell extracts from the yeast *Saccharomyces cerevisiae* (Lin *et al.*, 1985). The initial event in pre-mRNA splicing is the assembly of a large ribonucleoprotein complex on the pre-mRNA. This structure has been termed the spliceosome and varies in size from 40S in yeast (Brody and Abelson, 1985) to 50–60S in the HeLa system (Frendewey and Keller, 1985; Grabowski *et al.*, 1985; Bindereif and Green, 1986; Kaltwasser *et al.*, 1986). The assembly of the splicing complex proceeds in a stepwise fashion in the HeLa extract (Frendewey and Keller, 1985) and requires the U-type small nuclear ribonucleoproteins (snRNPs) (Frendewey and Keller, 1985; Grabowski *et al.*, 1985) and conserved sequence elements found near the splice junctions (Brody and Abelson, 1985; Frendewey and Keller, 1985; Grabowski *et al.*, 1985; Ruskin and Green, 1985b; Vijayraghavan *et al.*, 1986).

After the assembly of the splicing complex the pre-mRNA splicing reaction proceeds by a two-step pathway (Ruskin *et al.*, 1984; Padgett *et al.*, 1984). The first step is cleavage at the 5′ splice site with the concomitant formation of a 5′–2′ phosphodiester bond between the phosphate at the 5′ end of the intron and the 2′-OH of an adenosine residue lying within the intron near near to the 3′ splice site (Padgett *et al.*, 1984; Ruskin *et al.*, 1984; Konarska *et al.*, 1985). This reaction produces a two-molecule intermediate consisting of exon 1 (the 5′ exon) and the intron–exon 2 lariat RNA. The splicing reaction is completed by ligation of exons 1 and 2 with the release of the intron as a lariat RNA in a second coupled cleavage/ligation reaction at the 3′ splice site. This mechanism was initially discovered with HeLa extracts (Ruskin *et al.*, 1984; Padgett *et al.*, 1984), but the same pathway has been demonstrated in the *S. cerevisiae* system (Lin *et al.*, 1985). The two-step pathway has also been confirmed *in vivo* in rat (Zeitlin and Efstradiatis, 1984) and in *S. cerevisiae* (Rodriguez *et al.*, 1984; Domdey *et al.*, 1984).

Following the initial suggestion that the U1 snRNP may mediate pre-mRNA splicing (Lerner *et al.*, 1980; Rogers and Wall, 1980) evidence for the involvement of the U-type snRNPs in pre-mRNA splicing in higher eukaryotes was obtained (Yang *et al.*, 1981; Padgett *et al.*, 1983; Bozzoni *et al.*, 1984; Fradin *et al.*, 1984). The specific requirement of the individual U snRNPs has been demonstrated for U1 (Krämer *et al.*, 1984), U2 (Krainer and Maniatis, 1985; Black *et al.*, 1985) and the U4/U6 snRNP (Black *et al.*, 1986; Frendewey, Krämer and Keller, unpublished results). The U5 snRNP may also be involved in pre-mRNA splicing in higher eukaryotes (Chabot *et al.*, 1985). Small nuclear RNAs have been described in *S. cerevisiae* (Wise *et al.*, 1983) and have recently been shown to participate in splicing *in vitro* (Cheng and Abelson, 1986) and to be components of the yeast spliceosome (Pikielny and Rosbash, 1986).

In contrast to yeast and vertebrates, very little is known about the mechanism of pre-mRNA splicing in plants. Consensus sequences for the 5′ and 3′ splice junctions have been derived from plant introns (Slightom *et al.*, 1983; Rogers, 1985) from 20 and 30 plant intron sequences respectively. The plant consensus is very similar to that of vertebrates (Breathnach and Chambon, 1981; Mount, 1982; Padgett *et al.*, 1986) with the exception that the polypyrimidine stretch found immediately upstream of most vertebrate 3′ splice sites is not as pronounced in plant intron sequences (see Results). Plants also possess snRNAs that are very similar to those found in vertebrates (Krol *et al.*, 1983; Skuzeski and Jendrisak, 1985; Kiss *et al.*, 1985), but their role in pre-mRNA splicing has not been determined. These similarities suggest a pre-mRNA splicing mechanism which would resemble that seen in vertebrates. However, an attempt to splice the introns of the human growth hormone pre-mRNA *in vivo* in plants was not successful (Barta *et al.*, 1986).

As an initial investigation into the mechanism of pre-mRNA splicing in plants, we attempted to splice two different plant introns, one from a monocotyledonous and the other from a dicotyledonous plant, in a HeLa cell nuclear extract. We have found that both of the plant introns are accurately and efficiently

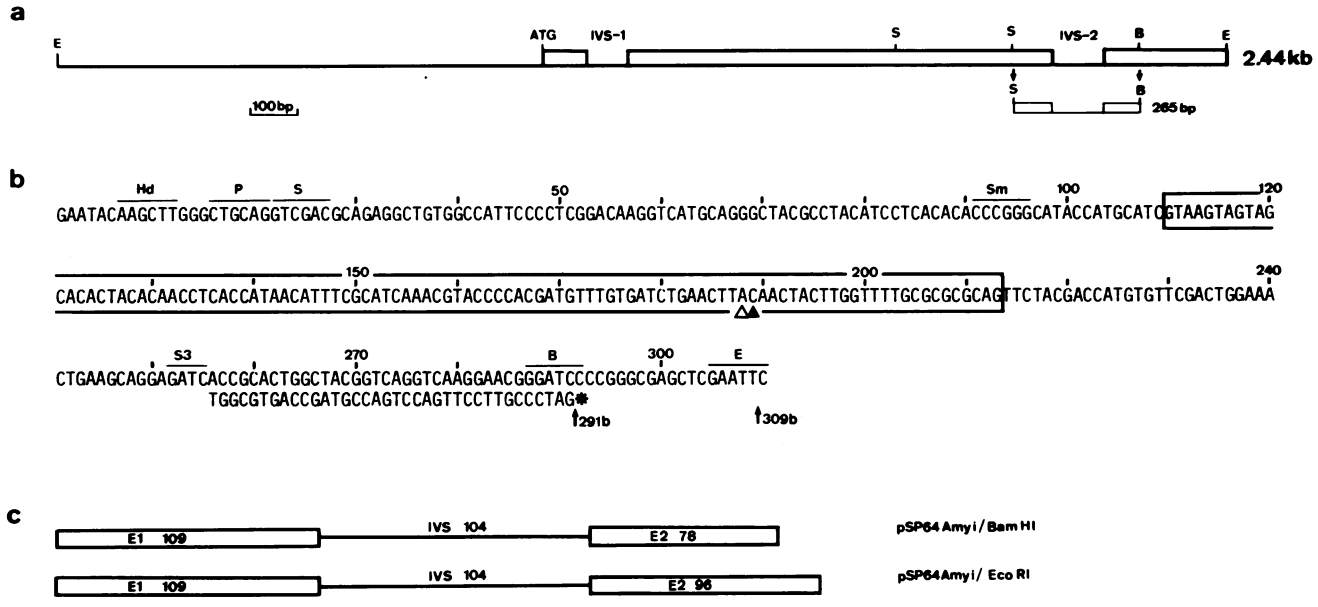

Fig. 1. Cloning and sequence of amylase 33 intron. **(a)** The 2.44-kb *Eco*RI fragment showing part of the genomic clone containing the amylase 33 gene, and the 265-bp *Sal*I–*Bam*HI fragment containing the second intron (IVS-2) which was cloned into pSP64. **(b)** Sequence of the cloned fragment. The first base corresponds to the first base transcribed by the SP6 polymerase. The intron is boxed. The arrows show the positions at which the DNA template were linearized in the *Bam*HI or the *Eco*RI sites and sizes of the corresponding pre-mRNAs are given. The sequence of the 36 base exon-2 specific primer is shown below its corresponding coding strand along with the position of the 5′ end label (*). The closed triangle shows the position of the longest cDNA produced by primer extension with isolated intron–exon 2 lariat RNA and the open triangle corresponds to the assumed branch point nucleotide (Figures 5 and 6). **(c)** Schematic representation of the pre-mRNAs produced from pSP64Amyi linearized with either *Bam*HI or *Eco*RI. Sizes of the first exons (E1), introns (IVS) and second exons (E2) are given in bases. Restriction sites used in cloning, sequencing and primer production are: B, *Bam*HI; E, *Eco*RI; Hd, *Hin*dIII; P, *Pst*I; S, *Sal*I; Sm, *Sma*I; S3, *Sau*3AI.

spliced in the HeLa *in vitro* system. The predicted splice junctions are used and lariat RNAs are formed by selection of the branch point at an adenosine near the 3′ end of the intron.

Results

In vitro splicing of the amylase intron

The introns of the amylase 33 gene of wheat were located by direct comparison of the DNA sequences from isolated cDNA and genomic clones (D.Baulcombe, in preparation). Part of the genomic structure of amylase 33 showing the locations of the two introns is shown in Figure 1a. The second intron (IVS-2) along with flanking exon sequences was cloned behind the phage SP6 promoter by transferring the 265-bp *Sal*I–*Bam*HI fragment into the *Sal*I and *Bam*HI restriction sites of pSP64 (Figure 1a). The DNA sequence of the inserted fragment of the resulting plasmid, pSP64Amyi, is shown in Figure 1b starting at the first nucleotide which would be transcribed by the SP6 polymerase and including the multiple cloning site sequence. As DNA template for *in vitro* transcriptions pSP64Amyi was linearized at either the *Bam*HI site or the *Eco*RI site giving transcripts of 291 and 309 nucleotides (nt) long respectively (Figure 1b and c). These RNA transcripts consisted of a chimaeric first exon (E1) of 109 nt of which the first 22 nt are derived from pSP64 DNA sequences, an intron (IVS) of 104 nt, and second exons (E2) of 78 nt and 96 nt for the *Bam*HI- and *Eco*RI-linearized plasmid respectively (Figure 1b and c).

The products of *in vitro* splicing reactions with pre-mRNA transcripts from pSP64Amyi are shown in Figure 2. In this experiment RNA transcripts from pSP64Amyi linearized with *Bam*HI (lanes 1–4) or *Eco*RI (lanes 5–8) were incubated in *in vitro* splicing reactions for 15 min (lanes 1 and 5), 30 min (lanes 2 and 6), 60 min (lanes 3 and 7) and 105 min (lanes 4 and 8). Following the splicing reactions the products were divided and loaded on a 9% polyacrylamide–urea gel (Figure 2a) and a 12% polyacrylamide–urea gel (Figure 2b). The DNA sequence marker was used as a guide in the identification of the 109 nt first exon (E1) (lanes 1–8) and the exon 1–exon 2 (E1– E2) spliced products (S): 187 nt (lanes 1–4) and 205 nt (lanes 5–8) for the *Bam*HI- and *Eco*RI-linearized templates respectively. RNAs were retarded relative to DNAs of the same length in the gel system used. The E1–E2 spliced products and the IVS– E2 lariat RNAs were also identified by their larger size in the reaction with the pre-mRNA derived from the *Eco*RI-linearized template which had a longer (18 nt) second exon (Figure 2a and b, cf. lanes 1–4 and 5–8). The E1 and IVS lariat RNA bands were the same size for both pre-mRNAs (Figure 2a and b, lanes 1–8). Comparison of lanes 1–4 and lanes 5–8 between Figure 2a and b showed the slower mobility of the lariat forms in the higher percentage gel. In the 9% gel the IVS band ran approximately half way between E1 and the E1–E2 spliced product (S) (Figure 2a) whereas in the 12% gel the IVS band ran in the same region of the gel as the E1–E2 spliced product (Figure 2b). Similarly, in the 12% gel the intron–exon 2 lariat (IVS–E2) ran higher in the gel than the pre-mRNA transcript (R) while in the 9% gel the IVS–E2 lariat ran with the pre-mRNA transcript. With increasing time the amount of E1–E2 spliced product increased while that of the intermediates (E1 and IVS–E2 lariat) increased in the 15, 30 and 60 min reactions (lanes 1–3 and 5–7) but had decreased in the 105 min reactions (lanes 4 and 8). Fainter bands are also visible in the regions of the IVS and IVS–E2 lariats. The nature of these minor products was not investigated further.

In vitro splicing of the legumin J intron

The intron of the legumin J storage protein gene from pea was identified by direct comparison of cDNA and genomic DNA se-

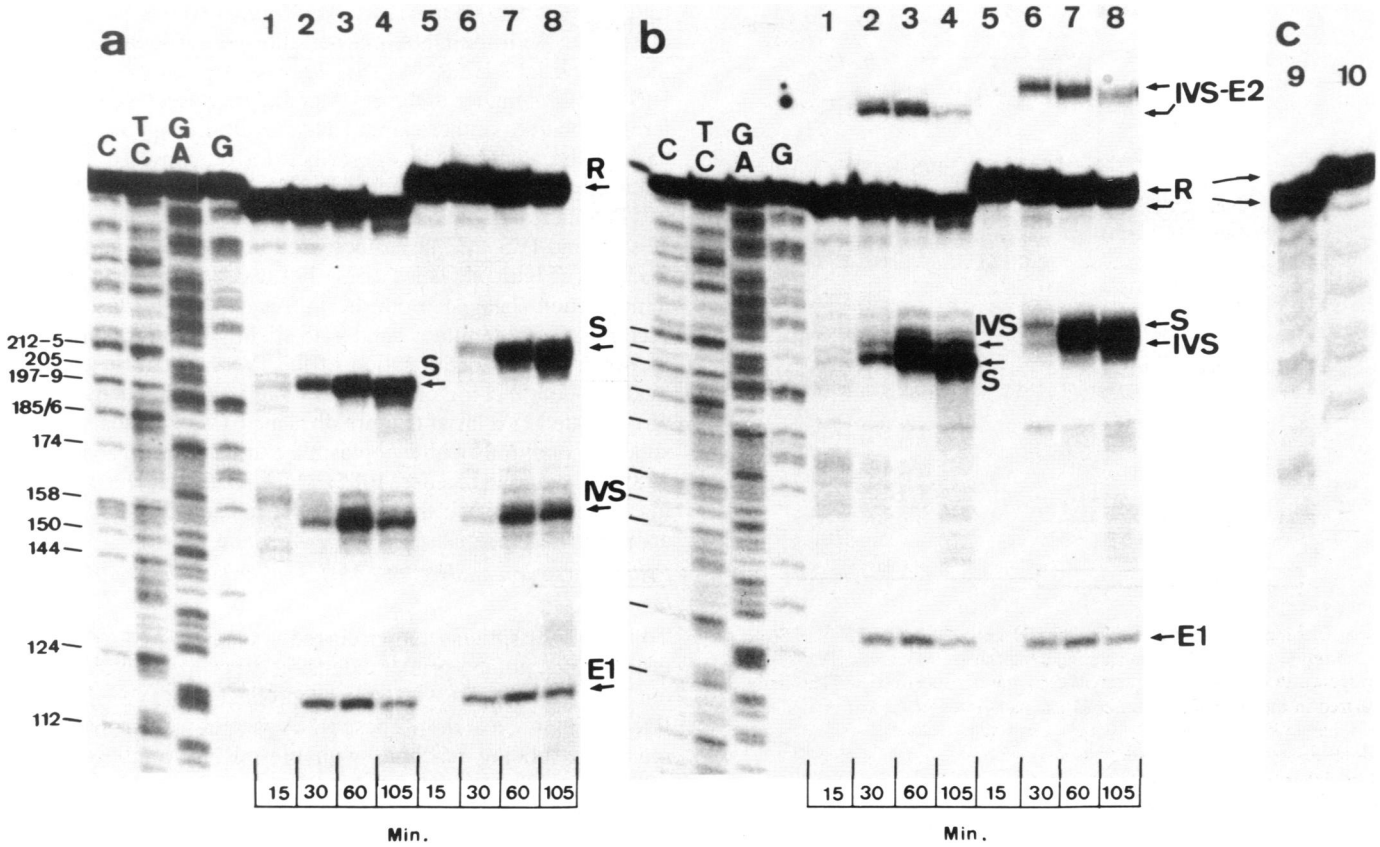

Fig. 2. Products of *in vitro* splicing of pre-mRNAs transcribed from pSP64Amyi. Splicing reactions were carried out on pre-mRNAs from pSP64Amyi linearized with *Bam*HI **(lanes 1–4)** or with *Eco*RI **(lanes 5–8)** for 15, 30, 60 and 105 min, divided, and separated on a 9% **(a)** and 12% **(b)** urea–polyacrylamide gel. The products of splicing of the pre-mRNAs (R) are shown with arrows: exon 1 (E1), intron lariat (IVS), exon 1–exon 2 spliced product (S) and intron–exon 2 lariat (IVS–E2). Marker is the sequenced *Eco*RI–*Pst*I DNA fragment of pSP64Amyi. The lines to the left of the marker sequence in **(b)** correspond to those in **(a)** where the sizes are given. Control splicing reactions without nuclear extract are shown for pre-mRNAs from pSP64Amyi linearized with *Bam*HI **(c, lane 9)** or with *Eco*RI **(c, lane 10)**, run on a 12% urea–polyacrylamide gel.

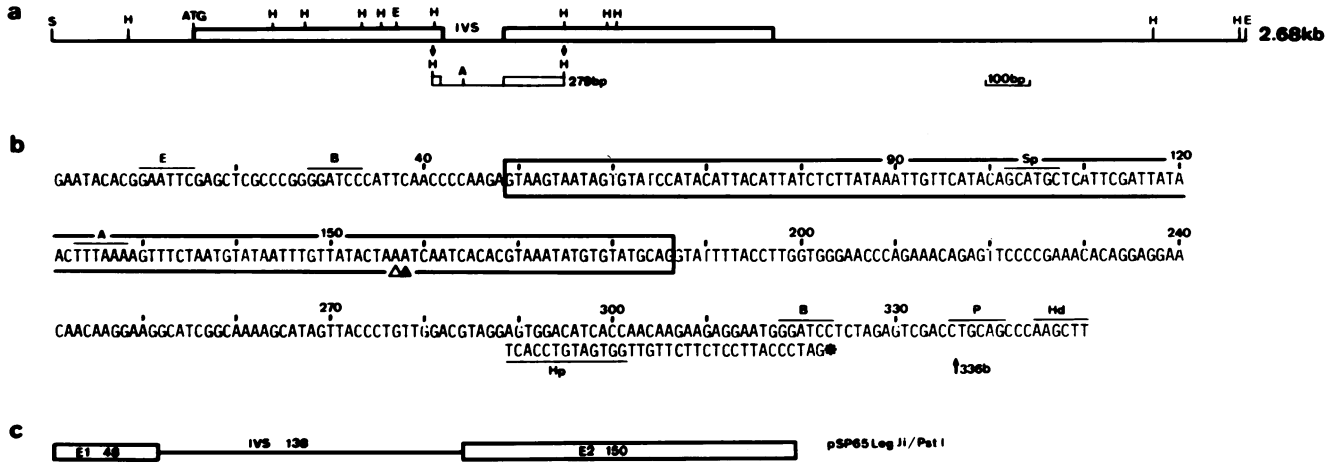

Fig. 3. Cloning and sequence of legumin J intron. **(a)** The 2.68-kb *Sal*I–*Eco*RI fragment showing part of the genomic clone containing the whole legumin J gene. The 279-bp *Hin*fI fragment containing the intron was cloned in pSP65 with *Bam*HI linkers. **(b)** Sequence of the cloned fragment. The first base corresponds to the first base transcribed by the SP6 polymerase. The intron is boxed. The arrow shows the position at which the DNA template is linearized (*Pst*I site) and gives the size of the corresponding pre-mRNA. The sequence of the 34 base exon-2 specific primer is shown below its corresponding coding strand along with the position of the 5′ end label (*). The closed triangle shows the position of the longest cDNA produced by primer extension with isolated intron–exon 2 lariat RNA and the open triangle corresponds to the assumed branch point nucleotide (Figures 5 and 6). **(c)** Schematic representation of the pre-mRNA produced from pSP65LegJi linearized with *Pst*I. Sizes of the first exon (E1), intron (IVS) and exon 2 (E2) are given in bases. Restriction sites used in cloning, sequencing and primer production are: A, *Aha*III; B, *Bam*HI; E, *Eco*RI; H, *Hin*fI; Hd, *Hin*dIII; Hp, *Hph*I; P, *Pst*I; Sp, *Sph*I.

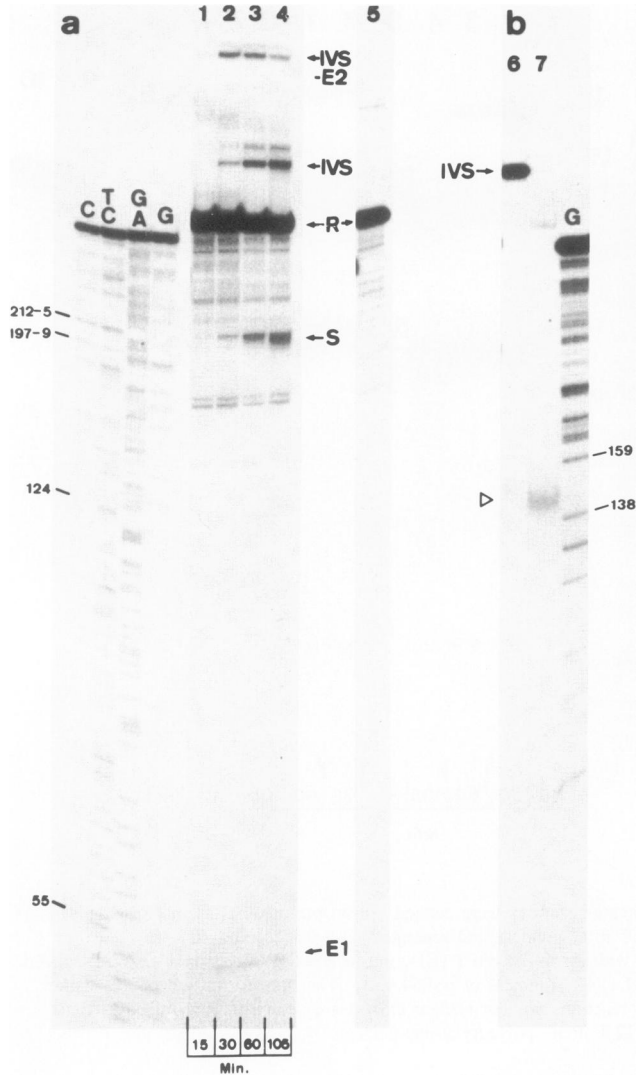

Fig. 4. **(a)** Products of *in vitro* splicing of pre-mRNA transcribed from pSP65LegJi and **(b)** debranching of lariat RNA. **(a)** Splicing reactions were carried out on the pre-mRNA shown in Figure 3c for 15, 30, 60 and 105 min **(lanes 1–4)**, and separated on a 12% urea–polyacrylamide gel. Products of splicing of the pre-mRNA (R) are indicated with arrows: exon 1 (E1), intron lariat (IVS), exon 1–exon 2 spliced product (S), and intron–exon 2 lariat (IVS–E2). Marker is the sequenced *Eco*RI–*Pst*I DNA fragment of pSP64Amyi. A control splicing reaction without nuclear extract is shown in **lane 5**. **(b)** IVS lariat RNA isolated from a scaled-up splicing reaction **(lane 6)** was debranched in a cytoplasmic extract and its new position is indicated by the open triangle **(lane 7)**. Marker is the G reaction of the above DNA marker.

quences (J.Gatehouse *et al.*, in preparation). The intron with flanking exon sequences was cloned behind the phage SP6 promoter by cloning a 279-bp *Hin*fI fragment (Figure 3a) with the aid of *Bam*HI linkers into the *Bam*HI site of pSP65. The DNA sequence of the inserted fragment from the resulting plasmid pSP65LegJi is shown in Figure 3b. The sequence starts at the first transcribed base and includes the multiple cloning site sequences. For *in vitro* transcription with SP6 polymerase pSP65-LegJi was linearized with *Pst*I which produced a pre-mRNA transcript of 336 nt in length (Figure 3b). This RNA consisted of a 48 nt chimaeric first exon (28 nt derived from pSP65 sequences), a 138 nt intron and a 150 nt second exon (Figure 3b and c).

A time course of *in vitro* splicing with the legumin J pre-mRNA transcripts is shown in Figure 4a. Comparison of the products of splicing with the DNA marker allowed the identification of the E1 and E1–E2 RNAs (S). The two lariat forms (IVS and IVS–E2) ran higher in the gel than the pre-mRNA (R). The different mobilities of these lariat RNAs in different percentage gels has also been observed as described for the amylase 33 intron (data not shown). The amounts of E1–E2 spliced product (S) and IVS increased with time while those of the splicing intermediates E1 and IVS–E2 lariat increased in the 15, 30 and 60 min reactions (Figure 4, lanes 1–3) but had decreased in the 105 min reaction (lane 4). Both the IVS and IVS–E2 lariats were accompanied by a minor band with slightly lower mobility than the main bands. Identification of the IVS lariat form was also made by debranching of isolated IVS lariat (main band) (Figure 4b). Isolated IVS lariat (Figure 4b, lane 6) was incubated under splicing conditions with cytoplasmic extract (Ruskin and Green, 1985a; Krämer and Keller, 1985) which resulted in the debranching of the lariat such that it ran as a linear RNA molecule of approximately 138 bases (Figure 4b, lane 7).

Primer extension analysis of E1–E2 spliced product and IVS–E2 lariat

To show that splicing has occurred accurately and to map the lariat branch points, primer extension experiments were carried out with isolated E1–E2 spliced product and IVS–E2 lariat RNAs (Figure 5a and b). These RNA species were isolated from scaled-up splicing reactions with pre-mRNA transcripts from pSP64Amyi linearized with *Eco*RI and pSP65LegJi linearized with *Pst*I. For the IVS–E2 lariats the major band only was isolated. These RNAs were hybridized with the 36-bp (amylase) and 34-bp (legumin) exon-2-specific, 5′ end-labelled primers respectively (Figures 1b and 3b) and primer extension with AMV reverse transcriptase carried out (Figure 5c and d). Primer extension products were loaded onto a gradient sequencing gel alongside sequenced DNA fragments as markers. The whole sequencing gel showing the labelled primers and the primer extension products is shown in Figure 5c and an enlargement showing the primer extension products is given in Figure 5d.

Primer extension with the E1–E2 spliced product from the pSP64Amyi pre-mRNA showed a number of cDNA bands (Figure 5c and d, lane 2) of which the longest ran between nucleotides 188 and 189, which after a correction of 1.5 nt (Green and Roeder, 1980) gives a size of 187 nt. The other major bands had sizes of 186, 146 and 144 nt. Examination of the sequence for the amylase 33 pre-mRNA and application of the GT-AG rule for splice junctions did not produce any possible combinations of splice sites which would give rise to products of 146 and 144 nt. Furthermore, the size of the isolated E1–E2 spliced product band (cf. Figure 2, lane 7) would not allow for the difference of 40 bases between these products and the full-sized E1–E2 spliced product. Thus, these additional primer extension bands probably represent specific degradation of the product RNAs or premature termination of the reverse transcriptase. Primer extension with the E1–E2 spliced product from the legumin J reaction showed a single cDNA band which ran between nucleotides 185 and 186 giving a corrected size of 184 bases (Figure 5d, lane 3). These values for the longest cDNAs, 187 and 184 bases, are the distances between the 5′ end-labelled nucleotide of the primer and the 5′ end of the RNAs and agree exactly with the expected values for accurate removal of the introns.

The primer extension products when the IVS–E2 lariat RNAs are used as templates are shown in Figure 5d (lanes 1 and 4). Due to the nature of the primer and the DNA sequence marker

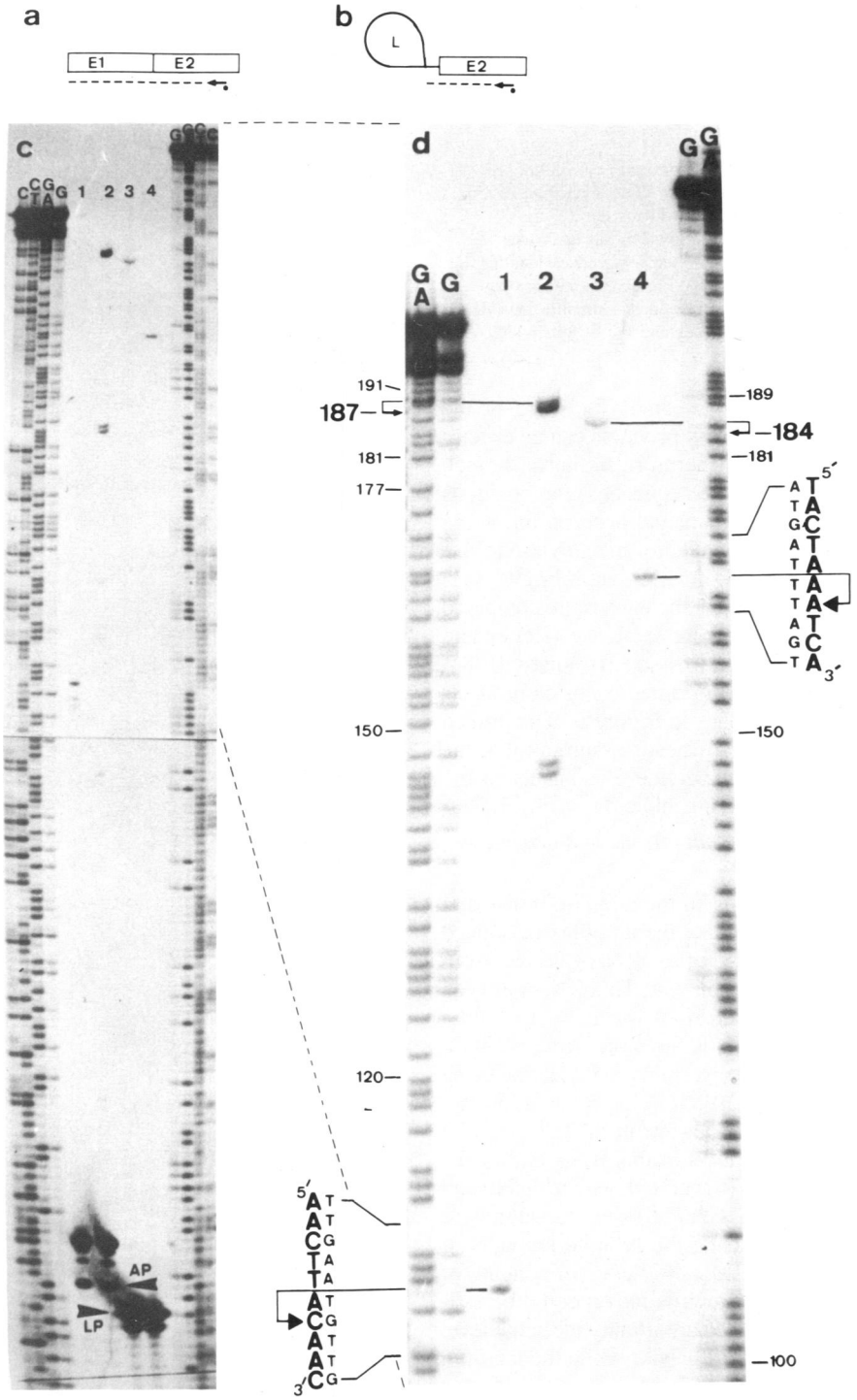

Fig. 5. Primer extension of the E1−E2 spliced products and mapping of the lariat branch points. **(a)** Schematic representation of primer extension using isolated E1−E2 spliced product and end-labelled exon 2-specific primer. The longest cDNA produced gives the exact distance of the 5′ end of the RNA from the end label (*) of the primer (small arrow)..**(b)** Schematic representation of mapping of the lariat branch point using isolated IVS−E2 lariat and exon 2-specific primer. The longest cDNA defines the distance between the end label (*) of the primer (small arrow) and the branch point because the reverse transcriptase will not extend the cDNA past the modified branch point nucleotide. **(c)** 6% gradient sequencing gel showing the products of primer extension using RNAs isolated from scaled up *in vitro* splicing reactions: IVS−E2 lariat RNA **(lane 1)** and E1−E2 spliced product RNA **(lane 2)** (amylase 33) and E1−E2 spliced product RNA **(lane 3)** and IVS−E2 lariat RNA **(lane 4)** (legumin J). The positions of the amylase 33 and legumin J exon 2-specific primers are shown by the arrowheads (AP and LP respectively) at the bottom of **lanes 1−4**. The DNA sequence marker on the left-hand side is that of the *Bam*HI−*Sma*I fragment of pSP64Amyi while that on the right-hand side is the *Bam*HI−*Sph*I fragment of pSP65LegJi. In both cases the exon 2-specific primers and the DNA markers are end-labelled at the same nucleotide (G of the *Bam*HI site). **(d)** Enlargement of **(c)** showing the positions of the primer extension products. The longest cDNAs produced with E1−E2 spliced product from amylase pre-mRNA **(lane 2)** and legumin J pre-mRNA **(lane 3)** are aligned with their corresponding marker sequences with thin lines. After a correction of 1.5 nucleotide positions (small arrows) reflecting the size and charge differences between the sequenced fragments and primer extension products, the sizes of the cDNA products are shown as 187 and 184 for amylase 33 and legumin J respectively. Similarly, primer extension bands obtained with IVS−E2 lariat RNAs for amylase 33 **(lane 1)** and legumin J **(lane 4)** are aligned with their corresponding marker sequences wtih thin lines. The sequence shown on the gel is given in small letters while that of the corresponding coding strand is given in large letters. The primer extension bands are exactly aligned on the sequence following a correction of 1.5 nucleotides (large arrows).

Animal	Y N Y T R A Y . .	18–40 .AG
Amylase	A A C T T A C . .	26 .AG
Legumin	T A C T A A A . .	30 .AG

Fig. 6. Comparison of branch point sequences from amylase 33 and legumin J introns to the animal consensus sequence. The adenosine residue in the animal consensus sequence which normally forms the 2′–5′ phosphodiester linkage to produce the lariat is shown by an open triangle. The closed triangles in the amylase 33 and legumin J sequences identify the positions of the longest primer extension cDNA (Figure 5) and the open triangles indicate the assumed branch point nucleotides. Numbers indicate the distance between the branch point nucleotides and the 3′ splice site. The sequences are given in DNA form.

where both are 5′ end-labelled at the same G residue of the *Bam*HI site the longest primer extension products can be exactly located on the DNA sequence, and therefore the lariat branch points can be accurately mapped. The sequences, the positions of the primer extension bands and the actual position following correction are shown on the left-hand side for the amylase IVS–E2 lariat and on the right-hand side for the legumin IVS–E2 lariat. The sequences in these regions of the introns are compared with the animal consensus branch point sequence (Keller and Noon, 1984; Ruskin *et al.*, 1984; Zeitlin and Efstradiatis, 1984) in Figure 6. The position of the longest primer extension products are also indicated by filled arrowheads in Figure 6. The branch points are indicated by the open arrowheads assuming that the reverse transcriptase stops at the nucleotide 3′ adjacent to the modified branch point nucleotide (Hagenbüchle *et al.*, 1978).

Efficiency of in vitro splicing of the plant introns and dependence on U snRNPs

To estimate the efficiency of splicing of the amylase intron and legumin intron in the HeLa cell extract the amounts of spliced products were compared with those of an RNA derived from the adenovirus 2 major late transcription unit. This RNA has been used extensively to study the mechanism of splicing in the HeLa system (Frendewey and Keller, 1985; Krämer and Keller, 1985). Pre-mRNA was synthesized from the plasmid pSP62Δil linearized with *Sca*I (Frendewey and Keller, 1985). This RNA transcript consisted of a first exon of 102 nt, an intron of 113 nt and a second exon of 38 nt. Spliced products of this RNA and of the RNAs derived from pSP64Amyi (linearized with *Bam*HI and *Eco*RI) and pSP65LegJi (linearized with *Pst*I) were run alongside one another (Figure 7). By comparing the relative amounts of E1 and of E1–E2 spliced products the efficiency of splicing of the plant introns relative to the adenovirus intron could be estimated. The amylase intron was spliced at virtually the same level of efficiency as the adenovirus control RNA, while the legumin J intron was spliced at a much lower level.

U snRNPs are required for splicing in the HeLa cell nuclear extract (Yang *et al.*, 1981; Padgett *et al.*, 1983; Bozzoni *et al.*, 1984; Fradin *et al.*, 1984; Krämer *et al.*, 1984). To show that the splicing of the two plant introns was U snRNP dependent, pre-mRNAs were incubated in a nuclear extract that had been depleted of U snRNPs by pre-treatment with an anti-Sm antiserum (Krämer *et al.*, 1984). A control reaction was also carried out with a nuclear extract which had been 'mock-depleted' with pre-immune serum. Splicing was observed for both plant introns in the 'mock-depleted' extract but no splicing was seen with the U snRNP depleted extract (results not shown).

Splice junction and branch point sequences of plant introns

The 5′ and 3′ splice junctions of 168 plant introns from a variety

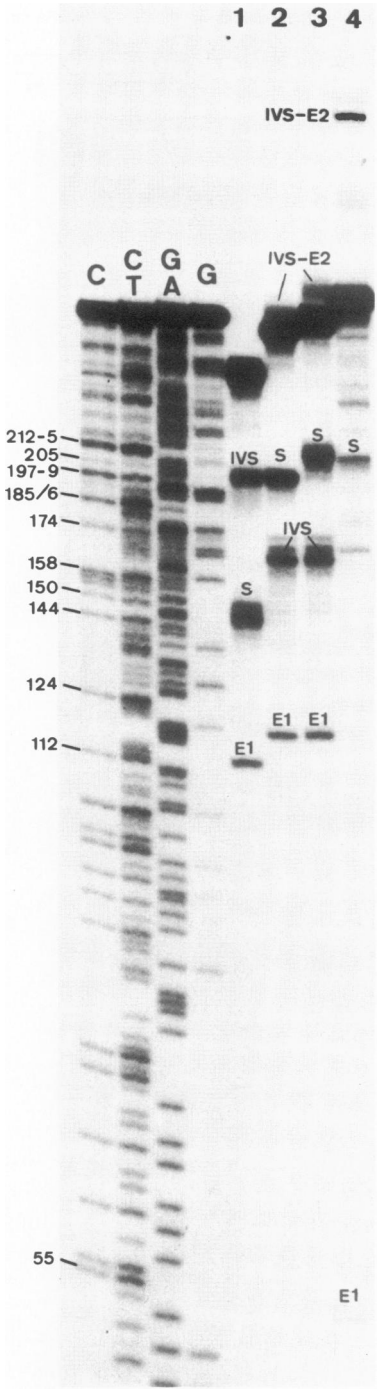

Fig. 7. Efficiency of splicing of plant introns in HeLa cell nuclear extracts. *In vitro* splicing products of pre-mRNAs derived from pSP62Δil (adenovirus 2 major late) **(lane 1)**, pSP64Amyi linearized with *Bam*HI **(lane 2)** or *Eco*RI **(lane 3)** and pSP65LegJi **(lane 4)** were separated on a 9% polyacrylamide–urea gel. Products of splicing are indicated: exon 1 (E1), intron lariat (IVS), exon 1–exon 2 spliced product (S), and intron–exon 2 lariat (IVS–E2). Marker is the sequenced *Eco*RI–*Pst*I DNA fragment from pSP64Amyi.

of nuclear-encoded genes from 15 different species (see Materials and methods) were compiled. The frequency of the appearance of nucleotides at specific positions was scored and the frequencies expressed as percentages were directly compared with a similar study with animal intron sequences (Mount, 1982) (Tables I and II). The consensus sequence for the 5′ splice site of plant

Table I. Nucleotide frequencies at the 5′ exon–intron splice junctions of plant introns

Position[a]	−3	−2	−1	:	+1	+2	+3	+4	+5	+6
Total	168	168	168		168	168	168	168	168	168
G	32	18	122		168	0	23	10	110	17
A	57	93	17		0	0	116	93	28	39
C	56	18	18		0	1	13	32	13	30
T[b]	23	39	11		0	167	16	33	17	82
%G	19(9)[c]	11(12)	73(73)		100(100)	0(0)	14(29)	6(12)	65(84)	10(8)
%A	34(40)	55(64)	10(9)		0(0)	0(0)	69(62)	55(68)	17(9)	23(17)
%C	33(43)	11(12)	11(6)		0(0)	1(0)	8(2)	19(9)	8(2)	18(12)
%T	14(7)	23(13)	7(12)		0(0)	99(100)	10(6)	20(12)	10(5)	49(63)
%Pu	53(50)	66(76)	83(82)		100(100)	0(0)	83(91)	61(79)	82(93)	33(25)
%Py	47(50)	34(24)	17(18)		0(0)	100(100)	17(9)	39(21)	18(7)	67(75)
Consensus	A/C	A	G	:	G	T	A	A	G	T

[a]Positions are numbered from the splice site (:).
[b]Sequences are given in DNA form.
[c]Numbers in parentheses are taken from a catalogue of animal and viral intron sequences (Mount, 1982) to allow direct comparison.

Table II. Nucleotide frequencies at the 3′ intron–exon splice junction of plant introns

Position[a]	−15	−14	−13	−12	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	:	+1
total	167	167	167	167	167	166	167	166	167	167	167	167	167	167	167		167
G	19	25	25	27	24	33	32	36	40	31	18	82	3	0	167		103
A	32	29	20	49	34	34	34	41	33	43	20	33	8	167	0		23
C	37	25	34	19	22	21	31	25	25	20	16	23	110	0	0		23
T[b]	79	88	88	72	87	78	70	64	69	73	113	29	46	0	0		18
%G	11(15)[c]	15(21)	15(10)	16(10)	14(10)	20(6)	19(7)	22(9)	24(7)	19(4)	11(5)	49(24)	2(1)	0(0)	100(100)		62(52)
%A	19(15)	17(10)	12(10)	29(15)	20(6)	20(15)	20(11)	25(19)	20(12)	26(3)	12(10)	20(25)	5(4)	100(100)	0(0)		14(22)
%C	22(19)	15(25)	20(31)	11(21)	13(24)	13(30)	19(33)	15(28)	15(36)	12(36)	10(28)	14(22)	66(65)	0(0)	0(0)		14(18)
%T	47(51)	53(44)	53(50)	43(53)	52(60)	47(49)	42(49)	39(45)	41(45)	44(57)	68(58)	17(29)	28(31)	0(0)	0(0)		11(8)
%Pu	31(30)	32(31)	27(29)	46(26)	35(16)	40(21)	40(18)	46(28)	44(19)	44(7)	23(15)	69(49)	7(4)	100(100)	100(100)		75(74)
%Py	69(70)	68(69)	73(71)	54(74)	65(84)	60(79)	60(82)	54(72)	56(81)	56(93)	77(85)	31(51)	93(96)	0(0)	0(0)		25(26)
Consensus	T	T	T	T[d]/Pu	T	T	T/Pu	T/Pu	T/Pu	T/Pu	T	G	C	A	G	:	G

[a]Positions are numbered from the splice site (:).
[b]Sequences are given in DNA form.
[c]Numbers in parentheses are taken from a catalogue of animal and viral intron sequences (Mount, 1982) to allow direct comparison.
[d]At these positions T is the most abundant nucleotide and the combined %G and %A are equivalent to or greater than the %T.

introns (Table I) is the same as that for animal introns: $\overset{A}{C}$AG:GTAAGT, although as seen from the values in Table I there is more variation in the plant sequences. The animal intron consensus sequence for the 3′ splice junction is $(\overset{T}{C})_{11}$NCAG:G (Mount, 1982). The plant intron 3′ consensus sequence derived from Table II is identical for the CAG:G sequence but at position −4 a G occurs while in the animal consensus this position can be occupied by any nucleotide (N). The plant 3′ consensus sequence is similar in the polypyrimidine stretch (positions −5 to −15) in that in each position Ts are the most abundant. However, the occurrence of purines is much higher in these positions in plant introns such that the percentage of pyrimidines is reduced from a range of 69–93% in animal introns to 54–77% in plant introns (Table II). The high level of purines is particularly evident in positions −6, −7, −8, −9 and −12 (Table II). Cytidine residues are also not as abundant in this region in plant introns as in the animal polypyrimidine stretch.

The plant intron sequences were also screened for putative branch points by searching for sequences similar to the highly conserved *S. cerevisiae* TACTAAC sequence, shown to be required for splicing (Langford and Gallwitz, 1983; Pikielney *et al.*, 1983) and to be the site of the branch point (Domdey *et al.*, 1984; Rodriguez *et al.*, 1984; Lin *et al.*, 1985), and for the less conserved vertebrate branch point sequence PyNPyTPuAPy (Keller and Noon, 1984; Ruskin *et al.*, 1984; Zeitlin and Efstradiatis, 1984). An initial investigation (results not shown) of the plant intron sequences within 50 nt of the 3′ splice site revealed a consensus sequence, CTPuAPy, which was very similar to the yeast and vertebrate sequences (above) and to fungal branch point consensus sequences (Kinnaird and Fincham, 1983; Käufer *et al.*, 1985). The distance between the possible adenosine branch point nucleotide and the 3′ splice site varied between 16 and 49 bases which compares with 18–40 bases for putative vertebrate branch point sequences (Reed and Maniatis, 1985; Ruskin *et al.*, 1985). This branch point consensus can only be taken tentatively since no plant branch points have been determined in homologous *in vivo* or *in vitro* experiments and therefore the nature of plant branch points is unknown. Also, the selection of the region within 50 nt of the 3′ splice junction may bias the analysis toward the vertebrate branch point distance.

Discussion

As a first step in the study of the mechanism of splicing of plant introns, nuclear pre-mRNA transcripts derived from the plasmids pSP64Amyi and pSP65LegJi, containing an intron from an amylase 33 gene from wheat and the intron from the legumin J storage

protein gene from pea respectively, were tested in the HeLa cell *in vitro* splicing system (Krainer *et al.*, 1984). In order to show successful splicing of pre-mRNAs *in vitro* it is necessary to identify the intermediate and final products of the reaction: exon 1, intron−exon 2 lariat, intron lariat and exon 1−exon 2 spliced product. We have identified such products by the following criteria: (i) the size of the observed products; (ii) the aberrant mobility of the two lariat forms in urea−polyacrylamide gels; (iii) the size of the splicing products from RNAs transcribed from DNA templates linearized at different sites to produce second exons of different lengths; (iv) the relative amounts of the products in a time course of the splicing reaction; (v) the size of RNAs produced by debranching lariat forms in a cytoplasmic extract; and (vi) by primer extension with isolated products (Figures 2, 4, 5 and 7). These analyses clearly show that both of the plant introns from the amylase and legumin genes are accurately spliced in the HeLa cell *in vitro* splicing system.

As the plant introns were spliced in a non-homologous system it was necessary to show that splicing was occurring at the expected 5′ and 3′ splice sites. Primer extension analysis provides a very accurate method of determining the size of RNA transcripts. The sizes of the longest primer extension cDNAs from isolated E1−E2 spliced products, 187 and 184 nt for amylase 33 and legumin J respectively, correspond exactly to the expected lengths of correctly spliced pre-mRNAs. The branch points of the amylase 33 and legumin J introns were mapped by primer extension with the major IVS−E2 lariat RNAs (Ruskin *et al.*, 1984; Krainer *et al.*, 1984). The minor RNAs seen in the direct gel analysis were not characterized further. We cannot rule out that these minor products are the result of the choice, at low frequency, of cryptic splice sites and/or other branch points. E2-specific primers and DNA markers both 5′ end-labelled at the same base were used in conjunction with the high resolution of the 1 m sequencing gels to align exactly the primer extension products with their corresponding sequences (Figures 5 and 6). Assuming that the longest primer extension product corresponds to a cDNA extending to the nucleotide lying directly 3′ to the branch point nucleotide the branch point sequences of the amylase 33 and legumin J introns were defined. In both cases the presumed branch point nucleotide was an adenosine. Alignment of the branch point sequences with the animal branch point consensus sequence (Figure 6) also showed reasonable homology.

With the exception of the intron sequences themselves and the conservation of the GT and AG dinucleotides at the 5′ and 3′ splice junctions little is known about splicing of plant mRNAs. Two studies have previously generated plant intron splice junction consensus sequences but these were based on relatively few sequences (20 intron sequences from three gene families of two species, Slightom *et al.*, 1983; and 30 sequences/six gene families/three species, Rogers, 1985). Publication of many plant gene sequences in the last two years has allowed us to derive splice junction consensus sequences from the analysis of 168 introns (Tables I and II). The plant consensus sequences for the 5′ splice junction and for the 3′ splice site region are virtually identical to the animal consensus sequences (Breathnach and Chambon, 1981; Mount, 1982). However, at position −4 the 3′ plant consensus has a G instead of any nucleotide and most of the plant introns do not contain a clear polypyrimidine stretch in positions −5 to −15 upstream of a 3′ splice site (Table II). This polypyrimidine stretch has been shown to be important in the formation of stable splicing complexes in the HeLa cell system (Frendewey and Keller, 1985; Ruskin and Green, 1985b). Indeed, the much lower efficiency of splicing of the legumin J intron in the HeLa extract (Figure 7) could reflect its low pyrimidine content (five out of 11 nucleotides in positions −5 to −15). In comparison the amylase intron which was spliced as efficiently as the adenovirus RNA control contained seven out of 11 pyrimidines including four consecutive uridines. No putative branch point consensus sequence could be found from the 30 sequences studied by Rogers (1985); however from the 168 introns studied here a consensus sequence was derived which was very similar to that of yeast, other fungi and vertebrates (Kinnaird and Fincham, 1983; Keller and Noon, 1984; Ruskin *et al.*, 1984; Zeitlin and Efstradiatis, 1984; Käufer *et al.*, 1985).

Plant genes containing introns have been transferred via the Ti-plasmid of *Agrobacterium tumefaciens* to other plant species, and were correctly processed as visualized by expression of the protein products (Murai *et al.*, 1983; Broglie *et al.*, 1984; Sengupta-Gopalan *et al.*, 1985; Beachy *et al.*, 1985). A similar approach, where the human growth hormone (hGH) gene (containing five introns) was transferred into tobacco and sunflower represents the only data where splicing of animal introns in plant systems has been examined (Barta *et al.*, 1986). An hGH RNA transcript was produced in the plant callus cells but appeared to be unspliced. However, when poly(A)$^+$ RNA was isolated from callus tissue and put into a HeLa cell *in vitro* splicing reaction, S1-mapping of the products apparently showed splicing of the first intron. Thus, the transcript which was not spliced in the plant cells was spliced when transferred to the animal system.

On the basis of the successful splicing of two plant introns in the HeLa cell splicing system it is difficult to draw firm conclusions about similar mechanisms of splicing in plants and animals. However, the strong similarities between the animal and plant splice junction sequences argue that the plant pre-mRNA splicing mechanism may resemble that seen in animals. Perhaps a more convincing argument for a relationship between plant and animal splicing systems is the identification in plants of U snRNAs that share significant sequence homologies and structural features (Krol *et al.*, 1983; Skuzeski and Jendrisdak, 1985; Kiss *et al.*, 1985) with the major U snRNAs shown to be involved in pre-mRNA splicing in vertebrates. The less well defined polypyrimidine stretch in the plant 3′ splice junction consensus sequence and the lack of splicing of the human growth hormone transcript in plant cells may reflect finer differences in the many factors necessary for splicing. Clearly an homologous *in vitro* splicing system from plants would be extremely useful for elucidating the mechanism of splicing of plant introns in the same way as has been done for animals.

Materials and methods

Materials

Restriction enzymes were obtained from Boehringer, Pharmacia or New England Biolabs and all were used in a buffer containing 33 mM Tris−acetic acid, pH 7.8, 62.5 mM potassium acetate, 10 mM magnesium acetate, 4 mM spermidine and 0.5 mM dithiothreitol (DTT). T4 DNA ligase, Klenow polymerase, SP6 polymerase, *Bam*HI 8-mer linkers and RNase inhibitor were obtained from Boehringer. T4 polynucleotide kinase was from Amersham or Biolabs. AMV reverse transcriptase was bought from Anglian Biotechnology. Radionucleotides for 5′ end-labelling and for *in vitro* transcription were from Amersham.

Cloning of two plant introns

The second intron along with flanking exon sequences of an amylase gene − amylase 33 from wheat (*Triticum aestivum* L.) was isolated from a 2.44-kb *Eco*RI fragment from a *Sal*I−*Bam*HI double digestion (Figure 1a). The purified 265-bp *Sal*I−*Bam*HI fragment was cloned into the *Sal*I and *Bam*HI restriction sites in the multiple cloning site of pSP64 (Melton *et al.*, 1984) producing the plasmid pSP64Amyi. The intron and adjoining exon sequences of a legumin J storage protein gene from pea (*Pisum sativum* L.) was isolated from an *Hin*fI digestion of a 1.9-kb *Eco*RI fragment (Figure 3a). The overhanging ends of the 279-bp

*Hin*fI fragment were filled-in with the Klenow fragment of polymerase I in the same buffer as described above for restriction endonuclease digestions, in the presence of dATP and dTTP. The blunt-ended fragment was ligated to 5′ end-labelled *Bam*HI 8-mer linkers, digested with *Bam*HI and the fragment re-isolated over an agarose gel. The purified fragment was cloned into the *Bam*HI site of pSP65 (Melton *et al.*, 1984) and the orientation determined using the *Aha*III site contained in the intron sequence (Figure 3a). The resulting plasmid was called pSP65LegJi. Both constructions were further checked by sequencing (Figures 1b and 3b).

Preparation of RNA transcripts

The plasmid pSP64Amyi was linearized at either the *Bam*HI or *Eco*RI restriction sites (Figure 1b and c) while pSP65LegJi was linearized at the *Pst*I restriction site (Figure 3b and c). *In vitro* transcription of these SP6 templates was carried out in a cap-primed reaction as described by Krämer and Keller (1985). RNA was synthesized from 2 μg of linearized plasmid DNA in a 40 μl reaction volume containing 40 mM Tris−HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT, 500 μM each of ATP and CTP, 100 μM GTP, 20 μM UTP, 2 mM G(5′)ppp(5′)G, 20 μCi [^{32}P]UTP, 10 units SP6 RNA polymerase and 63 units RNase inhibitor at 40°C for 1 h. Pre-RNA mRNA transcripts were purified and stored as described by Frendewey and Keller (1985).

In vitro splicing

Nuclear and cytoplasmic extracts were prepared from HeLa cells as described by Dignam *et al.* (1983). The standard splicing reaction contained in a 50 μl reaction volume, 0.4 mM ATP, 10 mM creatine phosphate, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 10 mM Hepes−KOH, pH 7.9, 50 mM KCl, 0.05 mM Na_2EDTA, 0.25 mM DTT, 0.15 mg/ml tRNA, 31.5 units of RNase inhibitor and 25 μl of nuclear extract. To each reaction 20 000 Cerenkov c.p.m. of ^{32}P-labelled pre-mRNA transcripts were added. The standard reactions were performed at 30°C for 60 min, and time course reactions at 30°C for 15, 30, 60 and 105 min. Proteinase K treatment, phenol extraction and precipitation of the RNA were as described by Hernandez and Keller (1983). Reaction products were separated on 8 M urea−polyacrylamide gels of various polyacrylamide concentrations and visualized by autoradiography.

In order to isolate the exon 1−exon 2 spliced product and intron−exon 2 lariat forms for primer extension, 0.5 × 10^6 Cerenkov c.p.m. of [^{32}P]pre-mRNA transcripts were added to a 3× scaled-up *in vitro* splicing reaction. The above RNA species were isolated from gels as described by Frendewey and Keller (1985).

DNA sequencing and primer extension

Sequencing of the DNA inserts containing the amylase 33 and legumin J introns was carried out by isolation of the *Hin*dIII−*Eco*RI fragments from pSP64Amyi and pSP65LegJi (Figures 1b and 3b). Following alkaline phosphatase treatment these fragments were 5′-labelled with T4 polynucleotide kinase and digested with *Pst*I. The *Eco*RI−*Pst*I fragments of pSP64Amyi and pSP65LegJi were isolated and sequenced by the base-specific chemical cleavage method (Maxam and Gilbert, 1980), and loaded on a 6%, 1 m long, gradient sequencing gel. This sequenced fragment from pSP64Amyi was also used as marker on the gels of the *in vitro* splicing reaction.

In order to produce primers and markers for primer extension experiments the *Sma*I−*Bam*HI fragment of pSP64Amyi and the *Bam*HI−*Bam*HI fragment of pSP65LegJi were isolated (Figures 1b and 3b). These two fragments were 5′ end-labelled as above and divided into two aliquots. For pSP64Amyi, one aliquot was directly sequenced while the second was digested with *Sau*3AI and the 36-bp fragment (labelled at the *Bam*HI site) was isolated as a primer. For pSP65-LegJi, one aliquot was digested with *Sph*I and the longer *Sph*I−*Bam*HI fragment isolated and sequenced. The second aliquot was digested with *Hph*I and the 34-bp fragment (labelled at the *Bam*HI site) was isolated as a primer. In this way exon 2-specific primers for both genes are produced which are 5′ end-labelled at the same site as a sequenced marker fragment. This allows the exact positioning of primer extension products by direct comparison with the corresponding DNA sequence.

Isolated exon 1−exon 2 spliced product or intron−exon 2 lariat RNAs were precipiated with primer, dissolved in 8 μl of H_2O and denatured at 90°C for 2 min. 2 μl of a 5× annealing buffer containing 2 M NaCl and 50 mM Pipes, pH 6.4 were added and the RNA and primer annealed at 50°C for 15 min followed by 15 min at room temperature. The solution was brought to 100 mM Tris−HCl, pH 8.3 (at 42°C), 8 mM $MgCl_2$, 145 mM KCl, 20 mM β-mercaptoethanol, 1 mM each of dATP, dCTP, dGTP and dTTP in a 100 μl volume to which 30 units of AMV reverse transcriptase was added. Primer extension was carried out at 42°C for 1 h, followed by phenol extraction and ethanol precipitation. RNA was degraded by alkaline hydrolysis by a 30 min, 65°C treatment in 0.3 M NaOH, 1 mM Na_2EDTA. The remaining cDNA was recovered by ethanol precipitation with 5 μg of tRNA carrier and loaded on a 6%, 1 m long gradient sequencing gel alongside the sequenced fragment marker. Primer extension bands were corrected by 1.5 nucleotide positions to compensate for the differences in the 3′ termini generated from primer extension and chemical sequencing reactions (Green and Roeder, 1980).

Analysis of plant intron sequences

The 168 introns which were used to generate the 5′ and 3′ splice junction consensus sequences (Tables I and II) and putative branch point consensus sequence were obtained from a variety of plant nuclear-encoded genes from 15 different species: maize − alcohol dehydrogenase, *Adh-l* and *Adh-2* (9 introns each) (Dennis *et al.*, 1985), sucrose synthetase, *shrunken* (15 introns) (Werr *et al.*, 1985), heat shock protein, *hsp70* (1 intron) (Rochester *et al.*, 1986), glutathione-S transferase, *gst1* (2 introns) (Shah *et al.*, 1986), *waxy* (13 introns) (Klösgen *et al.*, 1986) and actin, *MAcl* (3 introns) (Shah *et al.*, 1983); wheat − amylase genes, *Amy33* (2 introns), *Amy54* (3 introns), *Amy13* (2 introns), *Amy18* (2 introns) (D.Baulcombe, personal communication); pea − legumin genes, *LegA* (3 introns), *LegD* (3 introns) (Bown *et al.*, 1985), *LegJ* (1 intron) (N.Ellis, personal communication), small subunit of ribulose-1,5-bisphosphate carboxylase (Rubisco), *rbcS* (2 introns) (Corruzzi *et al.*, 1984); soybean − actin genes, *SAc1* and *SAc3* (3 introns each) (Shah *et al.*, 1982, 1983), five leghaemoglobin genes, *Lb* (gene 1), *Lba, Lbc1, Lbc2, Lbc3* (3 introns each) (Brisson and Verma, 1982; Hyldig-Nielsen *et al.*, 1982; Wiborg *et al.*, 1982), nodulin genes, *nod23* (1 intron) (Mauro *et al.*, 1985), *nod24* (4 introns) (Katinakis and Verma, 1985), conglycinin, *Gmga 17.1* (3 introns) (Schuler *et al.*, 1982), glycinin, *Ala* (1 intron) (Marco *et al.*, 1984); french bean − phaseolin (5 introns) (Slightom *et al.*, 1983); broad bean − legumin, *LeB4* (2 introns) (Bäumlein *et al.*, 1986); alfalfa − glutamine synthetase (11 introns) (Tischer *et al.*, 1986); potato − four patatin genes (6 introns each) (Bevan *et al.*, 1986; Pikaard *et al.*, 1986; Rosahl *et al.*, 1986); carrot − extensin (1 intron) (Chen and Varner, 1985); *Nicotiana tabacum* − Rubisco small subunit (3 introns) (Mazur and Chui, 1985); *N. plumbaginifolia* − mitochondrial ATP synthase, *atp2-1* (8 introns) (Boutry and Chua, 1985); *Petunia* (Mitchell) − Rubisco small subunit genes, *ssU8* (3 introns), *ssu11A* (2 introns) (Tumer *et al.*, 1986); *Antirrhinum majus* − chalcone synthase, *chs* (2 introns) (Sommer and Saedler, 1986); *Lemma gibba* − chlorophyll *a/b* protein (1 intron) (Karlin-Newmann *et al.*, 1985); and *Chlamydomonas reinhardtii* − Rubisco small subunit genes, *rbcs1* and *rbcs2* (3 introns each) (Goldschmidt-Clermont and Rahive, 1986). The full list of splice junction and putative branch point sequences and a detailed discussion of these data will be given elsewhere (J.Brown, in preparation).

Acknowledgements

We would like to thank Dr David Baulcombe and Dr Noel Ellis for generously making the amylase 33 and legumin J clones available. We are also grateful to Mrs E.Brutzer for her excellent technical assistance and Ms M.Frick for preparation of the HeLa cell extracts. We would also like to thank W.Keller for his support and encouragement. This research was supported by grants to Dr G.Feix and Dr W.Keller from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

References

Barta,A., Sommergruber,K., Thompson,D., Hartmuth,K., Matzke,M.A. and Matzke,A.J.M. (1986) *Plant Mol. Biol.*, **6**, 347−357.

Bäumlein,H., Wobus,U., Pustell,J. and Kafatos,F.C. (1986) *Nucleic Acids Res.*, **14**, 2707−2720.

Beachy,R.N., Chen,Z.-L., Horsch,R.B., Rogers,S.G., Hoffmann,N.J. and Fraley, R.T. (1985) *EMBO J.*, **4**, 3047−3053.

Bevan,M., Barker,R., Goldsbrough,A., Jarvis,M., Kavanagh,T. and Iturriaga,G. (1986) *Nucleic Acids Res.*, **14**, 4625−4638.

Bindereif,A. and Green,M.R. (1986) *Mol. Cell Biol.*, **6**, 2582−2593.

Black,D.L. and Steitz,J.A. (1986) *Cell*, in press.

Black,D.L., Chabot,B. and Steitz,J.A. (1985) *Cell*, **42**, 737−750.

Boutry,M. and Chua,N.-H. (1985) *EMBO J.*, **4**, 2159−2165.

Bown,D., Levasseur,M., Croy,R.R.D., Boulter,D. and Gatehouse,J.A. (1985) *Nucleic Acids Res.*, **13**, 4527−4537,

Bozzoni,I., Annesi,F., Beccari,E., Fragapane,P., Pierandrei-Amaldi,P. and Amaldi,F. (1984) *J. Mol. Biol.*, **180**, 1173−1178.

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349−383.

Brisson,N. and Verma,D.P.S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4055−4059.

Brody,E. and Abelson,J. (1985) *Science*, **228**, 963−967.

Broglie,R., Coruzzi,G., Fraley,R.T., Rogers,S.G., Horsch,R.B., Niedermeyer, J.G., Fink,C.L., Flick,J.S. and Chua,N.-H. (1984) *Science*, **224**, 838−843.

Chabot,B., Black,D.L., LeMaster,D.M. and Steitz,J.A. (1985) *Science*, **230**, 1344−1349.

Chen,J. and Varner,J.E. (1985) *EMBO J.*, **4**, 2145−2150.

Cheng,S.-C. and Abelson,J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2387−2391.

Coruzzi,G., Broglie,R., Edwards,C. and Chua,N.-H. (1984) *EMBO J.*, **3**, 1671−1679.

Dennis,E.S., Sachs,M.M., Gerlach,W.L., Finnegan,E.J. and Peacock,W.J. (1985) *Nucleic Acids Res.*, **13**, 727−743.

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.

Domdey,H., Apostol,B., Lin,R.J., Newman,A.J., Brody,E. and Abelson,J. (1984) *Cell*, **39**, 611–621.
Fradin,A., Jove,R., Hemenway,C., Keiser,H.D., Manley,J.L. and Prives,C. (1984) *Cell*, **37**, 927–936.
Frendewey,D. and Keller,W. (1985) *Cell*, **42**, 355–367.
Goldschmidt-Clermont,M. and Rahire,M. (1986) *J. Mol. Biol.*, in press.
Grabowski,P.J., Seiler,S.R. and Sharp,P.A. (1985) *Cell*, **42**, 345–353.
Green,M.R. and Roeder,R.G. (1980) *Cell*, **22**, 231–242.
Hagenbüchle,O., Santer,M., Steitz,J.A. and Mans,R.J. (1978) *Cell*, **13**, 551–563.
Hardy,S.F., Grabowski,P.J., Padgett,R.A. and Sharp,P.A. (1984) *Nature*, **308**, 375–377.
Hernandez,N. and Keller,W. (1983) *Cell*, **35**, 89–99.
Hyldig-Nielsen,J.J., Jensen,E.O., Paludan,K., Wiborg,O., Garrett,R., Jorgensen, P. and Marcker,K.A. (1982) *Nucleic Acids Res.*, **10**, 689–701.
Kaltwasser,G., Spitzer,S.G. and Goldenberg,C.J. (1986) *Nucleic Acids Res.*, **14**, 3687–3701.
Karlin-Neumann,G.A., Kohorn,B.D., Thornber,J.P. and Tobin,E.M. (1985) *J. Mol. Appl. Genet.*, **3**, 45–61.
Katinakis,P. and Verma,D.P.S. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4157–4161.
Käufer,N.F., Simianis,V. and Nurse,P. (1985) *Nature*, **318**, 78–80.
Keller,E.B. and Noon,W.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7417–7420.
Kinnaird,J.H. and Fincham,J.R.S. (1983) *Gene*, **26**, 253–260.
Kiss,T., Toth,M. and Solymosy,F. (1985) *Eur. J. Biochem.*, **152**, 259–266.
Klösgen,W.B., Gierl,A., Schwarz-Sommer,S. and Saedler,H. (1986) *Mol. Gen. Genet.*, **203**, 237–244.
Konarska,M.M., Grabowski,P.J., Padgett,R.A. and Sharp,P.A. (1985) *Nature*, **313**, 552–557.
Krainer,A.R. and Maniatis,T. (1985) *Cell*, **42**, 725–736.
Krainer,A.R., Maniatis,T., Ruskin,B. and Green,M.R. (1984) *Cell*, **36**, 993–1005.
Krämer,A. and Keller,W. (1985) *EMBO J.*, **4**, 3571–3581.
Krämer,A., Keller,W., Appel,B. and Lührmann,R. (1984) *Cell*, **38**, 299–307.
Krol,A., Ebel,J.-P., Rinke,J. and Lührmann,R. (1983) *Nucleic Acids Res.*, **11**, 8583–8594.
Langford,C.J. and Gallwitz,D. (1983) *Cell*, **33**, 519–527.
Lerner,M.R., Boyle,J.A., Mount,S.M., Wolin,S.M. and Steitz,J.A. (1980) *Nature*, **283**, 220–224.
Lin,R.J., Newman,A.J., Cheng,S.-C. and Abelson,J. (1985) *J. Biol. Chem.*, **260**, 14780–14792.
Marco,Y.A., Thanh,V.H., Tumer,N.E., Scallon,B.J. and Nielsen,N.C. (1984) *J. Biol. Chem.*, **259**, 13436–13441.
Mauro,V.P., Nguyen,T., Katinakis,P. and Verma,D.P.S. (1985) *Nucleic Acids Res.*, **13**, 339–349.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Mazur,B.J. and Chui,C.-F. (1985) *Nucleic Acids Res.*, **13**, 2373–2386.
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green, M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Murai,D.J., Sutton,D.W., Murray,M.G., Slightom,J.L., Merlo,D.J., Reichert, N.A., Sengupta-Gopalan,C., Stock,C.A., Barker,R.F., Kemp,J.D. and Hall, T.C. (1983) *Science*, **222**, 476–482.
Padgett,R.A., Mount,S.M., Steitz,J.A. and Sharp,P.A. (1983) *Cell*, **35**, 101–107.
Padgett,R.A., Konarska,M.M., Grabowski,P.J., Hardy,S.F. and Sharp,P.A. (1984) *Science*, **225**, 898–903.
Padgett,R.A., Grabowski,P.J., Konarska,M.M., Seiler,S. and Sharp,P.A. (1986) *Annu. Rev. Biochem.*, **55**, in press.
Pikaard,C.S., Mignery,G.A., Ma,D.P., Stark,V.J. and Park,W.D. (1986) *Nucleic Acids Res.*, **14**, 5564–5566.
Pikielney,C.W. and Rosbash,M. (1986) *Cell*, **45**, 869–877.
Pikielney,C.W., Teem,J.L. and Rosbash,M. (1983) *Cell*, **34**, 395–403.
Reed,R. and Maniatis,T. (1985) *Cell*, **41**, 95–105.
Rochester,D.E., Winer,J.A. and Shah,D.M. (1986) *EMBO J.*, **5**, 451–458.
Rodriguez,J.R., Pikielney,C.W. and Robash,M. (1984) *Cell*, **39**, 603–610.
Rogers,J.M. (1985) *Int. Rev. Cytol.*, **93**, 188–279.
Rogers,J. and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1877–1879.
Rosahl,S., Schmidt,R., Schell,J. and Willmitzer,L. (1986) *Mol. Gen. Genet.*, **203**, 214–220.
Ruskin,B. and Green,M.R. (1985a) *Science*, **229**, 135–140.
Ruskin,B. and Green,M.R. (1985b) *Cell*, **43**, 131–142.
Ruskin,B., Krainer,A.R., Maniatis,T. and Green,M.R. (1984) *Cell*, **38**, 317–331.
Ruskin,B., Greene,J.M. and Green,M.R. (1985) *Cell*, **41**, 833–844.
Schuler,M.A., Schmitt,E. and Beachy,R.N. (1982) *Nucleic Acids Res.*, **10**, 8225–8244.
Sengupta-Gopalan,C., Reichert,N.A., Barker,R.F., Hall,T.C. and Kemp,J.D. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3320–3324.
Shah,D.M., Hightower,R.C. and Meagher,R.B. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1022–1026.
Shah,D.M., Hightower,R.C. and Meagher,R.B. (1983) *J. Mol. Appl. Genet.*, **2**, 111–126.
Shah,D.M., Hironaka,C.M., Wiegand,R.C., Harding,E.I., Krivi,G.G. and Tiemeier,D.C. (1986) *Plant Mol. Biol.*, **6**, 203–211.
Skuzeski,J.M. and Jendrisak,J.J. (1985) *Plant Mol. Biol.*, **4**, 181–193.
Slightom,J.L., Sun,S.M. and Hall,T.C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1897–1901.
Sommer,H. and Saedler,M. (1986) *Mol. Gen. Genet.*, **202**, 429–434.
Tischer,E., DasSarma,S. and Goodman,H.M. (1986) *Mol. Gen. Genet.*, **203**, 221–229.
Tumer,N.E., Clark,W.G., Tabor,G.J., Hironaka,C.M., Fraley,R.T. and Shah,D.M. (1986) *Nucleic Acids Res.*, **14**, 3325–3342.
Vijayraghavan,U., Parker,R., Tamm,J., Iimura,Y., Rossi,J., Abelson,J. and Guthrie,C. (1986) *EMBO J.*, **5**, 1683–1695.
Werr,W., Frommer,W.-B., Maas,C. and Starlinger,P. (1985) *EMBO J.*, **4**, 1373–1380.
Wiborg,O., Hyldig-Nielsen,J.J., Jenson,E.O., Paludan,K. and Marcker,K.A. (1982) *Nucleic Acids Res.*, **10**, 3487–3493.
Wise,J.A., Tollervey,D., Maloney,D., Swerdlow,H., Dunn,E.J. and Guthrie,C. (1983) *Cell*, **35**, 743–751.
Yang,V.W., Lerner,M.R., Steitz,J.A. and Flint,S.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1371–1375.
Zeitlin,S. and Efstradiatis,A. (1984) *Cell*, **39**, 589–602.

Received on 6 August 1986; revised on 28 August 1986