A functional splice site in the 5' untranslated region of a zein gene

John W.S. Brown* and Gunter Feix

Institut für Biologie III, Albert Ludwigs Universität, D-7800 Freiburg, FRG

Received September 22, 1989; Revised and Accepted December 1, 1989

ABSTRACT

Zein genes, the genes coding for the zein storage proteins of maize, have a unique gene structure where at least two promoters lie upstream of the coding region. Between the P1 promoter (900 base pairs upstream of the coding region) and the translation initiation AUG codon are 18 short reading frames. A discrepancy between the signals obtained by S1-mapping and primer extension and the DNA sequence in the region of one of these signals suggests the presence of a 3' splice site lying 40 nucleotides upstream of the coding region. A splicing event removing all of the short reading frames from the mRNA transcribed from the P1 promoter would bring this mRNA into a translatable form. Further evidence for a functional 3' splice site has been obtained from sequencing of primer extension products and *in vitro* splicing of a hybrid intron in the HeLa cell *in vitro* splicing system.

INTRODUCTION

Zeins, the major storage proteins of maize, consist primarily of two size classes: 19 and 21 kilodaltons (kDa) and are encoded by a multigene family composed of between 50 and 100 members (see Refs. 1 and 2 for reviews). Analysis of zein genomic clones coding for 19kDa zein polypeptides (pMS1 and pMS2) and 21kDa zein polypeptides (pML1 and pML2) from the maize line A619 has shown these genes to have at least two promoter regions upstream of the coding sequence (3,4). One promoter, P2, lies approximately 50 base pairs (bp) upstream of the translation start while the second promoter, P1, lies 900–1000 bp 5' to the AUG initiation codon (3,4). Transcription from these promoters produces polyadenylated zein mRNAs of 800–900 bases and 1600–1800 bases respectively as shown by Northern analyses, R-loop analyses and S1-mapping (3–7). Although examples of genes with two promoters are well documented in animal systems (reviewed in Refs. 8 and 9), zein genes represent the only example to date of a plant gene with a complex promoter structure.

The genes contained on pMS1, pMS2 and pML1 are expressed to the protein level following microinjection into the nucleus of the unicellular alga *Acetabularia mediterranea* (7,10). The injection of 5' deletion mutants showed that the further upstream

P1 promoter was responsible for the activity of the zein genes in the algal cells and therefore mRNAs transcribed from the P1 promoter gave rise to immunologically detectable zein protein (10). However, the DNA sequence in the region between the P1 promoter and the AUG initiation codon contained eighteen AUG codons each with a short reading frame (Fig. 1a). A mechanism must therefore exist to allow the translation of the zein coding region from P1-derived RNA in spite of the preceeding open reading frames. Since expression experiments with zein mutants in *A. mediterranea* have shown that the presence of the region around P2 is essential for zein synthesis from P1-derived RNA (10), and because the sequence analysis of this region revealed attributes of a 3' splice site, processing of the P1-derived splice site was anticipated. This paper presents primer extension and *in vitro* splicing experiments used to test this hypothesis.

MATERIALS AND METHODS

Materials

Restriction enzymes were obtained from Boehringer, Pharmacia, or New England Biolabs and all were used in the buffer described previously (11). T4 DNA ligase, Klenow fragment of DNA polymerase I, SP6 RNA polymerase and RNase inhibitor were obtained from Boehringer. T4 polynucleotide kinase and radionucleotides for 5'-end labelling and for *in vitro* transcription were obtained from Amersham. AMV reverse transcriptase was bought from Anglian Biotechnology.

DNA sequencing, primer extension and S1-mapping

In order to isolate fragments for primer extension, S1-mapping and sequencing which were all labelled at the same nucleotide, the 830 bp Sal I–Bam HI fragment (positions –660 to +170) from pMS2 was isolated, alkaline phosphatased and 5'-end labelled with T4 polynucleotide kinase. One aliquot was then digested with Hph I (position +113) and the remainder with Hinc II (position –422). The resulting fragments were separated on a 6% polyacrylamide gel and the 62 bp Hph I–Bam HI fragment and the 598 bp Hinc II–Bam HI fragment, both labelled at the Bam HI site, were isolated for primer extension, and for S1-mapping and sequencing respectively. Fragments were sequenced by the base specific cleavage method (12).

* Present address: Department of Biological Sciences, University of Dundee, Dundee DD1 4HN, UK

Fig. 1. a) Schematic representation of the structure of the genomic clone, pMS1. The numbered lines represent the size and distribution of the 18 short reading frames (SRF) lying between the P1 promoter (P1) and the coding sequence (open box). P2–P2 promoter; E–Eco RI. b) DNA sequence of pMS1 in the regions of the P1 and P2 promoters. The closed triangles represent the P1 and P2 transcription start sites as determined by S1-mapping. The possible dinucleotide –GT and AG– borders of the proposed intron are partially blocked and potential branch point sequences are underlined.

Total RNA was extracted from A619 maize kernels at the 20 days after pollination stage. Poly A+ RNA was isolated on an oligo dT cellulose column. Hybridisation of poly A+ RNA (10 µg) to 5'-end labelled primer or fragment, and S1-mapping was carried out as described previously (7). For primer extension the hybridised RNA-DNA was ethanol precipitated and redissolved in a solution containing 100 mM Tris-HCl, pH 8.3 (at 42°C), 8 mM MgCl₂, 145 mM KCl, 20 mM 2-mercaptoethanol and 1 mM each of dATP, dCTP, dGTP and dTTP. Following addition of 50 units of AMV reverse transcriptase the reaction was incubated for 3 hours at 42°C. After degradation of RNA in the primer extension and S1-mapping reactions the remaining DNA was separated alongside the DNA sequence marker on a 6% sequencing gel.

Sequencing of primer extension products

For sequencing, ten primer extension reactions were set up using as primer an oligonucleotide complementary to the 5' untranslated region of pMS2 (positions –2 to –31). The reactions were pooled and primer extension products separated on an 8M urea, 17% polyacrylamide gel. Following autoradiography 5' end-labelled products were cut out of the gel and eluted overnight in a solution of 0.3M ammonium acetate, 10mM EDTA. Fragments were recovered by ethanol precipitation following the addition of 10 µg calf thymus carrier DNA and sequenced.

Construction of SP6 clones for *in vitro* transcription

In order to test the splicing of the proposed intron in the HeLa cell *in vitro* splicing system the zein 5' flanking sequence (positions –954 to +171) was cloned into pSP64 (13). A deletion mutant of the above clone, where positions –682 to –101 were deleted, was also constructed. Both clones contained the P1 and P2 promoter regions and the putative splice junction and branch point sequences (Fig. 1b).

From the sequence of the possible 5' and 3' splice sites (Fig. 1b) the 3' splice junction has clearly the better fit to animal and plant consensus sequences (14,15). In order to specifically test the zein 3' splice junction a hybrid intron was constructed by cloning the 5' end of a plant intron known to be spliced in the HeLa cell system in front of the 3' end of the proposed zein intron. The 97bp Bam HI–Aha III fragment from the plasmid, pSP65LegJi (containing an intron from a legumin J storage protein gene of pea—Ref. 11) was cloned in front of the 395bp Sna BI-Bam HI (positions –224 to +171) fragment from pMS1, into the Bam HI site of pSP64 (Fig. 4a) and this hybrid intron construct was checked by sequencing. *In vitro* transcription of the resulting plasmid, pSP64LegJ/Sli, linearised at the Eco RI site, would produce a 548 nucleotide (nt) long transcript made

up of a first exon (E1) of 54nt, a hybrid intron (IVS) of 262nt and a second exon (E2) of 232nt (Fig. 4c).

Preparation of pre-mRNA transcripts and *in vitro* splicing

In vitro transcription from linearised SP6 templates and purification of the RNA transcripts was carried out as described previously (11). HeLa cell nuclear extracts were prepared as described by Dignam et al. (16). The *in vitro* splicing reactions were performed exactly as previously described (11). The products of the splicing reactions were separated on 8M urea-polyacrylamide gels of different polyacrylamide concentrations and were visualised by autoradiography.

RESULTS

Structure of the 5' flanking region of zein genes

The structure and part of the 5' flanking region sequence of the zein gene, pMS1 (10), is shown schematically in Figure 1a. Upstream of the coding region are two promoter regions, P1 and P2, with transcription start points at –915 and –917 (P1) and –56 and –42 (P2) as envisaged by S1-mapping (4,7). Between the P1 and P2 promoters lie 18 AUG codons each with a short reading frame. The positions and the sizes of these reading frames (varying from 2 to 29 amino acids) are presented schematically in Figure 1a. Part of the DNA sequence of the 5' flanking region of pMS1, highlighting the P1 and P2 promoter regions relative to the translation initiation site is shown in Figure 1b.

Primer extension analysis of the S1 signals at the P2 promoter region

In order to ascertain whether both or only one of the two S1 signals in the P2 promoter region represented transcription start sites, primer extension and S1-mapping analyses were carried out in parallel. A direct comparison was made by using a primer which was labelled at the same nucleotide as the fragments used for S1-mapping and DNA sequencing (Fig. 2a). The products of the primer extension and S1-mapping analyses were separated on a 6% gradient sequencing gel alongside the DNA sequence marker (Fig. 2b).

The S1-mapping analysis showed two bands (Fig. 2b, lane S) lying 216 and 230 nucleotides from the labelled Bam HI site corresponding to positions –56 and –42 in the 5' flanking sequence of pMS2. The major primer extension products lay 230 and 231 nucleotides upstream from the Bam HI site (Fig. 2b, lane P) corresponding to positions –56 and –57 in the zein 5' flanking region. Shorter primer extension products were observed throughout the primer extension lane and were probably due to early termination of the reverse transcriptase. No primer

Fig. 2. Primer extension and S1-mapping of the P2 region. a) Schematic representation of part of the pMS2 gene showing the positions of the fragments used in the analyses: the 598bp Hinc II-Bam HI fragment (S1-mapping and DNA sequencing) and the 62bp Hph I-Bam HI fragment (primer extension) both labelled at the Bam HI site. b) S1-mapping and primer extension analysis. The products of primer extension (lane P) and S1-mapping (lane S) were separated alongside the DNA sequence of the Hinc II-Bam HI fragment (lanes G, A > C, G + A, T + C and C). The lanes P and S on the extreme left and right hand sides are shorter exposures of the main P and S lanes. The major products are labelled with arrows and are exactly aligned on the DNA sequence on the right. The signals are corrected by 1.5nt positions. Control lanes without maize RNA are shown in lanes 1 and 2.

A = AhaIII, B = BamHI, E = EcoRI, H = HindIII, S = SnaBI, Sm = SmaI

extension band corresponding to the lower S1 signal at position 216 (Fig. 2b, lane S) was apparent. These results were also observed with other primers—the 77bp BamHI-AluI fragment of pMS1 (positions +98 to +175), a 24-mer oligonucleotide complementary to the coding region of pMS1 (positions +173 to +197) and a 30-mer oligonucleotide complementary to the 5' untranslated region of pMS1 (results not shown).

Examination of the DNA sequence of the P2 promoter region (Fig. 1b) shows that the S1 signal at -42 lies next to an AG dinucleotide (positions -40 and -41) which is the highly conserved dinucleotide found at the 3' end of eukaryotic introns

(18). The DNA sequence lying directly upstream of this AG dinucleotide also exhibits other features of the 3' end of an intron in that it contains a number of pyrimidine nucleotides and potential branch point sequences (positions -60 to -64, -73 to -77 and -94 to -98) (Fig. 1b). The DNA sequence downstream of the P1 transcription start site shows a number of GT dinucleotides (the highly conserved dinucleotide at the 5' end of eukaryotic introns—Ref. 18).

Two longer primer extension products were observed 242 and 320 nucleotides upstream of the Bam HI site corresponding to positions -68 and -146 (Fig. 2b, lane P). Longer primer extension products were also been observed with the other primers (above) but bands corresponding to these signals have not been observed in S1 analyses. The lengths of the longer primer extension products varied depending on the sequence used as primer which may reflect the variation in sequence found among closely related zein genes.

Preparative primer extension was carried out with the end-labelled 30-mer oligonucleotide complementary to position -2 to -31 of the pMS1 5' untranslated region (Fig. 3). Primer extension bands of 55, 56 and 80 nt in length were isolated and sequenced. The first two corresponded to the major primer extension bands seen in Figure 2a. The third band was 13 nt longer than the longer primer extension product obtained with the pMS2 Bam HI-Hph I fragment (Fig. 2a). The sequences of these products are given in Figure 3. The sequence of the major band corresponding to the primer extension band at position -56 was colinear with the pMS1 sequence. The sequence of the second major band (-57) was colinear with the pMS1 sequence with the exception of a C at position -52. The sequence of the third band was incomplete due to lack of material and it was not possible to distinguish between A and G bases in the sequencing reactions such that in the coding strand C and T bases could not be distinguished and are labelled as pyrimidines (y). This sequence contained the primer sequence, an 8 base sequence which was colinear with the pMS1 sequence except for a pyrimidine instead of an A at position -36, and a further 24 bases of sequence which did not match the pMS1 sequence. The point at which the colinearity between this sequence and the pMS1 sequence breaks down is at position -39 which exactly corresponds to the proposed 3' splice site (Fig. 3).

In vitro splicing in the HeLa cell extract

The recently established *in vitro* splicing of plant introns in HeLa cell nuclear extracts was used as a functional test of the proposed intron in the P1-derived mRNA. Pre-mRNAs produced from the pSP64 clones of the pMS1 5' region and a deletion thereof did not show evidence of splicing in the HeLa cell *in vitro* splicing system (results not shown). The pre-mRNA produced from the hybrid intron plasmid, pSP64LegJ/S1i (Fig. 4c), was however successfully spliced in the HeLa cell *in vitro* splicing system (Fig. 5). The splicing of this pre-mRNA is directly compared to that of the pre-mRNA produced from the plasmid, pSP65LegJi (Fig. 4d and Ref. 11) containing the complete legumin J intron (Fig. 5).

Pre-mRNA transcripts from pSP65LegJi linearised with PstI (Fig. 5, lanes 1-4) and from pSP64LegJ/S1i linearised with EcoRI (Fig. 5, 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 onto a 6% polyacrylamide-urea gel (Fig. 5a) and a 12% polyacrylamide gel (Fig. 5b). The products of *in vitro* splicing

Fig. 3. Primer extension product sequences. The sequence of part of the 5' flanking sequence of pMS1 and the sequences of the 5'-end labelled (*) oligonucleotide primer and the -56, -57 and -81 primer extension products are compared. Single nucleotide differences are indicated with open triangles and the position of the proposed 3' splice site is arrowed.

Fig. 4. Construction of hybrid intron clone. a) The 97 bp Bam HI–Aha III fragment of pSP65LegJi (containing the legumin J 5' splice site) was cloned in front of the 395 bp Sna BI–Bam HI fragment of pMS1 (containing the putative 3' splice site) into pSP64. b) Sequence of the pre-mRNA transcript produced by *in vitro* transcription of the hybrid legumin J–pMS1 intron. The first base is the first base transcribed from the SP6 promoter. The intron is boxed and the junction of the legumin J and pMS1 sequences is shown by the arrowhead at position 131. Potential branch-point sequences are underlined by thin lines and the relative position of the translation ATG codon is doubly underlined. The small arrow shows the length of the transcript produced from the template linearised by Eco RI (548 bases) and the sizes of the exons and intron of this transcript are given in c). The sizes of the exons and the intron of the legumin J intron plasmid, pSP65LegJi are shown in d) for direct comparison.

of the pSP65LegJi transcript were used as a reference to analyse the products of splicing of the pSP64LegJ/S1i hybrid intron transcript. The first exon (E1) of the pSP65LegJi transcript is 48 nt long and is clearly seen in Figure 5b (lanes 2–4). The first exon of the hybrid intron transcript should be 54 nt long and is also seen in Figure 5b (lanes 7 and 8). As reported previously the labelled RNA bands run slower in the gels than the DNA sequence marker (11). Comparison of lanes 1–4 between Figures 5a and 5b showed the slower mobility of the intron lariat (IVS) and the intron-exon 2 lariat (IVS-E2) RNA forms in the higher percentage gel. In the 12% gel the IVS and IVS-E2 bands ran higher than the starting pre-mRNA transcript

(R) while in the 6% gel the IVS-E2 band ran slightly higher than the pre-mRNA. The pre-mRNA transcript of the hybrid intron construct (548nt) is much larger than that of pSP65LegJi (336nt). The difference in size is due to a larger intron and second exon (Fig. 4b and c) and this is reflected by the visualisation of the IVS-E2 lariat band running higher than the pre-mRNA transcript already in the 6% gel (Fig. 5a, lanes 6–8). Indeed, in the 12% gel this lariat RNA hardly enters the gel (Fig. 5b). The level of the splicing intermediates (E1 and IVS-E2) should increase and then decrease with time (11). This is most clearly seen with the IVS-E2 lariat RNAs (Fig. 5a, lanes 5–9 and Fig. 5b, lanes 1–4).

Fig. 5. *In vitro* splicing of the legumin J and legumin J-pMS1 hybrid intron. Splicing reactions were carried out on pre-mRNAs from pSP65LegJi linearised with Pst I (lanes 1–4) and pSP64LegJ/S1i linearised with Eco RI (lanes 5–8) for 15, 30, 60 and 105 minutes. The reactions were divided and separated on a 6% (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). Control splicing reactions without nuclear extract are shown for pre-mRNAs from pSP65LegJi (lane 9) and pSP64LegJ/S1i (lane 10). DNA sequence marker is run alongside the *in vitro* splicing products (lanes G and GA). W designates the well of the gel.

DISCUSSION

From the DNA sequence of the zein gene, pMS1, the mRNA produced from the P1 promoter contained 18 short reading frames between the 5' end of this mRNA and the initiator AUG codon. Some 5 to 10% of eukaryotic genes contain AUG codons or short reading frames in their 5' untranslated regions (19) which may influence the efficiency with which the mRNA is translated (20). It is possible that ribosomes may not recognise upstream AUG codons depending on their surrounding sequences (21), that translation is reinitiated after termination (22) or that the short reading frames in the 5' untranslated region have a role in translational control as in the yeast *GCN4* and *CPA1* genes (reviewed in Ref. 23). It is, however, unlikely that any of these mechanisms are involved in the zein system due to the number of AUG codons and short reading frames in the zein mRNA. Processing of the mRNA by cutting at a site in front of the initiator AUG is also unlikely because a recapping event would be required before translation of the mRNA could occur. The likeliest

explanation, therefore, of how the P1 mRNA could be translated, is that the P1 mRNA is spliced to remove the multiple short reading frames and bring the 5' cap to the initiator AUG (Fig. 6a).

S1-mapping has consistently produced two well-separated S1 signals in the P2 promoter region at positions –56 and –42 (4,7,17). An inherent difficulty in S1 and primer extension analyses when dealing with multigene families is that the RNA used for the experiments has been transcribed from many similar genes while the DNA probe is derived from a single cloned gene. Thus, it is not possible, from S1 analysis alone, to determine whether these two signals represent either 1) two single transcription start sites in two different subfamilies of genes, 2) two start sites in each gene, or 3) one transcription start site and one processing site. The S1 signals are unlikely to be due to the different subfamilies of 19kDa zein genes (z19ab, z19c and z19d) because of the sequence diversity in both the coding and 5' untranslated regions of these subfamilies (24). Furthermore, the same two signals were observed with RNA from sunflower callus tissue transformed with a zein gene coding for a 19 kD zein protein

Fig. 6. a) Schematic representation of the proposed splicing event in the mRNA transcript produced from the P1 promoter. The P2 mRNA is that transcribed from the P2 promoter beginning at position -56. The P1 pre-mRNA represents the transcript from the P1 promoter, the 5' end of which (dotted line) is spliced (dipped line) to position -39 to produce the P1 mRNA. b) Comparison of the 5' terminal sequences of P2 mRNA and P1-spliced mRNA based on a hypothetical splicing event in the P1 transcript from pMS1. The position of the splice site is arrowed and sequence homology is indicated by vertical lines.

via a Ti plasmid (25), showing that both signals could be produced from a single gene. If the two S1 signals represented two transcription start sites per gene then primer extension analyses must also show both signals. However, primer extension only corroborated the upstream S1 signal at position -56 but showed no correspondance to the S1 position at -42 (Fig. 2b) suggesting that this latter signal cannot represent a transcription start site.

The sequence of the major primer extension band (-56) was colinear with the pMS1 sequence while the sequence of the second major band (-57) was virtually colinear, differing by a single base change and being one nucleotide longer (Fig. 3). Clearly, due to the multigene nature of the zein gene family the sequences obtained from the primer extension products reflect the population of zein mRNAs extracted from the maize kernels. The single nucleotide difference in the -57 primer extension product has not been seen in any of the published genomic or cDNA 19kDa zein sequences (7,24,26-28) although it clearly represents a major mRNA species. A hypothetical splicing event using the first -GT dinucleotide downstream of the P1 promoter and the -AG dinucleotide at positions -40/-41 in pMS1 as the 5' and 3' ends of an intron respectively (Fig. 1b) would splice an 18nt exon to position -39 (Fig. 6b). The resultant spliced mRNA would have a 57nt 5' untranslated region and the spliced exon would be similar in sequence (11 out of 17nt) to the mRNA produced from the P2 promoter (Fig. 6b). It is possible, therefore, that the -57 primer extension product could represent a spliced mRNA although from the sequence it would have to be produced from 19kDa zein genes other than pMS1. The sequence of the longer primer extension product produced from the 5' untranslated region specific oligonucleotide primer showed colinearity with the pMS1 sequence only as far as position -39 which coincides exactly with the proposed 3' splice site. This signal may therefore represent a spliced zein mRNA species with a substantially different 5' end to the mRNA produced from the P2 promoter. Primer extension signals longer than the major signal at -56 were observed in all primer extension analyses but in none of the S1 analyses as would be expected if the 5' ends of the spliced mRNAs were not colinear with the S1-mapping DNA probe.

A number of plant introns have been successfully spliced in the HeLa cell *in vitro* splicing system (11,29-31). The splicing of the hybrid intron constructs containing the 5' end of the

legumin J intron cloned in front of the 3' end of the proposed intron of pMS1, in the HeLa cell nuclear extract was shown to be successful by the identification of the intermediate and final products of the splicing reaction (Fig. 5) (11). The successful splicing strongly supports the proposal that the zein sequence contains a functional 3' splice junction upstream of the coding region. The exact position of the 5' splice site in the zein sequence has not yet been determined in that the zein pre-mRNA was not spliced in the HeLa cell system. This can be explained by the 5' splice signal not being recognised in the HeLa system or that specific factors are needed for the splicing of the zein pre-mRNAs which are not present in the HeLa cell nuclear extract. In animal systems there are a number of examples of genes with two promoters and introns in their 5' untranslated regions (reviewed in Refs. 8,9). Examples such as the *Drosophila* alcohol dehydrogenase (32) and *Antennapedia* (33) genes, the mouse α -amylase (34) and *Thyl.2* glycoprotein (35) genes, the rat aldolase (36) and myosin light chain gene (37) produce mRNAs with different 5' sequences but with identical coding regions. In most cases, the differential transcription and alternative splicing which produces the mRNAs has been shown to be involved in the tissue specific and developmental specific expression of these genes. In plants there are as yet no known examples of alternative splicing and the zein system is the only example of a gene with more than one promoter. There are, however, five examples of plant genes with introns in their 5' untranslated regions: sucrose synthetase (maize) (38), *waxy* (maize) (39), *waxy* (barley) (40), phytochrome (oat) (41) and phytochrome (rice) (42). All five examples are of monocotyledonous origin and with the exception of the phytochrome genes are expressed almost exclusively in endosperm tissue which is also the site of zein gene expression.

The complexity of the zein multigene family has been seen in comparisons of cDNA and genomic DNA sequences where numerous point mutations, insertions, deletions and rearrangements have been observed (1,2,10,17,24,26-28). The variation in the sequences of the coding regions and flanking sequences is reinforced by the variation in the intergenic distances as envisaged by the differing lengths of the flanking sequences of cloned genomic zein genes (10,17,27,43). Thus, the positions of the promoters may have been altered relative to their structural genes due to insertions or deletions in their flanking regions, giving rise to the different zein-specific mRNAs. Indeed, a 600bp fragment of a zein 5' region containing the P1 promoter sequence has been found in a pMS2 clone, 2.8kb downstream of the pMS2 zein gene with no further zein coding sequence in the next 3kb to the end of the clone (43). The presence of a functional 3' splice site in front of the structural genes would maximise the possibility of splicing of the various precursor RNAs to a translatable form.

ACKNOWLEDGEMENTS

The excellent technical assistance of Mrs. E. Brutzer and Mrs. H. Jonas is gratefully acknowledged. This research was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

REFERENCES

1. Heidecker, G. and Messing, J. (1986) *Ann. Rev. Plant Physiol.* 37, 439-466.
2. Shotwell, M., Larkins, B. A. (1989). In Marcus, A. (ed): 'The biochemistry of Plants: A Comprehensive Treatise.' New York: Academic Press Vol XV, (in press).
3. Langridge, P. and Feix, G. (1983) *Cell* 34, 1015-1022.

4. Langridge, P., Eibel, H., Brown, J.W.S. and Feix, G. (1984) *EMBO J.* **3**, 2467–2471.
5. Langridge, P., Pintor-Toro, J.A. and Feix, G. (1982) *Mol. Gen. Genet.* **187**, 432–438.
6. Langridge, P., Pintor-Toro, J.A. and Feix, G. (1982) *Planta* **156**, 166–170.
7. Langridge, P., Brown, J.W.S., Pintor-Toro, J.A., Feix, G., Neuhaus, G., Neuhaus-Url, G. and Schweiger, H.-G. (1985) *Eur. J. Cell Biol.* **39**, 257–264.
8. Leff, S.E., Rosenfeld, M.G. and Evans, R.M. (1986) *Ann. Rev. Biochem.* **55**, 1091–1117.
9. Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987) *Ann. Rev. Biochem.* **56**, 467–495.
10. Brown, J.W.S., Wandelt, C., Feix, G., Neuhaus, G. and Schweiger, H.-G. (1986) *Eur. J. Cell Biol.* **42**, 161–170.
11. Brown, J.W.S., Feix, G. and Frendewey, D. (1986) *EMBO J.* **5**, 2749–2758.
12. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
13. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucleic Acids Res.* **12**, 7035–7056.
14. Mount, S.M. (1982) *Nucleic Acids Res.* **10**, 459–472.
15. Brown, J.W.S. (1986) *Nucleic Acids Res.* **14**, 9549–9559.
16. Dignam, J.D., Lebovitz, R. M. and Roeder, R.G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
17. Kriz, A.L., Boston, R.S. and Larkins, B.A. (1987) *Mol. Gen. Genet.* **207**, 90–98.
18. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* **50**, 349–383.
19. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
20. Peabody, D. S., Subramani, S. and Berg, P. (1986) *Mol. Cell Biol.* **6**, 2704–2711.
21. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
22. Peabody, D.S. and Berg, P. (1986) *Mol. Cell Biol.* **6**, 2695–2703.
23. Hinnebusch, A.G. (1988) *Trends in Genetics* **4**, 169–174.
24. Marks, M.D., Lindell, J.S. and Larkins, B.A. (1985) *J. Biol. Chem.* **260**, 16451–16459.
25. Matzke, M.A., Susani, M., Binns, A.N., Lewis, E.D., Rubenstein, I. and Matzke, A.J.M. (1984) *EMBO J.* **3**, 1525–1531.
26. Heidecker, G. and Messing, J. (1983) *Nucleic Acids Res.* **11**, 4891–4906.
27. Spena, A., Viotta, A. and Pirotta, V. (1983) *J. Mol. Biol.* **169**, 799–811.
28. Geraghty, D.E., Messing, J. and Rubenstein, I. (1982) *EMBO J.* **1**, 1329–1335.
29. Hartmuth, K. and Barta, A. (1986) *Nucleic Acids Res.* **14**, 7513–7528.
30. Van Santen, V.L. and Spritz, R.A. (1987) *Gene* **56**, 253–265.
31. Wiebauer, K., Herrero, J.-J. and Filipowicz, W. (1988) *Mol. Cell Biol.* **8**, 2042–2051.
32. Benyati, C., Spoerel, N., Haymerle, H. and Ashburner, M. (1983) *Cell* **33**, 125–133.
33. Scheuwly, S., Kuroiwa, A., Baumgartner, P. and Gehring, W.J. (1986) *EMBO J.* **5**, 733–
34. Hagenbüchle, O., Tosi, M., Schibler, V., Bovey, R., Wellaver, P.K. and Young, R.A. (1981) *Nature* **289**, 643–646.
35. Ingraham, H.A. and Evans, G.A. (1986) *Mol. Cell Biol.* **6**, 2923–2931.
36. Joh, K., Arai, Y., Mukai, T. and Hori, K. (1986) *J. Mol. Biol.* **190**, 401–410.
37. Periasamy, M., Strehler, E.E., Garfinkel, L.I., Gubits, R.M., Ruiz-Opazo, N. and Nadal-Ginard, B. (1984) *J. Biol. Chem.* **259**, 13595–13604.
38. Werr, W., Frommer, W.-B., Maas, C. and Starlinger, P. (1985) *EMBO J.* **4**, 1373–1380.
39. Klös gen, R.B., Gierl, A., Schwarz-Sommer, Z. and Saedler, H. (1986) *Mol. Gen. Genet.* **203**, 237–244.
40. Rohde, W., Becker, D. and Salamini, F. (1988) *Nucleic Acids Res.* **16**, 7185–7186.
41. Hershey, H.P., Barker, R.F., Idler, K.B., Murray, M.G. and Quail, P.H. (1987) *Gene* **61**, 339–348.
42. Kay, S.A., Keith, B., Shinozaki, K. and Chua, N.-H. (1989) *Nucleic Acids Res.* **17**, 2865–2866.
43. Quayle, T.J.A., Brown, J.W.S. and Feix, G. (1989) *Gene*, in press.