Sequence of a pseudogene in the legumin gene family of pea (*Pisum sativum* L.)

David Bown, Mark Levasseur, Ronald R.D.Croy, Donald Boulter and John A.Gatehouse

Department of Botany, Durham University, South Road, Durham DH1 3LE, UK

Received 4 April 1985; Revised and Accepted 29 May 1985

ABSTRACT

A second legumin gene, denoted Ψ Leg D, has been located on the pea genomic clone λ Leg 1, approx. 1.3 Kbases 3' of Leg A, in the same orientation. The complete sequence of Ψ Leg D shows that it is a pseudogene, having two stop codons near the 5' end of its predicted coding sequence, as well as deletions and frame shift errors when compared to Leg A. No transcripts from this gene could be detected in developing pea seeds. Leg A and Ψ Leg D are homologous over their coding sequences, and partially homologous in the intron sequences and the immediate 5' flanking sequences. Other flanking sequences of the two genes show no significant homology, apart from the presence of polyadenylation signals 3' to both coding sequences. The introns in the two genes occur in corresponding positions in the sequences, but a deletion in Ψ Leg D affects the 3' boundary of IVS-2. Hybridisation of Ψ Leg D to pea genomic DNA suggests that it does not represent a hitherto undetected sub-family of legumin genes.

INTRODUCTION

Legumin, a major seed storage protein in pea (*Pisum sativum* L.) is encoded by a small gene family containing 4 members encoding the "main" or "major" legumin polypeptides (1,2). Related gene families containing genes encoding "minor" legumin, or legumin-like, polypeptides are also present, again at small copy number (3,4). The isolation of a genomic clone containing a legumin gene, Leg A, and the complete coding and immediate 5' and 3' flanking sequences of Leg A have been previously reported (2). Leg A encodes a major legumin polypeptide, and is probably actively expressed *in vivo* since its sequence agrees with the determined legumin protein sequence and with the sequence of a legumin cDNA. Leg A is also transcriptionally active *in vitro* (5).

Non-functional genes, or pseudogenes are found in many gene families in eukaryotes. Several distinct types of pseudogene have been observed. One type appears to have been produced by reverse transcription of a mRNA and reinserion of this sequence into the genome, since it lacks all introns (6).

This type has been termed "processed" genes. A second type is intact apart from small errors in the reading frame (generally producing stop codons) or in intron-exon boundaries (7). A third type contains relatively large deletions and/or insertions. These latter two types of pseudogenes are thought to represent genes that were active at one stage of evolution, but then either suffered damage or became disadvantageous to the organism, and so were silenced. Fewer pseudogenes have been observed in plants than in animals, but both the soybean leghaemoglobin and the maize zeine gene families contain examples of the second or third type (8,9).

The present paper reports the complete sequence of a pseudogene in the legumin gene family of pea; this pseudogene is found in tandem with an active gene.

MATERIALS AND METHODS

Materials

The genomic clone λ Leg 1 containing legumin genes Leg A and Ψ Leg D was isolated as previously described (2). Restriction fragments of λ Leg 1 were subcloned into the plasmid vector pUC 8 by standard techniques; these plasmid subclones were then used as a basis for all further cloning and sequencing operations. Plasmid subclones were designated pDB 6.01, 6.03 and 6.06 as shown in fig. 1. M13 sequencing vectors mp8 and mp9 were the kind gift of Dr. J. Messing, as was the bacterial strain *E. coli* JM 101.

Restriction enzymes and DNA modifying enzymes used in this paper were routinely obtained from Bethesda Research Laboratories (UK) Ltd. (BRL). Deoxy- and dideoxynucleotides were from PL Biochemicals. The M13 sequencing primer used was from BRL and the hybridisation probe primer was from New England Biolabs. All radiochemicals were purchased from Amersham International Ltd.

Methods

Preparation of M13 subclones. Inserts from the λ Leg 1 subclones pDB 6.01, 6.03 and 6.06 (pDUB 22, pDUB 34 and pDUB 35 respectively) were isolated by restriction of the plasmids with the appropriate enzymes, separation of plasmid and insert by gel electrophoresis on a low melting temperature agarose (BRL LGT agarose) gel, excision of the insert bands and purification of the DNA therefrom by phenol extraction and ethanol precipitation after melting the agarose (10). Inserts were then treated with further restriction enzymes, where required, and ligated into M13 mp8 or mp9 replicative form that had been first treated with the appropriate restriction enzyme(s). The

ligation mixture was then used to transform *E. coli* JM 101, and the infected cells were plated with exponentially growing cells by standard protocols (11). Single stranded recombinant DNA species were prepared from putative recombinant phages (clear plaques) by standard methods (12).

Screening M13 subclones. M13 subclones were screened by using "dot-blot" assays. Approx. 0.5 μ g samples of single stranded M13 clones were spotted onto nitrocellulose sheets (type BA85; Schleicher and Schuell) by means of a Hybridot apparatus (BRL), according to the protocol supplied with the apparatus. The resulting dot-blots were vacuum dried for 90 min. at 85°C prior to use in hybridisation assays. Labelled DNA for use as hybridisation probes was prepared either by nick translation with α -(32 P) dCTP for double stranded probes (13), or by primer extension on M13 clone templates (14) to produce partially single stranded probes, again using α -(32 P) dCTP as radioactive label. The latter method was more useful in practice once M13 clones of the whole insert in both orientations had been obtained, as it allowed the sequencing direction of the subclones to be determined.

Sequencing. M13 subclones were sequenced by the dideoxy chain termination method (15), using α -(35 S) dATP as radioactive label (16). Sequencing reactions were analysed by electrophoresis of 6% polyacrylamide gels (17); after electrophoresis, gels were dried down and autoradiographed at -80°C using flashed X-ray film (Fuji RX).

Purification and analysis of nucleic acids. Total RNA was prepared from the cotyledons of developing pea seeds harvested at different developmental stages by the method of Hall et al. (18), and analysed by glyoxalation and agarose gel electrophoresis (19). The gels were blotted onto nitrocellulose, and the resulting filters were hybridised to labelled DNA probes as described by Thomas (20). Pea genomic DNA was prepared from pea leaves as described by Graham (21) and analysed by digestion to completion with restriction endonucleases, electrophoresis on agarose gel, transfer to nitrocellulose and hybridisation of the resulting filter to labelled DNA probes to a stringency of 0.1 xSSC at 65°C (22). Nitrocellulose filters were autoradiographed as above, but with the addition of an intensifying screen (duPont Lightning Plus).

RESULTS

Partial Restriction Map of λ Leg 1

A partial restriction map of the genomic clone λ Leg 1 is shown in Fig. 1, and the deduced positions of the sequences hybridising to a legumin cDNA probe is indicated. Ψ Leg D is situated approx. 1.3 Kbp 3' of the end of the coding

Fig. 1. Partial restriction map of the pea genomic clone λ Leg 1. Bars represent areas corresponding to the coding sequences plus introns of Leg A and Ψ Leg D. Arrows a, b and c indicate fragments used as probes on pea genomic DNA; 6.01, 6.03 and 6.06 indicate subclones used for sequencing. Restriction sites: B, Bam HI; E, Eco RI; H, Hind III; X, Xho I; S, Sst I.

sequence of Leg A. Subsequent sequencing showed that the two genes were in the same orientation; this could not be established from the restriction map alone as the two genes were not sufficiently similar at that level. No other fragments hybridising to legumin cDNA probes were present on this genomic clone.

Sequence of Ψ Leg D. The sequencing map of Ψ Leg D is given in fig. 2, and the complete nucleotide sequence from approx. 700 bp upstream of the apparent

Fig. 2. Sequencing map of Ψ Leg D and flanking sequences. Restriction sites: A, Alu I; C, Acc I; H, Hind III; P, Pst I; R, Rsa I; S, Sau 3AI. Duplicated sequencing runs carried out to check sequences are not noted.

start of coding sequence to approx. 200 bp downstream of the apparent end of the coding sequence is given in fig. 3; the sequence of Leg A is also given over those regions where significant homology between the two genes exists.

The most significant features of Ψ Leg D when compared to Leg A are those that render the former gene unlikely to function to produce a legumin polypeptide. Although the two genes contain coding sequences with strong homology at the 5' end, there are two in-frame stop codons in Ψ Leg D at positions corresponding to amino acids 6 and 29 in the polypeptide encoded by Leg A. In addition, there is an error in the reading frame of Ψ Leg D (single base deletion) at amino acid 150 of Leg A, and the Ψ Leg D coding sequence is interrupted by three deletions when compared to Leg A. The first of these deletions abolishes the 3' intron/exon boundary at IVS-2 and amino acids 159 - 205 of Leg A, although extra intron sequence appears to be present in Ψ Leg D (divergence of the two sequences commences approximately half way through IVS-2 in Leg A). Uncertainty about the potential 3' intron boundary in Ψ Leg D makes it impossible to rule out a potential coding sequence that would span this deletion, however. The second deletion removes amino acids 251 - 272 of Leg A. This deletion maintains the correct reading frame; it occurs in the region of Leg A that contains sequence repeats, but does not correspond to the unit length of the repeat, or a multiple of it. A third deletion removes amino acids 406 - 439 of Leg A, and introduces a further reading frame shift error, with an extra base in Ψ Leg D.

Apart from these differences, there is a close agreement of the two sequences, with a 1 : 1 correspondence of amino acids, apart from an extra amino acid in the leader sequence of Ψ Leg D, between positions -19 and -20 in Leg A and a deletion of amino acid -17 in the Leg A leader sequence. Of a total of 160 (out of 1554) base changes between the comparable parts of the coding sequences of Leg A and Ψ Leg D, there is a slight excess of active base changes as opposed to silent changes (84 active, 76 silent).

The intron sequences show varying degrees of homology; IVS-1 is most homologous, at 49/88 identical bases, IVS-2 is homologous at its 5' end (34/40 bases) but then diverges completely, as described above, and IVS-3 is divergent, with no homology apparent except in the first three and last 15 bases of the intron. IVS-1 and IVS-3 are both slightly shorter in Ψ Leg D than in Leg A; IVS-2 cannot be compared. Where the intron boundaries can be assigned, those of Ψ Leg D, as in Leg A, obey the Breathnach-Chambon rule (23).

The 3' flanking sequences of the two genes show a complete divergence immediately after the stop codon. Although Ψ Leg D has three potential poly-


```

      E       H C       N H       I       Q
LegD AGAGGAGAGGAGGAGAGGAGCAGAGAGGCGCCACTGCAGCCAAAAAGGCAAAAGCAGATGGCATGGAGACAATGGGCTTGAGGAAACCATTTGCACGGCCAAACTTCGGCAGAAC 1055
LegA AGAGGAGAGGAGGAGGAGAGGAGCAGAGAGGCGCGCGCAGCGCAAAAAAGGCAAAAGCAGAGGCAAAAGGCAATGGGCTTGAGGAAACCATTTGCACGCTGCTAACTTCGATTTGAACT 1247
A.A. R G E E E E E D K K E R G G S Q K G K S R R Q G D N G L E E T V U C T A K L R L N

      S       N       U       L F       I
LegD ATTGGTTCATCTTCATCACCATAACATCTACAACCTGTTGCTGGTAGAATCAAAACTGTACAGCCTTGACCTCCCACTTTTCAGATGGCTCAAACTAATGCTGAGCATGGATCTCTC 1175
LegA ATTGGCCCTCTTCATCACCAGACATCTACAACCTGTAAGTGGTAGAATCAAAACTGTACAGCCTTGACCTCCCACTTTTCAGGTTGCTCAAGGCTCAAACTAAGTGGATGGATCTCTC 1367
A.A. I G P S S S P D I Y N P E A G R I K T V T S L D L P V L R W L K L S A E H G S L

      <-----IVS-3----->
LegD CACAAGTACCATCATTTCTCTTTCTTTCTTTCTCTTTTATTACTTTT-----AATTTATTTCCATGACTTAATTTCTATGTCAAAACATTTTCATACAGATGCTATGTTCTGTG 1288
LegA CACAAGTATGTTTTTTCATCATTTAATTTGTTTTTCCATGAATCAATTTCTATGTGCAACTATGTTTGGAGATAATAGCTAACTCATTACAATCTTCATACAGATGCTATGTTGTTG 1487
A.A. H K <-----IVS-3-----> N A M F V

      C U       T       I       S       +
LegD CCTCCTACAACTCAACGCAAACTGGCTATATATACACACTGAAAGGAGCTGCAAGGCTACAATTTGTGAATGCAATGGCAACACTGTGTCTGATGGAG----- 1387
LegA CCTCCTACAACTGATGCAAGAGTATATATATACGATTGAAAGGAGCTGCAAGGCTACAAGTGAATGCAATGGCAACACTGTGTGATGAGAGCTAGAGCCGAGCTGTG 1487
A.A. P H Y N L N A N S I I Y A L K G R A R L Q V U N C N G N T V F A G E L E A G R A

      K       A       T
LegD -----DELETION-----AAAAGCTGCAATGCCAGGCTGCAGGGACATCTCCACG 1427
LegA TTGACAGTGCACAAACTATGCTGTGGCTGCAAGTCACTAAGGAGACAGGTTCTCATATGTAGCATTCAAGACCAATGATAGAGCTGGTATTTGCAAGACTGCAGGGACATCATCAGTT 1727
A.A. L T V P Q N Y A V A A K S L S D R F S Y V A F K T N D R A G I A R L A G T S S V

      L A M U       I       S       U       P
LegD CTAAATGCTATGCCAGTGTATATAGCCGCTCATTCACATTCGAGAGGAGTGAAGCTAGGCAAGTCCCAACATCTCTTCAATTTTCTAGTTCCTCGTGGTGCAGAAAC 1547
LegA ATAAATATCTGCCGTTGGATGTGGTTGCAGCTACATTCACATTCGAGAGGAGTGAAGCTAGGCAAGTCCCAACATCTCTTCAATTTTCTAGTTCCTCGTGGTGCAGAAC 1847
A.A. I N N L P L D V U A A T F N L Q R N E A R Q L K S N N P F K F L V P A R E S E N

      K
LegD AAAGCTTCTGCTAGAACAAACACTGCTTCGAGGCTTTTGTGTTGAGAGACATGTATCCACCCCAACTGGTAAATTAAGAGTACTTATGAATAAAAAGGTTTGTTTCTTTTGG 1667
LegA AGAGCTTCGGCTTAGATTGTCGCAACATCAATGAAGTAAATGAAGAAAACTAAGGCTTAGATGCTTTGTACTGTTGTAATTAACCTCGAGTCATGCTATCTTTTAGCGGAAACAG 1967
A.A. R A S A Amb>Stop.....*Poly A Site1.....vPoly A Site2.....

      .....vPoly A Site3.....
LegD AAACATGTATAGTACTATCTGAGTCACGAACCTTTTCGAAAGTGAATAAAAAGTAAAGTGAATTTGAGCATACAAGTTTATTTGATC (1757) .....

```

Fig. 3. Sequence of Ψ Leg D and flanking regions. The sequence of Leg A (2) is also given over those regions where the genes are homologous. The complete amino acid sequence encoded by Leg A is given; residues differing in Ψ Leg D are given above the sequence. Stop codons are noted. Δ = base deleted in Ψ Leg D; + = extra base in Ψ Leg D.

adenylation signals in its immediate 3' flanking sequence, none of these correspond to the polyadenylation signals in Leg A, and only the first is of the multiple overlapping type found in Leg A. In the 5' flanking sequence, the transcriptional start found in Leg A is present in Ψ Leg D, although there is a 7 base deletion in the latter between the transcription start and the start of the coding sequence. There is also significant sequence divergence between the two sequences in this region, which might affect the potential translational efficiency of a Ψ Leg D transcript. The first promoter sequence of Leg A, the "TATA" box is present in Ψ Leg D, in a form agreeing with the consensus sequence for this element (24). However, the sequence at the "CAAT" box is diverged from Leg A to such an extent that the Ψ Leg D sequence

no longer fits the "CAAT" consensus sequence (25), although homology in the "AGGA" box region (24) is significantly better (14/16 vs. 6/10). Sequence homology between the two genes is significant as assessed by a dot matrix comparison method at 7/10 nucleotide limit to approx. -145 bases relative to the Leg A transcription start; upstream beyond this the sequences diverge completely.

Hybridisation of ΨLeg D to mRNA. Total RNA was isolated from developing seeds glyoxalated, electrophoretically separated on agarose gel and transferred to nitrocellulose ("Northern" blot). When the Hind III fragment containing the coding sequence of ΨLeg D (a, fig. 1) was used as a probe on these blots, hybridisation only occurred to an mRNA species of approx. 2000 bases, corresponding to legumin mRNA (predicted size approx. 1800 bases); no additional bands which might correspond to ΨLeg D transcripts (predicted sizes approx. 1400 - 1600 bases) were observed (Fig. 4). The relative intensities of the hybridised bands from RNA preparations at different stages of cotyledon development were consistent with previous estimations (26), again suggesting that only mRNA species derived from Leg A type genes were being detected. When a fragment of 3' flanking sequence (Hind III - Sst I, probe c fig. 1) containing ΨLeg D sequence that would be transcribed but not translated was used as a probe, no hybridisation to any RNA species could be detected. Leg A and ΨLeg D are divergent in this region, and the probe is therefore specific for ΨLeg D (or very similar) transcripts. The absence of detectable levels of Leg D transcripts is thus confirmed (results not shown).

Hybridisation of ΨLeg D to Genomic DNA. Coding and 5' flanking sequences of ΨLeg D (Hind III fragments a and b, fig. 1) were hybridised separately to pea genomic DNA which had been cleaved with restriction enzymes, separated by electrophoresis on agarose gel, and transferred to nitrocellulose. The coding sequence probe, Hind III fragment a, hybridised to a matching fragment (1.5 kb) in genomic DNA at an intensity indicating the genomic fragment was present as a single copy in the genome. It also hybridised to the other legumin genomic fragments detected by Leg A- type cDNA probes in genomic DNA (2.5 kb, 4.1 kb) at lower intensity. In Eco RI and Bam HI digests of pea genomic DNA hybridisation only occurred to fragments previously detected by Leg A - type cDNA probes (results not presented). Similar results were obtained when the 5' flanking sequence, Hind III fragment b, was used as a probe, although in this case in the Hind III genomic DNA digest fragments of 0.8 kb (the matching fragment) 2.7 kb, 3.8 kb and 4.1 kb were detected.

Fig. 4. Hybridisation of Ψ Leg D probe (Hind III fragment a, fig. 1) to total RNA prepared from pea cotyledons. Tracks 1,2 : 10, 5 μ g (respectively) RNA from cotyledons 14 d.a.f. (days after flowering; tracks 3,4 : 10 5 μ g RNA from cotyledons 18 d.a.f.; tracks 5,6 : 10, 5 μ g RNA from cotyledons 12 d.a.f.

DISCUSSION

The gene described in this paper has been assumed to be a pseudogene on the basis of its "damaged" coding sequence; the presence of the stop codons near the 5' end of the coding sequence would prevent any functional polypeptides being produced from it. Further, the amino acid substitutions (compared to Leg A) encoded by Ψ Leg D are not observed in the protein sequence of legumin, or in the sequences of legumin cDNA species determined to the present; some of the predicted changes would be expected to have significant effects on the structure of the protein. In agreement with this assignment of pseudogene status, it is unlikely that significant amounts of processed mRNA are produced from Ψ Leg D, since no distinct RNA corresponding in size to a

predicted Ψ Leg D mRNA (approx. 1400 bases) could be detected in RNA preparations isolated from tissues where legumin genes are expressed at high levels. However, transcripts from this gene cannot be absolutely ruled out since it is still possible for very low levels of transcription to occur, or for the gene to be transcribed into a rapidly degraded RNA. The failure to detect expression of this gene may be a result of the changes to the "CAAT" box (compared to Leg A) which alter this promoter sequence so that it no longer fits the consensus sequence, and may thus be inactive. The "CAAT" box is known to be important for expression of eukaryotic genes *in vivo* (27, 28), although such effects have not been demonstrated for legumin genes. The homology between Leg A and Ψ Leg D does not extend far enough in the 5' direction to include the putative "enhancer" sequence suggested by Lycett et al. (2), nor is there strong evidence for any significant conserved sequences beyond -145 bases relative to the transcription start when the two sequences are compared by a dot-matrix test (data not presented).

The results of probing genomic DNA with Ψ Leg D probes suggest that this gene is not a representative of a hitherto undetected family of legumin-like genes, since only the Eco RI fragments of genomic DNA which hybridise to Leg A -type probes are detected by these probes. The results do not rule out the possibility that other members of the Leg A sub-family of genes have Leg D type genes attached to them, although neither of the other two genomic clones containing Leg A type genes so far isolated contains an analogous pseudogene sequence (R.R.D. Croy, unpublished data). If other Ψ Leg D genes are present, they may be active, although the mRNA hybridisation experiments make this an unlikely possibility. Ψ Leg D is nearly complete; it is not a pseudogene of the "processed" type, nor is it so truncated as to represent duplication of part of a functional gene. It is therefore possible that Leg A and Ψ Leg D were a tandem duplication of the same gene, and functioned as a coupled transcription unit, since they are located close enough to each other to come under the control of the same long-range control sequences, even though each gene has its own independent short-range promoter sequences. Since the introns and flanking sequences of Leg A and Ψ Leg D are less homologous than the exons it is likely that the two genes evolved independently while both were still functional. The sequence data support the conclusion that Ψ Leg D diverged from Leg A by the accumulation of deletions, rather than by Leg A gaining extra sequence by internal duplications leading to repeats, since the deletions in Ψ Leg D do not correspond to sequence repeats in Leg A. The third deletion in Ψ Leg D may have occurred via incorrect recombination,

since the sequence AGAGCT occurs very near the 5' and 3' ends of the deleted section in Leg A; it is possible that the other deletions occurred via similar mechanisms also. ¶Leg D thus seems to represent an "old" legumin gene that was silenced, either because of damage or because of selection against its encoded polypeptide.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. P. Preston for technical assistance, and Dr. A. Shirsat for help in preparing the restriction map shown in fig. 1. This work was supported by a grant from the Agriculture and Food Research Council.

REFERENCES

1. Croy, R.R.D., Lycett, G.W., Gatehouse, J.A., Yarwood, J.N. and Boulter, D. (1982) *Nature* (London) 295, 76-79.
2. Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984) *Nucleic Acids Res.* 12, 4493-4506.
3. Matta, N.K., Gatehouse, J.A. and Boulter, D. (1981) *J. Exp. Bot.* 32, 1295-1307.
4. Domoney, C. and Casey, R. (1985) *Nucleic Acids Res.* 13, 687-699.
5. Evans, I.M., Bown, D., Lycett, G.W., Croy, R.R.D., Boulter D. and Gatehouse, J.A. (1985) *Planta*, in press.
6. Vanin, E.F. (1984) *Biochim. Biophys. Acta* 782, 231-241.
7. Little, P.F.R. (1982) *Cell* 28, 683-684.
8. Brisson, N. and Verma, D.P.S. (1982) *Proc. Natl. Acad. Sci. USA*, 79, 4044-4059.
9. Spena, A., Viotti, A. and Pirrotta, V. (1983) *J. Mol. Biol.* 169, 799-811.
10. Kuhn, S., Anitz, H.J. and Starlinger, P. (1979) *Molec. Gen. Genet.* 167, 235-241.
11. Messing, J. (1983) *Methods Enzymol.* 101, 20-78.
12. Schreir, P.H. and Cortese, R. (1979) *J. Mol. Biol.* 129, 169-172.
13. Amersham International, Amersham. (1980) *Technical Bulletin* 80/3.
14. Hu, N. and Messing, J. (1982) *Gene* 17, 271-277.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
16. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3963-3965.
17. Sanger, F. and Coulson, A.R. (1978) *FEBS Letts.* 87, 107-110.
18. Hall, T.C., Ma, Y., Buchbinder, B.U., Pyne, J.W., Sun, S.M. and Bliss, F.A. (1978) *Natl. Acad. Sci. USA* 75, 3196-3200.
19. McMaster, G.K. and Carmichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4835-4838.
20. Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5202-5205.
21. Graham, D.E. (1978) *Anal. Biochem.* 85, 609-613.
22. Southern, E.M. (1979) *Methods Enzymol.* 68, 152-176.
23. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4853-4857.
24. Messing, J., Geraghty, D., Heidecker, G., Hu, N., Kridl, J. and Robenstein, I. (1983) In "Genetic Engineering of Plants", Kosuge, T., Meredith, C.P. and Hollaender, A. eds., pp. 211-227.

25. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids, Res. 8, 127-142.
26. Gatehouse, J.A., Evans, I.M., Bown, D., Croy, R.R.D. and Boulter, D. (1982) Biochem. J. 208, 119-127.
27. Mellon, P., Parker, U., Gluzman, Y. and Maniatis, T. (1981) Cell, 27, 279-288.
28. Shaw, C.H., Carter, G.H., Watson, M.D. and Shaw, C.H. (1984) Nucl. Acids. Res. 12, 7831-7846.