The complete nucleotide sequence of a legumin gene from pea (*Pisum sativum* L.)

Grantley W.Lycett*, Ronald R.D.Croy, Anil H.Shirsat and Donald Boulter

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

Received 1 March 1984; Revised and Accepted 2 May 1984

ABSTRACT

One of several genes coding for the major pea storage protein, legumin, has been completely sequenced. The sequence covers the whole of the transcribed region, plus 5' and 3' untranscribed sequences. The predicted protein sequence starts with a signal peptide and is followed by the legumin α polypeptide sequence of 36.44kd and the β polypeptide sequence of 20.19kd. Compared to other legume storage proteins, the α and β polypeptide sequences encoded by this legumin gene, which contain 3 met and 5 cys residues, are relatively rich in the sulphur amino acids. The coding sequence is interrupted by three introns which show boundary sequences typical of higher plant genes. The 5' end of the gene sequence contains a 'TATA box', a 'CAAT box' and a sequence showing some homology to an 'AGGA box'. An extra sequence, identical to the normal polyadenylation signal of the legumin message is seen in the 3' untranscribed region. The structure of the gene and the possible significance of secondary structures in the nascent RNA transcript in affecting the choice of polyadenylation site is discussed.

INTRODUCTION

Legumin, one of the major storage proteins of pea (*Pisum sativum*) seeds, is a hexameric protein. Each of the six constituent monomers consist of a 40 kilodalton (kd) acidic (α) polypeptide and a 20kd basic (β) polypeptide linked by disulphide bonds (1,2). These two polypeptides are generated by post-translational proteolysis of a 60kd prolegumin precursor polypeptide (3,4,5,6).

Legumin is found in the seeds of many leguminous (7) and non-leguminous plants. The important crop plants soyabean (*Glycine max*), broad bean (*Vicia faba*), oat (*Avena sativa*), and rice (*Oryza sativa*) have all been shown to contain proteins of similar subunit structure, which show sequence homology to pea legumin and which are also derived from precursors (2,8,9,10,11,12, 13). Legumin is important in leguminous crops such as peas, broad beans and soyabeans because it is a major storage protein and source of sulphur amino acids in seed meals (7). Sulphur amino acid levels are the main nutritional limitation of legume seeds and the production of lines which

produce seeds containing a high proportion of sulphur amino acid rich legumin is agronomically desirable.

The legumin group of proteins is therefore a candidate for in vitro crop improvement by genetic manipulation. Whilst it is known that heterogeneity at the level of the protein (14) and mRNA (15) is reflected by the presence of several genes (5), little is known about the structure of these genes except for what has been deduced from pea legumin cDNA sequences (5,15) and electron micrographic studies of soyabean legumin genes (17).

Knowledge of legumin gene structure may throw further light upon the control of seed protein synthesis, since there is good evidence that control is primarily at the transcriptional level in pea (18,19,20) and soyabean (21,22). Several sequences thought to be involved in transcription have been identified in animal systems (see 23), but it has been suggested that plant genes may differ in some minor respects (24,25). In order to obtain a better understanding of the legumin gene structure and control we have cloned several legumin genes from pea (16). Here we report the complete nucleotide sequence of one legumin gene, including the whole protein coding region, three introns and the 5' and 3' flanking sequences.

MATERIALS AND METHODS

Materials

Deoxyribonuclease I (DPFF) was obtained from Worthington Biochemicals (Millipore UK Ltd., London) Bovine alkaline phosphatase, endonuclease free DNA polymerase I and T4 polynucleotide kinase were from Boehringer Corporation Ltd. (London UK). Restriction endonucleases were from Boehringer Corporation Ltd., Bethesda Research Laboratories (Cambridge UK) or New England Biolabs (CP Laboratories Ltd., Bishop's Stortford, UK). Dideoxynucleotide triphosphates were from P-L Biochemicals Ltd., (Northampton UK) and γ -³²P ATP (5kCi mmol⁻¹) was from Amersham International Ltd. (Amersham UK).

Genomic clones

Full details of cloning and isolation of λ leg 1 and other legumin genomic clones from gene libraries of Pisum sativum, cv 'Feltham First' are described elsewhere, as is the construction of the sub-clones pDUB21 and pDUB24 (16).

DNA sequencing

Double stranded 5' end-labelled DNA fragments were prepared as described

by Maxam and Gilbert (26) except that 5' protruding ends were labelled by incubating phosphatase treated DNA for 45 min at 37°C with 130 μCi of $\gamma^{32}\text{P}$ ATP (3 kCi mmol^{-1}) and 30 units of polynucleotide kinase in 25 μl of 5 mM tris-HCl pH 7.6, 10 mM MgCl_2 , 5 mM dithiothreitol, 0.1 mM EDTA, 0.7 mM spermidine. The end-labelled fragments were sequenced by an enzymatic nick-translation method: the 'forwards-backwards' method of Seif *et al.* (27). The only modifications were that endonuclease free polymerase was used, DNase I (1 mg ml^{-1}) was added to the reaction mixtures to give more controlled nicking, and each backwards mixture was pooled with the corresponding forwards mixture prior to lyophilisation and analysis.

DNA sequences were analysed both by eye and by employing a standard computer programme (28). RNA secondary structure predictions were computed with the aid of a programme produced by Zuker and Stiegler (29).

S1 nuclease mapping

End labelled DNA fragments were prepared as above. Hybridisation and S1 treatment were based on the methods of Falvaloro *et al.* (30) and Pedersen *et al.* (31). The amount of labelled DNA fragments prepared from 1 μg of plasmid DNA (~ 40 ng) and 2 μg of poly A⁺ RNA from developing pea cotyledons were used for each hybridisation together with 20 μg *E.coli* tRNA carrier. Control reactions contained no poly A⁺ RNA. The nucleic acids were ethanol precipitated and resuspended in 20 μl of 80% deionised formamide, 0.4 M NaCl 40 mM PIPES (pH6.4), 1 mM EDTA. The DNA was denatured at 80°C for 5 min then incubated at 48.5°C for 6 hr. The samples were diluted with 200 μl of ice-cold 50 mM NaCl, 30 mM NaOAc (pH4.6), 1 mM ZnSO_4 , 5% glycerol containing 2500 units or 5000 units of S1 nuclease and incubated at 37°C for 60 min. The samples were then chilled and ethanol precipitated following the addition of a further 20 μg of tRNA carrier. Six μl of 80% formamide, 10 mM NaOH, 1 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue were added and the samples were loaded onto a normal sequencing gel alongside four sequencing tracks prepared as above.

RESULTS AND DISCUSSION

Cloning and Sequencing

The cloning and isolation of several pea genomic fragments coding for legumin in lambda vectors is described elsewhere (16). λ Leg 1 carries a 13.5kb EcoRI genomic fragment inserted into the vector λ gt WES λ B (32). Based on results from restriction mapping and southern blotting analysis using pea legumin cDNAs pDUB6 and pDUB8 (15), gene sequences from λ Leg 1

Figure 1

Restriction map of *legA* gene region. The figure shows: 1. the restriction sites and a scale of kilobases, 2. and 3. the regions covered by subclones pDUB24 and pDUB21, 4. the region shown in fig.2 with coding regions as heavy lines, 5. the sequencing strategy. Symbols for restriction enzyme cleavage sites are: A = *AccI*, B = *BamHI*, C = *HincII*, D = *NdeI*, G = *BglIII*, H = *HindIII*, L = *BglI*, N = *BstNI*, O = *XhoII*, P = *PstI*, V = *AvaI*, W = *AvaII*, X = *XhoI*.

were sub-cloned into pUC8 for the purpose of DNA sequencing. Two of these sub-clones pDUB21 and pDUB24 were restriction mapped and sequenced. A restriction map of the two sub-clones is shown in fig.1 with the sequencing strategy. A diagrammatic representation of the gene, designated as the *leg A* gene to distinguish it from another legumin gene found on the same λ clone (16), is included for clarity.

The sequence of the *leg A* gene (fig.2) extends into the 5' and 3' untranscribed regions, covers the whole protein coding region and contains three intervening sequences (IVS-1,2,3). The gene sequence shows strong homology to several previously sequenced cDNAs (5,15) and in particular shows 100% homology to pDUB8. Homology to pDUB6 is 99.4% and to pDUB3/1 98.3% excluding the large deletion in pDUB3 (15). Thus, whilst pDUB8 is identical to the *leg A* gene and may be ultimately derived from it, pDUB6 is quite different and must have come from a different gene. The degree of homology between these two cDNA clones was previously uncertain due to the small length of overlap between them. The *leg A* gene, like pDUB6 and pDUB8, contains a region of coding sequence consisting of three direct repeats (15). Since the most of this region is missing from the cDNA clone pDUB3, it has been proposed that there may be two classes of legumin gene in peas: one with the repeats and one without (15). The *leg A* sequence confirms that the repeats are present at the gene level and are not cDNA cloning artefacts.

[illegible]

Figure 2

The sequence of the legA gene. The predicted amino acid sequence is shown underneath the DNA sequence. Introns and other features discussed in the text are denoted [- - -]. Inverted repeats discussed in the text are shown by the symbols >>>> and <<<<. The termini of the mature α - and β -subunits are marked by brackets ().

Preliminary sequence data from a second but different legumin gene (leg C, data not presented) confirms that the repeats are present in more than one gene, but it is still not certain whether genes lacking the repeats exist in peas.

Predicted Protein Sequence

The protein sequence has been predicted from the gene sequence and is shown in fig.2. Much of the legumin sequence has been obtained previously from direct protein sequencing and from several different cDNA clones (5,15) but this is the first complete sequence including the N-terminal region of the α -subunit from pea. Comparison of our predicted sequence with the N-terminal α -polypeptide amino acid sequence determined by Casey et al. (33) shows a match of 23 out of 25 amino acids, the two differences being accounted for by two single base substitutions in the gene. As with other seed proteins of pea (34,35) and other plant species (36,37,38,39,40,41) there is a short peptide sequence rich in hydrophobic residues beyond the N-terminus of the mature protein, which is presumably a signal peptide (42). Unlike the clear evidence for pea vicilin leader sequences (43,44,45,34), the presence of signal peptides in pea legumin precursor polypeptides has not been demonstrated previously by cDNA cloning and sequencing, though one in vitro synthesis study may be interpreted as showing such a signal peptide (46). Leader sequences have been demonstrated in legumin precursors from other legumes, namely Vicia faba (47) and Glycine max (48) synthesized by 'in vitro' protein synthesis systems.

In order to predict the accurate size and composition of the protein subunits produced by the leg A gene, it is necessary to know the exact site of post-translational proteolysis. It was recently suggested by analogy with animal peptide prohormone processing that legumin might be post-translationally cleaved between the paired basic residues, five amino acids upstream from the N-terminus of the β -subunit, as well as adjacent to the N-terminus, leading to the removal of a five residue peptide (5). More recently, the C-terminal V8 peptide of an α -subunit of legumin was isolated and identified by its N-terminal sequence and amino acid composition. This peptide was shown to extend to the asparagine residue adjacent to the N-terminus of the β -peptide (15). Carboxypeptidase A digestion has confirmed that asparagine is the C-terminal residue (J. Gilroy pers.comm.). It is clear therefore, that the cleavage of at least some pea legumin precursors occurs at a single site, in contrast to a report of the removal of a linking peptide between the A and B subunits of soyabean legumin (49). This single

Table 1 Predicted amino acid composition of legA gene product

Amino acid	<u>α subunit</u>		<u>β subunit</u>		<u>$\alpha+\beta$ subunits</u>	
	<u>no.of</u> <u>residues</u>	<u>% of total</u>	<u>no.of</u> <u>residues</u>	<u>% of total</u>	<u>no.of</u> <u>residues</u>	<u>% of total</u>
Ala	12	(3.9)	22	(11.9)	34	(6.9)
Arg	35	(11.3)	13	(7.0)	48	(9.7)
Asn	20	(6.4)	18	(9.7)	38	(7.7)
Asp	18	(5.8)	6	(3.2)	24	(4.8)
Cys	3	(1.0)	2	(1.1)	5	(1.0)
Gln	33	(10.6)	4	(2.2)	37	(7.5)
Glu	43	(13.8)	9	(4.9)	52	(10.5)
Gly	27	(8.7)	10	(5.4)	37	(7.5)
His	7	(2.3)	3	(1.6)	10	(2.0)
Ile	14	(4.5)	7	(3.8)	21	(4.2)
Leu	15	(4.8)	21	(11.4)	36	(7.3)
Lys	13	(4.2)	9	(4.9)	22	(4.4)
Met	3	(1.0)	1	(0.5)	4	(0.8)
Phe	12	(3.9)	7	(3.8)	19	(3.8)
Pro	15	(4.8)	9	(4.9)	24	(4.8)
Ser	14	(4.5)	15	(8.1)	29	(5.8)
Thr	6	(1.9)	9	(4.9)	15	(3.0)
Trp	2	(0.6)	1	(0.5)	3	(0.6)
Tyr	8	(2.6)	5	(2.7)	13	(2.6)
Val	11	(3.5)	14	(7.6)	25	(5.0)

processing site in pea legumin resembles the internal cleavage points of vicilin in that it is adjacent to an asparagine residue preceded by an acidic residue (34,50).

There is considerable variation in the reported sulphur amino acid content of legumins from different pea strains (51). Since a single inbred strain of pea may contain a number of legumin genes (5) and several legumin protein types (14,52), the reported amino acid composition of the isolated legumin protein is undoubtedly a composite of several protein variants. As the low sulphur content of pea seeds is the main nutritional limitation, the isolation of genes coding for high sulphur legumin is an important first step toward crop improvement. The leg A gene codes for a protein having 5 cysteine residues (1.01 mol %) and 4 methionine residues (0.81 mol %), which is at the upper end of the range of variation found between legumins of

		DONOR BOUNDARY																
		EXON					INTRON											
legA	IVS-1	A	C	A	A	G	G	T	T	A	C	T	T	A	T	T		
legA	IVS-2	C	T	A	G	G	G	T	G	A	G	C	A	C	T	G		
legA	IVS-3	A	C	A	A	A	G	T	A	T	G	T	T	T	T	T		
P		$\frac{T}{A}$	$\frac{C}{A}$	X	$\frac{T}{A}$	G	G	T	A	$\frac{A}{T}$	G	T	$\frac{A}{T}$	$\frac{T}{A}$	$\frac{T}{A}$	$\frac{A}{T}$		
A		$\frac{A}{C}$	A	G	G	T	$\frac{A}{G}$	A	G	T								
↓																		
		ACCEPTOR BOUNDARY																
		INTRON														EXON		
legA	IVS1	A	T	A	T	A	C	C	A	A	T	T	A	C	A	G	G	T
legA	IVS2	A	T	C	T	A	T	G	T	T	T	G	A	C	A	G	A	G
legA	IVS3	A	C	A	A	T	C	T	T	C	A	T	A	C	A	G	A	A
P		$\frac{T}{A}$	T	T	$\frac{T}{A}$	$\frac{T}{A}$	$\frac{T}{A}$	A	$\frac{A}{T}$	$\frac{T}{A}$	T	$\frac{T}{A}$	$\frac{G}{A}$	$\frac{T}{C}$	A	G	G	$\frac{T}{A}$
A		←						$\frac{T}{C(15)}$	→		N	$\frac{C}{T}$	A	G	G	$\frac{G}{T}$		

Figure 3

The boundary sequences of the three introns are shown in comparison to: A, the consensus for animal genes (51); P, the consensus for plant genes. The plant consensus was taken from the data compiled by Slightom *et al.* (38).

different pea strains (51) although genes coding for higher sulphur containing legumin may yet be found. The predicted amino acid composition is shown in table 1.

It is not possible to definitely assign a particular gene product to the leg A gene but, interestingly, the predicted molecular weights of 36441 for the α -subunit and 20185 for the β -subunit match most closely with the L3 subunit pair of Matta *et al.* (14). As the predicted α subunit also matches reasonably well in terms of N-terminal sequence and amino acid sequence to the α^M subunit described by Casey *et al.* (33) in a different pea strain, the leg A gene product probably belongs to a major, widespread, legumin sub-family.

Intervening sequences

Comparison of the gene sequence with the sequences of several legumin cDNAs (3,15) indicates the presence of three intervening sequences, two within the sequence encoding the legumin α -subunit and one within the sequence encoding the β -subunit. All three introns obey the GT/AG boundary

'TATA Box'	
1	T C T A T A A A T T A
2	T $\frac{C}{G}$ T A T A $\frac{T}{A}$ A ₍₁₋₃₎ T A
'CAAT Box'	
1	G T C C A A T T C T
3	G G $\frac{C}{T}$ C A A T T C T
'AGGA Box'	
1	C A A G C T G C A G A A T G T C
2	$\frac{C}{T}$ A ₍₁₋₅₎ $\frac{G}{T}$ N G A ₍₂₋₄₎ $\frac{T}{C}$ $\frac{T}{C}$

Figure 4

The canonical sequences of the promotor region. 1, the sequence from the legA gene. 2, the consensus sequence for plant systems (25). 3, the consensus sequence for animal systems (23).

rule of Breathnach *et al.* (53). Slightom *et al.* (39) have compared a number of sequences for the intron boundaries of higher plant genes. The legumin intron boundaries are compared (Fig.3) with the plant consensus sequence and with that for animal genes (54). The legumin sequences are more in agreement with the plant consensus sequences in that the acceptor site is preceded by a run of A/Ts rather than C/Ts. The intervening sequences are also typical of plants (39) in being short (of lengths 88bp, 88bp and 99bp respectively) and A/T rich, with respective compositions of 71.6%, 68.2% and 71.7% A + T.

At least two of these introns are conserved within the sequence of the leg C gene of pea (data not presented). Introns have also been found in the soyabean legumin (glycinin) gene (17) and these appear to be in homologous positions to the leg A introns (49). However whilst the positions are conserved between the two species, the soyabean introns are much larger.

The 5' end of the gene

The perfect homology between the leg A gene and the cDNA clone pDUB8 suggests that the leg A gene is transcriptionally active and has all the necessary transcriptional control sequences. A search of the leg A sequence for the canonical sub-sequences observed at the 5' end of most genes from animal sources (see 23) reveals a 'TATA box' beginning at position -66 (fig.2). This shows good homology to the consensus for the 'TATA box' of higher plants

Figure 5

S1 nuclease mapping of the transcription start point. PolyA⁺ RNA was hybridised to DNA fragments 5' end-labelled at the XhoI site near to the initiation codon and extending to the PstI site directly upstream. Samples were treated with either 5000 units of S1 nuclease (lanes w and y) or 2500 units (lanes x and z). The control lanes y and z were loaded with ten times as much sample as lane w and x. Lanes W and X represent a sevenfold over exposure of lanes w and x respectively. A sequencing ladder prepared from the same end labelled DNA fragment was run alongside (lanes GAT and C).

(25), being preceded by the dinucleotide TC (fig.4). The transcriptional start point of several plant mRNAs lie within the sequence C/TATC/A 18-23 bp from the 'TATA box' (40) and a sequence CATC occurs 25bp downstream from the

Figure 6

A structure for the 3' end of the nascent message. The structure shown is that which gave the optimal free energy ($\Delta G = -14.5$ kcal) when computed by the RNA 9 folding programme. The arrows mark the points of polyA addition found in legumin cDNA clones (5,15).

'TATA box' in the leg A gene. The transcriptional start of the legumin message was located by an S1 nuclease digestion experiment (fig.5). Of the two main groups of bands which do not appear in the control, the major group bracket the CATC sequence while the more minor group centre upon an AT pair 10 bp further downstream. Consequently, on the basis of its position, it would seem that the 'TATA box' is part of the promotor of the leg A gene.

It has been observed by Messing *et al.* (25) that whilst sequences like the 'CAAT box' (23) are found upstream of the 'TATA box' in animal genes and have been found in some plant genes, the homology is often poor or no 'CAAT box' is apparent. A new consensus sequence known as the 'AGGA box' has therefore been proposed for plant genes (25). Examination of the leg A gene shows a good match for the 'CAAT box' at -126 (see fig.4). However the sequence adjacent to this on the 5' side shows partial homology to the 'AGGA box' (fig.4). It is also interesting to note that a sequence showing 80% homology to the consensus sequence for the adenovirus enhancer core element (55) occurs 55 bp upstream of the 'CAAT box' on the complementary strand (fig.2).

The 3' end of the gene

The 3' sequence of the leg A gene shows strong homology to the 3' end of the previously published cDNA sequences (5,15) and continues past the site of polyadenylation as found in the legumin cDNA sequences in pDUB1 (pRC 2.2.4),

pDUB3 (pRC 2.11.7) and pDUB6 (5.15). The most striking feature of the sequence is the existence of three sets of sub-sequences resembling polyadenylation signals. The polyadenylation signals of plant messages are known to be more complex than those of animals since there are normally at least two in each 3' untranslated region. Rather than the simple AATAAA (56) these often take the form of AATAA(A)TAAA, and sometimes a G residue is substituted for one of the nucleotides in the sequence (24,25). The three signals downstream of the legumin gene are respectively AATAAGAAAA, AATAAATAAAA, and AATAAATAAA. All of the polyadenylated legumin cDNAs so far examined have poly A tails commencing 19-20 bp downstream from the second of these signals, the third not being present on any of our cDNAs. This second signal differs from the others in two respects, firstly it is followed by the sequence ATTTCATGC (see fig.2) which is similar to a sequence that has been observed near the 3' end of many eukaryotic messages (23). Secondly, it is a site of considerable potential secondary structure. The structure in fig.6 was found to be the optimal structure for this region of the message according to the thermodynamic RNA folding programme RNA9 (29) with a ΔG value of -14.5kcal. It is very similar to that observed at the site of termination of the soyabean lectin message (40) in that the end of the message occurs at the end of a base paired region following the polyadenylation signal. Caution should be exercised in extending the significance of this finding, because although both of the structures resemble one proposed by Proudfoot and Brownlee (57) for two animal messages, such structures have not been found near the 3' end of several other messages. Also many other possible structures exist for the 3' end of the nascent legumin transcript including pairing of the region preceding the polyadenylation point with an almost perfect 12bp inverted repeat at position 2095 (fig.2). It does seem likely, however, that such a secondary structure may play a role in determining the choice of polyadenylation site.

ACKNOWLEDGEMENTS

We are very grateful to Margaret Richards for conscientious and expert technical assistance in sequencing the gene and to Tony Pickard and Phillipa Brown for their excellent assistance in the production and preliminary characterisation of the clones. We thank Dr. I.M. Evans for providing polyA⁺ RNA. We also thank Mrs. M. Raine for typing the final manuscript.

The work was supported by the SERC.

*To whom correspondence should be addressed

REFERENCES

1. Wright, D.J. and Boulter, D. (1974) *Biochem. J.* 141, 413-418.
2. Croy, R.R.D., Derbyshire, E., Krishna, T.G. and Boulter, D. (1979) *New Phytol.* 83, 29-35.
3. Evans, I.M., Croy, R.R.D., Hutchinson, P., Boulter, D., Payne, P.I. and Gordon, M.E. (1979) *Planta* 144, 455-462.
4. Croy, R.R.D., Gatehouse, J.A., Evans, I.M. and Boulter, D. (1980) *Planta* 148, 49-56.
5. Croy, R.R.D., Lycett, G.W., Gatehouse, J.A., Yarwood, J.N. and Boulter, D. (1982) *Nature* 295, 76-79.
6. Spencer, D. and Higgins, T.J.V. (1980) *Biochem. Int.* 1, 502-509.
7. Derbyshire, E., Wright, D.J. and Boulter, D. (1976) *Phytochemistry* 15, 3-25.
8. Tumer, N.E., Thanh, V.H. and Nielsen, N.C. (1981) *J. Biol. Chem.* 256, 8756-8760.
9. Gilroy, J., Wright, D.J. and Boulter, D. (1979) *Phytochemistry* 18, 315-316.
10. Maltashewski, G.J., Adeli, K., Altosaar, I., Shewry, P.R. and Mifflin, B.J. (1982) *FEBS Lett.* 145, 208-212.
11. Walburg, G. and Larkins, B.A. (1983) *Plant Physiol.* 72, 161-165.
12. Yamagata, H., Sugimoto, T., Tanaka, K. and Kasai, Z. (1982) *Plant Physiol.* 70, 1094-1100.
13. Zhao, W., Gatehouse, J.A. and Boulter, D. (1983) *FEBS Lett.* 162, 96-102.
14. Matta, N.K., Gatehouse, J.A. and Boulter, D. (1981) *J. Exp. Bot.* 32, 1295-1307.
15. Lycett, G.W., Delauney, A.J., Zhao, W., Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) *Plant Mol. Biol.* in press.
16. Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984) manuscript in preparation.
17. Fischer, R.L. and Goldberg, R.B. (1982) *Cell* 29, 651-660.
18. Gatehouse, J.A., Evans, I.M., Bown, D., Croy, R.R.D. and Boulter, D. (1982) *Biochem. J.* 208, 119-127.
19. Evans, I.M., Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) *Planta* (in press.)
20. Chandler, P.M., Higgins, T.J.V., Randall, P.J. and Spencer, D. (1983) *Plant Physiol.* 71, 41-54.
21. Goldberg, R.B., Hoschek, G., Ditta, G.S. and Breidenbach, R.W. (1981) *Dev. Biol.* 83, 218-231.
22. Meinke, D.W., Chen, J. and Beachy, R.N. (1981) *Planta* 153, 130-139.
23. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) *Nucl. Acids Res.* 8, 127-142.
24. Lycett, G.W., Delauney, A.J. and Croy, R.R.D. (1983) *FEBS Lett.* 153, 43-46.
25. Messing, J., Geraghty, D., Heidecker, G., Hu, N., Kridl, J. and Rubenstein, I. (1983) in *Genetic Engineering of Plants*, Kosuge, T., Meredith, C.P. and Hollaender, A. Eds., pp. 211-227.
26. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
27. Seif, I., Khoury, G. and Dhar, R. (1980) *Nucl. Acids Res.* 8, 2225-2240.
28. Sege, R.D., Soll, D., Ruddie, F.G. and Queen, C. (1981) *Nucl. Acids Res.* 9, 437-444.
29. Zuker, M. and Stiegler, P. (1981) *Nucl. Acids Res.* 9, 133-148.
30. Falvaloro, J., Treisman, R. and Kamen, R. (1980) *Methods Enzymol.* 65, 718-749.

31. Pedersen, K., Devereux, J., Wilson, D.R., Sheldon, E. and Larkins, B.A. (1982) *Cell* 29, 1015-1026.
32. Leder, P., Tiemeier, D. and Enquist, L. (1977) *Science* 196, 175-177.
33. Casey, R., March, J.F., Sharman, J.E. and Short, N.M. (1981) *Biochim. Biophys. Acta* 670, 428-432.
34. Lycett, G.W., Delauney, A.J., Gatehouse, J.A., Gilroy, J., Croy, R.R.D. and Boulter, D. (1983) *Nucl. Acids Res.* 11, 2367-2380.
35. Higgins, T.J.V., Chandler, P.M., Zurawski, G., Button, S.C. and Spencer, D. (1983) *J. Biol. Chem.* 258, 9544-9549.
36. Geraghty, D., Peifer, M.A., Rubenstein, I. and Messing, J. (1981) *Nucl. Acids Res.* 9, 5163-5174.
37. Hoffman, L.M., Ma, Y. and Barker, R.F. (1982) *Nucl. Acids Res.* 10, 7819-7828.
38. Edens, L., Heslinga, L., Klok, R., Ledebroer, A.M., Maat, J., Toonen, M.Y., Visser, C. and Verrips, C.T. (1982) *Gene* 18, 1-12.
39. Slightom, J.L., Sun, S.M. and Hall, T.C. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1897-1901.
40. Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983) *Cell* 32, 1023-1031.
41. Murai, N., 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.
42. Blobel, G. and Dobberstein, B. (1975) *J. Cell Biol.* 67, 835-851.
43. Gatehouse, J.A., Croy, R.R.D., Morton, H., Tyler, M. and Boulter, D. (1981) *Eur. J. Biochem* 118, 627-633.
44. Higgins, T.J.V. and Spencer, D. (1981) *Plant Physiol.* 67, 205-211.
45. Chrispeels, M.J., Higgins, T.J.V., Craig, S. and Spencer, D. (1982) *J. Cell Biol.* 93, 5-14.
46. Spencer, D. and Higgins, T.J.V. (1980) *Biochem. Int.* 1, 502-509.
47. Bassdner, R., Manteuffel, R., Müntz, K., Pöschel, M., Schmidt, P. and Weber, E. (1983) *Biochem. Physiol. Pflanzen* 178, 665-684.
48. Tumer, N.E., Richter, J.D. and Nielsen, N.C. (1982) *J. Biol. Chem.* 257, 4016-4018.
49. Nielsen, N.C. (1984) *Phil. Trans. Roy. Soc. Lond.* 304B, 273-407.
50. Gatehouse, J.A., Lycett, G.W., Delauney, A.J., Croy, R.R.D. and Boulter, D. (1983) *Biochem. J.* 212, 427-432.
51. Casey, R. and Short, M.N. (1981) *Phytochemistry* 20, 21-23.
52. Thomson, J.A., Schroeder, H.E. and Dudman, W.F. (1978) *Aust. J. Plant Physiol.* 5, 263-279.
53. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4853-4857.
54. Breathnach, R., and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
55. Hearing, P. and Shenk, T. (1983) *Cell* 33, 695-703.
56. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214.
57. Proudfoot, N.J. and Brownlee, G.G. (1974) *Nature* 252, 359-362.