Characteristics of two types of legumin genes in the field bean (*Vicia faba* L. var. minor) genome as revealed by cDNA analysis

Ulrich Wobus, Helmut Bäumlein, Ronald Bassüner, Ute Heim, Rudolf Jung, Klaus Müntz, Gerhard Saalbach and Winfriede Weschke

Zentralinstitut für Genetik und Kulturpflanzenforschung, GDR Academy of Sciences, DDR-4325 Gatersleben, GDR

Received 6 March 1986; revised version received 27 March 1986

Nucleotide sequence analysis of cDNA clones coding for field bean legumin precursor polypeptides revealed two different types, called A and B. Although homologous, both types differ in several sequence characteristics. Comparison with similar data from soybean and recent findings from pea leads to the following conclusions: (i) the two types of legumin genes described represent two subfamilies, A and B, which are probably of widespread occurrence; (ii) legumin genes or subunits can best be placed in either subfamily A or B by sequence homology, in addition B-type subunits contain generally fewer (or none at all in *V. faba*) Met residues as compared to A-type subunits; (iii) members of one subfamily from different species are more homologous than members of either subfamily within a species, therefore the two subfamilies must have arisen long before speciation of the genera *Glycine, Pisum* and *Vicia*; (iv) during speciation members of the B-subfamily diverged significantly more than members of the A-subfamily.

DNA sequence analysis *Storage protein gene* (Vicia faba) *Legumin* *Gene subfamily*

1. INTRODUCTION

The major seed storage proteins of *Vicia faba* belong to either the legumin or vicilin class of globulins. Legumin proteins consist of 6 nearly identical subunits. Subunits are derived from a single precursor polypeptide which, after formation of a disulphide bond, is cleaved into an acidic α- and a basic β-polypeptide chain, resulting in an α/β pair (cf. [1]). Chromatographic and electrophoretic analyses revealed 4 main specific α/β chain pairs or subunits belonging to two classes, named A and B [2,3]. This classification was initially based on differences in amino acid composition and on peptide maps [2] and later supported by N-terminal amino acid sequences of authentic β-polypeptides [4]. Here we present nucleotide sequence data of A-type and B-type cDNA clones which characterize two main subfamilies of legumin genes in the field bean genome.

2. EXPERIMENTAL

cDNA was synthesized on poly(A)$^+$ RNA from membrane-bound polysomes [5] by standard procedures and cloned into the *Pst*I site of pBR322. Storage protein specific clones were identified by hybrid-selected translation (cf. [6,7] for methods). Nucleotide sequences were determined both by the chemical degradation method [8] and the dideoxy sequencing procedure [9] according to the strategies depicted in fig.2. All sequences were determined at least twice with different DNA preparations.

3. RESULTS

3.1. *Identification of two types of legumin cDNA clones*

cDNA clones were first screened with poly(A)$^+$ RNA from a preparation already used for cDNA

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50 © 1986 Federation of European Biochemical Societies

synthesis [5] and 72 positive clones further checked by hybrid-selected translation. 19 of them bound mRNA coding for legumin precursor polypeptides. These precursors fall into two classes with apparent molecular masses of 67 and 64 kDa, respectively (fig.1); their legumin nature was previously identified by in vitro translation experiments using total poly(A)$^+$ RNA from cotyledons [6]. Incorporation of [^{35}S]methionine into only the slower moving polypeptide band [6,10] suggested its role as precursor of legumin A-subunits. Correspondingly, the faster moving band should represent legumin B-subunit precursor(s). This working hypothesis was verified by nucleotide sequence analysis.

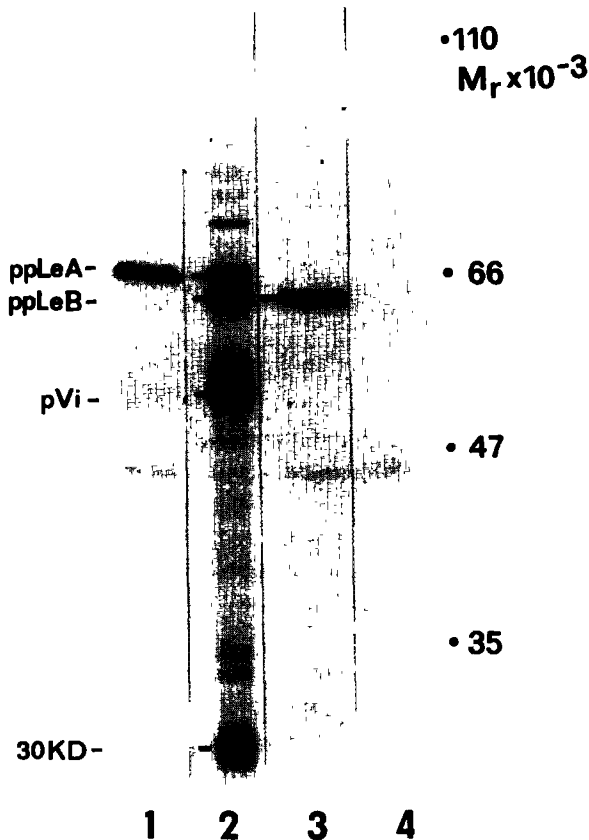

Fig.1. Cell-free translation of hybridization-selected mRNAs coding for *V. faba* legumin precursor polypeptides. The fluorograph was obtained after SDS gel electrophoresis (9.5% acrylamide) of the [^{3}H]leucine labeled translation products. For translation in the wheat germ system mRNA species were selected by hybridization to an A-type (lane 1) and a B-type (lane 3) cDNA clone. For comparison, translation products of total poly(A)$^+$ RNA are shown in lane 2; no mRNA added, lane 4. The primary translation products of storage globulins are designated ppLeA (preprolegumin A, M_r 67000), ppLeB (preprolegumin B, M_r 64000), pVi (previcilin, M_r 54000) and an unidentified polypeptide (30KD) of M_r 30000 (cf. [6]). The M_r scale is derived from brome mosaic virus RNA coded products (not shown).

3.2. *Nucleotide sequence analysis of two types of legumin cDNA clones*

Nucleotide sequences were determined from several clone inserts (fig.2) and combined for comparison into a single A- and B-type sequence, respectively (fig.3). Whereas the B-type clones pVfc68 and 70 are identical in the overlapping regions, A-type clones pVfc30, 53 and 77 differ from each other by a few base substitutions (see legend to fig.3). It remains unclear to what extent these differences are due to multiple genes, polymorphism, or cloning artefacts.

As is evident from fig.2, the combined sequences represent the C-terminal ends of the α-chain coding regions and the complete (B-type) or nearly complete (A-type) β-chain coding regions as well as the B-type 3′-untranslated region. Conceptual translation identifies for each type a sequence

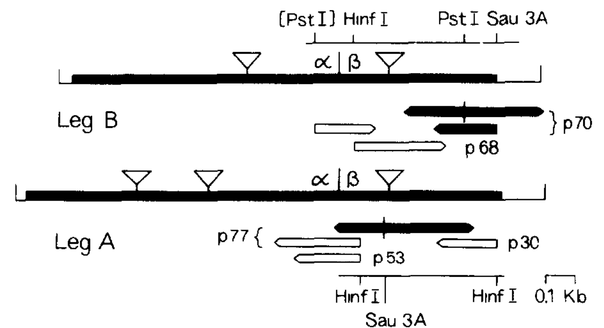

Fig.2. Map of the determined cDNA sequences aligned to a schematic outline of a B-type gene (LegB4 of *V. faba* [13]) and an A-type gene (LegA of *Pisum sativum* [16]), respectively. Open bars, sequences determined according to [8]; black bars, sequences recloned into M13mp phages and sequenced according to [9]. Restriction sites used are indicated. [PstI], site in vector DNA; ▽, location of introns; vertical line, border between the coding regions for the α- and β-chains, respectively; p68, etc., are abbreviated designations of the respective cDNA clones pVfc68, etc.

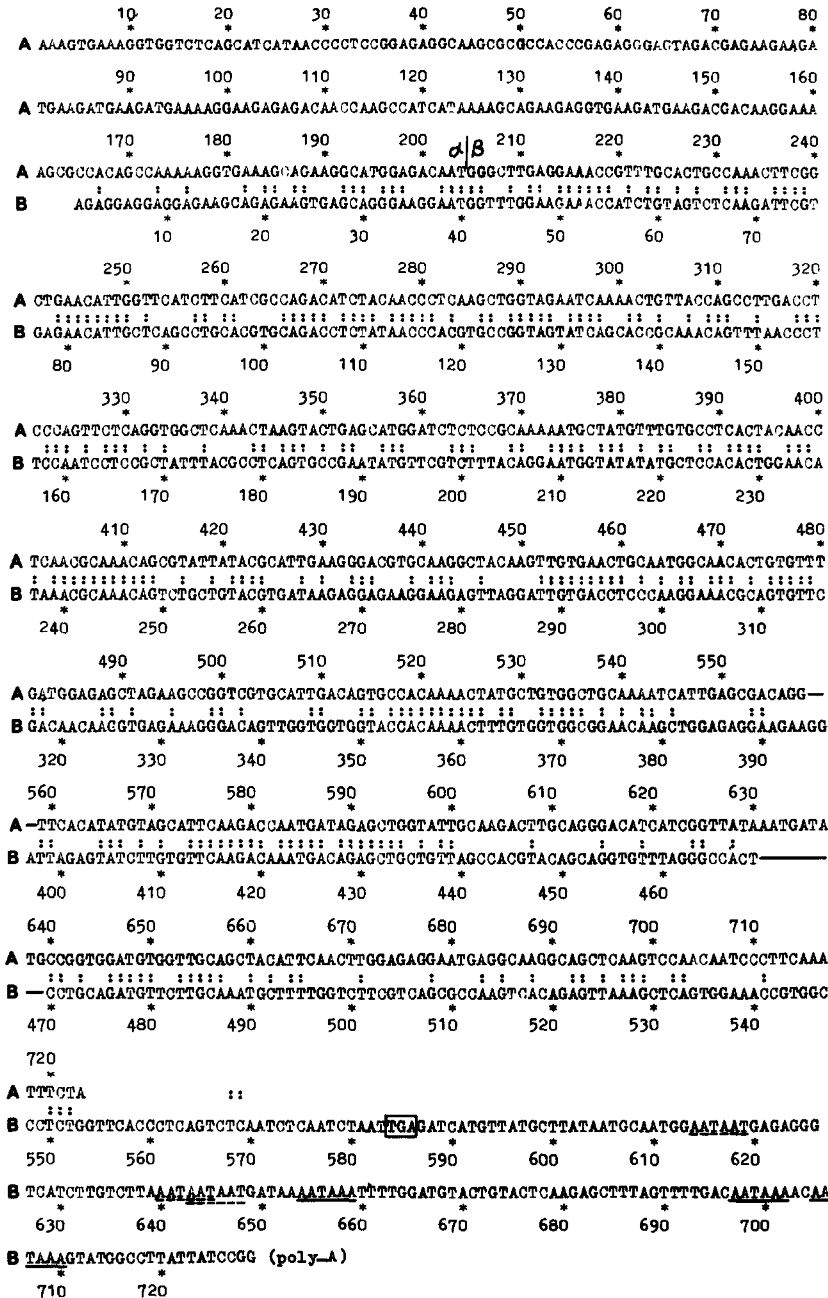

Fig.3. Aligned nucleotide sequences of pVfc77/30 (A-type legumin, marked A) and pVfc68/70 (B-type legumin, marked B). Clone p53 of fig.2 (not included in this figure) contains 4 base exchanges relative to p77, all in the α-chain coding part. p30 contains at position 624 an A instead of the G in p77 as given in the figure. Clones p68 and p70 are identical in the overlapping parts. Lines represent hypothetical deletions introduced to improve alignment. The border between α- and β-chain coding regions is indicated; the termination codon is boxed; putative polyadenylation signals are underlined.

matching exactly the published [4] 68 and 61 N-terminal amino acid residues for an A- and a B-type β-chain, respectively. B-type clone pVfc70 contains a complete 143 bp long coding sequence for the 3′-untranslated region of the respective mRNA with three AATAAA motifs and three closely related AATAAT sequences, two of them overlapping. This underlines the notion that plant genes often have multiple polyadenylation signals [11,12].

3.3. *Comparison of the β-chain coding regions*

Despite considerable sequence divergence, A- and B-type β-chain coding regions can be aligned nearly continuously (fig.3). Of the compared nucleotides 54% are identical. The nonidentical nucleotides occupy third codon positions less often (40% compared to 30% for both first and second codon positions) than in many other phylogenetically closely related genes. This results in a high number (63%) of replacement substitutions and leads to an amino acid homology of only 41%. However, in 55% of the cases, in which amino acid substitutions have occurred, the physical characteristics of the residues are maintained.

3.4. *The C-terminal region of the α-chains*

Close inspection of the 67 derived amino acid residues of the A-type α-chain C-terminus reveals two divergent repeats together with remnants of a third copy near the α/β border (fig.4). The B-type sequence extends for only 13 amino acid residues from the β-chain N-terminus into the α-chain region but can be supplemented (see fig.4) by data derived from a recently described complete genomic B-type sequence [13] that is identical to the cDNA B-type sequence in the comparable coding region (but different in the 3′-untranslated region). As is evident from fig.4, the B-type α-chain C-terminus does not possess repeats but instead a stretch of Glu residues which in turn shows similarity to a stretch of Glu and Asp residues in the second half of especially the first repeat in the A-type sequence. We note in addition that the α/β cleavage point Asn↓Gly is conserved as in all other described legumin-like precursor polypeptides (cf. [14–22]).

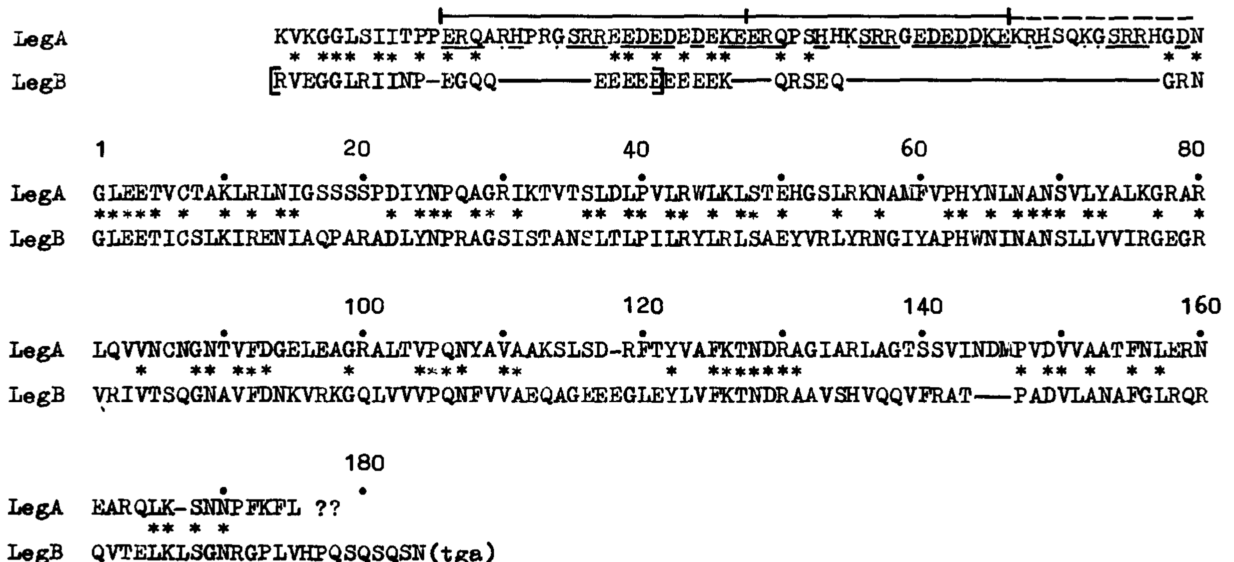

Fig.4. Aligned amino acid sequences of A- and B-type legumin subunits derived from the nucleotide sequences of fig.3. The residues in brackets are from [13] (see text for explanation). Insertions/deletions were introduced to improve alignment. Repeats in the A-type α-chain C-terminal region are overlined; residues present at comparable position in all three repeats are underlined; residues present in only two repeats are marked by dots. The C-terminal very end of the A-type β-chain is unknown due to the lack of the respective nucleotide sequence in clone pVfc30.

4. DISCUSSION

Legumin cDNA clones fall into two classes initially defined by the size of the prepropolypeptides translated in vitro from clone-selected mRNAs. Sequence analysis proved that the two classes represent genes coding for distinct groups of legumin subunits earlier described as type A and B [2]. Striking homologies between the two types at both the nucleotide (fig.3) and amino acid level (fig.4) point to a common ancestor gene and justify the consideration of genes of both types as members of a single legumin gene family. On the other hand, the degree of sequence divergence between the two types, which is most pronounced in the C-terminal parts of both the α- and β-chains, and the apparent lack of intermediate type sequences ([23]; unpublished) call for a subdivision into two main subfamilies named A and B according to the subunit nomenclature. Whereas the sequence data in this paper do not prove the existence of several genes of each type and thus the subfamily character, clear evidence for at least several B-type genes including pseudogenes is provided by the analysis of genomic clones and Southern blots (unpublished); two A-type β-polypeptides were found by cyanogen bromide cleavage [3].

The existence of two main subfamilies of legumin genes is not restricted to *V. faba*. Glycinins, the legumin-like globulins of the soybean *Glycine max*, can be separated into groups I and II (cf. [24]) with sequence characteristics reminiscent of *V. faba* A-type (group I) and B-type (group II) genes [14,17–21]. In the pea *Pisum sativum* all legumin gene sequences published to data [16,25–27] are of the A-type as deduced from the more than 90% homology to the *V. faba* A-type sequence (cf. fig.5). Recently, however, B-type genes have also been found [28,29]. Fig.5 provides a summarizing view of the differences between the β-chains and their coding regions. Three legume species representing 8 genes were included in the analysis. If several cDNA sequences differing by only a few nucleotides were available, only one was included although all of them may well represent different members of the respective subfamily.

From the data of fig.5 we can derive several conclusions: (i) The two subfamilies are always clearly separated not only within a species but also in an inter-species comparison. This means that legumin genes diverged into the two types already before speciation of the genera *Glycine*, *Pisum* and *Vicia*, and implies functional differences like in barley for which it is known that two subfamilies of B hordeins are differently expressed during seed development and under conditions of sulphur deficiency [30,31]. (ii) An intra-subfamily inter-species comparison reveals a significantly higher degree of divergence within the B-subfamily as compared to the A-subfamily suggesting a stronger selection

			P.sativ. A	P.sativ. B	V.faba A	V.faba B	G. max A	G. max A	G. max B	G. max B
			Leg Aβ	Leg Jβ	p77/30	p68/70	B 1a	Bx (1b)	B 3	B 4
P.sativ.	A	Leg Aβ		(56)	94	39	67	67	45	46
P.sativ.	B	Leg J β					(46)		(67)	(69)
V.faba	A	p 77/30	94			41	68	66	45	45
V.faba	B	p 68/70	54		54		38	38	63	59
G.max	A	B 1a	73		73	51		89	43	43
G.max	A	B x (1b)	73		73	51	87		40	39
G.max	B	B 3	51		54	61	53	52		86
G.max	B	B 4	52		54	60	52	52	88	

Below diagonal: % identical nucleotides; above diagonal: % identical amino acid residues

Fig.5. Percentages of identical nucleotides and amino acid residues between legumin β-chains and their coding regions. Data used for the calculations were from [16] for LegAβ, [18] for the B_{1a}-chain of glycinin A_2B_{1a}, [20] for the $B_{x(1b)}$-chain of probably glycinin $A_{1a}B_{1b}$, [19] for the B_3-chain of glycinin $A_5A_4B_3$ and [14] for the B_4-chain of glycinin A_3B_4. The differences between the sequences used here and that published in [17] for B_{1a} and [21] for B_3 and B_4 are irrelevant in this context. Amino acid sequences were aligned to maximize homology. Nucleotide sequences were aligned accordingly, the number of identical nucleotides as well as amino acid residues determined and divided by the total number of compared positions. No sequence data were available for LegJβ; values are from fig.6 in [29]. Comparisons with pVfc77/30 are slightly biased for higher values since the C-terminal end, which is always much more diverged than the N-terminal regions, is unknown. Values from intra-subfamily comparisons are on dotted ground. Essentially similar figures can be derived from a comparison of the N-terminal halfs of the α-chains and their respective coding sequences whereas the C-terminal parts are very variable and therefore much more difficult to compare.

pressure against changes in A-type subunits and/or a more recent origin of the A-type genes. (iii) Amino acid homologies are generally lower than nucleotide homologies. This fact, already noted by Schuler et al. [32], may be due to selection for overall structure rather than for specific sequences in proteins used as carbon and nitrogen source by the developing embryo [29] but selection at the DNA and/or RNA level may also play a role [19,23].

Whereas sequence homology is the most important feature in discriminating the two subfamilies, extensive analyses of soybean glycines reveal another distinguishing feature: group II (B-type) subunits contain fewer Met residues than group I (A-type) subunits (cf. [24]). In *V. faba*, B-type subunits are completely Met-free. This fact may be of practical importance in attempts to improve the nutritional quality of the seeds (cf. [1]). It may also be of physiological significance, for instance under conditions of sulphur deficiency (cf. [33]). Other subfamily specific features seem to be only relevant within closely related species or genera. Thus, the repeats near the C-terminus of the *V. faba* A-type α-chain seem to be A-type specific since all A-type precursor polypeptides run in a polyacrylamide gel as a single band with an apparent molecular mass of 3 kDa larger than the B-type precursor band, a size which accounts reasonably well for the additional 33 amino acid residues in the A-type repeat region. Pea A-type α-chains contain similar repeats, with the possible exception of clone pDUB3, which may, however, have lost its repeats during clone propagation in *E. coli* [26]. The two groups of glycinins mentioned above are also distinguished by the presence or absence of repeats near the α-chain C-terminus but here the repeats are characteristic of the B-type group II subunits (cf. [29]).

ACKNOWLEDGEMENTS

We thank Ms G. Jüttner for competent technical assistance, E. Bauschke, G. Eitner, M. Hartmann and H. Liebscher for gifts of [^{32}P]dATP and enzymes and R. Manteuffel for help with immunological methods in clone identification experiments. Critical remarks from F.C. Kafatos, S. Case, I. Schubert and a referee helped to improve the manuscript.

REFERENCES

[1] Müntz, K., Horstmann, C. and Schlesier, B. (1986) Biol. Zbl. 105, 107–120.
[2] Horstmann, C. (1983) Phytochemistry 22, 1861–1866.
[3] Horstmann, C. (1984) Kulturpflanze 32, S109–S116.
[4] Otto, A., Kraft, R. and Etzold, G. (1984) Kulturpflanze 32, S219–S221.
[5] Püchel, M., Bäumlein, H., Silhengst, P. and Wobus, U. (1981) Abh. Akad. Wiss. DDR, Abt. Math. Naturw. Tech. N5, 227.
[6] Bassüner, R., Manteuffel, R., Müntz, K., Püchel, M., Schmidt, P. and Weber, E. (1983) Biochem. Physiol. Pflanzen 178, 665–684.
[7] Bassüner, R., Wobus, U. and Rapoport, T.A. (1984) FEBS Lett. 166, 314–320.
[8] Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[9] Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[10] Bassüner, R., Wobus, U. and Rapoport, T.A. (1984) Kulturpflanze 32, S71–S75.
[11] Lycett, G.W., Delauney, A.J. and Croy, R.R.D. (1983) FEBS Lett. 153, 43–46.
[12] Messing, J.D., Geraghty, D., Heidecker, G., Hu, N.-T., Tridl, J. and Rubinstein, I. (1983) in: Genetic Engineering of Plants (Kosuge, T. et al. eds) pp.211–227, Plenum, New York.
[13] Bäumlein, H., Pustell, J., Wobus, U. and Kafatos, F.C. (1986) Nucleic Acids Res., in press.
[14] Fukazawa, C., Momma, T., Hirano, H., Harada, K. and Udaka, K. (1985) J. Biol. Chem. 260, 6234–6239.
[15] Müntz, K., Bassüner, R., Lichtenfeld, C., Scholz, G. and Weber, E. (1985) Physiol. Vég. 23, 75–94.
[16] Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984) Nucleic Acids Res. 12, 4493–4506.
[17] Marco, Y.A., Thanh, V.H., Tumer, N.E., Scallon, B.J. and Nielsen, N.C. (1984) J. Biol. Chem. 259, 13436–13441.
[18] Momma, T., Negoro, T., Udaka, K. and Fukazawa, C. (1985) FEBS Lett. 188, 117–122.
[19] Momma, T., Negoro, T., Hirano, H., Matsumoto, A., Udaka, K. and Fukazawa, G. (1985) Eur. J. Biochem. 149, 491–496.
[20] Negoro, T., Momma, T. and Fukazawa, C. (1985) Nucleic Acids Res. 13, 6719–6731.
[21] Scallon, B., Thanh, V.H., Floener, L.A. and Nielsen, N.L. (1985) Theor. Appl. Genet. 70, 510–519.
[22] Simson, A.E., Tenbarge, K.M., Scofield, S.R., Finkelstein, R.R. and Crouch, M.L. (1985) Plant. Mol. Biol. 5, 191–202.

[23] Wobus, U., Bäumlein, H., Bassüner, R., Grafe, R., Jung, R., Müntz, K., Saalbach, G. and Weschke, W. (1986) Biol. Zbl. 105, 121–125.
[24] Nielsen, N.C. (1984) Phil. Trans. R. Soc. Lond. B304, 287–296.
[25] Croy, R.R.D., Lycett, G.W., Gatehouse, J.A., Yarwood, J.N. and Boulter, D. (1982) Nature 295, 76–79.
[26] Lycett, G.W., Delauney, A.J., Zhao, W., Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) Plant Mol. Biol. 3, 91–96.
[27] Bown, D., Levasseur, M., Croy, R.R.D., Boulter, D. and Gatehouse, J.A. (1985) Nucleic Acids Res. 13, 4527–4538.
[28] Domoney, C. and Casey, R. (1985) Nucleic Acids Res. 13, 687–699.
[29] Casey, R., Domoney, C. and Ellis, N. (1986) Oxford Surveys Plant Mol. Cell Biol., in press.
[30] Rahmann, S., Shewry, P.R., Forde, B.G., Kreis, M. and Miflin, B.J. (1983) Planta 159, 366–372.
[31] Rahman, S., Kreis, M., Forde, B.G., Shewry, P.R. and Miflin, B.J. (1984) Biochem. J. 223, 315–322.
[32] Schuler, M.A., Doyle, J.J. and Beachy, R.N. (1983) Plant Mol. Biol. 2, 119–127.
[33] Chandler, P.M., Spencer, D., Randall, P.J. and Higgins, T.J.Y. (1984) Plant Physiol. 75, 651–657.