FEBS Letters 383 (1996) 46–50 FEBS 16830

Legumin encoding sequences from the Redwood family (Taxodiaceae) reveal precursors lacking the conserved Asn-Gly processing site

Claudia Wind, Klaus-Peter Häger*

University of Bayreuth, Department of Plant Ecology and Systematics, Universitätsstr. 30, D-95440 Bayreuth, Germany

Received 21 November 1995; revised version received 14 February 1996

Abstract **We have cloned and sequenced two different cDNAs encoding legumins from Japanese red cedar (*Cryptomeria japonica*, Taxodiaceae). The derived amino acid sequences show between 34% and 55% identity when compared with legumins from angiosperms and from Pinaceae, respectively. The predicted precursors are unusual in that they contain potential glycosylation signals, and we have found the corresponding β-polypeptides actually to be glycosylated. As most outstanding feature one of the precursors is lacking the Asn-Gly processing site which has been assumed to be highly conserved in legumin gene evolution. Legumin encoding sequences amplified from genomic DNA suggest that these unusual precursors are widespread if not ubiquitous in the Taxodiaceae family. From previous reports on legumin precursors with divergent processing sites, on the proteases involved in legumin precursor processing and from the results presented here it is concluded that the Asn-Gly processing site has been acquired rather than conserved during legumin gene evolution.**

Key words: Gymnosperm; Taxodiaceae; Legumin; Posttranslational processing site; Evolution

1. Introduction

Legumins, also designated 11S globulins, are the predominant seed storage proteins in a wide variety of dicot and in some monocot plants. They function as a source of carbon skeletons and nitrogen to support the initial growth of the seedling, and their structure, biosynthesis, and intracellular transfer are well conserved throughout the species investigated [1]. A most invariable character of legumin precursors hitherto known is the Asn-Gly junction located at that site where cleavage occurs to form the α- and β-polypeptides constituting mature legumin subunits, and it was assumed that this site has been highly conserved during evolution [2–6].

The proteases involved in legumin precursor processing actually do not show pronounced specifity for Asn-Gly cleavage [7–10]; this has been concluded from in vitro studies using artificial substrates. However, little attention has been drawn to precursors found to have divergent processing sites. We have characterized two legumin cDNAs from Japanese red cedar, and one of the precursors deduced is lacking the Asn-Gly processing site. For the first time evidence is given that these unusual legumin precursors are common in a plant family and implications for the evolution of legumin precursor processing are discussed.

*Corresponding author. Fax: (49) (921) 552786.

2. Materials and methods

2.1. Plant materials

Seeds from Japanese red cedar were harvested at midstage development (end of July, 1994) from an individual tree located in the arboretum of the Island of Mainau, Lake Constance. Leaves were collected in the Botanical Gardens of Munich and Bayreuth. Plant materials were immediately frozen in liquid nitrogen and stored at −70°C.

2.2. Construction and screening of a Japanese red cedar cDNA library

RNA was isolated from Japanese red cedar endosperm according to the method described in [11] and poly(A)$^+$ RNA was obtained by two successive runs on oligo(dT)-cellulose columns (type 7, Pharmacia). 5 μg of this mRNA were used as a template for cDNA synthesis (Pharmacia cDNA synthesis kit), and first-strand synthesis was monitored and yields estimated by incorporating [α-^{32}P]dCTP (3000 Ci/mMol, NEN). Double-stranded cDNA was ligated with *NotI*/*Eco*RI linkers, size fractionated on 1% agarose gels and cDNAs >500 bp were ligated into λ nm1149 and packaged [12]. Screening was performed by plaque hybridization at 50°C in 3×SSPE using a random-prime labelled fragment of the legumin gene GbLeg1 [13]; filters were washed at 55°C with 1×SSC. Several hybridizing cDNAs were cloned into the plasmid vector Bluescript pKS$^+$ (Stratagene) and nested sets of deletions were created using exonuclease III [12]. Double-stranded sequencing was done using the T7 sequencing kit of Pharmacia.

2.3. PCR amplification and cloning of legumin gene fragments

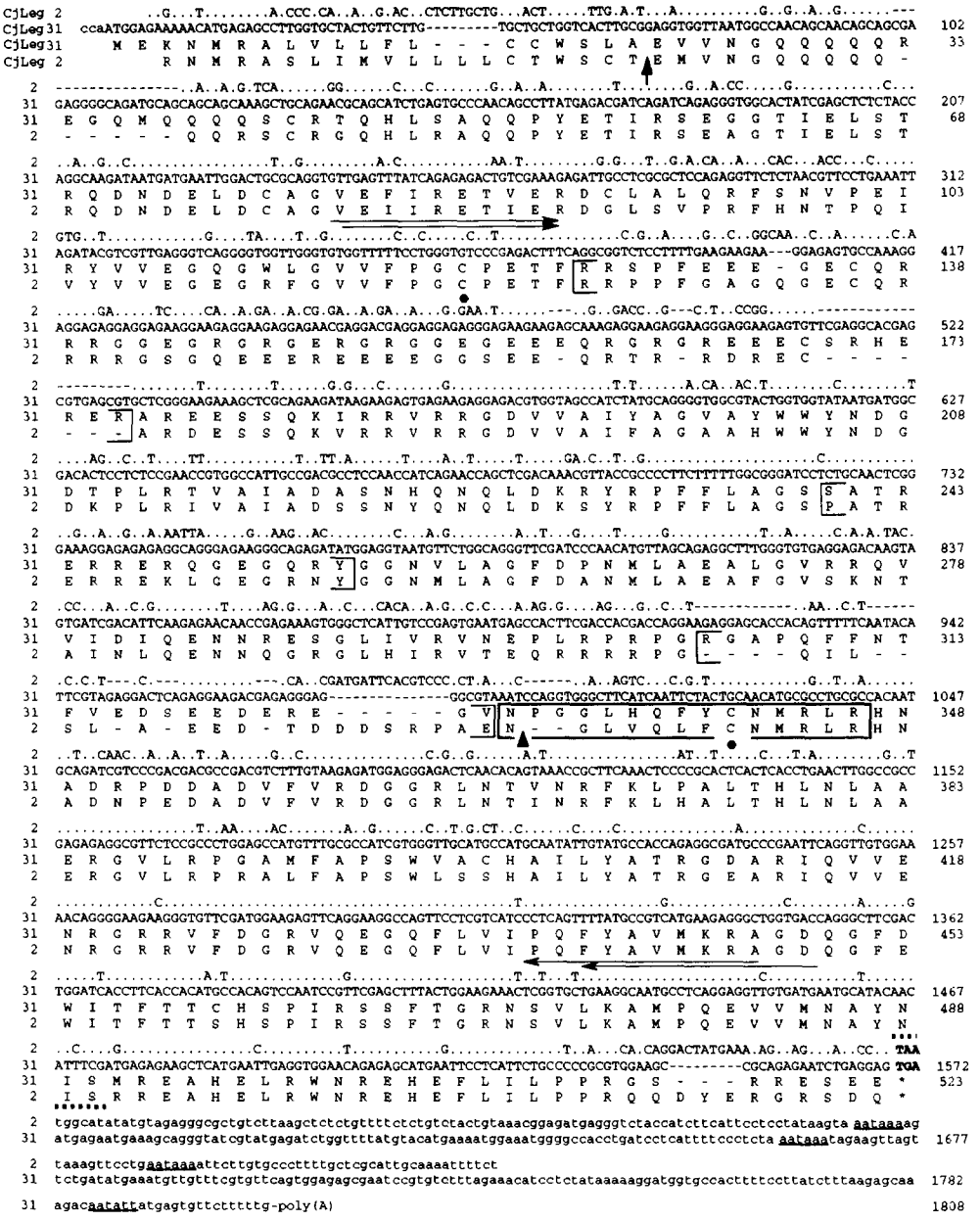
Genomic DNA was isolated from leaves essentially as described in [14]. PCR amplification of legumin gene fragments from the taxa given was performed using alternative sets of primers, essentially differing in degeneracy: 5′-RTKGTNTT-YCCNGGRTGYCCNGA-GACTT-3′ and 5′-ACHTCCTSHGGCATYGCYTTSARHACHGA-3′ for *Metasequoia*, *Sequoia* and *Sequoiadendron* and 5′-TKGTHTTYCCYGGRTGYCCHGAGACTT-3′ and 5′-CATYAC-NACYTCCTSHGGCATYSCCTTCA-3′ for *Cryptomeria*, *Glyptostrobus*, *Taiwania* and *Cunninghamia*. Amplifications were carried out in 25 μl of 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 1.5–3.0 mM $MgCl_2$, 2 μM of respective primers each, 0.2 mM of each dNTP, 80–200 ng genomic DNA and 1.5 units Taq polymerase. After initial template denaturation for 3 min at 94°C, 30 PCR cycles were performed, each cycle consisting of 0.5 min denaturation at 96°C, 1 min annealing at 55°C and 2 min extension at 72°C. Finally, an additional polymerization step was added at 72°C for 3 min. PCR products were purified on agarose gels, phosphorylated and blunt ended [12] and ligated into the *Eco*RV site of pKS$^+$ vectors. Recombinants and subclones of them, containing suitable restriction fragments, were sequenced using T3, universal (Pharmacia) and internal sequencing primers according to the method referred above.

3. Results and discussion

3.1. Isolation and characterization of Japanese red cedar legumin cDNAs

A cDNA library was constructed using poly(A)$^+$ RNA isolated from Japanese red cedar endosperm at midstage maturation. About 10^5 recombinant phage were screened by hybridization under moderate stringency, using a radiolabelled fragment of the *Ginkgo biloba* legumin gene GbLeg1 recently

SSDI S0014-5793(96)00216-5



F g. 1. Nucleotide and deduced amino acid sequences of Japanese red cedar legumin cDNAs. Numbering given on the right refers to CjLeg31, g:ps (dashes) introduced to maximize alignment were included. Coding and non-coding sequences are denoted by upper and lower case letters, respectively. Coding nucleotides differing from the reference are shown, identities are indicated by dots. Horizontal arrows above the sequence mark positions of PCR primers specified in section 2.3, stop codons are given in bold and potential polyadenylation signals are underlined. The vertical arrow and the arrowhead indicate putative co- and posttranslational processing sites, respectively, and the boxed region corresponds to those compared in Fig. 3. Segments in square brackets refer to Variable and Hypervariable Regions VR I, VR II and HVR and glycosylation signals are marked by a dotted line. Conserved cysteine residues probably involved in interpolypeptide linkage are labelled with filled circles.

described [13]. The coding sequence of the probe corresponds to the region spanning amino acid positions 347 to 487 in the alignment shown in Fig. 1. From the hybridizing clones detected, the inserts of 20 of them were analyzed by agarose gel electrophoresis after excising them with *Not*I. Clones containing inserts of 1.6–1.8 kb in length, suggesting full size legumin cDNAs, could be subdivided into two groups differing in restriction fragment patterns: while the inserts of six clones were recovered intact, those from three clones were cleaved to fragments of about 0.4 and 1.3 kb each when excised with *Not*I. From each group, one clone was chosen for further analysis: CjLeg31 (without a *Not*I site) and CjLeg2 (with a *Not*I site).

The aligned nucleotide sequences of both inserts, together with the respective deduced amino acid sequences, are given in Fig. 1. Open reading frames of 1530 bp and 1483 bp code for different legumin precursors which are discussed below. CjLeg31 represents a full length clone containing two potential translation initiation codons located at positions 4–6 and 16–18, respectively. We have assigned the first of them as the initiating codon since its vicinity is very similar to the AACAATGGC translation initiation consensus from plants [15].

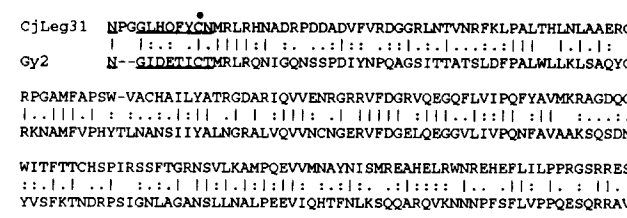


Fig. 2. Alignment of the *Glycine* legumin Gy2 β-polypeptide [5] and the homologous region of the CjLeg31 precursor. Preceding Asn residues flanking the cleavage site are depicted and regions homologous to the endoprotease recognition site are underlined. Gaps (dashes) and the filled circle are used as specified in Fig. 1. ı marks identical amino acids, : and · indicate similarity values of ≥0.5 and ≥0.1, respectively, resulting from a homology analysis using the 'Bestfit' program (Wisconsin GCG package).

Obviously, CjLeg2 is slightly truncated in its 5′ region. The 3′ trailers of CjLeg31 and CjLeg2 are 236 and 159 bp in length and contain several AATAAA polyadenylation signals; the AATATT motif 16 bp upstream from the CjLeg31 poly(A) tail may also function as a poly(A) addition signal. Such deviations from the signal consensus have been reported previously with the 3′-most signal in angiosperm mRNAs [15]. The coding regions show 77% nucleotide identity relative to one another; obviously, duplications and successive divergence have given rise to a small legumin gene family present in the Japanese red cedar genome.

3.2. The predicted structures of Japanese red cedar legumin precursors

The amino acid sequences derived from both cDNAs clearly verify their identity as legumin encoding transcripts. Comparisons with legumin precursors from *Picea glauca*, *Pinus strobus* (EMBL accession numbers X63192, Z11486) and *Pseudotsuga menziesii* [16,17] reveal between 50% and 55% sequence identity; nearly half of the changes are represented by conservative replacements. When compared with legumin precursors from angiosperms, the values for the highest and lowest identities are 41% and 34% for CjLeg2/Magnolia B14 [18] and CjLeg31/Vicia VfLeb4 [19], respectively.

The deduced amino acid sequences suggest that Japanese red cedar legumins are synthesized as precursors to be processed similarly as it is known from those of angiosperms. The N-terminal regions of the predicted polypeptides show hydrophobic domains with the characteristics of typical signal peptides. Using the algorithm of von Heijne [20], we have assigned the cotranslational processing sites located between Ala 22/Glu 23 and Thr 22/Glu 23, respec-tively (Fig. 1). The resulting propolypeptides show low levels of cysteine (1.3–1.8 Mol%) and methionine (ca. 1.5 Mol%) whereas they are rich in arginine (13.0–13.6 Mol%), consistent with the role of legumins as a nitrogen source.

From comparisons of angiosperm legumins, three regions displaying considerable heterogeneity in sequence and size are known [5]. Corresponding heterogeneity can be recognized in the precursors presented here, and the areas concerned are homologous in position with the Variable Region I and the Hypervariable Region HVR as defined by Nielsen [5]. Interestingly, the sequences are quite well conserved in the region corresponding to Variable Region II, the position of which is also marked in Fig. 1. Most likely, the residues Cys 120 and Cys 341 are involved in disulfide linkage of mature α- and β-polypeptides; cysteine residues at homologous position are strictly conserved in angiosperm legumins [5] and have been shown to participate in interpolypeptide linkage [21].

For reasons discussed below the most probable site for posttranslational cleavage is between Asn 332/Pro 333 and Asn 332/Gly 335, respectively (Fig. 1). This predicts mature α- and β-polypeptides with molecular weights of 34.8 kDa and 21.5 kDa (CjLeg31) and 32.5 kDa and 21.8 kDa (CjLeg2), respectively, and these values closely fit to those obtained from SDS-PAGE analysis (not shown). The pI values calculated (CjLeg31, α: 5.2, β: 10.4; CjLeg2, α: 5.7, β: 11.1) reveal pairs of acidic and basic polypeptides typically known from legumin subunits of angiosperms.

A remarkable feature common to both Japanese red cedar legumins is the presence of the putative glycosylation signal Asn-Ile-Ser (position 488–490, Fig. 1), suggesting glycosylation in the C-terminal region of their β-polypeptides. Indeed, glycoprotein staining using concanavalin A as the first ligand proved glycosylation of the legumin β-polypeptides (not shown). This finding is unusual in that legumins are generally considered devoid of covalently bound carbohydrates [1] with the exception of a lupin legumin the α-polypeptide of which has been shown to be glycosylated [22]. The physiological function of seed storage protein glycosylation remains still unclear.

3.3. Legumin encoding sequences from Taxodiaceae reveal precursors lacking the conserved Asn-Gly processing site

From comparisons of legumin precursors from a wide variety of flowering plants a most conserved region, immediately following the Hypervariable Region HVR mentioned above, has been detected. It has been described by the consensus sequence N-G-L/I-E/D-E-T-I/F-C-S/T [2,3], and it has been regularly observed that the cleavage site for posttranslational processing is between the Asn and the Gly that begin this region of homology. This region, which is believed to function as an endoprotease recognition site and in particular the Asn-

Consensus cleavage region	. . N - - G $^{L}_{I}$ $^{E}_{D}$ E T $^{I}_{F}$ C $^{S}_{T}$ . .	[2, 3]
Glycine max	. . N - - G I D E T I C T M R L R . .	[5]
Pisum sativum	. . N - - F L E E T V C T L K L H . .	[24]
Dioscorea caucasica	. . N - - V F A D N I C S F R I R . .	a)
Ginkgo biloba	. . N - - N V E E L Y C S M R L R . .	[23]
Pseudotsuga menziesii	. . N - - D V E E V V C A L R V K . .	[17]
Pinus pinaster	. . N - - A V E E L V C P ? R V S . .	[25]
Taxus baccata	. . N P S G L H Q F Y C N M R L R . .	b)
Calocedrus decurrens	. . N P S G L Q Q L I C N M R L R . .	b)
Cryptomeria japonica	. . N P G G L H Q F Y C N M R L R . .	c)
Metasequoia glyptostroboides	. N P S G L Q Q Q L C N M R L R . .	c)
" "	. . N P G G L H Q F Y C N M R L R . .	c)
Sequoia sempervirens	. . N P G G L H Q F Y C N M R L R . .	c)
Sequoiadendron giganteum	. . N P S G S Q Q Q L C N M R L R . .	c)
Cunninghamia lanceolata	. . N P S G L Q Q S Y C N M R L R . .	c)
" "	. . N P G G L Q Q L L C N L R L R . .	c)
Glyptostrobus pensilis	. . N P G G L H Q F Y C N M R L R . .	c)
Taiwania cryptomerioides	. . N P G G L H Q L Y C N M R L R . .	c)
" "	. . N P A G L H Q L Y C N M R L R . .	c)

Fig. 3. Alignment of legumin precursor regions homologous to the consensus cleavage region yet lacking the Asn-Gly cleavage site. The consensus and the *Glycine* legumin are included as references. Data from *Pisum* and *Pinus* result from N-terminal amino acid sequencing of β-polypeptides and Asn residues preceding these sequences are predicted. Sources of sequences are indicated [2–4,17,23–25], (a) Fischer, H. and Chen, L., pers. comm., (b) own data, unpublished and (c) PCR derived taxodiaceous sequences, this paper.

Gly junction to be cleaved, have been assumed to be highly conserved during legumin gene evolution [2–6].

From the two Japanese red cedar legumin precursors described here, CjLeg2 contains a region which is homologous in sequence and position to the endoprotease recognition site mentioned and in particular it shows the Asn-Gly junction to be cleaved (within the boxed region of Fig. 1). The same region can be recognized in the CjLeg31 precursor, however it is significantly divergent in that it lacks the conventional Asn-Gly cleavage site. It should be emphasized that this region is located at the same position as the endoprotease recognition sites from angiosperm legumin precursors: for comparison, the β-polypeptide of a prototype legumin from a legume has been aligned with the corresponding sequence of the CjLeg31 precursor (Fig. 2). Evidence from studies carried out both in vitro and in vivo [4,6] establish that posttranslational processing is a prerequisite for legumin hexamer formation and therefore vital for packaging and the realization of a storage function [4]; thus, the CjLeg31 precursor should be processed as well. Its cleavage as suggested in section 3.2 predicts a β-polypeptide which is extended in its N-terminal region by two residue positions relatively to those hitherto described (see also Fig. 2). Similar insertions within cleavage regions have been observed for legumin precursors from species of *Gnetum*, *Ephedra* and *Welwitschia* (EMBL accession numbers Z50779, Z50777, Z50780) and of *Taxus* and *Calocedrus* (own results, unpublished). More essentially, the CjLeg31 legumin precursor is a representative of the legumins lacking the Asn-Gly processing site (Fig. 3).

Notably, most of these unusual precursors are found in gymnosperms, and we have also investigated whether respective genomic sequences can be detected in further members of the Taxodiaceae, too. Considering the gymnosperm legumin sequences known to us, degenerate PCR primers were constructed; the primer positions relative to the sequences reported here are marked in Fig. 1. Corresponding legumin gene fragments, containing the respective introns [13] and varying in length between 1.4 kb and 1.6 kb, were amplified from genomic DNA of the Taxodiaceae species given in Fig. 3. The PCR products were purified, cloned and sequenced as described in section 2.3. Three of the polypeptides deduced, originating from species of *Sequoia* and *Cunninghamia*, revealed legumin precursors containing the conventional Asn-Gly processing site (not shown); the other amplificates from *Sequoia* and *Cunninghamia* as well as those from the remaining species investigated predict precursors which are very similar to each other and to the CjLeg31 precursor described above. Their putative processing regions, which are homologous in sequence and position to that from CjLeg31 (see boxed region in Fig. 1) and to those from legumin precursors of angiosperms, are shown in Fig. 3. The most outstanding feature common to all of them is the lack of the Asn-Gly processing site. Obviously, these otherwise unusual legumin precursors are widespread if not ubiquitous in the Taxodiaceae family.

We suggest that these precursors undergo posttranslational cleavage as well and we assume that it may occur at the Asn-Pro junction common to all of them. The close vicinity to the Hypervariable Region HVR located at the surface of the protein would ensure accessibility for the maturation proteases as it has been noted for the conventional cleavage site as well [5]. Our assumption is also in agreement with previous reports on the maturation proteases involved, the properties of which have been studied in vitro: a processing enzyme from castor bean seeds has been shown to be capable of converting prolegumin and two other vacuolar seed proproteins from pumpkin into their respective mature forms [7]. Common to these precursors are only the processing sites being N-terminally flanked by Asn, there is no further homology at or surrounding the junction to be cleaved. An endoprotease from soybean has been characterized and it has been shown that the Gly residue of the conventional legumin precursor processing site is not indispensible for cleavage [8], e.g. the Asn-Asn bond of the peptide GNNVEEL is specifically cleaved by the enzyme [9]. The peptide corresponds to the processing region of a legumin precursor from *Ginkgo*, and the cleavage observed is in accordance with the in vivo processing of this precursor [9,13,23]. The specifity of a soybean maturation protease, different from that refered above, has been investigated using octapeptides or mutated prolegumins with altered amino acids at or adjacent to Asn-Gly bonds. Elimination or replacements of the Asn residue prevent cleavage by the protease, whereas conservative changes on the C-terminal side of the cleavage site seemed not to hinder proteolysis [6]. Obviously, additional features of the secondary structure are important for the specifity of cleavage [8].

Summarizing, these findings and our results presented here provide a base to reflect on the evolution of legumin precursor processing. The almost invariable Asn-Gly processing site of angiosperm legumin precursors mainly known from crop plants gave rise to the assumption that it has been highly conserved during evolution. On the other hand the processing proteases involved are significantly less specific than may be deduced from the preponderance of the Asn-Gly junction to be cleaved. Thus, its invariability does not result from selective constraints on these proteins caused by the properties of the processing proteases. It is more obvious that the processing protease rather than the junction to be cleaved has been conserved during evolution. This suggests that ancient legumin precursors and in particular those from the common ancestors of the angiosperm and gymnosperm lineage were more variable with respect to the processing site than is known from those of present-day angiosperms. Since gymnosperms show a considerably conservative mode of evolution in many respects, their legumins may be regarded as ancient stages of legumin evolution. Accordingly, most of the known legumin precursors lacking the Asn-Gly processing site are from gymnosperms. Altogether we conclude that the Asn-Gly processing site has been acquired rather than conserved during evolution of legumins and of the genes encoding them.

Acknowledgements: This study was supported by DFG grant Hä 1925/1-2 and by a GRAKO fellowship ('proteins') to C.W. We are indebted to H. Fischer and L. Chen for making available unpublished data.

References

[1] Shotwell, M.A. and Larkins, B.A. (1989) in: The Biochemistry of Plants (Marcus, A., Ed.) Vol. 15, pp. 297–345, Academic Press, New York.
[2] Borroto, K. and Dure III, L. (1987) Plant Mol. Biol. 8, 113–131.
[3] vonder Haar, R.A., Allen, R.D., Cohen, E.A., Nessler, C.L. and Thomas, T.L. (1988) Gene 74, 433–443.
[4] Dickinson, C.D., Hussein, E.E.A. and Nielsen, N.C. (1989) Plant Cell 1, 459–469.

[5] Nielsen, N.C., Dickinson, C.D., Cho, T.-J., Thanh, V.H., Scallon, B.J., Fischer, R.L., Sims, T.L., Drews, G.N. and Goldberg, R.B. (1989) Plant Cell 1,313–328.
[6] Nielsen, N.C., Jung, R., Nam, Y.-W., Beaman, T.W., Oliveira, L.O. and Bassüner, R. (1995) J. Plant Physiol. 145, 641–647.
[7] Hara-Nishimura, I., Inoue, K. and Nishimura, M. (1991) FEBS Lett. 294, 89–93.
[8] Muramatsu, M. and Fukazawa, C. (1993) Eur. J. Biochem. 215, 123–132.
[9] Arahira, M. and Fukazawa, C. (1994) Plant Mol. Biol. 25, 597–605.
[10] Scott, M.P., Jung, R., Müntz, K. and Nielsen, N.C. (1992) Proc. Natl. Acad. Sci. USA 89, 658–662.
[11] Cerff, R. and Kloppstech, K. (1982) Proc. Natl. Acad. Sci. USA 79, 7624–7628.
[12] Sambrook, J., Fritsch E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (2nd edn.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[13] Häger, K.-P., Braun, H., Czihal, A., Müller, B. and Bäumlein, H. (1995) J. Mol. Evol. 41, 457–466.
[14] Doyle, J.J. and Doyle, J.L. (1987) Phytochem. Bull. 19, 11–15.
[15] Elliston, K. and Messing, J. (1988) in: Architecture of Eukaryotic Genes (Kahl, G., Ed.) pp. 21–56, VCH, Weinheim.
[16] Leal, I. and Misra, S. (1993) Plant Mol. Biol. 21, 709–715.
[17] Leal, I. and Misra, S. (1994) Plant Mol. Biol. 26, 1019.
[18] Fischer, H., Haake, V., Horstmann, C. and Jensen, U. (1995) Eur. J. Biochem. 229, 645–650.
[19] Bäumlein, H., Wobus, U., Pustell, J. and Kafatos, F.C. (1986) Nucleic Acids Res. 6, 2707–2720.
[20] von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.
[21] Staswick, P.E., Hermodson, M.A. and Nielsen, N.C. (1984) J. Biol. Chem. 259, 13431–13435.
[22] Duranti, M., Guerrieri, N., Cerletti, P., and Vecchio, G. (1992) Eur. J. Biochem. 206, 941–947.
[23] Häger, K.-P., Jensen, U., Gilroy, J. and Richardson, M. (1992) Phytochem. 31, 523–525.
[24] March, J.F., Pappin, D.J.C. and Casey, R. (1988) Biochem. J. 250, 911–915.
[25] Allona, I., Casado, R. and Aragoncillo, C. (1992) Plant Sci. 87, 9–18.