FEBS Letters 397 (1996) 352–356 FEBS 17817

Molecular cloning of the mitogenic mannose/maltose-specific rhizome lectin from *Calystegia sepium*

Els J.M. Van Damme[a,*], Annick Barre[b], Peter Verhaert[c], Pierre Rougé[b], Willy J. Peumans[a]

[a]*Laboratorium voor Fytopathologie en Plantenbescherming, Katholieke Universiteit Leuven, Willem de Croylaan 42, 3001 Leuven, Belgium*
[b]*Institut de Pharmacologie et Biologie Structurale, UPR CNRS 9062, Université Paul Sabatier, 35 Chemin des Maraîchers, 31062 Toulouse Cedex, France*
[c]*Laboratorium voor Ontwikkelingsfysiologie en Moleculaire Biologie, Katholieke Universiteit Leuven, Naamsestraat 59, 3000 Leuven, Belgium*

Received 23 August 1996; revised version received 15 October 1996

Abstract cDNA clones encoding the mitogenic mannose/maltose-specific lectin from the rhizomes of hedge bindweed (*Calystegia sepium*) have been isolated and sequenced. Comparison of the deduced amino acid sequence and the molecular weight of the lectin subunit as determined by mass spectrometry indicated that the mature protein comprises the entire open reading frame of the cDNA, which implies that the primary translation product contains no signal peptide and is not proteolytically processed. Searches in the databases revealed sequence homology with the previously described lectins from the taxonomically unrelated Moraceae species *Artocarpus integrifolia* and *Maclura pomifera*.

Key words: Hedge bindweed; *Calystegia sepium*; cDNA cloning; Convolvulaceae; Lectin

1. Introduction

Recent advances in the biochemistry and molecular biology of plant lectins have provided evidence that this heterogeneous group of proteins can be subdivided into a limited number of families of evolutionarily and structurally related proteins. According to the data available most lectins belong to either the legume lectins [1], the chitin-binding lectins composed of hevein domains [2], the type 2 ribosome-inactivating proteins (RIP) [3] or the monocot mannose-binding lectins [4]. Several lectins, however, do not fit into one of these four subgroups, e.g. the lectins from *Amaranthus hypochondriacus* [5] and *Artocarpus integrifolia* seeds [6]. The apparent lack of sequence homology with members of the four major lectin families implies that they belong to a distinct group. Evidently, when insufficient sequence information is available no such firm conclusions can be drawn. This is very well illustrated by the sialic acid-binding lectins from *Sambucus nigra* and *S. sieboldiana*, which had not been recognized as type 2 RIP until molecular cloning of their genes demonstrated a high sequence similarity with these proteins [7,8].

*Corresponding author. Fax: (32) (16) 322976.

Abbreviations: Calsepa, lectin from *Calystegia sepium*; HCA, hydrophobic cluster analysis; LECCalsepa, cDNA encoding the lectin from *Calystegia sepium*; MALDI-TOF-MS, matrix-assisted laser desorption ionization time of flight mass spectrometry; RIP, ribosome-inactivating protein

The nucleotide sequences reported in this paper have been submitted to the Genbank/EMBL Data library and are available under accession number U56820.

In a recent paper we reported the isolation and characterization of a novel lectin from *Calystegia sepium* (called *Calystegia sepium* agglutinin or Calsepa) [9]. Based on its molecular structure, partial amino acid sequence and carbohydrate-binding specificity it was suggested that Calsepa differed from all other plant lectins. To corroborate the presumed unique properties of this new lectin its cDNA was cloned. Sequence analysis confirmed that Calsepa does not belong to one of the four major lectin families but also demonstrated that it has reasonable homology with the previously described lectins from the (taxonomically far distant) Moraceae species *Artocarpus integrifolia* (jackfruit) and *Maclura pomifera* (osage orange). Moreover, analysis of the cDNA revealed that the mature lectin polypeptide comprises the entire open reading frame. It can be concluded, therefore, that the primary sequence of Calsepa contains no N-terminal signal sequence and undergoes no proteolytic processing.

To avoid confusion, the novel *Calystegia sepium* lectin will be referred to as Calsepa and its corresponding gene as LEC-Calsepa.

2. Materials and methods

2.1. Plant material

Rhizomes of the hedge bindweed (*Calystegia sepium*) were collected locally in December. Rhizome apices (top 2 cm) were dissected manually, frozen in dry ice and stored at −70°C.

2.2. Isolation of Calsepa

Calsepa was isolated from whole rhizomes as described previously [9].

2.3. Mass spectrometry

Mass spectrometry was performed using a MALDI-TOF (matrix-assisted laser desorption ionization time of flight) instrument, a VG TofSpec by Micromass Ltd. (Manchester, UK). Calibration was done externally using horse heart cytochrome *c* singly and doubly charged ions (respectively [M+H]+12361.10 and [M+2H]2+6181.05). The matrix employed was sinapinic acid (Aldrich). The instrument was operated in the linear mode with an accelerating voltage of 24000 V. The energy of the pulsed UV laser was adjusted to near threshold values for desorption/ionization in order to obtain optimal resolution at the observed masses. Approximately 20 shots were averaged to obtain a spectrum.

2.4. Analytical methods

Extracts and purified Calsepa was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using 12.5–25% (w/v) acrylamide gradient gels as described by Laemmli [10].

2.5. RNA isolation, construction and screening of cDNA library

Total cellular RNA was prepared from rhizome apices stored at −70°C as described by Van Damme and Peumans [11]. Poly(A)-rich RNA was enriched by chromatography on oligo-deoxythymidine cel-

PII S0014-5793(96)01211-2

lulose. A cDNA library was constructed with poly(A)-rich RNA from rhizome apices using the cDNA synthesis kit from Pharmacia (Uppsala, Sweden). cDNA fragments were inserted into the *Eco*RI site of PUC18 (Pharmacia). The library was propagated in *Escherichia coli* XL1 Blue (Stratagene, La Jolla, CA, USA).

Recombinant lectin clones were screened using a ^{32}P end-labeled degenerate oligonucleotide probe (17-mer, 5'-CCRTTRTTN-CCCCANGG-3') derived from the amino acid sequence PWGNNG of a cyanogen bromide cleavage fragment of Calsepa [9]. In a later stage, cDNA clones encoding Calsepa were used as probes to screen for more cDNA clones. Hybridizations were done overnight as reported previously [12]. Colonies that produced positive signals were selected and rescreened at low density using the same conditions. Plasmids were isolated from purified single colonies on a miniprep scale using the alkaline lysis method as described by Mierendorf and Pfeffer [13] and sequenced by the dideoxy method [14]. DNA sequences were analyzed using programs from PC Gene (Intelligenetics, Mountain View, CA, USA) and Genepro (Riverside Scientific, Seattle, WA, USA).

2.6. Northern blot

RNA electrophoresis was performed according to Maniatis et al. [15]. Approximately 3 μg of poly(A)-rich RNA was denatured in glyoxal and DMSO and separated in a 1.2% (w/v) agarose gel. Following electrophoresis the RNA was transferred to Immobilon N membranes (Millipore, Bedford, MA, USA) and the blot hybridized using a random-primer-labeled cDNA insert. Hybridization was performed as reported by Van Damme et al. [16]. An RNA ladder (0.16–1.77 kb) was used as a marker.

2.7. Molecular modelling

Amino acid sequence alignments were performed on a MicroVAX 3100 (Digital, Evry, France) using the ialign program of PIR/NBRF (Washington, DC, USA). Hydropathic profiles of Kyte and Doolittle [17] were calculated with a window of 9 amino acid residues using the MacProMass software (S. Vemuri and T.D. Lee, 1989) run on a Macintosh LC. Secondary structure features (β-sheets, α-helices and β-turns) were predicted according to Chou and Fasman [18] and the GOR [19] methods, using MacVector (Kodak) run on the Macintosh LC. The method of Cid et al. [20], based on a hydrophobicity scale, was also used to predict the secondary structure of Calsepa. Hydrophobic cluster analysis (HCA) [21,22] was performed to delineate the hydrophobic clusters along the amino acid sequences of the lectins from *Artocarpus integrifolia* (called jacalin), *Maclura pomifera* and *Calystegia sepium*. HCA plots were generated on the Macintosh LC using the HCA-Plot2 program (Doriane, Paris, France).

```
lpsfcfpiiiqei MAVPMDTISG PWGNNGGNFW SFRPVNKINQ IVISYGGGGN  40

   NPIALTFSST KADGSKDTIT VGGGGPDSIT GTEMVNIGTD EYLTGISGTF  90

   GIYLDNNVLR SITFTTNLKA HGPYGQKVGT PFSSANVVGN EIVGFLGRSG 140

   YYVDAIGTYN RHK  153
```

Fig. 1. Deduced amino acid sequence of the *Calystegia sepium* lectin. Since the methionine at position 14 is probably the first amino acid the residues preceding this methionine are shown in lower case. The sequences corresponding to the sequences of the CNBR fragments are underlined.

3. Results

3.1. Molecular cloning of LECCalsepa

Screening of a cDNA library constructed with poly(A)-rich RNA from rhizome apices using a synthetic oligonucleotide derived from the amino acid sequence of the CNBR fragments of Calsepa yielded multiple positive clones. Sequence analysis of the cDNA clones encoding Calsepa revealed that they contain an open reading frame of 498 bp encoding a 166 amino acid precursor with one putative initiation codon at position 14 of the deduced amino acid sequence (Fig. 1). Translation starting with this methionine residue results in a protein of 153 amino acids with a calculated molecular mass of 16062 Da and an isoelectric point of 6.07. A comparison of the deduced amino acid sequence of LECCalsepa and the previously reported sequences of two CNBr cleavage fragments of Calsepa demonstrates a perfect match between residues D6–V25 and V75–I92 of the cDNA and the two fragments (Fig. 1). In addition, the amino acid composition of Calsepa and the composition calculated from the cDNA are virtually identical, which strongly suggests that the correct reading frame has been determined (Table 1).

Multiple LECCalsepa cDNAs have been sequenced. Although most of the sequences were identical one cDNA encodes a lectin polypeptide in which A125 is substituted by a valine residue.

Table 1
Comparison of the amino acid composition of Calsepa calculated from the deduced amino acid sequence and the compostion determined by analysis of the purified protein

Amino acid	Number of residues calculated from the deduced sequence	Number of residues determined by analysis of Calsepa[a]
Ala	6	6
Arg	4	5
Asp/Asn	21	21
Cys	0	0
Glu/Gln	5	5
Gly	26	26
His	2	2
Ile	13	12
Leu	6	6
Lys	6	6
Met	3	1–2
Phe	7	7
Pro	7	7
Ser	12	12
Thr	15	15
Trp	2	(2)
Tyr	7	7
Val	11	12
Total	153	(153)

[a]For the amino acid analysis, the number of residues was the nearest integer calculated from the observed mol % (including 1.3% tryptophan) and assuming 153 total residues (assumed values indicated in parentheses).

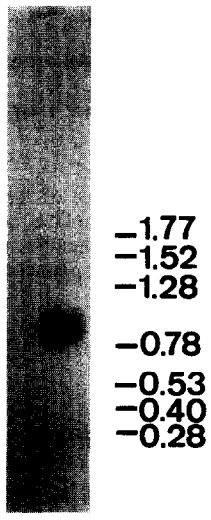

Fig. 2. Northern blot of poly(A)-rich RNA isolated from *Calystegia sepium* rhizomes. The blot was hybridized using the labelled cDNA insert LECCalsepa. Numbers on the right show RNA size (kb).

No signal peptide could be traced in the deduced amino acid sequence [23], indicating that Calsepa is not synthesized on the endoplasmic reticulum. Since the mature protein is blocked, the N-terminal amino acid residue could not be determined. However, the molecular mass of the (unglycosylated) Calsepa subunits as determined by mass spectrometry (16092 ± 40 Da) suggest that (i) the mature protein starts with the first methionine (amino-terminally blocked) and (ii) no proteolytic post-translational processing occurs.

3.2. Northern blot analysis

Northern blot analysis was performed to determine the total length of the lectin mRNA from *Calystegia sepium*. Hybridization of the blot using the synthetic oligonucleotide as a probe yielded one band of approximately 800 nucleotides (Fig. 2). This result was identical when hybridization was performed using the random-primer-labeled cDNA clone LECCalsepa. The size of the RNA is consistent with the length of the cDNA clones which were analyzed.

3.3. Sequence homology and structural similarities between Calsepa and the lectins from the Moraceae species Artocarpus integrifolia and Maclura pomifera

A search in the databases revealed that Calsepa exhibits reasonable sequence homology with the lectins from the taxonomically unrelated Moraceae species *Artocarpus integrifolia* and *Maclura pomifera*. As shown in Fig. 3 Calsepa can be aligned with the β- and α-chains of jacalin, the percentages of identity and homology being 30% and 49%, respectively. However, a few insertions and deletions have to be introduced to maximize the homologies between the amino acid sequences. HCA further indicates that Calsepa also exhibits structural homologies with the Moraceae lectins (Fig. 4). In this respect, some highly conserved hydrophobic residues of jacalin occur at homologous positions in Calsepa and various hydrophobic clusters appear as totally or partly conserved in both HCA plots. These structural homologies suggest a very similar folding for both proteins, despite some deletions or insertions of a few amino acid residues occurring along the amino acid sequence of Calsepa when compared to that of jacalin. Due to the insertions and deletions between the amino acid sequences, the hydropathic profiles of both proteins are slightly different (results not shown).

4. Discussion

The present paper describes the molecular cloning of Calsepa, a recently described novel mitogenic mannose/maltose lectin from the rhizomes of hedge bindweed (*Calystegia sepium*), a typical representative of the plant family Convolvulaceae. Based on the results of mass spectrometry and the deduced amino acid sequence of the cDNA it was shown that the mature Calsepa subunit comprises the complete open reading frame between the start and stop codons. It can be concluded, therefore, that the primary translation product of the lectin mRNA contains no signal peptide and does not undergo any proteolytic processing. According to the currently accepted ideas of protein synthesis and targetting the apparent absence of a signal peptide implies that Calsepa is not synthesized on the endoplasmic reticulum and most probably is a cytoplasmic protein. To the best of our knowledge, Calsepa is the first example of a plant lectin that is synthesized without a signal peptide. It should be mentioned, however, that the lectin from the fungus *Arthrobotrys oligospora*, which is a predominant cytoplasmic protein in the mycelium, is also synthesized without a signal peptide [24,25].

As has been reported previously, the N-terminal sequences of the cyanogen bromide cleavage fragments of Calsepa had no significant homology with any other protein. However, a search in the databases using the complete deduced amino acid sequence of LECCalsepa revealed that Calsepa has reasonable homology with the lectins from *Artocarpus integrifolia*

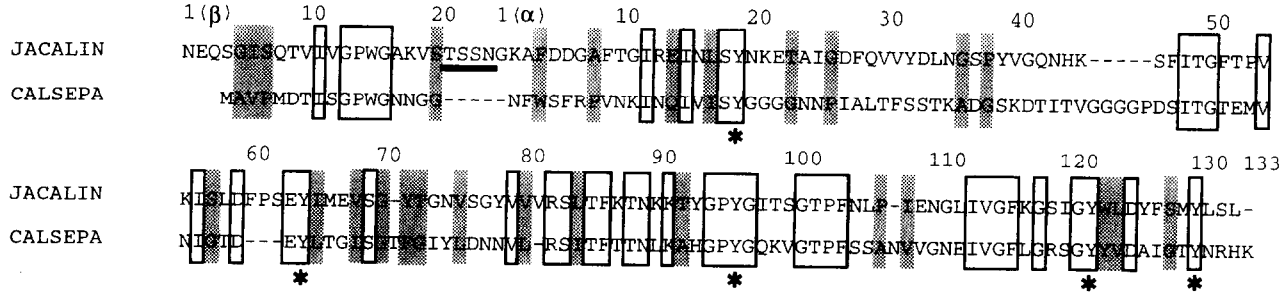

Fig. 3. Alignment of the amino acid sequences of the β- and α-chains of jacalin and Calsepa. Identical residues (open boxes) and homologous residues (gray boxes) have been indicated. The linker tetrapeptide TSSN connecting the β- and α-chains of jacalin is underlined. Conserved tyrosine residues are indicated by asterisks (*).

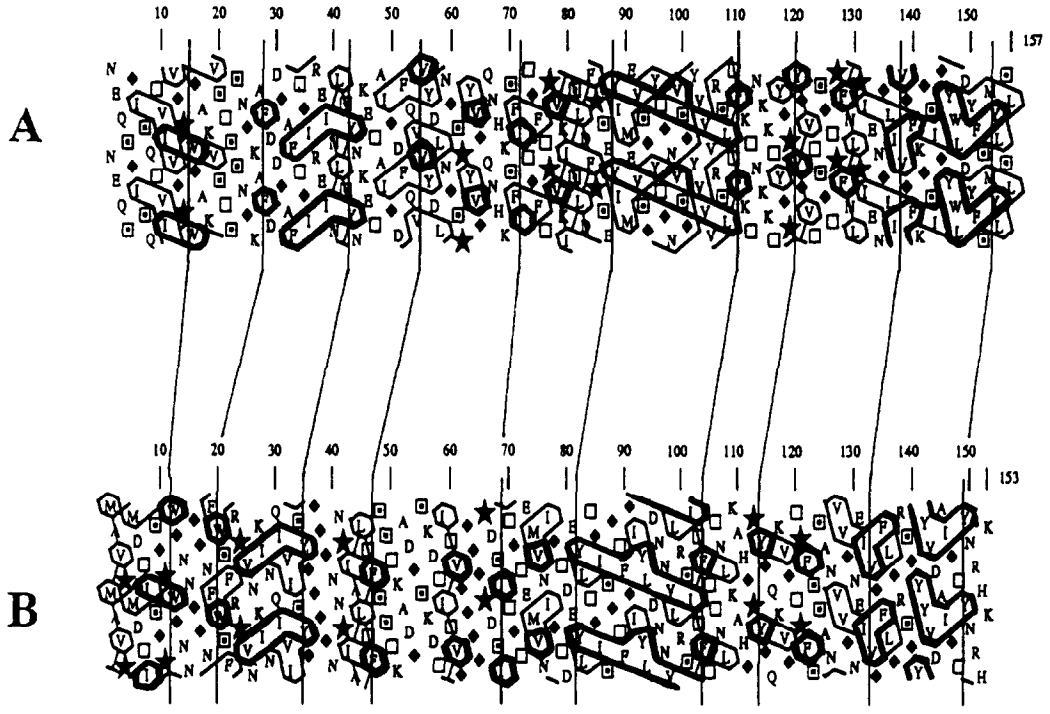

Fig. 4. Comparison of the HCA plots of the β- and α-chains of jacalin (A) and Calsepa (B). Some structurally conserved hydrophobic residues common to both sequences are indicated by joined vertical bars to facilitate the identification of the structurally conserved regions common to both lectins. These structurally conserved regions, which mainly correspond to hydrophobic clusters, are circled in bold.

and *Maclura pomifera*. Molecular cloning of jacalin made it possible to unravel its complex post-translational processing. The primary translation product of jacalin mRNA is a pre-prolectin comprising a signal peptide of 21 residues followed by a propeptide of 39 residues, a β-peptide of 20 residues, a linker tetrapeptide TSSN and an α-peptide of 133 residues [6]. Pre-projacalin is processed into projacalin by cotranslational removal of the signal peptide. Subsequently, the propeptide is removed from the projacalin by proteolytic cleavage between the propeptide and the N-terminus of the β-chain. Finally, the resulting polypeptide is processed into the β- and α-chains by the excision of the linker tetrapeptide TSSN. The biosynthesis and post-translational processing of the *Maclura pomifera* lectin are still unclear because the gene has not been cloned yet. However, considering that both lectins have a very similar polypeptide composition and a high sequence homology, one can assume that the post-translational processing of the *Maclura pomifera* lectin is comparable to that of jacalin. As shown in Fig. 5, the primary translation product of the Calsepa mRNA is very different from pre-projacalin since it lacks the signal peptide, the N-terminal propeptide and the linker tetrapeptide TSSN. However, a closer examination of the aligned sequences of the lectins shows that the Calsepa subunit corresponds to a polypeptide built up of the (peptide bond-linked) homologues of both the β- and α-chains of jacalin. It appears, therefore, that the final processing of the jacalin precursor, which separates the β- and α-chains by the excision of a linker tetrapeptide, does not occur in Calsepa. Possibly, the deletion in Calsepa of the linker tetrapeptide prevents the proteolytic cleavage between the β- and α-chain homologues.

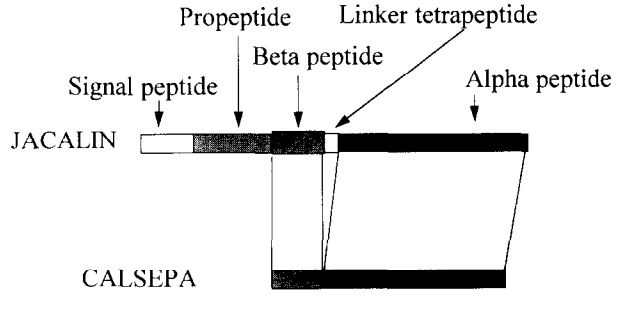

Fig. 5. Schematic representation of the structure of Calsepa and jacalin.

Although jacalin has been crystallized [26,27], no three-dimensional model has been reported yet. Circular dichroism spectra of jacalin and the *Maclura pomifera* lectin indicate that they are both β-sheet proteins [28,29], a presumption which was confirmed by the structure prediction of the α-chains of jacalin with the GOR method for the α-chains [30]. The results of our structure predictions of Calsepa by three different methods [18–20] indicate that this lectin also is mainly built up of β-sheets (Fig. 6). Moreover, the occurrence of β-sheets along the amino acid sequence of Calsepa is confirmed by the HCA plot (Fig. 4).

The three-dimensional structure of the carbohydrate-binding site of the Moraceae lectins has not been resolved yet. However, it has been demonstrated that tyrosine, tryptophan and histidine residues are involved in the binding of sugars [28,29,31]. Most of the tyrosine residues occurring along the amino acid sequence of Calsepa (Fig. 3) are well exposed and could therefore participate in the carbohydrate-binding site of this lectin.

The results of the molecular cloning and modelling clearly indicate that Calsepa is evolutionarily related to the Moraceae lectins. Although no general conclusions can be drawn as yet, the occurrence of related lectins in the taxonomically far dis-

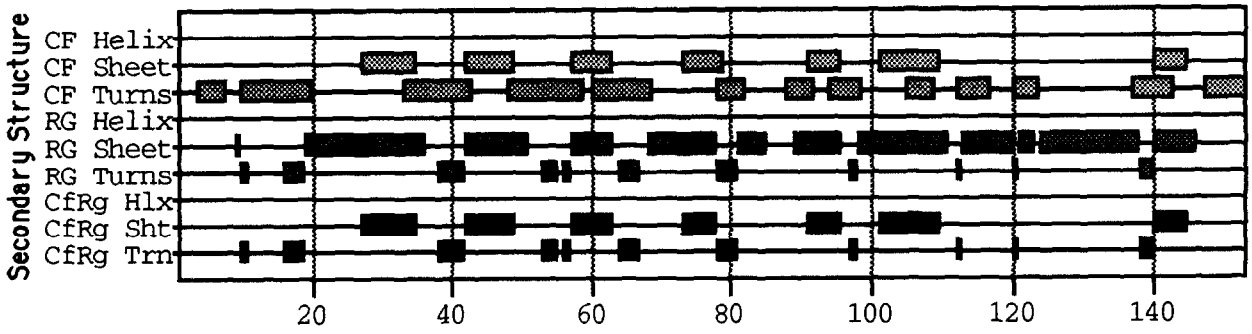

Fig. 6. Structure prediction of α-helices, β-sheets and β-turns occurring along the amino acid sequence of Calsepa according to the Chou and Fasman (CF) [18] and the Garnier et al. (RG) [19] methods.

tant families Moraceae and Convolvulaceae raises the question whether these lectins possibly represent a fifth subgroup of plant lectins.

Acknowledgements: This work was supported in part by grants from the Catholic University of Leuven (OT/94/17), CNRS and the Conseil régional de Midi-Pyrénées (A.B., P.R.), and the National Fund for Scientific Research (Belgium, Fonds voor Geneeskundig Wetenschappelijk Onderzoek Grant 2.0046.93). W.P. is Research Director, E.J.M.V.D. Postdoctoral Fellow and P.V. a Research Associate of this fund.

References

[1] Sharon, N. and Lis, H. (1990) FASEB J. 4, 3198–3208.
[2] Raikhel, N.V., Lee, H.-I. and Broekaert W.F. (1993) Annu. Rev. Plant Physiol. Plant Mol. Biol. 44, 591–615.
[3] Barbieri, L., Batelli, G.B. and Stirpe, F. (1993) Biochim. Biophys. Acta 1154, 237–282.
[4] Van Damme, E.J.M., Smeets, K. and Peumans, W.J. (1995) in: Lectins, Biomedical Perspectives (Pusztai, A. and Bardocz, S., Eds.), pp. 59–80. Taylor and Francis, London.
[5] Raina, A. and Datta, A. (1992) Proc. Natl. Acad. Sci. USA 89, 11774–11778.
[6] Yang, H. and Czapla, T.H. (1993) J. Biol. Chem. 268, 5905–5910.
[7] Van Damme, E.J.M., Barre, A., Rougé, P., Van Leuven, F. and Peumans, W.J. (1996) Eur. J. Biochem. 235, 128–137.
[8] Kaku, H., Tanaka, Y., Tazaki, K., Minami, E., Mizuno, H. and Shibuya, N. (1996) J. Biol. Chem. 271, 1480–1485.
[9] Peumans, W.J., Winter, H.C., Bemer, V., Van Leuven, F., Goldstein, I.J., Truffa-Bachi, P. and Van Damme, E.J.M. (1996) Glycoconjugate J. (in press).
[10] Laemmli, U.K. (1970) Nature 227, 680–685.
[11] Van Damme, E.J.M. and Peumans, W.J. (1993) in: Lectins and Glycobiology (Gabius, H.-J. and Gabius, S., Eds.), pp. 458–468, Springer Verlag, Berlin.
[12] Van Damme, E.J.M., Kaku, H., Perini, F., Goldstein, I.J., Peeters, B., Yagi, F., Decock, B. and Peumans, W.J. (1991) Eur. J. Biochem. 202, 23–30.
[13] Mierendorf, R.C. and Pfeffer, D. (1987) Methods Enzymol. 152, 556–562.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[16] Van Damme, E.J.M., Smeets, K., Torrekens, S., Van Leuven, F., Goldstein, I.J. and Peumans, W.J. (1992) Eur. J. Biochem. 206, 413–420.
[17] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[18] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[19] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[20] Cid, H., Bunster, M., Arriagada, E. and Campos, M. (1982) FEBS Lett. 150, 247–254.
[21] Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P. (1987) FEBS Lett. 224, 149–155.
[22] Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.P. (1990) Biochimie 72, 555–574.
[23] von Heijne, G. (1986) Nucleic Acids Res. 11, 4683–4690.
[24] Rosén, S., Kata, M., Persson, Y., Lipniunas, P.H., Wikström, M., Van den Hondel, C.A.M.J.J., van den Brink, J.M., Rask, L., Heden, L.-O. and Tunlid, A. (1996) Eur. J. Biochem. 238, 822–829.
[25] Rosén, S. (1996) Dissertation, University of Lund.
[26] Dhanaraj, V., Patanjali, S.R., Surolia, A. and Vijayan, M. (1988) J. Mol. Biol. 203, 1135–1136.
[27] Banerjee, R., Dhanaraj, V., Mahanta, S.K., Surolia, A. and Vijayan, M. (1991) J. Mol. Biol. 221, 773–776.
[28] Young, N.M., Johnston, R.A.Z., Szabo, A.G. and Watson, D.C. (1989) Arch. Biochem. Biophys. 270, 596–603.
[29] Mahanta, S.K., Sastry, M.V.K. and Surolia, A. (1990) Biochem. J. 265, 831–840.
[30] Young, N.M., Johnston, R.A.Z. and Watson, D.C. (1991) FEBS Lett. 282, 382–384.
[31] Chowdhury, S., Ahmed, H. and Chatterjee, B.P. (1991) Biochimie 73, 563–571.