Volume 257, number 1, 81–85 FEB 07725 October 1989

The amino acid sequence of *Erythrina corallodendron* lectin and its homology with other legume lectins*

Rivka Adar, Michael Richardson+, Halina Lis and Nathan Sharon

Department of Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel and +*Department of Biological Sciences, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, England*

Received 1 September 1989

The primary sequence of *Erythrina corallodendron* lectin was deduced from analysis of the peptides derived from the lectin by digestion with trypsin, chymotrypsin, *Staphylococcus aureus* V8 protease, elastase and lysylendopeptidase-C, and of fragments generated by cleavage of the lectin with dilute formic acid in 6 M guanidine hydrochloride. Purification of the individual peptides was achieved by gel filtration, followed by reverse phase HPLC. The glycosylation site (Asn^{17}-Leu^{18}-Thr^{19}) was deduced from analysis of the glycopeptide isolated from a pronase digest of the lectin before and after deglycosylation of the glycopeptide with endoglycosidase F. Comparison of the sequence of 244 residues thus obtained with those of 9 other legume lectins revealed extensive homologies, including 39 invariant positions and 60 partial identities. These data provide further evidence for the conservation of the lectin gene in leguminous plants.

Primary sequence; Sequence homology; Lectin, legume; Glycosylation site

1. INTRODUCTION

Erythrina is a family of deciduous leguminous trees and shrubs widely spread in the tropics and subtropics. Since 1980, lectins from some 20 species of this family have been isolated in different laboratories, a dozen of which by us [1,2]. All *Erythrina* lectins studied are specific for galactose and *N*-acetylgalactosamine and show pronounced preference for *N*-acetyllactosamine. They are glycoproteins (3–10% carbohydrate) of molecular masses in the range of 56000–68000 Da and are composed of two identical or nearly identical subunits. Whenever examined, the carbohydrate of the lectins was found to consist of glucosamine, mannose, L-fucose and xylose, predominantly in the form of the asparagine-linked heptasaccharide Manα3(Manα6)(Xylβ2)Manβ4GlcNAcβ4 (Fucα3)GlcNAc [3]. The N-terminal amino acid sequences, determined on 10 *Erythrina* lectins for up to 15 amino acids have been nearly identical [2].

Correspondence address: N. Sharon, Department of Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel

Abbreviations: TPCK, L-1-tosylamide-2-phenylethyl-chloromethyl ketone; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; ConA, concanavalin A; DBL, *Dolichos biflorus* lectin; ECorL, *Erythrina corallodendron* lectin; LCL, *Lens culinaris* (lentil) lectin; PHA-L, leukoagglutinin from *Phaseolus vulgaris*; PNA, peanut agglutinin; PSL, *Pisum sativum* (pea) lectin; SBA, soybean agglutinin; SL, sainfoin (*Onobrychis viciifolia* lectin)

* Dedicated to the memory of Professor Edgar Lederer (1908–1988)

We have now established the nearly complete sequence (the first 244 amino acids) of *Erythrina corallodendron* lectin (ECorL), as well as the glycosylation site. Comparison of this sequence with those of 9 other legume lectins reveals, as predicted [4], extensive homologies.

2. MATERIALS AND METHODS

2.1. *Preparation of ECorL*

The lectin extracted from the seeds of *Erythrina corallodendron* was purified by fractional precipitation with ammonium sulfate and affinity chromatography on a column of lactose coupled to divinylsulphone-activated Sepharose [2].

2.2. *Enzyme digestion*

The following digestions were performed, all at an enzyme/substrate ratio of 1:50. (i) ECorL (2 mg/ml) was heat-denatured at pH 2 [5] and digested with lysylendopeptidase-C (Calbiochem) in phosphate buffer, pH 7.8, for 6 h at 37°C [6]. (ii) Samples of ECorL, 10–15 mg in 1 ml of 0.2 M *N*-ethylmorpholine HCl buffer, pH 8.5, were digested separately with trypsin (TPCK-treated, Sigma) and chymotrypsin (α-chymotrypsin, type VIII, TPCK-treated, Sigma), for 5–10 h at 37°C. (iii) ECorL, 7 mg/ml of 0.1 M sodium bicarbonate, pH 8.1, was digested with *Staphylococcus aureus* V8 protease (Miles) at 37°C. Half of the sample was removed after 24 h, while the remainder was kept for 5 days at room temperature. (iv) The lectin, 10 mg in 2 ml of 0.2 M Tris-HCl, pH 8.8, was denatured by boiling for 10 min and digested with elastase (Sigma) for 75 min at 37°C.

2.3. *Chemical cleavage*

ECorL, 10 mg/ml, was cleaved with 75% formic acid in 6 M guanidine hydrochloride for 72 h at 37°C.

2.4. *Peptide separation and purification*

For the separation of the peptides in the lysylendopeptidase-C digest, aliquots (100 μg/well) were loaded onto 8–25% gradient

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

polyacrylamide slab gels containing 0.1% SDS and electrophoresed as described [7]. The separated peptides were electrophoretically transferred to a polyvinylidene difluoride membrane (PVDF, Millipore). Strips of the membranes were stained with Coomassie blue and bands corresponding to the peptides were cut out from the membrane [8].

Peptides in the tryptic, chymotryptic and elastase digests of ECorL were fractionated on columns (1 × 200 cm) of Biogel P-6 in 0.1 M ammonium bicarbonate pH 8.1 as in [9], while fragments resulting from acid hydrolysis were fractionated on a column (1 × 200 cm) of Biogel P-30 in 70% formic acid [10]. Elution was followed by measuring the absorbance at 230 nm. Fractions corresponding to the peaks were collected, lyophilized and dissolved in 0.1% trifluoroacetic acid for further purification by reverse-phase HPLC. The peptides in the *S. aureus* V8 digest were purified directly by reverse-phase HPLC.

HPLC was carried out on a Vydac analytical reverse-phase column (25 cm × 4.6 mm; 218TP54 HP Genenchem, San Francisco, CA) in a Waters 600E multisolvent delivery system using variable gradients of 0–50% acetonitrile (HPLC grade, Merck) in 0.1% trifluoroacetic acid [9,11]. Peptides were detected by measuring the absorbance at 214 nm, and collected manually directly from the UV monitor.

2.5. *Preparation of glycopeptide*

ECorL (500 mg, containing 3% neutral sugar) was denatured and digested with 10 mg of pronase (Calbiochem) as described [5]. The digest was lyophilized, the dry material dissolved in 7 ml of 0.01 M acetic acid and insoluble material removed by centrifugation. The supernatant was applied in two portions to a Sephadex G-50 column (2 × 160 cm) using 0.01 M acetic acid as a solvent. Fractions were collected, monitored by absorbance at 230 nm, and examined for their neutral sugar content by the phenol-sulfuric method [12], using mannose as standard. Sugar-containing fractions were pooled, lyophilized and rechromatographed on the same column, under the same conditions as described above.

2.6. *Deglycosylation*

An aliquot of the rechromatographed material (575 μg of mannose) was lyophilized, dissolved in 100 μl of 20 mM potassium phosphate buffer, pH 6.5, 25 mM EDTA and treated with 1 U of endoglycosidase F (Boehringer-Mannheim) for 20 h at 37°C. The enzyme-treated, as well as a sample of the untreated material, were analysed by reverse-phase HPLC using a gradient of 0–7% acetonitrile in 0.1% trifluoroacetic acid.

2.7. *Sequence determination*

Peptides derived from the various digests were subjected to microsequence analysis using the 4-*N,N*-dimethylaminoazobenzene-4′-isothiocyanate (DABITC)/phenylisothiocyanate (PITC) double-coupling method, followed by thin-layer chromatographic identification of the amino acid derivatives liberated [13,14]. The glycopeptide purified by HPLC, its endoglycosidase digestion product and the PVDF membranes containing peptides from the lysylendopeptidase-C digest were sequenced on a gas phase Applied Biosystem automatic sequencer model 470 A.

3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

The sequence of 244 amino acids, together with the details of the overlapping peptides and fragments from which it was deduced is shown in fig.1. Digestion with trypsin and *S. aureus* V8 protease yielded peptides from which most of the sequence of ECorL could readily be established. Digestions with chymotrypsin, lysylendopeptidase-C and elastase gave good overlaps that were helpful in determining the missing residues.

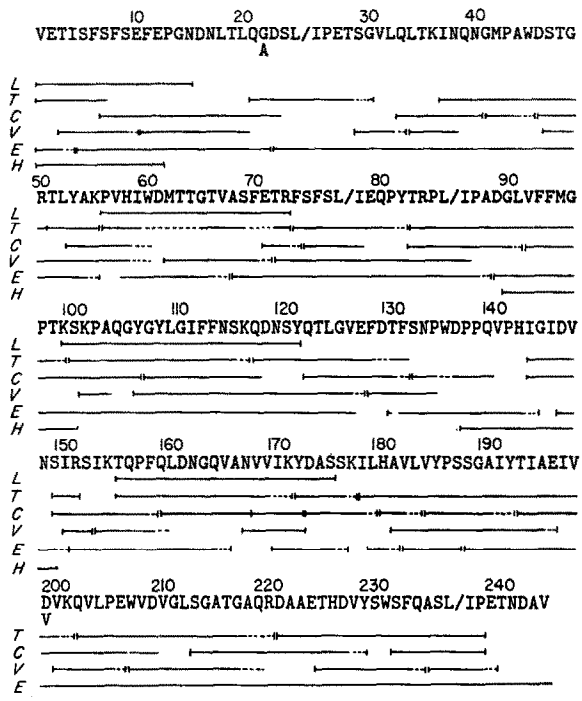

Fig.1. Alignment of ECorL and peptides used for sequence determination. L, peptides obtained from digestion with lysylendopeptidase-C; T, tryptic peptides; C, chymotryptic peptides; E and V, peptides obtained by digestion with elastase and *S. aureus* V8 protease, respectively; H, peptides obtained by hydrolysis with dilute formic acid in 6 M guanidine hydrochloride. Solid lines indicate regions of peptides sequenced by the manual DABITC/PITC method (except for the lysylendopeptidase-C peptides which were sequenced by the automatic sequencer). Dashed lines indicate residues which were not sequenced or yielded an unsatisfactory result.

Cleavage with dilute formic acid in 6 M guanidine HCl yielded 3 main fragments; one of them, resulting from hydrolysis of peptide bond Asp^{136}-Pro^{137}, facilitated the elucidation of the sequence from amino acid 137 onwards. Each of the residues shown in fig.1 was identified at least twice in the sequence analysis using the manual DABITC/PITC double-coupling method or the automatic sequencer.

Microheterogeneity was observed in positions 22 (with Ala replacing Gly) and 199 (with Val replacing Asp), suggesting differences in the sequence of the two subunits of ECorL in these positions.

3.2. *Glycosylation site*

Digestion of ECorL by pronase, followed by gel filtration on Sephadex G-50, afforded a crude glycopeptide preparation in good yield (77% based on the neutral sugar content of the starting material). Fractionation of this preparation on reverse-phase HPLC led to the isolation of 5 main (glyco)peptides (fig.2A). HPLC of the crude glycopeptide after deglycosylation by endoglycosidase F (fig.2B), showed a single change in the enzyme-treated sample, i.e. a shift of peptide no.2 to a new position (peptide no.6),

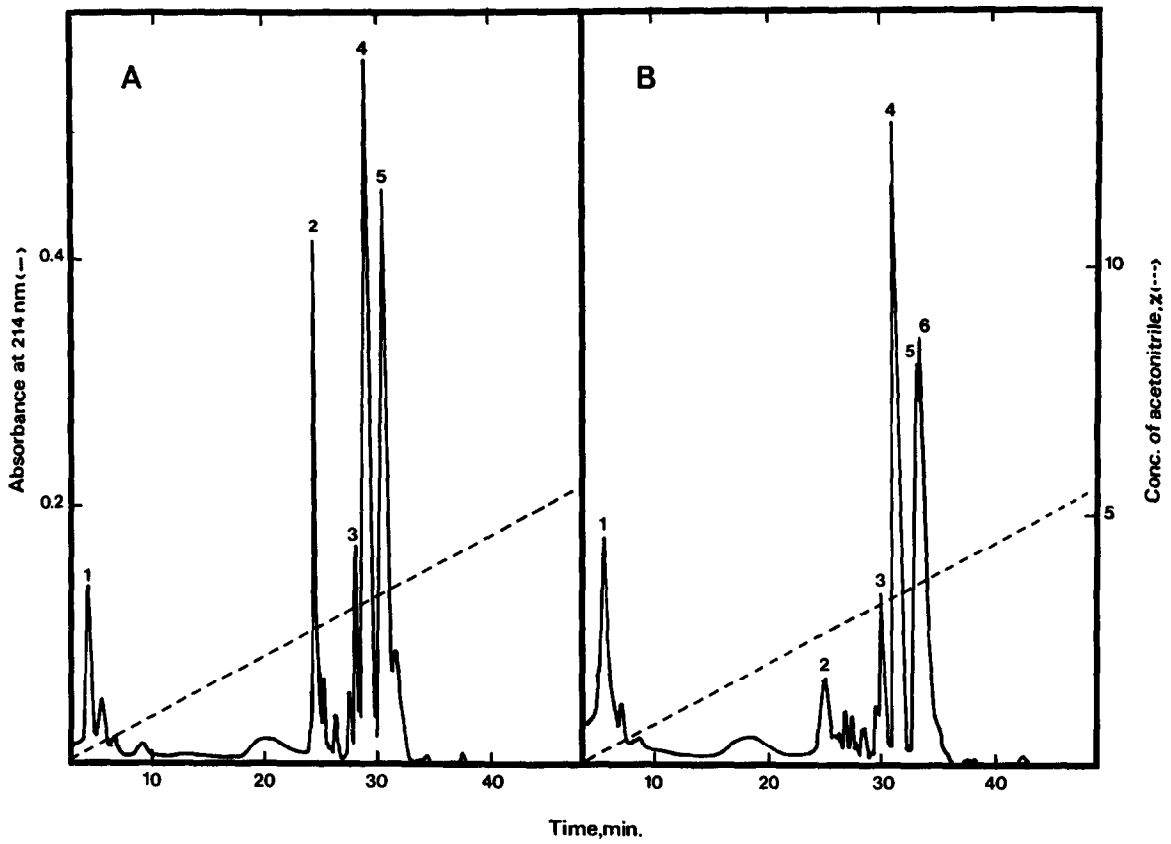

Fig.2. Reverse-phase HPLC of the crude glycopeptide preparation isolated from the pronase digest of ECorL before (A) and after (B) deglycosylation with endoglycosidase F, respectively. Chromatography was performed as described in section 2, except that a gradient of 0–7% acetonitrile in 0.1% trifluoroacetic acid was used.

indicating that peptide no.2 contains the glycosylation site. Analysis on the automatic sequenator of peptides 2 and 6 demonstrated that the sequence around the glycosylation site is Glu-Pro-Gly-Asn-Asp-Asn-Leu-Thr (corresponding to amino acids 12–19) and that Asn[17] is glycosylated. This conclusion is based on the finding that in the intact glycopeptide (no.2), analysis stopped after Asp[16], whereas in the deglycosylated glycopeptide (no.6, which contains a single *N*-acetylglucosamine residue) there was a 'hole' after this position [15,16], followed by Leu[18]-Thr[19].

3.3. *Characteristics of the sequence*

The molecular mass of a single subunit of ECorL calculated from the sequence of 244 amino acids (and including the carbohydrate group) corresponds to 27935 Da. This value is in excellent agreement with previous estimates for the subunits (28000–30000 Da) based on SDS-PAGE and analytical ultracentrifugation [2], indicating that only a very small number of amino acids is still missing. The amino acid composition as derived from the sequence (table 1) is in agreement with the results of previous amino acid analysis [2].

Table 1

The amino acid composition of a single subunit of ECorL

Amino acid	Residues/mol	
	Analysis [2]	Sequence (this work)
Ala	20	17
Arg	5.5	5
Asn		11
Asp	30	16
Gln		14
Glu	29	11
Gly	20.5	17
His	4.5	4
Ile	14.5	13
Leu	18	13
Ile/Leu	–	4
Lys	9.5	9
Met	2	3
Phe	14	14
Pro	19.5	17
Ser	23.5	23
Thr	20.5	19
Trp	6.5	5
Tyr	9	9
Val	16.5	20

3.4. *Homologies*

Comparison of the amino acid sequence of ECorL with that of other legume lectins (fig.3) reveals a high degree of homology. The amino acids in 39 positions are invariant in all the lectins listed. It is notable that these highly conserved amino acids include 4 which correspond to residues previously identified in ConA [17] as being important in the binding of Ca^{2+} and Mn^{2+},

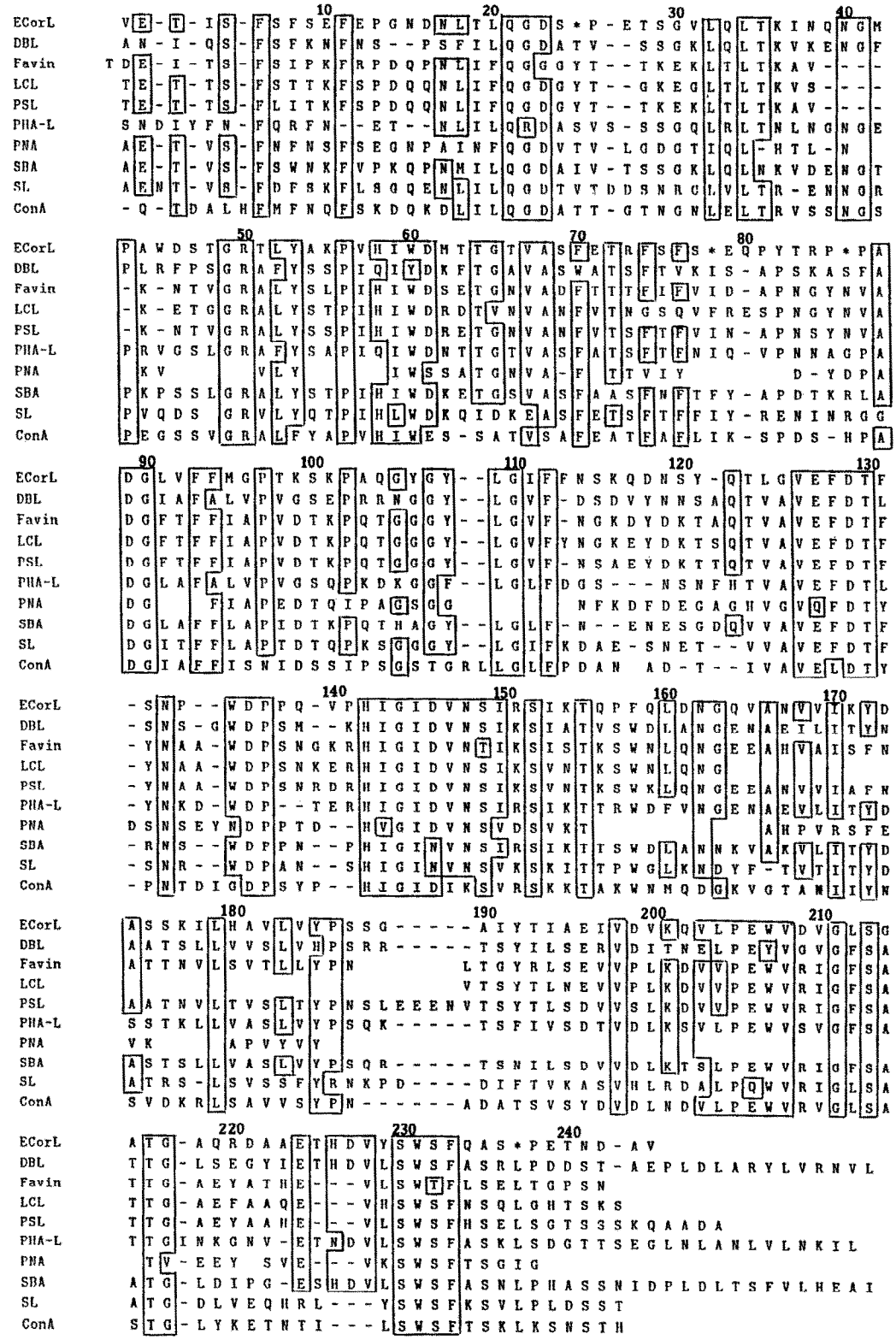

Fig.3. Comparison of amino acid sequences of ECorL, DBL [22,23], favin [24], LCL [24], PSL [24], PHA-L [18], PNA [25], SBA [24] and SL [24]. The sequence marked as ConA is that of pro-ConA, which exhibits circular homology with the mature lectin [26,27]. The identical residues have been boxed. –, Deletions inserted in all sequences to maximize homology; *, leucine or isoleucine.

i.e. $Asp^{10(129)}$, $Asp^{19(136)}$, $His^{24(142)}$, and $Ser^{34(152)}$. (The numbers without parentheses refer to the ConA sequence; those in parentheses are of pro-ConA after alignment with ECorL as given in fig.3.) Another metal binding residue, $Glu^{8(127)}$, is nearly invariant, except for peanut agglutinin in which glutamine was found instead. Similarly, the amino acids comprising the 3-dimensional structure of the hydrophobic cavity of ConA [17,18] are invariant ($Val^{89(208)}$, $Phe^{111(233)}$ and $Phe^{212(93)}$ or highly conserved (e.g. $Ser^{113(235)}$) in homologous positions in all the lectins. On the other hand, the residues which constitute the monosaccharide binding site in ConA [17] appear to be poorly conserved, for instance, $Asp^{208(89)}$ is the only residue maintained in this site in the other lectins (except for leukoagglutinin from phaseolus vulgaris in which it is replaced by a valine residue). It is also of interest that *N*-glycosylation triplets (Asn-X-Ser/Thr) are absent from some of the lectins; when present, they are located at different positions in the primary structure and are not always glycosylated. Thus, on ECorL, there is a single occupied glycosylation site at Asn^{17}; in PHA-L there are two occupied sites at positions 10 and 62 [18,19] and in soybean agglutinin there are 3 glycosylation sites at positions 40, 75 and 114, only one of which (position 75) is occupied [20]. In favin, the sugar moiety is located at Asn^{173} [21].

The similarity in the primary structure between ECorL and other legume lectins listed in fig.3 further supports the proposal that all these lectins share a common evolutionary origin. Their structures have been highly conserved in evolution, presumably for ensuring the maintenance of an important physiological function(s) yet to be determined.

Acknowledgements: We wish to thank Ms S. Haalman for the preparation of the lectin and Ms R. Samuel for valuable secretarial help. R.A. wishes to acknowledge a FEBS Fellowship for a short-term visit in the laboratory of Dr Michael Richardson. This work was supported by the US-Israel Binational Science Foundation Grant No.3727.

REFERENCES

[1] Iglesias, J.L., Lis, H. and Sharon, N. (1982) Eur. J. Biochem. 123, 247–252.
[2] Lis, H., Joubert, F.J. and Sharon, N. (1985) Phytochemistry 24, 2803–2809.
[3] Ashford, D., Dwek, R.A., Welply, J.K., Amatayakul, S., Homans, S.W., Lis, H., Taylor, G.N., Sharon, N. and Rademacher, T. (1987) Eur. J. Biochem. 166, 311–320.
[4] Foriers, A., Wuilmart, C., Sharon, N. and Strosberg, D. (1977) Biochem. Biophys. Res. Commun. 4, 980–986.
[5] Lis, H. and Sharon, N. (1978) J. Biol. Chem. 253, 3468–3476.
[6] Richardson, M. and Granum, P.E. (1985) FEBS Lett. 182, 479–484.
[7] Fling, S.P. and Gregerson, D.S. (1986) Anal. Biochem. 155, 83–88.
[8] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[9] Richardson, M., Campos, F.A.P., Moriera, R.A., Ainouz, I.L., Begbie, R., Watt, M.B. and Pusztai, A. (1984) Eur. J. Biochem. 144, 101–111.
[10] Yarwood, A., Richardson, M., Sousa-Cavada, B. and Rouge, P. (1985) FEBS Lett. 184, 104–109.
[11] Richardson, M., Yarwood, A. and Rouge, P. (1987) FEBS Lett. 216, 145–150.
[12] Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A. and Smith, F. (1956) Anal. Chem. 28, 350–356.
[13] Chang, J.Y., Creaser, E.H. and Hughes, G.J. (1977) J. Chromatogr. 140, 125–128.
[14] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[15] Welinder, K.G. (1976) FEBS Lett. 72, 19–23.
[16] Miller, J.B., Hsu, R., Heinrikson, R. and Yachnin, S. (1975) Proc. Natl. Acad. Sci. USA 72, 1388–1391.
[17] Becker, J.W., Reeke, G.N., Wang, J.L., Cunningham, B.A. and Edelman, G.M. (1975) J. Biol. Chem. 250, 1513–1524.
[18] Hoffman, L.M. and Donaldson, D.D. (1985) EMBO J. 4, 883–889.
[19] Sturm, A. and Chrispeels, M.J. (1986) Plant Physiol. 81, 320–322.
[20] Becker, J.W., Cunningham, B.A. and Hemperly, J.J. (1983) in: Chemical Taxonomy, Molecular Biology and Function of Plant Lectins (Goldstein, I.J. and Etzler, M.E. eds) pp.31–45, A.R. Liss, New York.
[21] Cunningham, B.A., Hemperly, J.J., Hopp, T.P. and Edelman, G.M. (1979) Proc. Natl. Acad. Sci. USA 76, 3218–3222.
[22] Schnell, D.J. and Etzler, M.E. (1987) J. Biol. Chem. 262, 7220–7225.
[23] Schnell, D.J. and Etzler, M.E. (1988) J. Biol. Chem. 263, 14648–14653.
[24] Strosberg, A.D., Buffard, D., Lauwereys, M. and Foriers, A. (1986) in: The Lectins: Properties, Functions and Applications in Biology and Medicine (Liener, I.E. et al. eds) pp.249–264, Academic Press.
[25] Lauwereys, M., Foriers, A., Sharon, N. and Strosberg, A.D. (1985) FEBS Lett. 181, 241–244.
[26] Hemperly, J.J. and Cunningham, B.A. (1983) Trends Biochem. Sci. 8, 100–102.
[27] Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R. and Burgess, J. (1986) J. Cell Biol. 102, 1284–1297.