Biochemical and Biophysical Research Communications **276,** 23–28 (2000)
doi:10.1006/bbrc.2000.3420, available online at http://www.idealibrary.com on IDEAL®

# Complete Amino Acid Sequence of Japanese Chestnut Agglutinin

Keiichi Nomura,*[,1] Sachiko Nakamura,† Mihoyo Fujitake,‡ and Tetsu Nakanishi*

**Department of Plant Resource Science, Faculty of Agriculture, and †Graduate School of Science and Technology, Kobe University, 1 Rokkodai-chou, Nada-ku, Kobe 657-8501, Japan; and ‡Mass Spectrometry Laboratory, Osaka University of Pharmaceutical Sciences, 4-20-1 Nasahara, Takatsuki, Osaka 569-1094, Japan*

Received August 3, 2000

**The complete amino acid sequence of Japanese chestnut (*Castanea crenata* Sieb. et Zucc.) agglutinin (CCA) was determined. Analysis by SIMS of the acidic peptide obtained by pepsin digestion revealed that the N-terminal amino acid sequence should be Acetyl-Met-Glu-Glu. Prior to sequence analysis, redetermination of cysteine residues indicated the presence of one cysteine residue per subunit. The complete sequence was determined by endoproteinase Arg-C and *Achromobacter* protease I digestion, and CNBr cleavage. CCA consists of 309 amino acid residues with a high content of glycine (16.5 mol%) and one cysteine residue. The calculated molecular mass was 33,387 Da including the N-terminal acetyl group. C-terminal sequence analysis of intact CCA gave only one sequence, HMEYF, indicating that no heterogeneous CCA formed by posttranslational cleavage at the C-terminal region, as occurs in some legume lectins. Analysis of the sequence of CCA itself revealed that CCA could be divided into two structural domains, the N-domain and the C-domain, almost at the center. These domains share about 35% identical residues, so CCA has a repeat sequence. Also, both domains show a homology to jacalin-related lectins with 27–38% identity. These results suggest that the structure of CCA resembles two molecules of jacalin-related lectin.** 

***Key Words:*** **Japanese chestnut; lectin; amino acid sequence; jacalin-related lectin; β-prism.**

The plant lectins have been grouped according to their specificity for monosaccharides as assessed by the hapten inhibition method. Recently, Van Damme *et al.* proposed new criteria for classification of plant lectins based on their structure, function and evolution (1). According to this, plant lectins are divided into four large and three small families, respectively: legume lectins, chitin-binding lectins, ribosome-inactivating proteins, monocot mannose-binding lectins; and jacalin-related lectins, amaranthin lectin, Cucurbitaceae pholem lectins. Many lectins belonging to the larger families have been well-characterized, for example, with X-ray crystallography (1–4). On the other hand, the latter three families had been considered as small groups since they were found in restricted plant families. Among them, it is considered that the jacalin-related lectin family may become a major in plant lectins because some members have been discovered in taxonomically unrelated plant families.

The physiological roles of plant lectins have not been clearly demonstrated, but the dominant view is that they are defense proteins (5, 6). In particular, mannose-binding lectins are assumed to be biologically important proteins, because mannose is a scare sugar in plants but is widely distributed in microorganisms, insects and animals. Thus, mannose-binding lectins are considered to be a promising plant resource for generating genetically modified plants. In fact, transgenic plants expressing the snowdrop lectin (GNA) gene exhibit increased resistance against sucking insects (7). In order to promote one lectin to such levels of application, systematic investigations are required into the relationships between structure and function and into the mechanism of expression.

Previously, we isolated a mannose/glucose-binding lectin (CCA) from the seeds of the Japanese chestnut, which is the first lectin derived from the plant family Fagaceae (8). Some physiological properties, such as the molecular weight and amino acid composition, were different from other mannose-binding plant lectins that were legume lectins or monocot lectins. CCA consists of 6 or 8 identical subunits without a disulfide

Abbreviations used: CCA, *Castanea crenata* agglutinin; AARE, acylamino acid releasing enzyme; AP-I, *Achromobacter* protease I; Cm-Cys, carboxymethylated-cysteine; EPase RC, endoproteinase Arg-C; Gu-HCl, guanidine hydrochloride; 2-ME, 2-mercaptoethanol; RCM-, reduced and *S*-carboxymethylated; SIMS, secondary ion mass spectrometry.

[1] To whom correspondence should be addressed. Fax: +81-78-803-5828. E-mail: knomurak@kobe-u.ac.jp.



0006-291X/00 $35.00

bond and has a molecular mass of 37 kDa as determined by SDS-PAGE. In addition, the N-terminus of CCA is blocked. Thus, it is not clear to which plant lectin family CCA belongs.

The aim of the present study was to determine the complete amino acid sequence of CCA including blocked N-terminal region. Based on the homology to other jacalin-related lectins, the structural futures were discussed.

## MATERIALS AND METHODS

*Materials.* Japanese chestnut (*Castanea crenata* Sieb. et Zucc.) agglutinin (CCA) was purified from fresh seeds of cv. "Kunimi" which was harvested in the experimental farm of Kobe University in Kasai, Hyogo Prefecture, Japan, as stated in our previous paper (8).

*Chemicals.* Endoproteinase Arg-C (EPase RC) was obtained from Roche Diagnostics K.K. *Achromobacter* protease I (AP-I) and acylamino releasing enzyme (AARE) were purchased from Wako Pure Chemical Industries Ltd. and Takara Co., respectively. Papain and other chemical reagents were purchased from Nakarai Tesque Co.

*Isolation of acidic peptides.* About 20 mg of CCA was digested with pepsin in 2 ml of 0.01 N HCl at 37°C for 24 h. The mixture was lyophilized, then dissolved in 2 ml of Milli-Q water and applied on a Dowex 50 × 2 column (1 × 4 cm) equilibrated with water. Fifteen ml of through and wash fractions were collected and lyophilized, and the material was dissolved in 0.5 ml of 0.1% TFA. Then, acidic peptides were purified by RP-HPLC equipped with a Cosmosil C8 column (Nacalai Tesque Co.).

*Mass spectrometry.* The secondary ion mass spectrum (SIMS) was measured with a Hitachi M-4000H double-focusing mass spectrometer using glycerol as the matrix, in positive-ion mode with CSK ion bombardment (6 kV). Ten nmol of acidic peptide was used in the analysis.

*Re-determination of cysteine residue.* In order to re-determine the cysteine residue, performic acid oxidation and reduced *S*-carboxymethylation were carried out under various conditions. Performic acid oxidation was carried out according to the method of Moore (9) at 4°C for 4 h and at 0°C overnight. Reduced *S*-carboxymethylation was carried out with monoiodoacetate according to the method of Crestfield *et al.* (10) in the presence of 50 mM 2-mercaptoethanol and various concentrations of urea or guanidine-hydrochloride (Gu-HCl).

*Enzymatic digestions and chemical cleavage.* Prior to fragmentation, CCA was treated with monoiodoacetate to modify the cysteine residue. EPase RC digestion was performed in 50 mM HEPES buffer, pH 7.8, containing 10 mM 2-ME, 5 mM $CaCl_2$, 20 mM methylamine and 3 M urea. AP-I digestion was carried out in 0.1 M Tris-HCl buffer, pH 8.5, containing 3 M urea. In both cases, the ratio of enzyme to substrate was 1 to 100 by weight and reactions were carried out at 37°C for 24 h. The RCM-CCA was cleaved with 100 molar excess of CNBr in 70% of formic acid at 37°C for 24 h.

Separation of peptides was carried out by RP-HPLC (Beckman) equipped with a Cosmosil C8 column. The mobile phase used was 0.1% (v/v) TFA and elution was performed with an adequate linear concentration gradient of acetonitrile with a flow rate of 1 ml/min.

*Amino acid analysis and sequence analysis.* Amino acid analysis of peptides was carried out with 30 μl of 3-*N*-mercaptoethanesulfonic acid to determine the content of tryptophan residues in evacuated and sealed tubes at 120°C for 12 h. The hydrolysates were analyzed with a Hitachi 835 amino acid analyzer. The N-terminal sequence of peptide was determined by a Shimadzu PPSQ-10 protein sequencer using 100–150 nmol of peptides. The C-terminal sequence of intact CCA (about 5 nmol was used) was determined on a Procise C 494 protein sequencer (PE Applied Biosystems) using the method of Boyd *et al.* (11).

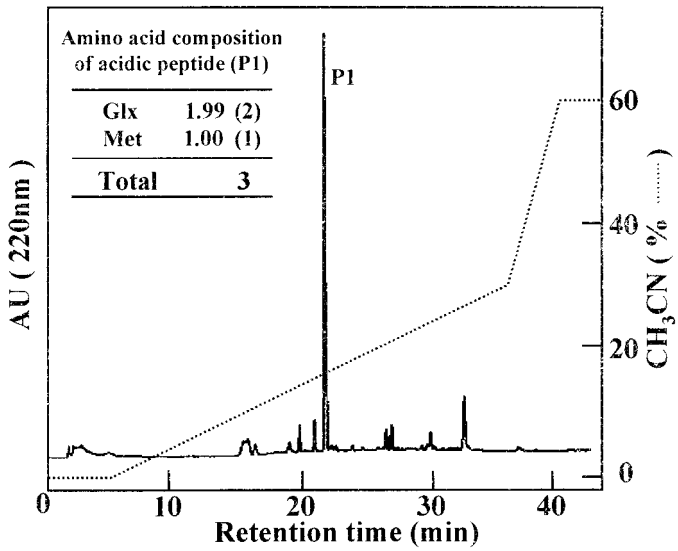


**FIG. 1.** Separation of acidic peptides derived from peptic digestion of CCA by RP-HPLC. The table (inset) shows the amino acid composition of P1.

*Nomenclature of peptides.* The peptides are numbered in the order of elution from HPLC. In this paper, P, R, K, and M represent the peptides derived from the pepsin, EPase RC, AP-I digestions and CNBr cleavage, respectively.

*Sequence alignment of CCA.* Sequence homology of the N- and C-domains of CCA was sought in the protein database (SWISS-PROT) on the web site at http://www.ncbi.nlm.nih.gov/blast. On the five jacalin-related lectins, which were found by homology search, multiple sequence alignment was carried out using MALTALIN on the web site at http://w3.toulouse.inra.fr/lgc/multalin.

## RESULTS AND DISCUSSION

*N-terminal sequence analysis.* The blocked N-terminal residue and partial N-terminal sequence of CCA were determined by analysis of the acidic peptide which was isolated from the pepsin digest of intact CCA using a Dowex 50 × 2 column. The through fraction was applied to a Cosmosil 5C8 column. Only one major peptide, P1, was then obtained (Fig. 1). The amino acid composition of P1 was Met 1 and Glx 2 (table inset in Fig. 1) and the sequence could not determined directly. The molecular mass of P1 was determined to be 450 ($MH^+$) by SIMS (Fig. 2A). This value corresponds to the sum of acetyl groups, one methionine and two glutamic acids minus three water molecules. By considering some fragment ions (Fig. 2B), the N-terminal sequence should be Ac-Met-Glu-Glu. After treatment with AARE, the sequence of Glu-Glu was obtained (data not shown).

*Redetermination of cysteine residues.* Although we previously reported that CCA has no cysteine residues, some nonreproducible results were obtained during preliminary fragmentation without pretreatment. Therefore, the cysteine residues were redetermined

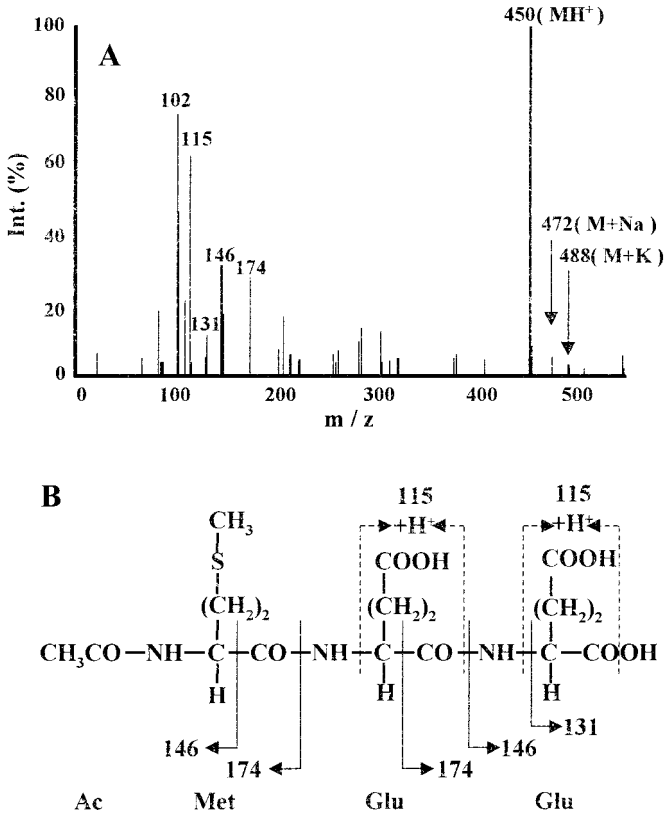


**FIG. 2.** (A) Secondary ion mass spectrum of acidic peptide, P1. The apparent mass value, 450, of a major peak corresponds exactly to the sum of the acetyl group, Met, two Glu minus three water molecules. (B) A possible assignment of some fragment ions. A mass of *m/z* 102 may be derived from immonium ion ($+H_2N$ = CRH) of Glu.

under various conditions (Table 1). Performic acid oxidation performed at 0°C for 4 h gave a result of 0.48 mol cysteine per mol of subunit, while oxidation at 4°C for 12 h gave 0.69 mol/mol. On the other hand, the results of reduction and *S*-carboxymethylation depended on the denaturants used and their concentrations. In the case of urea, 0.59 mol/mol of Cm-Cys could be detected in the presence of 10 M. In contrast to this, Gu-HCl is a more effective denaturant because 0.77 and 1.11 mol/mol of Cm-Cys were detected in the presence of 6 and 8 M Gu-HCl, respectively. Thus, it was concluded that CCA contains one cysteine residue per subunit. This cysteine does not form an intersubunit S–S bond, because the molecular weight of the subunit was the same in the presence and in the absence of reductant (8). These results suggest that this cysteine exists at a buried site and that CCA may maintain the structure by strong hydrophobic interactions which could not be easily broken by urea.

*Sequence determination.* C-terminal analysis was carried out on intact CCA and gave only one sequence, HMEYF. This indicates that CCA is a single chain lectin and does not show heterogeneity by posttranslational proteolysis, although this has been reported about some legume lectins (2).

The complete amino acid sequence of CCA was deduced from the peptide analyses summarized in Fig. 3. The main fragmentation of sequence determination was performed with EPase RC digestion and CNBr cleavage of RCM-CCA. EPase RC digestion gave 12 fragments (R1-R12). Among them, R1-4 and R7-9 were completely sequenced. Only R10 could not be sequenced directly, but was sequenced after digestion of AARE. This is led to the conclusion that R10 is an N-terminal peptide. In the case of CNBr cleavage, peptides other than M1 show heterogeneity in RP-HPLC because of the equilibrium of homoserine and homoserine lactone. According to the results of amino acid analysis, peptides showing the same composition were connected then subjected to sequence analysis. The sequence of M3 coincided with the sequences of AARE treated R1 and acidic peptide, since the N-terminal acetyl-Met may be removed by CNBr cleavage. This supports the nation that the N-terminal blocked amino acid is a methionine. The complete sequence was determined by obtaining the overlap peptide generated by AP-I digestion. Although AP-I digestion gave 12 fragments, di- and tri-peptides and N-terminal fragments could not be obtained.

The present study reveals that the subunit of CCA is a single polypeptide chain consisting of 309 amino acid residues, and that its molecular mass is 33,387 Da which agrees with a previously estimated value of 37 kDa by SDS-PAGE within experimental error. In addition, the presence of one cysteine residue, Cys45, was in agreement with the results of the redetermination of cysteine.

One attractive feature was found by the homology matrix of CCA itself (data not shown). When CCA is divided almost at the center into two domains (N-domain and C-domain), these domains show 35% identity with each other; CCA has an internal repeat sequence. Such large repeat sequences were observed in the B-chain of ricin (of the ribosome-inactivating protein family) and bluebell fetuin-binding lectin (SCAfet, of the monocot mannose-binding lectin family). The

**TABLE 1**

Number of Cysteine Residues in a Subunit of CCA

| Methods | Conditions | Detected Cys (mol/mol subunit) |
|---|---|---|
| Reduction[a] and *S*-carboxymethylation | 8 M urea | 0.46 |
| | 10 M urea | 0.59 |
| | 6 M Gu-HCl | 0.77 |
| | 8 M Gu-HCl | 1.11 |
| Performic acid oxidation | 0°C, 4 h | 0.48 |
| | 4°C, 12 h | 0.69 |

[a] Reduction was carried out with 50 mM 2-ME.

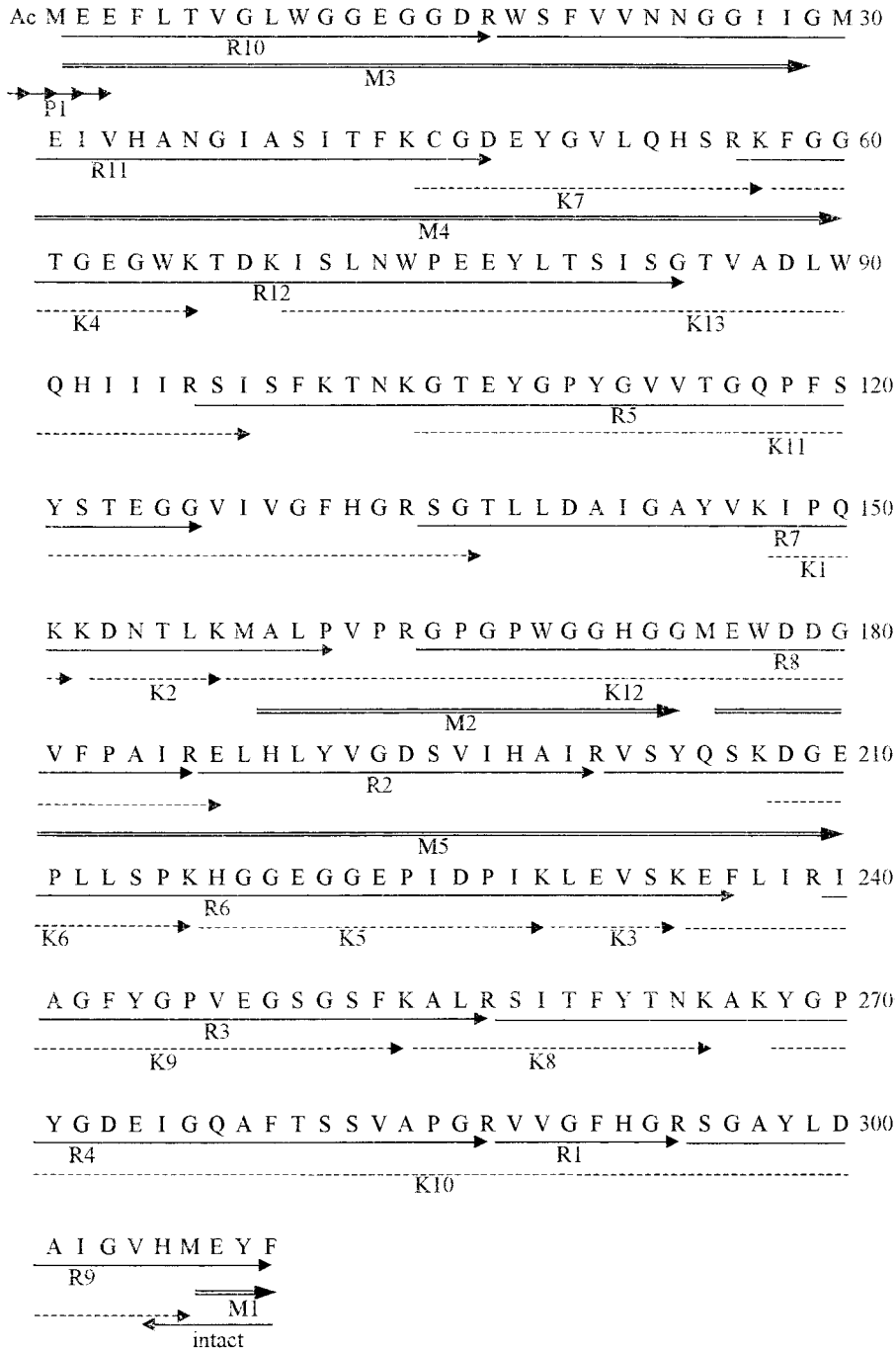


**FIG. 3.** Complete amino acid sequence of CCA. All arrows indicate determined sequences of peptides. Solid lines, double lines, and dotted lines represent peptides obtained by EPase RC digestion, CNBr cleavage, and AP-I digestion, respectively.

ricin B-chain shows a 32% identity, which is assumed to be caused by a gene duplication and has two sugar binding sites (12, 13). The SCAfet exhibits 55% identity and contains three putative carbohydrate-binding sites (14, 15). As to CCA, at least two binding sites may exist in one subunit, according to the homology between these domains, though kinetic studies have never been conducted.

*Multiple alignment of CCA and other jacalin-related lectins.* A homology search of the whole sequence of CCA indicated that G64-I148 and W177-M306 in CCA exhibited 49 and 35% identity with F4-L131 and G50-L133 in the $\alpha$-chain of MPA (*Maclura promifera* lectin) (16), respectively. Furthermore, regarding the same positions, similar identities were observed between CCA and the $\alpha$-chain of jacalin (*Artocarpus integrifolia*


**FIG. 4.** Sequence alignment of the N- and C-domains of CCA with jacalin-related lectins. Conserved residues in all sequences and in N- and C-domains are boxed with a solid line and dotted lines, respectively. Hyphens are gaps inserted to achieve maximum alignment. The arrows above the sequence are estimated $\beta$-sheet regions in heltuba. The bold number on the right represents the corresponding Greek key motifs in heltuba.

lectin) (16). MPA and jacalin are major members of the jacalin-related lectin family. These results support the theory that CCA has a repeat sequence, the N- and C-domains, as described above, and suggest that they have homology to jacalin-related lectins. Therefore, multiple alignment was carried out in relation to the N- and C-domains in CCA and five jacalin-related lectins, MPA, jacalin, KM+ (*Artocarpus integrifolia*) (17), calsepa (*Calystegia sepium*) (18) and heltuba (*Helianthus tuberosus*) (19) (Fig. 4).

The common structural feature of jacalin-related lectins is based that they are considered to be a threefold symmetric $\beta$-prism fold on the basis of the results of X-ray crystallographic studies (20–22) and a molecular building study (17). The $\beta$-prism fold of jacalin-related lectins consists of three Greek key motifs. According to the results on heltuba, Greek keys 1 and 3 show the high conservation because the loop regions that are present in both motifs are important for sugar binding. Whereas the Greek key 2 is a variant region though the role has not been cleared. The results of alignment exhibited similar tendencies. In the comparison of the N- and C-domains, identical residues were 20 and 21 in Greek key 1 and 3, respectively, but only 13 residues were identical in Greek key 2. Since the low reactive cysteine residue, Cys45, was present in Greek key 2, it may be estimated that this motif exists in a buried site and contributes to maintain the conformation.

To date, a jacalin-related lectin family further divided into two subgroups: a galactose-specific subgroup and a mannose-binding subgroup. The former comprises jacalin and MPA which are tetrameric two-chain lectins, $(\alpha\beta)_4$). They bind to galactose with a hydrogen bond between the $\alpha$-amino group of the $\alpha$-chain, which are generated by posttranslational cleavage, and O4 of galactose. While, the later includes KM+, calsepa and

heltuba which are homo-tetramer or -dimer, these one chain jacalin-related lectins cannot bind to galactose because they have not been subjected to such modifications. Posttranslational cleavage may be an important factor contributing to the difference of carbohydrate specificity. CCA consists of a single polypeptide chain and shows specificity for mannose. Thus, it is assumed that the structure of CCA resembles two linked mannose-binding jacalin-related lectins.

Based on the information obtained above, cDNA cloning of CCA is now progressing to determine whether or not CCA has a signal sequence.

## REFERENCES


1. Van Damme, E. J. M., Peumans, W. J., Barre, A., and Rougé, P. (1998) Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* **17,** 575–692.
2. Sharon, N., and Lis, H. (1990) Legume lectins—A large family of homologous proteins. *FASEB J.* **4,** 3198–3208.
3. Sharma, V., and Surolia, A. (1997) Analyses of carbohydrate recognition by legume lectins: Size of the combining site loops and their primary specificity. *J. Mol. Biol.* **267,** 433–445.
4. Bouckaert, J., Hamelryck, T., Wyns, L., and Loris, R. (1999) Novel structures of plant lectins and their complexes with carbohydrates. *Curr. Opin. Struct. Biol.* **9,** 572–577.
5. Chrispeels, M. J., and Raikhel, N. V. (1991) Lectins, lectin genes, and their role in plant defense. *Plant Cell* **3,** 1–9.
6. Peumans, W. J., and Van Damme, E. J. M. (1995) Lectins as plant defense protein. *Plant Physiol.* **109,** 347–352.
7. Hilder, V. A., Boulter, D., *et al.* (1995) Expression of snowdrop lectin in transgenic tabacco plants results in added protection against aphids. *Transgenic Res.* **4,** 18–25.
8. Nomura, K., Ashida, H., Uemura, N., Kushibe, S., Ozaki, T., and Yoshida, M. (1998) Purification and characterization of a mannose/glucose-specific lectin from *Castanea crenata. Phytochemistry* **49,** 667–673.
9. Moore, S. (1963) On the determination of cysteine as cysteic acid. *J. Biol. Chem.* **238,** 235–237.
10. Crestfield, A. M., Moore, S., and Stein, W. H. (1963) The preparation and enzymatic hydrolysis of reduced and *S*-carboxymethylated proteins. *J. Biol. Chem.* **238,** 622–627.
11. Boyd, V. L., Bozzini, M., Zon, G., Noble, R. L., and Mattaliano, R. J. (1992) Sequencing of peptides and proteins from the carboxy terminus. *Anal. Biochem.* **206,** 344–352.
12. Lamb, F. I., Roberts, L. M., and Lord, J. M. (1985) The primary structure of cloned cDNA coding for preproricin. *Eur. J. Biochem.* **148,** 265–270.
13. Rutenber, E., and Roberts, J. D. (1991) Structure of ricin B-chain at 2.5 Å resolution. *Proteins, Struct. Funct. Genet.* **10,** 260–269.
14. Wright, L. M., Van Damme, E. J., Barre, A., Allen, A. K., Van Leuven, F., Reynolds, C. D., Rounge, P., and Peumans, W. J. (1999) Isolation, characterization, molecular cloning and molecular modeling of two lectins of different specificities from bluebell (*Scilla campanulata*) bulbs. *Biochem. J.* **340,** 299–308.
15. Wright, L. M., Reynolds, C. D., Rizkallah, P. J., Allen, A. K., Van Damme, E. J. M., Donovan, M. J., and Peumans, W. J. (2000). Structural characterization of the native fetuin-binding protein *Scilla campanulata* agglutinin: A novel two-domain lectin. *FEBS Lett.* **468,** 19–22.
16. Young, N. M., Johnston, R. A. Z., and Watson, D. C. (1991) The amino acid sequences of jacalin and the *Maclura pomifera* agglutinin. *FEBS Lett.* **282,** 382–384.
17. Rosa, J. C., Oliveira, P. S. L., Garratt, R., Beltramini, L., Resing, K., Roque-Barreira, M.-C., and Greene, L. J. (1999) KM+, a mannose-binding lectin from *Artocarpus integrifolia:* Amino acid sequence, predicted tertiary structure, carbohydrate recognition, and analysis of the $\beta$-prism fold. *Protein Sci.* **8,** 13–24.
18. Van Damme, E. J. M., Barre, A., Verhaert, P., Rouge, P., and Peumans, W. J. (1996) Molecular cloning of the mitogenic mannose/maltose-specific rhizome lectin from *Calystegia sepium. FEBS Lett.* **397,** 352–356.
19. Van Damme, E. J. M., Barre, A., Mazard, A.-M., Verhaert, P., Horman, A., Debray, H., Rouge, P., and Peumans, W. J. (1999) Characterization and molecular cloning of the lectin from *Helianthus tuberosus. Eur. J. Biochem.* **259,** 135–142.
20. Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A., and Vijayan, M. (1996). A novel mode of carbohydrate recognition in jacalin, a *Moraceae* plant lectin with a $\beta$-prism fold. *Nat. Struct. Biol.* **3,** 596–603.
21. Lee, X., Thompson, A., Zhang, Z., Ton-that, H., Biesterfildt, J., Ogata, C., Xu, L., Johnston, R. A. Z., and Young, N. M. (1998). Structure of the complex of *Maclura pomifera* agglutinin and the T-antigen disaccharide, Gal$\beta$1,3GalNAc. *J. Biol. Chem.* **273,** 6312–6318.
22. Bourne, Y., Zamboni, V., Barre, A., Peumans, W. J., Van Damme, E. J. M., and Rougé, P. (1999) *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. *Structure* **7,** 1473–1482.