This is to certify that the

dissertation entitled

THE MOLECULAR BASIS OF POST-TRANSCRIPTIONAL AND POST-TRANSLATIONAL MODIFICATIONS OF GRAMINEAE LECTINS

presented by

Thea Ann Wilkins

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Botany

Natasha Raikhel
Major professor

Date 12/15/89

# THE MOLECULAR BASIS OF POST-TRANSCRIPTIONAL AND POST-TRANSLATIONAL MODIFICATIONS OF GRAMINEAE LECTINS

By

Thea Ann Wilkins

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Botany and Plant Pathology

1989

ABSTRACT

# THE MOLECULAR MECHANISMS OF POST-TRANSCRIPTIONAL AND POST-TRANSLATIONAL MODIFICATIONS OF GRAMINEAE LECTINS

By

Thea Ann Wilkins

The lectins of the Gramineae are dimeric vacuolar proteins which specifically bind to oligomers of N-acetylglucosamine or chitin. The lectins of wheat, barley and rice are highly related but exhibit discrete differences in antigenicity, biochemical properties and distribution. cDNA clones encoding wheat germ agglutinin (isolectin B) and rice lectin were isolated and used as tools to investigate the expression of the Gramineae lectins at the cellular and molecular levels.

The expression of rice lectin in developing embryos is distinctive relative to expression of lectins in wheat and barley. Rice lectin is synthesized from two mRNA species derived from a single gene present in one to two copies per haploid genome. The two mRNA species are presumably generated by a mechanism of alternative polyadenylation site selection during the post-transcriptional processing of the pre-mRNA. The differential expression and accumulation of the individual rice lectin mRNA species is also regulated at the post-transcriptional level during embryogenesis. *In situ* localization of rice lectin mRNA in root caps,

peripheral cell layers of the coleorhiza, radicle, scutellum and throughout all cell layers of the coleoptile.

The functional role of the N-linked glycan in the processing and transport of Gramineae lectins to vacuoles was investigated by introducing cDNA clones encoding a wild-type or glycosylation-minus barley lectin preproprotein into tobacco under the transcriptional control of the cauliflower mosaic virus 35S promoter. The correct synthesis, assembly, processing and transport of active barley lectin to vacuoles in transgenic tobacco indicates that monocots and dicots possess a similar mechanism for the processing and transport of these vacuolar proteins. Although the proprotein N-linked glycan is not essential for targeting of barley lectin to vacuoles, the presence of the glycan modulates the rate of post-translational processing and transport of barley lectin proproteins. These results constitute the demonstration of a role for N-linked glycans in plants.

To David and Natasha

ACKNOWLEDGEMENTS

Achievement of this degree has been a long and arduous journey and no one is more deserving of my appreciation and gratitude than Natasha Raikhel. Natasha's contributions to my success are immeasurable and will continue to be a source of inspiration to me. I would like to extend my sincere appreciation to Natasha for providing a cheerful and stimulating environment and the opportunity to obtain my degree. I would also like to thank Andrew Hanson for his advice and guidance in pursuit of my career.

I would also like to express my gratitude to Dr. Elizabeth Weretilnyk and Sebastian Bednarek for many helpful discussions, but mostly for their encouragement, understanding and friendship. I would especially like to acknowledge Sebastian for the difference he has made in my life. Finally, my husband, David, is particularly deserving of my appreciation for his patience and understanding, but particularly for the sacrifices he made while I was pursuing my degree.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# Chapter 1

## Introduction

Lectins are a class of proteins of nonimmune origin that noncovalently bind specific carbohydrates (Lis and Sharon, 1986). Although lectins have been found to be ubiquitous in nature, lectins have been primarily derived from plant sources. In fact, plant lectins have been vital components in biomedical research as a result of their abundance and diversity in carbohydrate-binding specificity. The purification and molecular analysis of many plant lectins has contributed significantly to our knowledge regarding their biosynthesis, expression, and distribution. For instance, studies on the biosynthesis of Concanavalin A led to the discovery of a novel mechanism of protein maturation (Bowles *et al.*, 1986). Accumulating evidence (reviewed in Sharon and Lis, 1989) indicates that lectins function as cell recognition molecules in mediating cell-cell interactions. Such interactions would thereby enable the lectins to regulate both normal cellular and pathological processes as a consequence. Despite 100 years of lectin research (Sharon and Lis, 1987), however, the endogenous function of many lectins, especially plant lectins, remains elusive.

Many species of the Gramineae synthesize a lectin which specifically binds oligomers of N-acetylglucosamine (GlcNAc) or chitin. The Gramineae

lectins can be further subdivided into three subtypes based upon discrete structural and antigenic differences (Peumans and Stinissen, 1983). The three subtypes belong to two Gramineae subfamilies and include 1) the cereal lectins, represented by wheat and barley, 2) *Brachypodium* lectins, and 3) rice lectin. The immunologically-related lectins of the Gramineae are embryo-specific vacuolar proteins, but do not function as storage proteins. Mature Gramineae lectins are 36 kd dimers comprised of two identical 18 kd subunits processed from 23 kd precursors (Peumans and Stinissen, 1983). However, the lectin of cultivated rice is unique in that approximately 90% of the 18 kd subunit is endoproteolytically cleaved into two smaller polypeptides (Stinissen *et al.*, 1984).

The Gramineae lectins accumulate in specific cell-layers of embryonic tissues (i.e. root caps, radicle, coleorhiza, coleoptile, scutellum). Distribution of the lectins in these embryonic tissues is species-specific. For instance, wheat germ agglutinin (WGA) accumulates in the outer cell-layers of the coleoptile, rye lectin is localized to both the inner and outer layers, whereas rice lectin is present throughout all cell-layers of the coleoptile. Conversely, barley lectin is not apparently expressed in the coleoptile (Mishkind *et al.*, 1983). The specific accumulation of the Gramineae lectins in tissues of embryos which establish direct contact with the environment during germination has long been interpreted as a defensive role against fungal infections. Both circumstantial and empirical evidence for the potential role of

the Gramineae lectins in plant defense is rapidly accumulating. The Gramineae lectins exhibit extensive homology to other chitin-binding proteins shown to be plant defense-related proteins (Broglie *et al.*, 1986; Parsons *et al.*, 1989) or possess antifungal properties (Broekaert *et al.*, 1989). Very exciting is the recent demonstration that WGA possesses insecticidal activity against the cowpea weevil *in vitro* (Murdock *et al.*, 1989). The deleterious effect on development of this insect by WGA is believed to be mediated by the nonspecific binding of WGA to chitin in the peritrophic membrane of the insect midgut.

Altering the levels of expression, distribution and subcellular site of accumulation of the Gramineae lectins would facilitate the identification of the endogenous functions of these lectins. This dissertation describes the isolation of cDNA clones encoding WGA (isolectin B) and rice lectin and their utilization to characterize the expression of the Gramineae lectins at the molecular and cellular levels in developing embryos and transgenic tobacco. The results demonstrated that the expression of rice lectin at the post-transcriptional and post-translational levels is distinctive and more complex relative to WGA or barley lectin. The discovery of a COOH-terminal propeptide which is post-translationally modified by the addition of an N-linked high mannose glycan has provided the missing link to elucidating the series of events involved in the post-translational processing of modified Gramineae lectin proproteins to mature subunits. Site-directed mutagenesis

of the COOH-terminal glycan of barley lectin has provided significant insight into the molecular mechanism of protein targeting in plants. The N-linked glycan of the barley lectin propeptide modulates the rate of post-translational processing of transport of barley lectin, but is not essential for targeting of barley lectin to vacuoles. The research summarized in this dissertation provides the background for future endeavors to bioengineer the Gramineae lectins for specific applications in plant resistance programs.

## REFERENCES

Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R. and Burgess. (1986) *J. Cell Biol.* **102,** 1284-1297.

Broekaert, W.F., Van Parijs, J., Leyns, F., Joos, H. and Peumans, W.J. (1989) A chitin-binding lectin from stinging nettle rhizomes with antifungal properties. *Science* **245,** 1100-1102.

Broglie, K.E., Gaynor, J.J. and Broglie, R.M. (1986) Ethylene-regulated gene expression: Molecular cloning of the genes encoding an endochitinase from *Phaseolus vulgaris. Proc. Natl. Acad. Sci. USA* **83,** 6820-6824.

Lis, H. and Sharon, N. (1986) Lectins as molecules and as tools. *Ann. Rev. Biochem.* **55,**35-67.

Mishkind, M.L., Palevitz, B.A., Raikhel, N.V., and Keegstra, K. (1983) Localization of wheat germ agglutinin-like lectins in various species of the Gramineae. *Science* **220,** 1290-1292.

Murdock, L.L., Huesing, J.E., Nielsen, S.S., Pratt, R.C. and Shade, R.E. (1989) Biological effects of plant lectins on the cowpea weevil. *Phytochemistry,* In press.

Parsons, T.J., Bradshaw, Jr., H.D. and Gordon, M.P. (1989) Systemic accumulation of specific mRNAs in response to wounding in poplar trees. *Proc. Natl. Acad. Sci. USA* **86,** 7895-7899.

Peumans, W.J. and Stinissen, H.M. (1983) *Gramineae* lectins: Occurrence, molecular biology and physiological function. In *Chemical Taxonomy, Molecular Biology, and Function of Plant Lectins*, I.J. Goldstein and M.E. Etzler, eds. (Alan R.Liss, Inc., New York), pp. 99-116.

Sharon, N. and Lis, H. (1987) A century of lectin research (1888-1988). *Trends Biochem. Sci.* **12,**488-491.

Sharon, N. and Lis, H. (1989) Lectins as cell recognition molecules. *Science* **246,** 227-246.

Stinissen, H.M., Peumans, W.J. and Chrispeels, M.J. (1984) Subcellular site of lectin synthesis in developing rice embryos. *EMBO J.* **3,** 1979-1985.

## Chapter 2

### Isolation and characterization of a cDNA clone encoding wheat germ agglutinin


**ABSTRACT**

Two sets of synthetic oligonucleotides coding for amino acids in the amino- and carboxyl-terminal portions of wheat germ agglutinin were synthesized and used as hybridization probes to screen cDNA libraries derived from developing embryos of tetraploid wheat. The nucleotide sequence for a cDNA clone recovered from the cDNA library was determined by dideoxynucleotide chain-termination sequencing in vector M13. The amino acid sequence deduced from the DNA sequence indicated that this cDNA clone (pNVR1) encodes isolectin 3 of wheat germ agglutinin. Comparison of the deduced amino acid sequence of clone pNVR1 with published sequences indicates isolectin 3 differs from isolectins 1 and 2 by 10 and 8 amino acid changes, respectively. In addition, the protein encoded by pNVR1 extends 15 amino acids beyond the carboxyl terminus of the published amino acid sequence for isolectins 1 and 2 and includes a potential site for N-linked glycosylation. Utilizing the insert of pNVR1 as a hybridization probe, we have demonstrated that the expression of genes for wheat germ agglutinin is modulated by exogenous abscisic acid. Striking homology is observed between wheat germ agglutinin and chitinase, both of which are proteins that bind chitin.

## INTRODUCTION

Lectins, sugar-binding proteins derived mainly from plant sources, have been of great value as specific probes for investigating the distribution and function of carbohydrates on the surfaces of animal cells (Goldstein and Hayes, 1978; Lis and Sharon, 1981). In recent years, however, the notion has become widely accepted that the ability of lectins to distinguish discrete sugars did not arise fortuitously during evolution (Lis and Sharon, 1981), and as a result, there has been increased interest in the synthesis and biochemistry of this group of proteins. Wheat germ agglutinin (WGA), the first cereal lectin characterized in detail, binds specifically to the sugar *N*-acetylglucosamine and to chitin, a polymer of *N*-acetylglucosamine residues (Allen, *et al.*, 1973; Nagata and Burger, 1974). In the hexaploid wheat *Triticum aestivum*, WGA exists as three closely related isolectins derived from the A, B, and D genomes (Rice and Etzler, 1975; Peumans, *et al.*, 1982). Comparison of the amino acid sequences for isolectin 1 (A genome) and isolectin 2 (D genome) indicates that these proteins differ at four residues(Wright, *et al.*, 1984; Wright and Olafsdottir, 1986). The amino acid sequence for isolectin 3 (B genome), the least abundant form, is not yet available. These three isolectins randomly associate into functional dimers *in vivo* (Rice and Etzler, 1975) and are immunologically indistinguishable (Raikhel and Pratt, 1987).

In wheat plants, WGA is found in the embryos and adventitious roots (Raikhel and Pratt, 1987; Mishkind, *et al.*, 1982; Raikhel, *et al.*, 1984). During

embryogenesis, WGA expression is under temporal control (Raikhel and Quatrano, 1986). Accumulation of WGA is tissue-specific and cell-type specific in various organs of the embryo (e.g., coleoptile, coleorhiza, and radicle) (Raikhel and Pratt, 1987; Mishkind, *et al.*, 1982). In other species of Gramineae, a lectin immunologically related to WGA is synthesized during seed development and in the roots of adult plants (Mishkind, *et al.*, 1983; Stinissen, *et al.*, 1985). Furthermore, the accumulation of lectin is modulated by the hormone abscisic acid (Raikhel and Quatrano, 1986; Triplett and Quatrano, 1982). Although biochemical, immunological, and microscopic studies have helped to characterize the composition and distribution of WGA (Allen, *et al.*, 1973; Nagata and Burger, 1974; Rice and Etzler, 1975; Peumans, *et al.*, 1982; Wright, *et al.*, 1984; Wright and Olafsdottir, 1986; Mishkind, *et al.*, 1982; Raikhel, *et al.*, 1984), the genes for WGA have not been isolated.

We are interested in investigating the molecular mechanisms that regulate the developmental tissue-specific expression of WGA genes. To isolate clones for WGA, cDNA libraries from developing grains of the tetraploid wheat *Triticum durum* (AABB) were used. Here, we report the isolation and the nucleotide sequence of a cDNA clone that we presume to encode isolectin 3. Using this clone as a hybridization probe, we present evidence that the expression of WGA genes is modulated by abscisic acid. Because of the common ability of WGA and chitinase to bind chitin, we searched for amino acid homology using the recently published sequence for chitinase from *Phaseolus vulgaris* (Broglie, *et*

*al.*, 1986). We found strong homology between the amino terminus of chitinase and four regions of WGA. The significance of this similarity is addressed.

## MATERIALS AND METHODS

***Plant material.*** Wheat *T. aestivum* L. (AABBDD) cv. Marshall was obtained from the Minnesota Crop Improvement Association (St. Paul, MN). Plants were grown as previously described (Raikhel, *et al.*, 1984), and embryos were collected at 20 days after bloom (anthesis) according to Raikhel and Quatrano (Raikhel and Quatrano, 1986). Abscisic acid treatment involved culturing isolated embryos in the dark at 27°C for 3 days on filter paper containing growth medium (Triplett and Quatrano, 1982) with and without $10^{-4}$ M abscisic acid (Sigma).

***Materials.*** Two cDNA libraries, derived from mRNA isolated from developing wheat grains of *T. durum* (AABB) cv. Mexicali at 3 and 4 weeks post-anthesis, were provided by C. Brinegar (ARCO Plant Cell Research Institute, Dublin, CA). Two sets of degenerate synthetic oligonucleotides were prepared for amino acid regions in isolectin 1 (Wright and Olafsdottir, 1986): for the sequence Asn-Met-Glu-Cys-Pro-Asn-Asn in the amino-terminal region (residues 9-15), probe 2, TTR TAC CTY ACR GGN TTR TT; and for the sequence Cys-Thr-Asn-Asn-Tyr-Cys-Cys in the carboxyl terminal region (residues 141-147), probe 1, ACR TGN TTR TTR ATR ACR AC. The oligonucleotide mixtures were synthesized on an

Applied Biosystems (Foster City, CA) 380 DNA synthesizer by a solid-phase method (Matteucci and Caruthers, 1980) and separated by electrophoresis on a 20% polyacrylamide gel containing 8 M urea in TBE, pH 8.3 (0.89 M Tris/0.089 M boric acid/2.7 mM EDTA). The oligonucleotides were eluted in 0.5 M ammonium acetate/10 mM magnesium acetate/0.1% $NaDodSO_4$, and then end-labeled with $^{32}P$ using T4 polynucleotide kinase (Maniatis, *et al.*, 1982).

***Isolation and screening of cDNA clones.*** The cDNA libraries, in *Escherichia coli* strain DH5$\alpha$ (Hanahan, 1985), were plated directly onto nitrocellulose filters laid on agar plates containing Luria broth medium with ampicillin at 50 ug/ml (Hanahan and Meselson, 1983). After colonies were established, the bacteria were lysed, and the filters were probed with oligonucleotide probes 1 and 2 as follows. The temperature of hybridization ($T_H$) for each oligonucleotide was calculated using the formula $T_H = T_D - 12°C$, where $T_D = 2°C$ x (the number of A·T base pairs) plus 4°C x (the number of G·C base pairs). Replicate filters were prehybridized in 6x SSC (0.9 M sodium chloride and 0.09 M sodium citrate, pH 7.0) (1x SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7) plus 0.25% nonfat milk, and hybridized in the same buffer containing the labeled oligonucleotide probes and 0.1% $NaDodSO_4$ at 36°C (probe 1) or 38°C (probe 2). After hybridization, filters were washed three times in 6 x SSC/0.25% nonfat milk/0.1% $NaDodSO_4$ at room temperature for 10 min, followed by a 2-min wash at 46°C (probe 1) or at 48°C (probe 2). Filters were dried and

autoradiographed for 16-18 hr. Colonies that produced positive signals were selected and rescreened using the same probes under the same conditions.

***DNA sequence determination.*** Inserts from recombinant plasmids were purified by electrophoresis in low-melting-point agarose. Excised cDNA inserts or appropriate restriction fragments were then subcloned into M13mp18 or M13mp19. Dideoxynucleotide chain-termination sequencing from single-stranded M13 templates was accomplished using a Bethesda Research Laboratories M13 sequencing kit with the exception that dGTP was replaced by 7-deaza-2'-deoxy-guanosine triphosphate (Boehringer Mannheim).

***RNA blot analysis.*** Total RNA was isolated as described (Williamson, *et al.*, 1985) and poly(A)$^+$ RNA was purified by chromatography on oligo(dT)-cellulose (Maniatis, *et al.*, 1982). Poly(A)$^+$ RNA was electrophoresed in adjacent lanes (1 ug per lane) on 2% agarose gels containing 6% formaldehyde and then transferred to nitrocellulose (Thomas, 1980). Filters were hybridized with inserts labeled by the random primer method of Feinberg and Vogelstein (1983) and washed under stringent conditions as described in Thomas (1980).

## RESULTS

***Isolation of cDNA clones.*** Two synthetic oligonucleotides, each consisting of 20 nucleotides complementary to the 5' and 3' ends of the coding portion of

isolectin 1 mRNA (Wright and Olafsdottir, 1986), were used for isolation of cDNA clones specific for WGA. These two sequences corresponded to amino acids 9-15 (probe 2) and 141-147 (probe 1). Because of the degeneracy of the sequences involved, probe 2 was a mixture of 64 sequences, and probe 1 was a mixture of 128 sequences. One cDNA clone, pNVR1 [1.0 kilobase (kb)], was selected by hybridization to both probes on the assumption that this insert contains sequences spanning the coding region delimited by the oligonucleotide probes. A second clone, pNVR2 (0.7 kb), was recognized by probe 1 only and is presumably truncated at the 5' end. The restriction map and partial sequence of pNVR2 indicate that it is a shorter version of pNVR1. When the insert from clone pNVR1 was labeled by the random primer method (Feinberg and Vogelstein, 1983) and used as probe to rescreen the cDNA libraries, no additional cDNA clones were retrieved.

***Nucleotide sequence.*** The cDNA insert of pNVR1 was subcloned into M13mp18 and M13mp19 according to the strategy shown in Figure 1, and its nucleotide sequence was determined as described. The nucleotide sequence of the cDNA clone and the deduced amino acid sequence are shown in Figure 2. Clone pNVR1 contains a 558-nucleotide open reading frame encoding a 186-amino acid polypeptide rich in cysteine and glycine but lacking an ATG start codon at the 5' end. Protein sequence analysis indicates that the amino terminus of WGA is blocked (Wright, *et al.*, 1984; Wright and Olafsdottir, 1986) so that the first

**Figure 1. Restriction map and sequencing strategy for WGA cDNA clone pNVR1.** Open bar, cloned cDNA; arrows, length and direction of the sequenced restriction fragments. Scale of the map is in kb pairs.

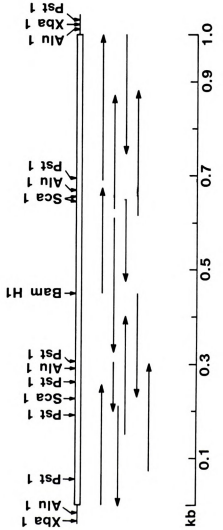
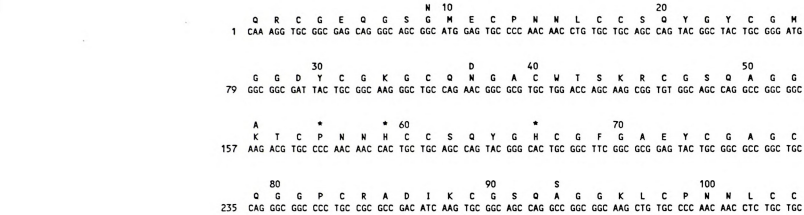

Xba 1
Alu 1
Pst 1
Pst 1
Pst 1
Sca 1
Pst 1
Alu 1
Pst 1
Bam H1
Sca 1
Alu 1
Pst 1
Alu 1
Xba 1
Pst 1
Kb
0.1
0.3
0.5
0.7
0.9
1.0

**Figure 2. Nucleotide sequence of WGA cDNA clone pNVR1.** The deduced amino acid residues are shown above the nucleotide triplets. The differences between the deduced amino acid sequence and the published amino acid sequence of isolectin 2 are designated by the amino acid codes above the deduced sequence. The additional differences with isolectin 1 are designated by asterisks. Proline at position 56 is substituted with threonine, and the histidines at positions 59 and 66 are substituted with glutamine and tyrosine, respectively. Previously described differences at positions 150 (Wright, 1987) and 134 (arrows) have been resolved (C. Wright, personal communication). Four termination codons (single underline) and a putative polyadenylation signal (double underline) are indicated. Fifteen amino acids that extend beyond the carboxyl terminus of the published sequence for WGA are designated by squares. A potential glycosylation site is indicated by dots above the squares.


N 10 20
Q R C G E Q G S G M E C P N N L C C S Q Y G Y C G M
1 CAA AGG TGC GGC GAG CAG GGC AGC GGC ATG GAG TGC CCC AAC AAC CTG TGC TGC AGC CAG TAC GGC TAC TGC GGG ATG
30 D 40 50
G G D Y C G K G C Q N G A C W T S K R C G S Q A G G
79 GGC GGC GAT TAC TGC GGC AAG GGC TGC CAG AAC GGC GCG TGC TGG ACC AGC AAG CGG TGT GGC AGC CAG GCC GGC GGC
A * * 60 * 70
K T C P N N H C C S Q Y G H C G F G A E Y C G A G C
157 AAG ACG TGC CCC AAC AAC CAC TGC TGC AGC CAG TAC GGG CAC TGC GGC TTC GGC GCG GAG TAC TGC GGC GCC GGC TGC
80 90 S 100
Q G G P C R A D I K C G S Q A G G K L C P N N L C C
235 CAG GGC GGC CCC TGC CGC GCC GAC ATC AAG TGC GGC AGC CAG GCC GGC GGC AAG CTG TGC CCC AAC AAC CTC TGC TGC
S 110 G 120 S 130
S Q W G Y C G L G S E F C G E G C Q N G A C S T D K
313 AGC CAG TGG GGG TAC TGC GGC CTC GGT TCC GAG TTC TGC GGC GAG GGC TGC CAG AAC GGC GCT TGC AGC ACC GAC AAG
140 150
P C G K D A G G R V C T N N Y C C S K W G S C G I G
391 CCG TGT GGC AAG GAC GCC GGC GGC AGG GTT TGC ACT AAC AAC TAC TGC TGT AGC AAG TGG GGA TCC TGT GGC ATC GGT
160 170 A
P G Y C G A G C Q S G G C D G V F A E A I A T N S T
469 CCC GGC TAC TGC GGT GCA GGC TGC CAG AGC GGC GGC TGC GAT GGT GTC TTC GCC GAG GCC ATC GCC ACC AAC TCC ACT
L L A E
547 CTT CTC GCA GAA TGATGATCTTGCTAATGGTAGTATTGCAACGACGCATAATCCGTGGCAGTTCCATTGCCACGTACGGTTTTCCCTTCACTTACTTT
645 TAGTACTAGTACTTAATAATTCTCTAGCTTGCAATATGACATGCAGGTTACTGCAGCAGAAACAAAATATTGCTGTAGTACATGCATGGAAATATTACCGAGA
748 GAAGTGTTGTGTGGCAATATAGAGTGTACTATAGGCGCCACAAATTAGTGTTCTTGTTATGACCTGCTGTCAGGATCGATGCATGCATCGCTGTTGTAATGTT
851 GGAGTACTTCGTGATATCGTTGCAATATATTACCATGAGTCTCACATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

residue (glutamine) of the published sequence may not be the amino terminus of mature WGA. It is, therefore, presumably fortuitous that the cDNA clone pNVR1 and the published amino acid sequence for WGA initiate with the same amino acid. The hydropathy plot (Kyte and Doolittle, 1982) of the polypeptide encoded by clone pNVR1 shows the polypeptide to be comprised mostly of hydrophilic amino acids (Figure 3). The polypeptide encoded by pNVR1 extends 15 amino acids beyond the carboxyl terminus of the amino acid sequence published for isolectins 1 and 2 (Figure 2, squares) (Wright, *et al.*, 1984; Wright and Olafsdottir, 1986). The carboxyl-terminal segment contains the most hydrophobic portion of the entire protein (Figure 3). A potential site for N-linked glycosylation occurs at residues 180-182 (Asn-Ser-Thr) (Figure 2, dots above squares). The 3'-untranslated region contains four in-frame termination codons (TGA, TGA, TAA, and TAG, underlined in Figure 2) and a potential polyadenylation signal (AATAAT, double-underlined in Fig.2), and terminates in a poly(A) tail that begins 229 nucleotides downstream from that signal.

***Comparison with published sequences of isolectin 1 and 2.*** The amino acid sequence deduced from the cDNA nucleotide sequence (Figure 2) was compared with published protein sequence data. Re-evaluation of the discrepancies at positions 134 and 150 (Figure 2, arrows) has indicated a low yield of lysine in addition to glycine for residue 134 (C. Wright, personal communication) and has confirmed the presence of tryptophan at residue 150

**Figure 3. Hydropathy plot of the protein encoded by cDNA pNVR1.** Ordinate, hydropathic index (Kyte and Doolittle, 1982); abscissa, amino acid position. The additional 15 amino acids at the carboxyl terminus are right of the broken line.

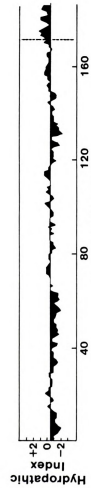
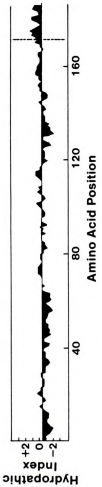

Hydropathic Index
+2
0
-2
40
80
120
160
Amino Acid Position

**Figure 3. Hydropathy plot of the protein encoded by cDNA pNVR1.** Ordinate, hydropathic index (Kyte and Doolittle, 1982); abscissa, amino acid position. The additional 15 amino acids at the carboxyl terminus are right of the broken line.


Hydropathic Index
+2
0
-2
40
80
120
160
Amino Acid Position

(Wright, 1987). The deduced amino acid sequence of pNVR1 was found to differ from the published sequence of isolectin 1 (Wright and Olafsdottir, 1986) at 10 positions and isolectin 2 at 8 positions (Wright, *et al.*, 1984) (Table 1).

***RNA blot analysis.*** Embryos isolated from hexaploid wheat at 20 days post-anthesis were cultured in the presence and absence of abscisic acid (Figure 4). Equal amounts of poly(A)$^+$ RNA from the embryos were fractionated by agarose-formaldehyde gel electrophoresis and transferred to nitrocellulose filters. A 1.1-kb mRNA was detected (Figure 4) after hybridization with pNVR1 insert labeled by the random primer method (Feinberg and Vogelstein, 1983). The autoradiograph showed that the level of RNA in abscisic acid-treated embryos was several times higher than the level in embryos cultured in the absence of abscisic acid.

***Nucleotide and amino acid homology between WGA and chitinase.*** The deduced amino acid sequence of cDNA clone pNVR1 was used to search for homology with chitinase, an enzyme that catalyzes the hydrolysis of 1,4-B linkages of *N*-acetylglucosamine polymers in chitin. The amino acid homology matrix between clone pNVR1 and chitinase from *P. vulgaris* is shown in Figure 5. This matrix was generated using the analysis program of Pustell and Kafatos (1984) with parameters set so that each letter within the matrix represents a match of 50% or greater over a span of 21 amino acids. Extensive homology

Table 1. Amino acids at positions in which there are differences between the residues of isolectins 1 and 2 and the protein encoded by pNVR1

| Amino acid | Isolectin 1 | Isolectin 2 | pNVR1 |
|---|---|---|---|
| 56 | Thr | Pro | Pro |
| 59 | Gln | His | His |
| 66 | Tyr | His | His |
| 93 | Ala | Ser | Ala |
| 9 | Asn | Asn | Gly |
| 37 | Asp | Asp | Asn |
| 53 | Ala | Ala | Lys |
| 109 | Ser | Ser | Tyr |
| 119 | Gly | Gly | Glu |
| 123 | Ser | Ser | Asn |
| 171 | Ala | Ala | Gly |

**Figure 4. RNA blot analysis of WGA mRNA levels.** Poly(A)$^+$ RNA (1 ug), isolated from embryos excised at 20-day post-anthesis and cultured in the presence (lane 1) and absence (lane 2) of abscisic acid, was separated on a 2% agarose, 6% formaldehyde gel. After transferring the RNA to nitrocellulose, the filter was hybridized to a $^{32}$P-labeled DNA insert from pNVR1 under stringent conditions. Positions of DNA $M_r$ markers were obtained from the ethidium bromide-stained portion of the gel.

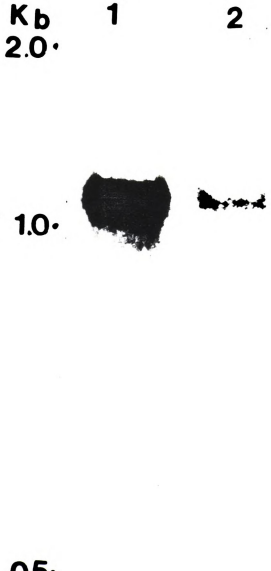
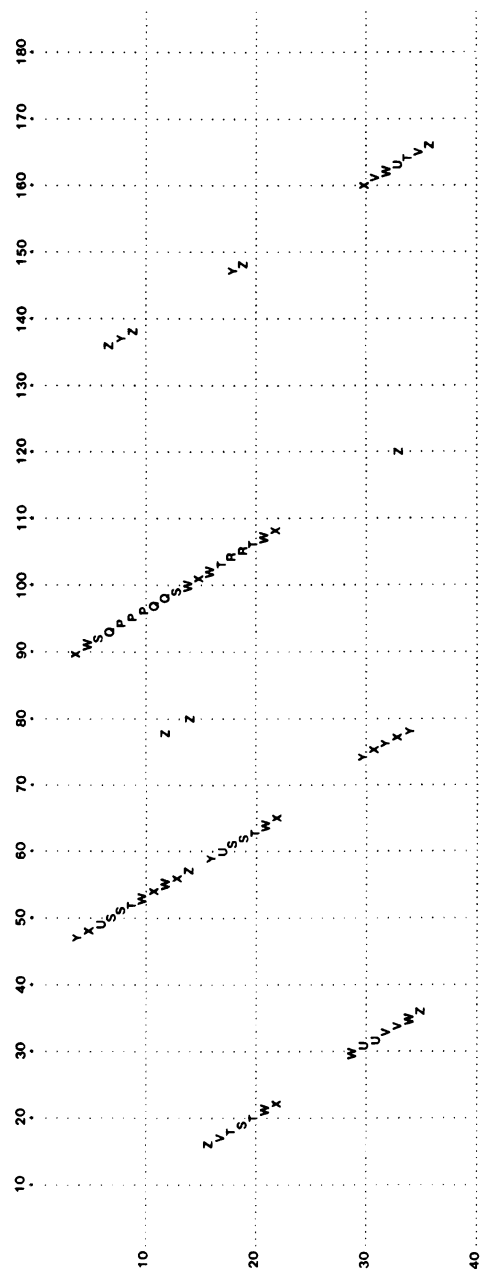

Kb
1
2
2.0
1.0
0.5

**Figure 5. Amino-acid homology matrix of WGA (*x* axis) and chitinase from *P. vulgaris* (*y* axis).** Homology matrices were plotted with the Pustell and Kafatos sequence analysis program (1984) using the following parameters: range = 10, scale factor = 0.75, minimum value = 50, compression = 1. Each letter represents homology at that point in the matrix where A = 100%, B = 98%,..., Z = 50%. Only homologies in the first 40 amino acids of chitinase are plotted; the remainder of the protein shows no homology with WGA.


10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180

between the amino terminus of chitinase and four regions of WGA is apparent.

## DISCUSSION

In this paper we present the amino acid sequence of WGA as deduced from a cDNA clone designated pNVR1. That this clone encodes WGA has been verified by comparison of the deduced amino acid sequence with the sequence determined by direct amino acid sequencing of the purified protein (Wright, *et al.*, 1984; Wright and Olafsdottir, 1986). The length of the polypeptide derived from the deduced amino acid sequence is 186 amino acids, and the calculated $M_r$ is 18,754. Nevertheless, pNVR1 does not represent the complete coding sequence for WGA. First, the initiating methionine codon is absent from the cDNA. Second, because WGA is synthesized on and translocated across the rough endoplasmic membrane (Mansfield, *et al.*, 1988) an amino-terminal signal peptide would be expected (Chrispeels, 1984). Third, there may be one or more amino acids at the amino terminus that have not been detected by peptide sequencing because of blockage of the amino terminus (Wright, *et al.*, 1984; Wright and Olafsdottir, 1986). The size of the mRNA recognized by clone pNVR1 predicts a full-length cDNA of 1.1 kb.

The DNA sequence of pNVR1 encodes a protein that extends 15 amino acids beyond the carboxyl terminus of the published amino acid sequence and includes a potential site for N-linked glycosylation. Isolated WGA is not a glycoprotein, but the precursor form is glycosylated (Mansfield, *et al.*, 1988).

The site of glycosylation probably lies in the 15 amino acid carboxyl-terminal sequence because the only possible site for glycosylation resides in this region. The glycosylated precursor is known to be processed (Mansfield, *et al.*, 1988) and to accumulate in protein bodies or vacuoles (Mishkind, *et al.*, 1982; Raikhel, *et al.*, 1984). The WGA precursor in the endoplasmic reticulum-associated fraction is 5 kDa larger than the mature WGA (Mansfield, *et al.*, 1988). The difference in molecular mass between the precursor and mature WGA may be a consequence of the extra 15 amino acids and glycosylation of the carboxyl terminus. The hydropathy plot of the amino acid sequence derived from pNVR1 clearly indicates that the carboxyl terminus of the cloned WGA sequence consists of hydrophobic amino acids, which is consistent with the possibility that it is removed post-translationally. Removal of carboxyl-terminal residues was seen during maturation of napin, a rapeseed storage protein (Ericson, *et al.*, 1986). It was recently shown that the lectin concanavalin A (Con A), which is not a glycoprotein, is synthesized as a glycosylated precursor (Herman, *et al.*, 1985). Normal transport of this protein is dependent on the presence of the glycan (Faye and Chrispeels, 1987). It is interesting that WGA precursor is a biologically active lectin (Mansfield, *et al.*, 1988), whereas precursor for Con A does not have lectin activity (Herman, *et al.*, 1985). In other words, the loss of the pro-WGA carboxyl-terminal domain does not relate to its ability to bind *N*-acetylglucosamine. Alternatively, cleavage of the carboxyl terminus may occur during the purification of WGA such that the mature protein actually contains 186

amino acids *in vivo*.

Clone pNVR1 mRNA contains four termination codons and a 3'-untranslated region. A potential polyadenylation signal (AAUAAU) is found in the noncoding region followed by a poly(A) tail. Whereas the consensus sequence for the polyadenylation signal is very highly conserved in animal systems (AAUAAA), plant mRNAs frequently deviate from this theme (Dean, *et al.*, 1986). The deduced amino acid sequence confirms extensive interdomain homology. The 7-amino acid sequence Gly-Cys-Gln-Asn-Gly-Ala-Cys is found at residues 34-40 and again at residues 120-126. Short repeated stretches of Tyr-Cys-Gly, Ala-Gly-Gly, Gly-Cys-Gln, Cys-Cys-Ser, or Cys-Gly-Gly are found throughout the polypeptide.

Amino acid sequence studies on wheat isolectins 1 (A genome) and 2 (D genome) (Rice and Etzler, 1975; Peumans, *et al.*, 1982; Wright, *et al.*, 1984; Wright and Olafsdottir, 1986) indicated that they differ distinctly in their histidine content: two histidines in isolectin 2 and no histidine in isolectin 1 (8). Because clone pNVR1 was isolated from a cDNA library derived from the tetraploid wheat *T. durum* (AABB), the cDNA clone cannot encode isolectin 2 derived from the D genome. Furthermore, pNVR1 probably does not encode isolectin 1 from the A genome. Isolectin 1 does not contain any histidine, whereas pNVR1 encodes a protein containing two histidine residues. Thus, pNVR1 probably represents isolectin 3 derived from the B genome. Although the amino acid compositions of isolectins 2 and 3 are very similar, eight discrete differences were identified

between them. At least four of these differences (residues 9, 53, 93, and 119) are authentic. The x-ray crystallographic data for these four positions in isolectin 2 are definitive, and there is no evidence for heterogeneity in peptide preparations (Wright, *et al.*, 1984). The discrepancies at the remaining four positions (37, 109, 123, and 171) between the deduced amino acids and isolectin 2 could be because of inaccuracies resulting from cross-contamination of the isolectins during fractionation.

Abscisic acid treatment of developing wheat embryos has been shown to affect temporal expression of WGA (Raikhel and Quatrano, 1986; Triplett and Quatrano, 1982). Using clone pNVR1 as a hybridization probe, we found that abscisic acid treatment of excised wheat embryos modulates mRNA levels for WGA, which is consistent with known effects of abscisic acid on lectin levels (Triplett and Quatrano, 1982). Similar results were reported by Williamson *et al.* (Williamson, *et al.*, 1985) for the abundant embryo storage protein. It is possible that abscisic acid regulation is based upon changes in the rates of mRNA transcription, turnover, or processing. It also needs to be mentioned that clone pNVR1 may be hybridizing to the mRNAs for isolectins 1 and 2, as well as to the mRNA for isolectin 3 on the RNA blot. Given the similarity of the isolectin sequences, the high-stringency conditions used for hybridization may not have prevented cross-hybridization with mRNAs from related isolectins.

WGA and chitinase are two chitin-binding proteins that are thought to have antimicrobial activity (Mirelman, *et al.*, 1975). Recently, however, evidence

was presented to show that antifungal activity of WGA can result from contamination by chitinase (Schlumbaum, *et al.*, 1986). Comparison of amino acid sequences demonstrated a striking homology between the amino terminus of chitinase (Broglie, *et al.*, 1986) and four regions of the WGA molecule. The amino acid residues of WGA directly involved in primary sugar-binding sites are tyrosine-73, serine-62, and glutamic acid-115 (Wright, 1984). These three residues are found in the regions of homology between chitinase and WGA. One may speculate that these regions of homology account for the similarity in chitin-binding activity of these proteins and , subsequently, in copurification. Additionally, the sequence homology between WGA and chitinase may be of functional significance.

## REFERENCES

Allen, A.K., Neuberger, A. and Sharon, N. (1973) The purification, composition and specificity of wheat-germ agglutinin. *Biochem. J.* **131,** 155-162.

Broglie, K.E., Gaynor, J.J. and Broglie, R.M. (1986) Ethylene-regulated gene expression: Molecular cloning of the genes encoding an endochitinase from *Phaseolus vulgaris. Proc. Natl. Acad. Sci. USA* **83,** 6820-6824.

Chrispeels, M.J. (1984) Biosynthesis, processing and transport of storage proteins and lectins in cotyledons of developing legume seeds. *Philos. Trans. R. Soc. London. Ser.* **B 304,** 309-322.

Dean, D., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. *Nucleic Acids Res.* **14,** 2229-2240.

Ericson, M.L., Rodin, J., Lenman, M., Glimelius, K., Josefson, L.G. and Rask, L. (1986) Structure of the rapeseed 1.7S storage protein, napin, and its precursor. *J. Biol. Chem.* **261,** 14576-14581.

Faye, L. and Chrispeels, M.J. (1987) Transport and processing of the glycosylated precursor of Concanavalin A in jack-bean. *Planta* **170,** 217-224.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132,** 6-13.

Goldstein, I.J. and Hayes, C.E. (1978) The lectins: Carbohydrate-binding proteins of plants and animals. *Adv. Carbohydr. Chem. Biochem.* **35,** 127-340.

Hanahan, D. (1985) Techniques for transformation of *E. coli*. In *DNA Cloning*, ed. Glover, D.M. (IRL, Oxford), **Vol. 1,** pp. 109-135.

Hanahan, D. and Meselson, M. (1983) Plasmid screening at high colony density. *Methods Enzymol.* **100,** 333-342.

Herman, E.M., Shannon, L.M. and Chrispeels, M.J. (1985) Concanavalin A is synthesized as a glycoprotein precursor. *Planta* **165,** 23-29.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105-132.

Lis, H. and Sharon, N. (1981) Lectins as molecules and tools. *Annu. Rev. Biochem.* **55,** 35-67.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Mansfield, M.A., Peumans, W.J. and Raikhel, N.V. (1988) Wheat-germ agglutinin is synthesized as a glycosylated precursor. *Planta* **173,** 482-489.

Matteucci, M.D. and Caruthers, M.H. (1980) *Tetrahedron Lett.* **21,** 219-722.

Mirelman, D., Galun, E., Sharon, N. and Loyan, R. (1975) Inhibition of fungal growth by wheat germ agglutinin. *Nature* (London) **256,** 414-416.

Mishkind, M.L., Palevitz, B.A., Raikhel, N.V. and Keegstra, K. (1983) Localization of wheat germ agglutinin-like lectins in various species of the Gramineae. *Science* **220,** 1290-1292.

Mishkind, M.L., Raikhel, N.V., Palevitz, B.A. and Keegstra, K. (1982) Immunocytochemical localization of wheat germ agglutinin in wheat. *J. Cell Biol.* **92,** 753-764.

Nagata, Y. and Burger, M.M. (1974) Wheat germ agglutinin. Molecular characteristics and specificity for sugar binding. *J. Biol. Chem.* **10,** 3116-3122.

Peumans, W.J., Stinissen, H.M. and Carlier, A.R. (1982) A genetic basis for the origin of six different isolectins in hexaploid wheat. *Planta* **154,** 562-567.

Pustell, J. and Kafatos, F.C. (1984) A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. *Nucleic Acids Res.* **12,** 643-655.

Raikhel, N.V. and Pratt, L. (1987) Wheat germ accumulation in coleoptiles of different genotypes of wheat. Localization by monoclonal antibodies. *Plant Cell Rep.* **6,** 146-149.

Raikhel, N.V. and Quatrano, R.S. (1986) Localization of wheat germ agglutinin in developing wheat embryos and those cultured in abscisic acid. *Planta* **168,** 433-440.

Raikhel, N.V., Mishkind, M.L. and Palevitz, B.A. (1984) Characterization of a wheat germ agglutinin-like lectin from adult wheat plants. *Planta* **162,** 55-61.

Raikhel, N.V., Mishkind, M. and Palevitz, B.A. (1984) Immunocytochemistry in plants with colloidal gold conjugates. *Protoplasma* **121,** 25-33.

Rice, R.H. and Etzler, M.E. (1975) Chemical Modification and hybridization of wheat germ agglutinins. *Biochemistry* **14,** 4093-4099.

Schlumbaum, A., Mauch, F., Vogeli, U. and Boller, T. (1986) Plant chitinases are potent inhibitors of fungal growth. *Nature* **324,** 365-367.

Stinissen, H.M., Chrispeels, M.J. and Peumans, W.J. (1985) Biosynthesis of lectin in roots of germinating and adult cereal plants. *Planta* **164,** 278-286.

Thomas, P.S. (1980) Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77,** 5201-5205.

Triplett, B.A. and Quatrano, R.S. (1982) Timing, localization, and control of wheat germ agglutinin synthesis in developing wheat embryos. *Dev. Biol.* **91,** 491-496.

Williamson, J.D., Quatrano, R.S. and Cuming, A.C. (1985) $E_m$ polypeptide and its messenger RNA levels are modulated by abscisic acid during embryogenesis in wheat. *Eur. J. Biochem.* **152,** 501-507.

Wright, C.S. (1984) Structural comparison of the two distinct sugar binding sites in wheat germ agglutinin isolectin II. *J. Mol. Biol.* **178,** 91-104.

Wright, C.S. (1987) Refinement of the crystal structure of wheat germ agglutinin isolectin 2 at 1.8 A resolution. *J. Mol. Biol.* **194,** 501-529.

Wright, C.S. and Olafsdottir, S. (1986) Structural differences in the two major wheat germ agglutinin isolectins. *J. Biol. Chem.* **16,** 7191-7195.

Wright, C.S., Gavilanes, F. and Peterson, D.L. (1984) Primary structure of wheat germ agglutinin isolectin 2. Peptide order deduced from x-ray structure. *Biochemistry* **23,** 280-287.

## Chapter 3

## Expression of rice lectin is governed by two temporally and spatially regulated mRNAs in developing embryos


### ABSTRACT

Two cDNA clones encoding rice lectin have been isolated and characterized to investigate the expression of rice lectin at the molecular and cellular levels. The two cDNA clones code for an identical 23 kd protein which is processed to the mature polypeptide of 18 kd by co-translational cleavage of a 2.6 kd signal sequence and selective removal of a 2.7 kd COOH-terminal peptide which contains a potential N-linked glycosylation site. In addition, the mature 18 kd lectin is post-translationally cleaved between residues 94 and 95 to yield polypeptides of 10 kd and 8 kd, corresponding to the $NH_2$- and COOH-terminal portions of the mature subunit, respectively. Northern blot analysis established that rice lectin is encoded by two mRNA transcripts (0.9 kb and 1.1 kb). On Southern blots, the rice lectin cDNAs hybridize specifically to single restriction fragments between 10 kb and 20 kb. *In situ* hybridization showed localization of the 1.1 kb rice lectin mRNA in root caps and specific cell-layers of the radicle, coleorhiza, scutellum, and coleoptile. Northern blot analysis demonstrated that both the 0.9 kb and 1.1 kb mRNAs are present in developing rice embryos. The two lectin mRNAs are differentially expressed temporally such that the 1.1 kb lectin mRNA accumulates to levels two-fold higher than the 0.9 kb mRNA.

## INTRODUCTION

Plant lectins are a class of proteins that bind and cross-link specific carbohydrates. Because of their unique carbohydrate-binding properties, lectins are widely used as tools in medical cell biology (Lis and Sharon, 1986). Historically, plant lectin research has focused on the isolation and characterization of new lectin species to broaden the spectrum of specific carbohydrate-binding moieties. Although the function of lectins in plants remains obscure, dissecting the regulation of expression of lectin genes at the molecular level should facilitate elucidation of the protein function *in vivo*.

Many of the Gramineae synthesize N-acetylglucosamine (GlcNAc)-binding lectins with similar immunological properties (Peumans and Stinissen, 1983). These lectins accumulate in a cell-type specific manner in various organs of developing embryos and young seedlings. Rice lectin, initially purified and characterized by Tsuda (1979) from rice bran, is a dimeric protein composed of two glycine- and cysteine-rich 18 kd subunits that lack covalently-bound sugar residues. In the cultivated rice species *Oryza sativa* L., the majority of the 18 kd subunits undergo a proteolytic cleavage event which yields two subunits of 8 kd and 10 kd (Stinissen *et al.*, 1984). This lectin is synthesized as a 23 kd monomeric precursor on the rough endoplasmic reticulum (RER) and is subsequently assembled into dimers within the lumen of the RER (Stinissen *et al.*, 1984). Assembled dimers are only transiently associated with the RER before being transported to and deposited in vacuoles/protein bodies (Stinissen

*et al.*, 1984). Rice lectin accumulates in specific cell-layers of the scutellum, coleorhiza, radicle, root cap, and throughout cell-layers of the coleoptile of embryos (Mishkind *et al.*, 1983). We are interested in the molecular mechanisms regulating cell-specific expression of the Gramineae lectins. Two cDNA clones encoding rice lectin have been isolated and used to examine the expression of rice lectin in developing embryos. In this paper, we present evidence that the two cDNA clones represent two distinct mRNA transcripts. Each lectin mRNA transcript exhibits a distinct pattern of temporal expression in developing embryos. Moreover, the cell-type specific expression of rice lectin mRNAs is developmentally and spatially regulated.

## MATERIALS AND METHODS

***Plant material.*** Developing rice (*Oryza sativa* L. cv. Lemont) embryos were collected from spikes harvested at 5, 10, 20, 30, and 40 days post-anthesis (DPA) from plants maintained under greenhouse conditions. Embryos used for *in situ* hybridization experiments were processed immediately, while the bulk of collected embryos (10, 30, 40 DPA) were quick frozen in liquid nitrogen and stored at -80° C for RNA isolation.

Young seedlings of the rice cultivars Nato or IR36 were germinated and grown in Baccto professional potting mix in a growth chamber with a 12-hr light period at 27°C and a 12-hr dark period at 21°C, with the 70% humidity. Shoots of 10-day old seedlings were collected and frozen in liquid nitrogen for isolation

of total DNA.

***Screening of a λgt10 cDNA library for rice lectin.*** A λgt10 cDNA library constructed from poly(A)$^+$ RNA isolated from spikes of rice *Oryza sativa* L. cv. Nato was provided by Susan Wessler and Ron Okagaki (University of Georgia, Athens, GA). Approximately 160,000 recombinant phage were grown on *Escherichia coli* C600hfl at a density of 40,000 per 150-mm Petri plate and replicated onto nitrocellulose filters as described in Maniatis *et al.* (1982). The nitrocellulose filters were hybridized with $^{32}$P-random primer-labeled cDNA insert (Feinberg and Vogelstein, 1983) from clone WGA-B (clone pNVR1 described in Raikhel and Wilkins, 1987) for 18 hr in 6x SSC, 5x Denhardt's solution, 0.2% SDS, and sonicated salmon sperm DNA at 5 ug/ml. Post-hybridization washes included three 15 min washes at room temperature and two 15 min washes at 60$^0$C in 3x SSC,0.1% SDS. Positive phage were plaque-purified to homogeneity (Maniatis *et al.*, 1982) under high stringency screening conditions (Mansfield, *et al.*, 1989) using $^{32}$P-labeled insert from WGA-B (Raikhel and Wilkins, 1987).

***DNA nucleotide sequence analysis.*** Inserts, designated cRL852 and cRL1035, were purified from selected phage by electrophoresis in low-melting-point agarose (Struhl, 1985) and cloned into pUC119 (Vieira and Messing, 1987) in both orientations for subsequent DNA sequence determination. A sequential

series of overlapping deletions from both strands of the cDNA were generated by T4 DNA Polymerase (Dale and Arrow, 1987) from full-length single-stranded DNA templates (Vieira and Messing, 1987). Single-stranded deletion templates were sequenced by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) using $^{35}$S-dATP and 7-deaza-dGTP instead of dGTP (Mizusawa *et al.*, 1986). Computer alignment of overlapping deletions, amino acid and sequence analysis were performed using Microgenie software [Beckman].

A fortuitous deletion encompassing the terminal 165 bp of 3'-untranslated region unique to cRL1035 was retrieved for use as a clone-specific probe. This partial cDNA clone was maintained in pUC119 and given the designation cRL165.

***Amino acid sequence determination of $NH_2$- and COOH-terminal amino acid residues of rice lectin.*** Rice lectin was purified from 10 g of mature rice embryos (cv. IR36) via affinity chromatography on immobilized N-acetylglucosamine (Selectin 1, Pierce) according to the procedure detailed in Mansfield *et al.* (1988). To enhance resolution of the rice lectin during SDS-PAGE on a 15% polyacrylamide gel (Laemmli, 1970), the purified protein was S-carboxyamidated at 37°C for 30 min in the presence of 240 mM iodoacetimide (Raikhel *et al.*, 1984) prior to electrophoresis. Individual subunits (8 kd and 10 kd) of rice lectin were visualized by staining the gel briefly (10 min) in Coomassie blue, followed by destaining in 30% methanol,7.5% acetic acid. The

8 kd and 10 kd polypeptides were excised from the gel, electroeluted in Laemmli (1970) buffer, and lyophilized. SDS was removed from protein by the organic extraction method of Konigsberg and Henderson (1983). Removal of salts from the protein was accomplished by dialysis against 15% acetic acid at 4°C in the dark for two days prior to lyophilization.

Amino acid sequence determinations were performed at the Protein Chemistry Facility, University of California, Irvine. Approximately 200 picomoles of gel-purified rice lectin was applied to a Model 477 Sequenator equipped with a 120 on-line PTH-amino acid analyzer (Applied Biosystems, Inc.) for determination of $NH_2$-terminal amino acid residues. The terminal amino acids of the COOH-terminus were determined by carboxypeptidase Y digestion of 500 picomoles of rice lectin via the procedure of Hayashi (1977). The identification and quantitation of free amino acids in digestion mixtures were accomplished by HPLC analysis using precolumn derivatization with o-phthaldialdehyde.

***Northern blot analysis.*** Total RNA was isolated from 50 to 150 mg of developing rice embryos via the hot phenol method of Finkelstein and Crouch (1986) with the addition of 1% 2-mercaptoethanol to the homogenization buffer. Northern blots were prepared from 25 ug RNA for each developmental stage of embryos and hybridized with random-primer-labeled cRL852 insert under stringent conditions (Raikhel *et al.*, 1988). Blots were exposed to Kodak XAR-5 film with intensifying screens at -80°C for 10 to 15 hours. Autoradiograms were scanned

with a Gilford densitometer.

***Gene reconstruction analysis.*** Total DNA was isolated from 10-day old rice (cv. IR36 or Nato) seedlings according to Shure *et al.* (1983) and restricted to completion with *Eco*RI, *Hind*III, *Kpn*I-, *Sma*I-, or *Xba*I. Two ug of digested DNA ($3.3 \times 10^6$ genome equivalents) and 0.5-, 1.0-, and 3.0-copy equivalents of the cRL1035 cDNA clone were fractionated by agarose gel electrophoresis and transferred to nitrocellulose (Maniatis *et al.*, 1982). Gene copy reconstructions were based upon a rice genome size of $5.47 \times 10^5$ kb per haploid genome (Francis *et al.*, 1985). Hybridization and post-hybridization washes of the reconstruction blot were performed as described for Northern blots with the exception that random-primer radiolabeled insert from cRL1035 was used as a probe.

***In situ hybridization.*** Inserts from the cDNA clones cRL852, cRL1035, and cRL165 were subcloned into the *Eco*RI site of Bluescript (M13+) [Stratagene] for *in vitro* transcription of RNA transcripts. Sense and antisense transcripts labeled with $^{35}$S-rUTP were synthesized from the $T_3$ or $T_7$ promoters, hydrolyzed to approximately 100-300 bases, and prepared for *in situ* hybridization as described in Raikhel *et al.* (1989). Frozen tissue sections (8 to 10 um) from developing rice embryos were processed and hybridized with radiolabeled sense or antisense RNA transcripts from cRL852 or cRL1035 as detailed in Raikhel *et*

*al.* (1989). *In situ* hybridization with clone cRL165 was performed identically as with the full-length clones as described above with the exception that hybridization and post-hybridization washes were conducted at 42°C.

## RESULTS

***Two cDNA clones encoding rice lectin differ in the length of their 3'-untranslated region.*** Relying upon molecular and immunological similarities to rice lectin (Peumans and Stinissen, 1983), the cDNA clone WGA-B (Raikhel and Wilkins, 1987; Raikhel *et al.*, 1988) encoding isolectin B of wheat germ agglutinin (WGA) was used as a heterologous probe to isolate cDNA clones encoding rice lectin. The complete nucleotide sequence and the deduced amino acids of two cDNA clones, designated cRL852 and cRL1035, are presented in Figure 1A. The two cDNA clones are identical at the nucleotide and amino acid levels with the exception that cRL1035 contains an additional 183 bp of 3'-untranslated region extending beyond the 3'-terminus of cRL852 (denoted by a closed arrowhead in Figures 1A and 1B). The cDNA clone cRL1035 contains two putative polyadenylation signals (underlined in Figure 1A), AATAAA and an extended AATAAATAAA, located 86 and 150 nucleotides downstream from the coding region, respectively. The polyadenylation signal located proximal to the coding region (AATAAA), is common to both cDNA clones and is positioned 64 nucleotides upstream from the second polyadenylation signal. The AATAAATAAA polyadenylation signal distal to the coding region is unique to

**Figure 1. Nucleotide sequence and deduced amino acids of two cDNA clones, cRL852 and cRL1035, encoding rice lectin.**

**(A)** The deduced amino acid sequence for the rice lectin preproprotein is depicted as single letter codes positioned above the nucleotide sequence. Nucleotides are numbered on the left. The first amino acid (methionine) of the preproprotein is indicated by -28. The respective $NH_2$- and COOH-terminal residues of the mature lectin are indicated by arrows positioned above residues Q+1 and G173. An open triangle represents the endoproteolytic cleavage site between residues N94 and G95 of the mature 18 kd subunit of rice lectin. The 26 amino acids which extend beyond the COOH-terminus of the mature protein are depicted by the boxed residues. The potential N-linked glycosylation site at residue N183 is denoted by an asterisk. The final nucleotide of cRL852 is indicated by the solid arrowhead at position 852. Nucleotides beyond position 852 represent 3'-untranslated region unique to cRL1035. Underlined AT-rich sequences at positions 825 and 889 indicate presumptive polyadenylation signals. The extended polyadenylation signal at position 889 apparently represents two overlapping polyadenylation consensus sequences.

**A**

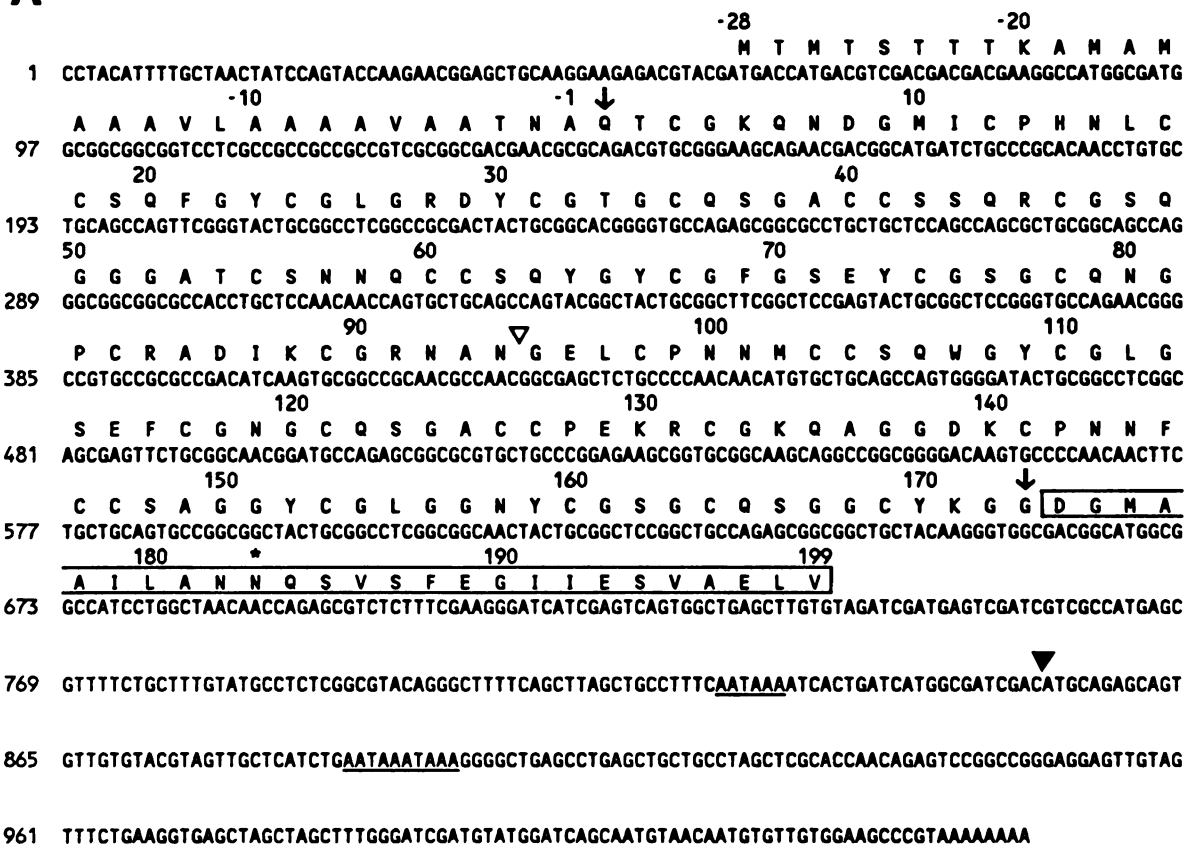
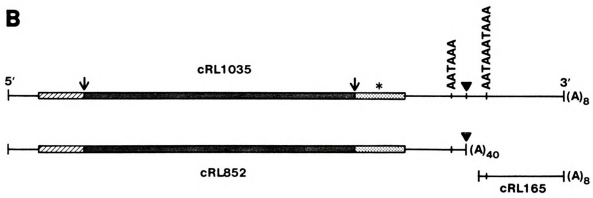

-28 -20
M T M T S T T T K A M A M
1 CCTACATTTTGCTAACTATCCAGTACCAAGAACGGAGCTGCAAGGAAGAGACGTACGATGACCATGACGTCGACGACGACGAAGGCCATGGCGATG
-10 -1 10
A A A V L A A A A V A A T N A Q T C G K Q N D G M I C P H N L C
97 GCGGCGGCGGTCCTCGCCGCCGCCGCCGTCGCGGCGACGAACGCGCAGACGTGCGGGAAGCAGAACGACGGCATGATCTGCCCGCACAACCTGTGC
20 30 40
C S Q F G Y C G L G R D Y C G T G C Q S G A C C S S Q R C G S Q
193 TGCAGCCAGTTCGGGTACTGCGGCCTCGGCCGCGACTACTGCGGCACGGGGTGCCAGAGCGGCGCCTGCTGCTCCAGCCAGCGCTGCGGCAGCCAG
50 60 70 80
G G G A T C S N N Q C C S Q Y G Y C G F G S E Y C G S G C Q N G
289 GGCGGCGGCGCCACCTGCTCCAACAACCAGTGCTGCAGCCAGTACGGCTACTGCGGCTTCGGCTCCGAGTACTGCGGCTCCGGGTGCCAGAACGGG
90 100 110
P C R A D I K C G R N A N G E L C P N N M C C S Q W G Y C G L G
385 CCGTGCCGCGCCGACATCAAGTGCGGCCGCAACGCCAACGGCGAGCTCTGCCCCAACAACATGTGCTGCAGCCAGTGGGGATACTGCGGCCTCGGC
120 130 140
S E F C G N G C Q S G A C C P E K R C G K Q A G G D K C P N N F
481 AGCGAGTTCTGCGGCAACGGATGCCAGAGCGGCGCGTGCTGCCCGGAGAAGCGGTGCGGCAAGCAGGCCGGCGGGGACAAGTGCCCCAACAACTTC
150 160 170
C C S A G G Y C G L G G N Y C G S G C Q S G G C Y K G G D G M A
577 TGCTGCAGTGCCGGCGGCTACTGCGGCCTCGGCGGCAACTACTGCGGCTCCGGCTGCCAGAGCGGCGGCTGCTACAAGGGTGGCGACGGCATGGCG
180 * 190 199
A I L A N N Q S V S F E G I I E S V A E L V
673 GCCATCCTGGCTAACAACCAGAGCGTCTCTTTCGAAGGGATCATCGAGTCAGTGGCTGAGCTTGTGTAGATCGATGAGTCGATCGTCGCCATGAGC
769 GTTTTCTGCTTTGTATGCCTCTCGGCGTACAGGGCTTTTCAGCTTAGCTGCCTTTCAATAAAATCACTGATCATGGCGATCGACATGCAGAGCAGT
865 GTTGTGTACGTAGTTGCTCATCTGAATAAATAAAGGGGCTGAGCCTGAGCTGCTGCCTAGCTCGCACCAACAGAGTCCGGCCGGGAGGAGTTGTAG
961 TTTCTGAAGGTGAGCTAGCTAGCTTTGGGATCGATGTATGGATCAGCAATGTAACAATGTGTTGTGGAAGCCCGTAAAAAAAA

**Figure 1. Nucleotide sequence and deduced amino acids of two cDNA clones, cRL852 and cRL1035, encoding rice lectin.**

**(B)** Organization of the cDNA clones cRL852, cRL1035, and cRL1035-specific clone cRL165. The clone-specific cDNA cRL165 corresponds to nucleotides 878 to 1043 of cRL1035 3'-untranslated regions are also indicated on the map. The solid box represents the coding region while the hatched and stippled boxes refer to the signal sequence and the COOH-terminal extension peptide, respectively. Arrows demarcate the $NH_2$- and COOH-terminal of mature rice lectin, respectively. The asterisk denoted the potential N-linked glycosylation site contained within the COOH-terminal domain.


B
AATAAA
AATAAATAAA
5′
cRL1035
*
3′
(A)8
(A)40
cRL852
cRL165
(A)8

cRL1035 and represents an apparent fusion of two overlapping polyadenylation signal consensus motifs (AATAAA). Both clones, cRL852 and cRL1035, contain 57 nucleotides of 5'-untranslated region preceding an ATG initiation codon, which is followed by an open reading frame of 681 bases with a translation termination codon (TAG) at position 739. The protein encoded by the open reading frame of both clones encompasses 227 amino acid residues with calculated $M_r$ 22,798. Comparison of partial amino acid sequence data (Chapot *et al.*, 1986) with the deduced amino acid sequence from the two cDNA clones confirms that the clones encode rice lectin. Moreover, *in vitro* translation products synthesized from RNA transcripts generated for each clone are immunoprecipitable with polyclonal antiserum raised against WGA (data not shown).

The polypeptide encoded by the two lectin cDNA clones contains an alanine-rich array of 28 amino acid residues ($M_r$ 2,643) at the $NH_2$-terminus (hatched box, Figure 1B) exhibiting the predicted tripartite organization of eukaryotic signal sequences (von Heijne, 1983). Predicated upon the organization, the proteolytic processing of the signal sequence presumably occurs in rice lectin between an alanine (A-1) and glutamine residue (Q+1, arrow in Figure 1A,B) of the deduced amino acid sequence. $NH_2$-terminal amino acid analysis of mature rice lectin indicated that the terminus is blocked. These data are congruent with earlier reports predicting that the $NH_2$-terminal residue of rice lectin is a glutamine which is presumably modified by cyclization to pyrrolidone carboxylic acid (Chapot *et*

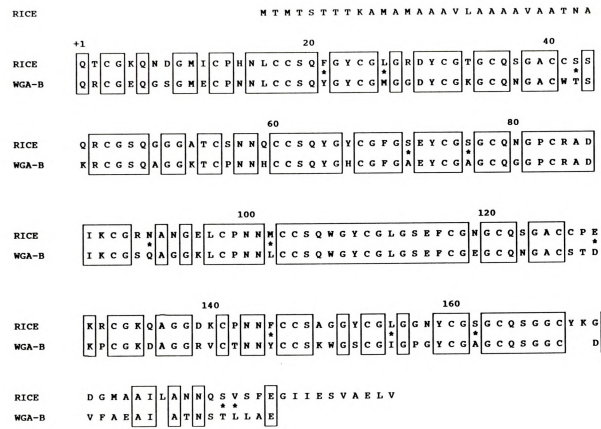
*al.*, 1986), a residue which is resistant to Edman degradation. Initiating with glutamine Q+1, the cDNAs encode a protein comprised of 199 amino acids with calculated M, 20,172. However, amino acid sequence analysis of the COOH-terminal amino acids indicates that mature rice lectin terminates at the glycine residue G173 (arrow in Figure 1A,1B). Thus, determination of terminal amino acid residues revealed that the mature polypeptide of rice lectin is comprised of 173 amino acids with M, 17,512 (demarcated by arrows in Figure 1A,B and a solid box in Figure 1B). The amino acid composition indicates that the mature subunit of rice lectin is a glycine- and cysteine-rich polypeptide. Cysteine (23%) and glycine (19.7%) together account for almost 43% of the mature polypeptide amino acid composition.

In addition to the signal sequence and the mature rice lectin subunit, the cDNAs encode proteins with an additional 26 amino acids (M, 2,678) extending beyond the COOH-terminus of mature rice lectin (boxed residues in Figure 1A, stippled box in Figure 1B). This COOH-terminal extension is a relatively hydrophobic domain and contains a potential N-linked glycosylation site at asparagine residue N179 (asterisk in Figures 1A,B). Rice lectin is therefore synthesized as a preproprotein that requires the proteolytic removal of the signal sequence and post-translational processing of a COOH-terminal domain to yield the mature polypeptide. In vacuoles, the mature 18 kd subunit polypeptide undergoes additional post-translational processing to yield two smaller polypeptides of approximately 10 kd and 8 kd (Stinissen *et al.*, 1984). To

resolve the relationship between these polypeptides and the protein encoded by the cDNAs, both polypeptides were purified and subjected to $NH_2$-terminal and COOH-terminal amino acid sequence analyses. Results from these analyses indicate that the mature subunit of rice lectin is proteolytically cleaved between amino acids residues 94 and 95 as deduced from the cDNA clones (open arrowhead, Figure 1). The resultant 10 kd and 8 kd polypeptides correspond to the $NH_2$- and COOH-terminal portions of the mature 18 kd protein, respectively.

A comparison of amino acids from rice lectin and isolectin B of wheat germ agglutinin (WGA-B) is presented in Figure 2. Rice lectin exhibits 73% identity with WGA-B (boxed amino acids in Figure 2) within the coding region of the mature subunits spanning from glutamine Q+1 to glycine G171 in WGA-B or glycine G173 in rice lectin. The overall homology between the two lectins increases to 79.5% when conserved amino acid changes (asterisks in Figure 2) are included in the comparison. Both rice lectin and WGA-B require the post-translational processing of COOH-terminal domains to produce the mature 18 kd subunit. Alignment of the 26 amino acid COOH-terminal domain from the proprotein of rice lectin and the 15 amino acid COOH-terminal domain from pro-WGA-B for maximal homology shows a 46.7% overall amino acid conservation, indicating that this region is less conserved than the coding region of the mature protein.

**Figure 2. Comparison of amino acid sequences between rice lectin and isolectin B of wheat germ agglutinin (WGA-B).** The complete deduced amino acid sequence of rice lectin was aligned for maximal homology to the available amino acid sequence of WGA-B (Raikhel and Wilkins, 1987). Identical amino acids are depicted by boxed residues whereas conserved amino acid changes between the two lectins are denoted by asterisks.


RICE
M T M T S T T T K A M A M A A A V L A A A A V A A T N A
+1
20
40
RICE
Q T C G K Q N D G M I C P H N L C C S Q F G Y C G L G R D Y C G T G C Q S G A C C S S
WGA-B
Q R C G E Q G S G M E C P N N L C C S Q Y G Y C G M G G D Y C G K G C Q N G A C W T S
60
80
RICE
Q R C G S Q G G G A T C S N N Q C C S Q Y G Y C G F G S E Y C G S G C Q N G P C R A D
WGA-B
K R C G S Q A G G K T C P N N H C C S Q Y G H C G F G A E Y C G A G C Q G G P C R A D
100
120
RICE
I K C G R N A N G E L C P N N M C C S Q W G Y C G L G S E F C G N G C Q S G A C C P E
WGA-B
I K C G S Q A G G K L C P N N L C C S Q W G Y C G L G S E F C G E G C Q N G A C S T D
140
160
RICE
K R C G K Q A G G D K C P N N F C C S A G G Y C G L G G N Y C G S G C Q S G G C Y K G G
WGA-B
K P C G K D A G G R V C T N N Y C C S K W G S C G I G P G Y C G A G C Q S G G C D G
RICE
D G M A A I L A N N Q S V S F E G I I E S V A E L V
WGA-B
V F A E A I A T N S T L L A E

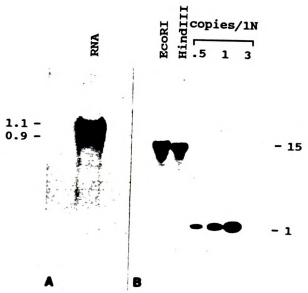
***Rice lectin is encoded by two different mRNAs.*** To explore the relationship between the two cDNA clones encoding rice lectin, a Northern blot containing total RNA from developing rice embryos (10 to 20 days post-anthesis (DPA)) was probed with $^{32}$P-labeled insert from cRL852 or cRL1035. Two mRNA species of approximately 1.1 kb and 0.9 kb were identified (Figure 3A). Therefore, the two cDNA clones correspond to the two mRNAs and did not arise from cloning artifacts. To discriminate expression due solely to the 1.1 kb mRNA species, a clone-specific probe (cRL165) encompassing 165 bp of the 3'-untranslated region unique to cRL1035 was constructed (see Figure 1B). The specificity of cRL165 as a clone-specific probe for the cDNA cRL1035 and the 1.1 kb lectin mRNA transcript was confirmed by Southern and Northern blot analysis, respectively (data not shown).

To determine the number of rice lectin genes responsible for the two mRNAs, Southern blots containing restricted genomic DNA were hybridized with $^{32}$P-labeled inserts from cDNA clones cRL852, cRL1035, or cRL165. Figure 3B shows a representative Southern blot depicting single restriction fragments of 11 kb and 15 kb detected in *Eco*RI- and *Hind*III-digested genomic DNA (cv. IR36), respectively, probed with radiolabeled insert from cRL1035. Single restriction fragments between 10 kb and 20 kb were also detected in *Eco*RI-, *Hind*III, *Kpn*I-, *Sma*I- and *Xba*I-restricted genomic DNA from the rice cultivars IR36 or Nato (data not shown). The identification of single restriction fragments by both the full-length and the clone-specific cDNAs suggests that rice lectin is

**Figure 3. Northern blot and gene reconstruction analysis of rice lectin.**

**(A)** Northern blot containing 25 ug total RNA isolated from developing rice embryos (cv. IR36) and hybridized with radiolabeled insert from cRL852. The size of the lectin mRNAs in kb are indicated to the left of the blot.

**(B)** Southern blot containing 3.3 X $10^6$ genome equivalents of rice genomic DNA (cv. IR36) restricted with *Eco*RI of *Hind*III and probed with radiolabeled cRL1035. Reconstruction lanes represent 0.5-, 1.0-, and 3.0-gene copy equivalents of cRL1035 per haploid genome. Approximate sizes in kb are positioned to the right.


RNA
EcoRI
HindIII
Copies/1N
.5 1 3
1.1 –
0.9 –
– 15
– 1
A
B

encoded by a single gene.

Gene reconstruction experiments (Figure 3B) were performed with restricted rice genomic DNA and purified insert from cRL1035 to ascertain the number of copies of the rice lectin gene present in the genome. Comparison of relative intensities of cDNA cRL1035 titered at 0.5-, 1-, and 3-copies per haploid genome and the intensity of genomic DNA restriction fragments from cv. IR36 (Figure 3B) demonstrated that the gene for rice lectin is present in 1 to 2 copies per haploid genome. In summary, Northern and Southern analyses indicate that the expression of rice lectin is determined by two mRNAs possibly derived from a single gene present in low copies in the genome of the diploid rice species *Oryza sativa* L.

***The differential expression of the two rice lectin mRNAs are temporally and spatially regulated.*** Using *in situ* hybridization in conjunction with Northern blot analysis, the cDNA clones encoding rice lectin were employed to investigate the expression of rice lectin during embryo development. *In situ* hybridization experiments were specifically performed to determine if the cell-specific accumulation of rice lectin in the diverse tissue types of developing embryos (i.e. radicle versus coleoptile) results from the spatial regulation of the two rice lectin mRNAs. Tissue sections of developing embryos harvested at 5, 10, 20 and 30 DPA were hybridized with $^{35}$S-labeled antisense RNA transcripts generated from cRL852 to localize both the 0.9 kb and 1.1 kb lectin mRNAs *in situ*. As

illustrated in Figure 4, rice lectin mRNAs localized by cRL852 antisense transcripts in embryos 20 DPA are confined to root caps, several cell-layers at the periphery of the coleorhiza and radicle, and in all cell-layers of the coleoptile. Not shown is the presence of lectin mRNA in the peripheral cell-layers of the scutellum. Moreover, the tissue-specific accumulation of lectin mRNA in the diverse tissue-types of the embryos is discernible as early as 5 DPA and persists throughout embryogenesis and seed maturation (data not shown). Sense RNA transcripts were used as controls to ascertain backgrounds levels due to the non-specific binding of probes to tissue sections. No hybridization was observed in control sections probed with sense RNA transcripts (data not shown).

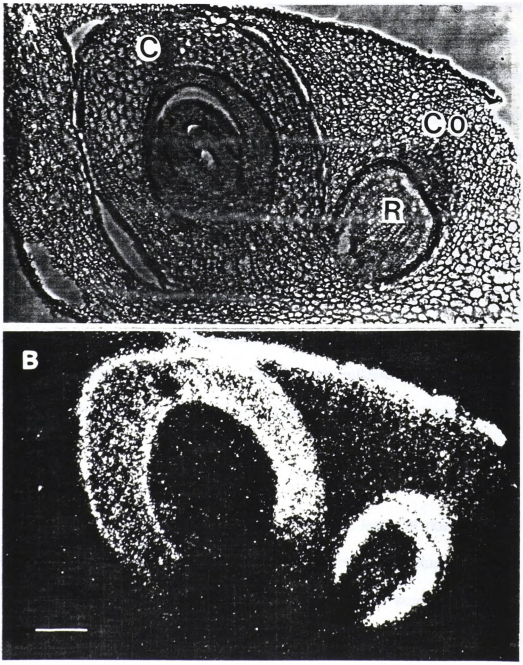
*In situ* hybridization was performed in parallel using radiolabeled antisense RNA transcripts synthesized from the clone-specific cDNA, cRL165, to detect the specific accumulation of the 1.1 kb lectin mRNA during embryo development. The results revealed that the 1.1 kb mRNA shows the same tissue-specific distribution of lectin mRNA localized by cRL852 antisense RNA transcripts in all developmental stages examined (data not shown). Thus, the 1.1 kb lectin mRNA accumulates in specific cell-layers of the root caps, coleorhiza, scutellum, radicle, and coleoptile of mature and developing embryos as early as 5 DPA.

Northern blot analysis was employed to resolve both the 0.9 kb and 1.1 kb lectin mRNA species and to ascertain what fraction of lectin mRNA can be attributed to each mRNA during embryogenesis. Northern blots prepared from

**Figure 4. Localization of rice lectin mRNAs in a developing embryo.** *In situ* hybridization of embryo sections harvested at 20 DPA and hybridized with $^{35}S$-labeled antisense transcript generated from the rice lectin cDNA clone cRL852 as described in Methods.

**(A)** Bright field photomicrograph detailing the cellular organization of a developing rice embryo. C, R, Co, refer to the coleoptile, radicle, and coleorhiza, respectively.

**(B)** Dark field photomicrograph depicting the tissue-specific distribution of rice lectin mRNA in a developing rice embryo. The white grains represent formation of RNA/RNA hybrids. Bar=50 um.

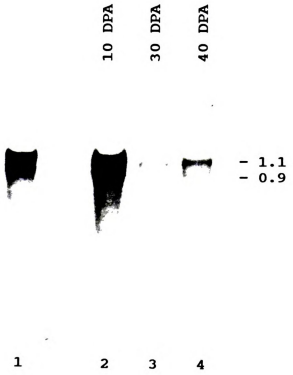

A
C
Co
R
B

total RNA isolated from embryos harvested at 10, 30 and 40 DPA were hybridized with radiolabeled insert from cRL852 and are depicted in Figure 5. Both lectin mRNAs are present at high levels in developing embryos harvested at 10 DPA (Figure 5, lane 2). Lane 1 of Figure 5 contains one-half of the total RNA contained in lane 2 to enhance resolution of the two lectin mRNAs. Concomitant with the onset of desiccation and seed maturation at around 20 DPA, relative levels of lectin mRNA present in total RNA decreased to low levels (Figure 5, lanes 3 and 4). The dramatic decline in lectin mRNA levels between 10 and 30 DPA embryos is accompanied by a temporal change in the relative levels of the individual mRNA species. In embryos at 10 DPA, the 0.9 kb and 1.1 kb lectin mRNA transcripts are present at similar levels. By 30 DPA, however, the 1.1 kb lectin mRNA accumulates to levels approximately two-fold higher than the 0.9 kb mRNA. Thus, both mRNAs are expressed and differentially regulated in the radicles and coleoptiles of developing embryos.

## DISCUSSION

Two cDNA clones encoding the embryo-specific lectin of rice were isolated and used to characterize the expression of rice lectin in developing embryos at the molecular and cellular levels. The results from these studies have resolved the relationship between the different molecular forms of rice lectin. Previous studies demonstrated that the 23 kd precursor of rice lectin synthesized on the RER is post-translationally processed to an 18 kd polypeptide (Stinissen *et al.*,

**Figure 5. Differential expression of rice lectin mRNA during embryo development.** A Northern blot of total RNA from developing embryos was hybridized with $^{32}$P-labeled insert from cDNA clone cRL852. Lane 1 contains 12.5 ug of total RNA from embryos harvested at 10 DPA. Lanes 2, 3, and 4 contain 25 ug of total RNA from embryos collected at 10, 30, and 40 DPA, respectively. Sizes (kb) of each lectin mRNA are indicated to the right.


10 DPA
30 DPA
40 DPA
- 1.1
- 0.9
1
2
3
4

1984). However, the nature of these processing events which result in the conversion of the 23 kd protein to the 18 kd form remained unidentified. Characterization of two cDNAs, which encode identical 23 kd preproproteins, indicates that processing of both $NH_2$- and COOH-terminal sequences are required to generate the 18 kd rice lectin subunit. As predicted for proteins destined for entry into the endomembrane system of the secretory pathway, the deduced amino acid sequence of both rice lectin cDNA clones contain a 2.6 kd signal sequence which is co-translationally cleaved by an endopeptidase within the lumen of the RER. Mapping of COOH-terminal amino acid residues also implicates the selective removal of a COOH-terminal domain of 2.7 kd containing a potential N-linked glycosylation site to yield the 18 kd polypeptide. Similar COOH-terminal N-glycopeptides have been described for isolectin B of WGA (Raikhel and Wilkins, 1987; Mansfield *et al.*, 1988) and tobacco $\beta$-1,3-glucanase (Shinshi *et al.*, 1988). The primary sequence of the COOH-terminal domains of WGA-B, rice lectin, and $\beta$-1,3-glucanase are not conserved. Interestingly, however, secondary structure predictions of these COOH-terminal domains revealed that these domains are amphipathic $\alpha$-helices. The possible functional role of the COOH-terminal domains of WGA-B and rice lectin in the targeting of these proteins to vacuoles is currently under investigation.

Unlike cereal lectins, the majority of the 18 kd mature subunit of rice lectin is post-translationally processed to two smaller polypeptides of 10 kd and 8 kd (Stinissen *et al.*, 1984). Determination of the $NH_2$-terminal and COOH-terminal

residues of these two polypeptides established the internal cleavage of the 18 kd subunit occurs between asparagine N94 and glycine G95. Moreover, the data show that the 10 kd and 8 kd polypeptides correspond to the $NH_2$- and COOH-terminal portions of the 18 kd subunit, respectively.

DNA sequence and Northern blot analyses established that the two rice lectin cDNA clones, which differ solely in the length of their 3'-untranslated regions, represent two mRNAs which code for the same polypeptide. The occurrence of two transcripts displaying heterogeneous 3'-untranslated regions but encode identical polypeptides is not unprecedented in plants (Messing *et al.*, 1983; Dean *et al.*, 1986; Shinshi *et al.*, 1988). However, these observations have been limited to analysis of genomic sequences or the isolation of multiple related cDNAs exhibiting 3' heterogeneity. Unlike animal systems (Breitbart *et al.*, 1987), a possible correlation between multiple mRNA transcripts and tissue-specific or developmentally regulated genes has not been established in plants. *In situ* hybridization experiments were therefore performed to determine if the cell-specific expression of rice lectin in particular tissues of developing embryos is due to the spatial regulation of the two rice lectin mRNAs. The 1.1 kb lectin mRNA was specifically localized to root caps, peripheral cell-layers of the radicle, coleorhiza and scutellum, and throughout the cell-layers of the coleoptile of developing embryos. In the absence of a probe specific for the 0.9 kb lectin mRNA, however, *in situ* hybridization experiments were unable to localize the expression of this mRNA species to specific embryonic tissues. Furthermore,

sensitivity limits encountered with *in situ* hybridizations mask potential quantitative changes in the temporal expression of the two lectin mRNAs.

Northern blot hybridizations demonstrate conclusively that both lectin mRNAs are temporally and spatially regulated in developing rice embryos. Both lectin mRNAs are actively expressed and accumulate in a cell-specific manner commensurate with the deposition of the protein in developing embryos (Mishkind *et al.*, 1983). The expression of rice lectin in specific cell-layers of embryonic tissues represents the net expression of both the 0.9 kb and 1.1 kb mRNAs. The differential expression of the individual lectin mRNAs is regulated by more than one mechanism. On one level, the expression of rice lectin is developmentally regulated such that both lectin mRNAs are present at high levels during early embryogenesis but decline to low levels during the maturation of the embryo. In addition to the developmental regulation of rice lectin, the two mRNAs are differentially expressed and accumulate to different levels. Although both lectin mRNAs are expressed to the same levels during the rapid synthesis and accumulation of rice lectin at around 10 DPA (Peumans and Stinissen, 1983), the 1.1 kb lectin mRNA is the more prevalent species during the latter stages of embryo development.

Hybridization of the lectin cDNA clones to single restriction fragments on Southern blots suggests that rice lectin is encoded by a single gene. However, this analysis does not preclude the possibility that several genes are involved in the expression of the two rice lectin mRNA transcripts. Operating on the

premise that rice lectin is encoded by a single gene, the two lectin mRNAs may result from alternative polyadenylation site selection during post-transcriptional processing of the pre-mRNA. The temporal regulation of the two rice lectin mRNAs observed during embryogenesis may be the consequence of preferential polyadenylation site selection, differential stability, or selective export of the transcripts from the nucleus.

The lectins of the Gramineae exhibit differential localization in specific cell-layers of the embryo (Mishkind *et al.*, 1983). The most limited distribution of the Gramineae lectins is observed in barley (2n=2x=14) embryos, where the lectin accumulates only in root tissues. In wheat (2n=6x=42), WGA is present not only in embryonic root tissues, but is also localized in the outermost cell layer of the coleoptile. At least three isolectins are responsible for this expression of WGA in hexaploid wheat embryos. Rice (2n=2x=24) embryos exhibit the broadest distribution of lectin with expression in root tissues and throughout the cell-layers of the coleoptile. Thus, the net tissue-specific expression of rice lectin is distinctive from the other lectins of the Gramineae. These results have extended our understanding of the regulation of rice lectin and have facilitated our studies to elucidate the molecular mechanisms regulating the cell-layer specific expression of Gramineae lectins in embryonic tissues.

## REFERENCES

Breitbart, R.E., Andreadis, A., and Nadal-Ginard, B. (1987) Alternative splicing: A ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Ann. Rev. Biochem.* **56,** 467-495.

Chapot, M.-P., Peumans, W.J., and Strosberg, A.D. (1986) Extensive homologies between lectins from non-leguminous plants. *FEBS Lett* **195,** 231-234.

Dale, R.M.K., and Arrow, A. (1987) A rapid single-stranded cloning, sequencing, insertion, and deletion strategy. *Meth. Enzymol.* **155,** 204-231.

Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H., and Bedbrook, J. (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. *Nuc. Acids Res.* **14,** 2229-2240.

Feinberg, A.P., and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132,** 6-13.

Finkelstein, R.R., Crouch, M.L. (1986) Rapeseed embryo development in culture on high osmoticum is similar to that in seeds. *Plant Physiol.* **81,** 907-912.

Francis, D., Kidd, A.D., and Bennett, M.D. (1985) DNA replication in relation to DNA *c* values. In The Cell Division Cycle in *Plants*, J.A. Bryant and D. Frances, eds (Cambridge: Cambridge University Press), pp. 61-82.

Hayashi, R. (1977) Carboxypeptidase Y in sequence determination of peptides. *Methods Enzymol.* **47,** 84-93.

Konigsberg, W.M., and Henderson, L. (1983) Removal of sodium dodecyl sulfate from proteins by ion-pair extraction. *Meth. Enzymol.* **91,** 254-259.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680-685.

Lis, H., and Sharon, N. (1986) Lectins as molecules and as tools. *Ann. Rev. Biochem.* **55,** 35-67.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual.* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Mansfield, M.A., Peumans, W.J., and Raikhel, N.V. (1988) Wheat-germ agglutinin is synthesized as a glycosylated precursor. *Planta* **173,** 482-489.

Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J., and Rubenstein, I. (1983) Plant gene structure. In *Basic Life Sciences*, **Vol. 26,** Genetic Engineering of Plants: An Agricultural Perspective, T. Kosuge, C.P. Meredith and A. Hollaender, eds (New York: Plenum Press), pp. 211-227.

Mishkind, M.L., Palevitz, B.A., Raikhel, N.V., and Keegstra, K. (1983) Localization of wheat germ agglutinin-like lectins in various species of the Gramineae. *Science* **220,** 1290-1292.

Mizusawa, S., Nishimura, S., and Seela, F. (1986) Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nuc. Acids Res.* **14,** 1319-1324.

Peumans, W.J., and Stinissen, H.M. (1983) Gramineae lectins: Occurrence, molecular biology and physiological function. In *Chemical Taxonomy, Molecular Biology, and Function of Plant Lectins*, I.J. Goldstein and M.E. Etzler, eds (New York: Alan R. Liss, Inc.), pp. 99-116.

Raikhel, N.V., Mishkind,M.L., and Palevitz, B.A. (1984) Characterization of a wheat germ agglutinin-like lectin from adult wheat plants. *Planta* **162,** 55-61.

Raikhel, N.V., Bednarek, S.Y., and Lerner, D.R. (1989) *In situ* RNA hybridization in plant tissues. In *Plant Molecular Biology Manual*, S.B. Gelvin and R.A. Schilperoort, eds (Boston: Kluwer Academic Publishers), **B 9,** 1-32.

Raikhel, N.V., Bednarek, S.Y., and Wilkins, T.A. (1988) Cell-type specific expression of a wheat-germ agglutinin gene in embryos and young seedlings of *Triticum aestivum*. *Planta* **126,** 406-414.

Raikhel, N.V., and Wilkins, T.A. (1987) Isolation and characterization of a cDNA clone encoding wheat germ agglutinin. *Proc. Natl. Acad. Sci. USA* **84,** 6745-6749.

Sanger, F., Nicklen, S., and Coulsen, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **56,** 5463-5467.

Shinshi, H., Wenzler, H., Neuhaus, J-M., Felix, G., Hofsteenge, J., and Meins, F. Jr. (1988) Evidence for N- and C-terminal processing of a plant defense-related enzyme: Primary structure of tobacco prepro-$\beta$-1,3-glucanase. *Proc. Natl. Acad. Sci. USA* **85,** 5541-5545.

Shure, M., Wessler, S., and Fedoroff, N. (1983) Molecular identification and isolation of the *waxy* locus in maize. *Cell* **35,** 225-233.

Stinissen, H.M., Peumans, W.J., and Chrispeels, M.J. (1984) Subcellular site of lectin synthesis in developing rice embryos. *EMBO J.* **3,** 1979-1985.

Struhl, K. (1985) A rapid method for creating recombinant DNA molecules. *BioTechniques* **3,** 452-453.

Tsuda, M. (1979) Purification and characterization of a lectin from rice bran. *J. Biochem.* **86,** 1451-1461.

Vieira, J., and Messing, J. (1987) Production of single-stranded plasmid DNA. *Meth. Enzymol.* **153,** 3-11.

von Heijne, G. (1982) Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133,** 17-21.

## Chapter 4

### Role of propeptide glycan in post-translational processing and transport of barley lectin to vacuoles in transgenic tobacco


## ABSTRACT

Mature barley lectin is a dimeric protein comprised of two identical 18 kd polypeptides. The subunits of barley lectin are initially synthesized as glycosylated proproteins which are post-translationally processed to the mature protein preceding or concomitant with deposition of barley lectin in vacuoles. To investigate the functional role of the glycan in processing and intracellular transport of barley lectin to vacuoles, the sole N-linked glycosylation site residing within the COOH-terminal propeptide of barley lectin was altered by site-directed mutagenesis. cDNA clones encoding wild-type (*wt*) or glycosylation-minus (*gly*) barley lectin preproproteins were placed under the transcriptional control of the cauliflower mosaic virus 35S promoter and introduced into *Nicotiana tabacum* cv. Wisconsin 38. Barley lectin synthesized from both the *wt* or *gly* constructs is processed and correctly targeted to vacuoles of tobacco leaves. Localization of barley lectin in vacuoles processed from the nonglycosylated *gly* proprotein indicates that the high mannose glycan of the barley lectin proprotein is not essential for targeting barley lectin to vacuoles. However, pulse-chase labeling and monensin experiments demonstrated that the glycosylated *wt* proprotein and the nonglycosylated *gly* proprotein are differentially processed to the mature protein and transported from the Golgi complex with distinctive

and characteristic kinetics. These results implicate an indirect functional role for the glycan in post-translational processing and transport of barley lectin to vacuoles.

## INTRODUCTION

Many proteins entering the endomembrane system of the secretory pathway are modified in the lumen of the rough endoplasmic reticulum (RER) by the covalent attachment of high-mannose oligosaccharide sidechains (glycans) to selected asparagine (N) residues. The N-linked high-mannose glycans may be subsequently modified to complex glycans as the glycoprotein traverses through the Golgi complex. Inhibition of glycosylation by site-directed mutagenesis or the drug tunicamycin apparently does not affect the synthesis, intracellular transport, or function of some glycoproteins (reviewed in Olden *et al.*, 1985). However, the N-linked glycans of other glycoproteins have been shown to influence protein folding (Machamer and Rose, 1988; Matzuk and Boime, 1988), oligomerization (Matzuk and Boime, 1988), stability (reviewed in Olden *et al.*, 1985), and protein targeting (Kornfeld, 1986).

Studies exploring the functional role of N-linked oligosaccharides in plants have been limited. Proteins modified by N-linked glycosylation may be localized within a subcellular compartment or in the cell wall. The glycans of the vacuolar protein phytohemagglutinin (PHA) and the secreted $\alpha$-amylase of rice, however, are not required for transport and targeting of these proteins to their respective

compartments (Bollini, *et al.*, 1985; Voelker, *et al.*, 1989; Akazawa and Hara-Nishimura, 1985). In fact, many vacuolar and secretory proteins are not glycoproteins, suggesting that N-linked oligosaccharide side-chains do not generally function as sorting signals. A functional role for the glycan of the vacuolar protein concanavalin A (ConA), however, has been implicated in the intracellular processing and transport of this protein (Faye and Chrispeels, 1987). Mature ConA is not a glycoprotein, although it is synthesized as a glycosylated precursor (proConA) (Herman, *et al.*, 1985). The mature ConA polypeptide is generated by the excision of an internal glycopeptide from proConA and subsequent ligation of the two resultant polypeptides (Bowles, *et al.*, 1986). Inhibition of N-linked glycosylation with the inhibitor tunicamycin significantly impedes transport of proConA from the ER/Golgi compartment to vacuoles (Faye and Chrispeels, 1987).

The post-translational processing of Gramineae lectins, which are soluble vacuolar proteins, is distinctive from PHA and ConA. The mature lectins of wheat, barley, and rice are 36 kd dimers assembled from two identical 18 kd subunits (Rice and Etzler, 1974; Peumans, *et al.*, 1982a; Peumans, *et al.*, 1983). Similar to ConA, mature Gramineae lectins are not glycoproteins. However, the lectin subunits are initially synthesized as glycosylated proproteins in wheat (Mansfield, *et al.*, 1988), barley (Lerner and Raikhel, 1989) and rice (Wilkins and Raikhel, 1989; unpublished results). The sole N-linked glycosylation site (Asn-X-Ser/Thr) resides within the propeptide located at the COOH-terminal of these

proproteins. Endo-$\beta$-N-acetylglucosaminidase H (Endo H) studies have demonstrated that the oligosaccharide side-chain of these proproteins is a high mannose glycan with a molecular weight of approximately 2 kd (Lerner and Raikhel, 1989; Smith and Raikhel, 1989; Wilkins and Raikhel, unpublished results). The COOH-terminal N-glycopeptide of the proprotein is post-translationally removed prior to or concomitant with deposition of the mature protein in vacuoles. The transient glycosylation of the Gramineae lectin proproteins provides a unique opportunity to investigate the molecular mechanisms which mediate the maturation and targeting of mature lectins to vacuoles. In this study, we have examined the synthesis, assembly, processing, and subcellular localization of barley lectin in transgenic tobacco. In addition, the functional role of the barley lectin propeptide glycan was assessed by introducing a mutant barley lectin cDNA into tobacco. The N-linked glycosylation site within the COOH-terminal propeptide in the mutant barley lectin cDNA was modified by site-directed mutagenesis to prevent the co-translational N-glycosylation of the barley lectin proprotein. The results establish that both the wild-type and mutant barley lectin are expressed, correctly processed, and transported to vacuoles of tobacco leaves. However, the kinetics of post-translational processing through the RER/Golgi complex is distinctive for the wild-type or mutant barley lectin proproteins. Moreover, expression of wild-type and mutant barley lectin in the presence of the inhibitor monensin has provided

additional insight into the post-translational processing and sorting of barley lectin to vacuoles.

## MATERIALS AND METHODS

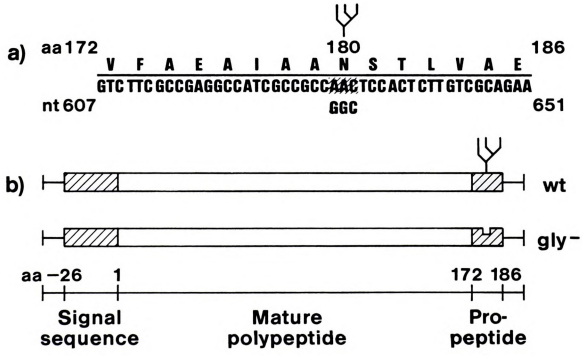
***Modification of barley lectin cDNA flanking regions.*** The EcoRI sites flanking the 972 bp cDNA clone (pBLc3) encoding barley lectin (Lerner and Raikhel, 1989) were blunt-ended using DNA Polymerase I Klenow fragment as described in Maniatis, *et al.*, 1982. Following the addition of XbaI phosphorylated linkers (Maniatis *et al.*, 1982), the cDNA was purified from low-melting point agarose and subcloned (Struhl, 1985) into pUC118 (Vieira and Messing, 1987). Restriction mapping of the cDNA revealed that the EcoRI sites originally flanking the barley lectin cDNA were restored by the addition of XbaI linkers.

***Site-directed mutagenesis.*** The N-linked glycosylation site at $Asn_{180}$-Ser-$Thr_{182}$ in the COOH-terminal propeptide of the barley lectin proprotein (Lerner and Raikhel, 1989) was altered by converting the $Asn_{180}$ (AAC) to a $Gly_{180}$ (GGC) residue by the site-directed mutagenesis method of Kunkel, *et al.* (1987) (see Figure 1A). Site-directed mutagenesis of the barley lectin propeptide was performed using Bio-Rad's Muta-Gene phagemid *in vitro* mutagenesis kit with a mutagenic 16-base synthetic oligonucleotide spanning amino acids $Ala_{178}$ to $Thr_{182}$ (Lerner and Raikhel, 1989) and uracil-containing single-strand DNA prepared in the *dut ung E. coli* strain CJ236. Mutants encoding the altered

**Figure 1. Alteration of the N-linked glycosylation site of barley lectin by site-directed mutagenesis and organization of the wild-type (*wt*) and mutant (*gly*) barley lectin cDNAs introduced into tobacco.**

(A) The 15 amino acid COOH-terminal propeptide of barley lectin (aa 172 through aa 186) and the corresponding nucleotide sequence (nt 607 to 651). The N-linked glycosylation site ($Asn_{180}$-Ser-$Thr_{182}$) is depicted by attachment of a high mannose glycan tree to Asn (N) residue 180. The N-linked glycosylation site at $N_{180}$ (shaded codon) was converted to a Gly residue (GGC) by site-directed mutagenesis to generate a barley lectin mutant which cannot be glycosylated.

(B) The structure of *wt* and *gly* barley lectin cDNA clones subcloned into the plant expression vector pGA643 (An, *et al.*, 1988).


a)
aa 172
180
186
V F A E A I A A N S T L V A E
GTC TTC GCC GAG GCC ATC GCC GCC AAC TCC ACT CTT GTC GCA GAA
nt 607
GGC
651
b)
wt
gly−
aa −26
1
172 186
Signal sequence
Mature polypeptide
Pro-peptide

tripeptide $Gly_{180}$-Ser-$Thr_{182}$ were identified and selected by $^{35}S$-dideoxy sequencing (Sanger *et al.*, 1977) of single-strand DNA prepared from phagemids in the *dut⁺ung⁺ E. coli* strain MV1193.

***Plant transformation.*** Both mutated (*gly*) and wild-type (*wt*) barley lectin cDNAs were excised from pUC118 with XbaI and subcloned (Struhl, 1985) into the binary plant expression vector pGA643 (An, *et al.*, 1988). These binary vector constructs were mobilized from the *E. coli* strain DH5$\alpha$ into *Agrobacterium tumefaciens* LBA4404 by triparental mating (Hooykaas, 1988) using the *E. coli* strain HB101 harboring the wide-host range mobilizing plasmid pRK2013. Transconjugates were selected on minimal nutrient plates (An, *et al.*, 1988) containing streptomycin (200 ug/ml), kanamycin (25 ug/ml) and tetracycline (5 ug/ml).

*Agrobacteria* cells containing the *wt* and *gly* barley lectin cDNAs were introduced into tobacco plants (*Nicotiana tabacum* cv. Wisconsin 38) by the leaf disc transformation method of Horsch, *et al.* (1988). The leaf discs were cocultivated with the *Agrobacteria* for 48 hrs on MS104 plates prior to transfer to MS selection media (Horsch, *et al.*, 1988). After several weeks, shoots were transferred to MS rooting media (Horsch, *et al.*, 1988). At least three independent transformants, maintained as axenic cultures, were subsequently analyzed for each construct.

***Nucleic acid analysis.*** Total DNA was isolated from leaf tissue of untransformed and transgenic tobacco plants according to Shure *et al.* (1983). DNA (12 ug) was restricted with HindIII and fractionated on 1.0% agarose gels prior to transfer to nitrocellulose (Maniatis, *et al.*, 1982). Nitrocellulose filters were hybridized with $^{32}P$ random-primer-labeled (Feinberg and Vogelstein, 1983) BLc3 barley lectin cDNA (Lerner and Raikhel, 1989) as described previously (Raikhel, *et al.*, 1988). For gene reconstruction experiments, tobacco genomic DNA was restricted with EcoRI and BLc3 titered at 0.5-, 1.0-, 3.0-, and 5.0-copy equivalents per tobacco (*N. tabacum*) genome ($4.8 \times 10^9$ bp per haploid genome; Zimmerman and Goldberg, 1977). Gene reconstruction blots were hybridized with a HindIII-SalI restriction fragment from a pGA643 construct containing BLc3 radiolabeled by the random-primer method (Feinberg and Vogelstein, 1983). Filters were exposed to Kodak X-OMAT AR film at -70°C with intensifying screens.

Total RNA was isolated from leaves of untransformed and transgenic tobacco plants as described previously (Wilkins and Raikhel, 1989). Total RNA (25 ug) from each construct was resolved in a 2% agarose/6% formaldehyde gel, transferred to nitrocellulose, and hybridized with the BLc3 cDNA labeled with $^{32}P$ as described above.

***Protein extraction, affinity chromatography, immunoblots, and ELISA.*** Barley lectin was purified from acid-soluble protein extracts by affinity chromatography

on immobilized N-acetylglucosamine columns from transgenic tobacco leaves (500 mg) essentially as described in Mansfield *et al.* (1988) with the exception that the homogenization buffer consisted of 50 mM HCl containing 1 mM phenylmethylsulfonyl fluoride (PMSF). The affinity-purified lectin was carboxyamidated (Raikhel, *et al.*, 1984), fractionated by SDS-PAGE (Mansfield, *et al.*, 1988), and electroblotted onto nitrocellulose (Towbin, *et al.*, 1979). Barley lectin was detected using anti-WGA polyclonal antiserum (Mansfield, *et al.*, 1988) and protein A-alkaline phosphatase as described in Blake, *et al.* (1984) using nitroblue tetrazolium as the substrate.

Extracts of acid soluble proteins were assayed using double-bind ELISA (Raikhel, *et al.*, 1984) to quantitate the amount of barley lectin in transgenic tobacco leaves. Crude extracts were prepared by homogenization of tobacco leaves (1.0 g) in 2 ml 50 mM Tris-acetate, pH 5.0, 100 mM NaCl, 1 mM PMSF. The extracts were clarified by centrifugation at 10 krpm for 10 min to remove cellular debris and insoluble material. Barley lectin was detected in crude extracts using guinea pig anti-WGA antiserum and rabbit anti-WGA IgGs conjugated to alkaline phosphatase (Raikhel, *et al.*, 1984). A standard curve, constructed from affinity-purified WGA (E-Y Labs), was used to estimate the level of barley lectin in tobacco leaves. Total protein in the crude extracts was determined by the method of Bradford (1976).

***Vacuole isolation and enzyme assays.*** Protoplasts for vacuole isolation were prepared from leaves of axenic cultured plants. Leaves were digested overnight in an enzyme medium composed of 0.5 M mannitol and 3 mM MES, pH 5.7 containing the same enzymes as described below. Vacuoles were isolated from tobacco protoplasts by ultracentrifugation as described in Guy *et al.* (1979) with the exception that the isolation buffer was 0.5 M sorbitol and 10 mM HEPES, pH 7.2 and the Ficoll step gradient consisted of 10% and 5% Ficoll. Vacuoles stained with neutral red were collected from the 0%/5% interface and adjusted to 10% Ficoll and subjected to further purification on a second Ficoll gradient. Vacuoles were collected from the 0%/5% Ficoll interface on a second gradient by flotation of the vacuoles during centrifugation. The vacuoles recovered were counted in a hemocytometer, frozen in liquid nitrogen, and stored at -80°C for biochemical analysis.

Vacuolar-specific enzyme activities of $\alpha$-mannosidase (Boller and Kende, 1979) and acid phosphatase (Shimomura, *et al.*, 1988) were assayed in protoplast and vacuole fractions by monitoring the release of p-nitrophenol spectrophotometrically from the appropriate substrates. Catalase activity (Aebi, 1974) was measured in protoplast and vacuole fractions as an extravacuolar enzyme marker.

***Immunocytochemistry.*** Leaf tissue from axenic tobacco plants was excised and trimmed into 2 $mm^2$ pieces. Fixation and immunocytochemistry was performed essentially as described in Mansfield *et al.* (1988).

***Radiolabeling of tobacco protoplasts, Endo H digestion, and monensin.*** Protoplasts for labeling were prepared from fully expanded leaves of axenically cultured tobacco plants. Leaves were digested overnight in an enzyme mixture comprised of 0.5% cellulase (Onozuka R10), 0.25% macerozyme R10, and 0.1% BSA in MSA media (An, *et al.*, 1988) supplemented with 0.5 M mannitol. The yield of protoplasts was quantitated using a hemocytometer counting chamber.

For pulse-labeling experiments, 1 X $10^5$ leaf protoplasts (per well) were incubated in a 24-well Falcon tissue culture plate in 500 ul MSA media in 0.5 M mannitol supplemented with 48 uCi of $^{35}S$-Trans label (ICN $^{35}S$ *E. coli* hydrolysate labeling reagent containing $\geq$ 70% L-methionine and $\leq$ 15% L-cysteine; 1000-1200 Ci/mmole). The culture plates were incubated in the dark at room temperature with gentle shaking. Two wells or a total of 200,000 protoplasts were labeled for each experiment. Pulse-chase experiments were performed by supplementing the media with 1 mM L-methionine and 0.5 mM L-cysteine 8 to 10 hr after pulse-labeling protoplasts as described above. Following labeling, protoplasts were pooled and collected by centrifugation at 2 krpm for 15 sec at 4°C. The resulting protoplast pellet was suspended in 100 ul of 50 mM Tris-acetate, pH 5.0, 100 mM NaCl and lysed at room temperature for 10 min with

gentle agitation following the addition of 100 ul of 1.2 mM dithiothreitol and 1.2% (v/v) Triton X-100 in Tris-acetate/NaCl. Samples were frozen in liquid $N_2$ and stored at -70°C.

Endo-$\beta$-N-acetylglucosaminidase H (Endo H) digestion of radiolabeled barley lectin was performed at 37°C for 23 hr in 50 mM Tris-acetate, pH 5.5, 100 mM NaCl, 1 mM PMSF with 4 mU Endo H immediately following affinity-purification of barley lectin from protoplasts pulse-labeled for 12 hr as described above.

For monensin experiments, 500 mM monensin in absolute ethanol was added directly to each well containing tobacco protoplasts to a final concentration of 50 uM monensin, 0.1% ethanol. Absolute ethanol was added to a final concentration of 0.1% in controls. A total of 800,000 *wt* or *gly* protoplasts were pretreated in the presence of ethanol or monensin for 1 hr prior to the addition of $^{35}$S-trans label and pulse-labeled for 12 hr. A fraction of the pulse-labeled protoplasts (200,000) were processed as described above. The remaining 600,000 protoplasts were pooled and gently homogenized in 150 ul of 100 mM Tris pH 7.8, 1 mM EDTA, 12% sucrose (w/w) and separated into soluble and organelle fractions on Sepharose 4B columns (8.0 cm X 1.0 cm) according to Stinissen *et al.* (1985). The Sepharose 4B elution profile of total radioactivity associated with organelles and soluble proteins concurred with previous studies (Stinissen, *et al.*, 1984; Stinissen, *et al.*, 1985; Mansfield, *et al.*, 1988). In addition, NADH cytochrome C reductase activity (Lord, 1983) was

primarily associated with the organelle fractions. The samples were adjusted to 0.5% Triton X-100 and stored at -70°C. Following collection of protoplasts by centrifugation (see above), the culture media was recovered from a total of 800,000 protoplasts and contaminating intact protoplasts removed by gravity filtration through a Isolab quick-sep column fitted with a paper filter and a Whatman GF/C glass fiber filter (1.2 um exclusion). Proteins contained in the culture media were precipitated with ammonium sulfate at 60% saturation at 4°C for at least 2 hr. Precipitated proteins were collected by centrifugation for 10 min at 15 krpm. The protein pellet was resuspended in 200 ul of 50 mM Tris-acetate, pH 5.0, 100 mM NaCl and stored at -70°C. $^{35}$S-labeled barley lectin was purified by affinity chromatography, carboxyamidated and analyzed by SDS-PAGE as described above. The SDS-PAGE gels were treated for fluorography as detailed in Mansfield, *et al.* (1988).

## RESULTS

***Inactivation of N-linked glycosylation site of barley lectin proprotein by site-directed mutagenesis.*** The cDNA clone pBLc3 (Lerner and Raikhel, 1989) encodes the 23 kd preproprotein of barley lectin. The preproprotein is comprised of a 2.5 kd signal sequence, the 18 kd mature protein, and a 1.5 kd COOH-terminal propeptide (Figure 1A). In barley embryos, the proprotein is modified by the addition of a 2 kd high-mannose oligosaccharide sidechain to the sole N-linked glycosylation site located within the COOH-terminal propeptide

at $Asn_{180}$-Ser-$Thr_{182}$ (Figure 1B). To further investigate the assembly, post-translational processing and transport of barley lectin to vacuoles, the cDNA encoding barley lectin was introduced into tobacco and the post-translational processing of monocot barley lectin examined in this heterologous dicot system. The barley lectin cDNA was subcloned into the binary plant expression vector pGA643 (An, *et al.*, 1988) under the transcriptional control of the cauliflower mosaic virus (CaMV) 35S promoter. *Agrobacteria*-mediated transformation of tobacco (*Nicotiana tabacum* cv. Wisconsin 38) was accomplished via the leaf disc method of Horsch, *et al.* (1988). Both the constructs and kanamycin-resistant tobacco transformants containing the barley lectin cDNA are designated by the code *wt* (Figure 1B).

The glycosylated COOH-terminal propeptide is transiently associated with barley lectin proprotein but not with the nonglycosylated mature protein localized in vacuoles. Mature barley lectin is generated by the cleavage of the N-linked glycosylated propeptide from the proprotein preceding or concomitant with the deposition of barley lectin in vacuoles. To assess the functional role of the N-linked high mannose glycan in the assembly, processing, and targeting of barley lectin to vacuoles, site-directed mutagenesis was performed to alter the N-linked glycosylation site within the COOH-terminal propeptide. The N-linked glycosylation site was altered by converting $Asn_{180}$ (AAC) to a $Gly_{180}$ (GGC) residue using a 16-base mutagenic synthetic oligonucleotide spanning the glycosylation site at $Asn_{180}$-Ser-$Thr_{182}$ (Figure 1A). The mutant barley lectin cDNA

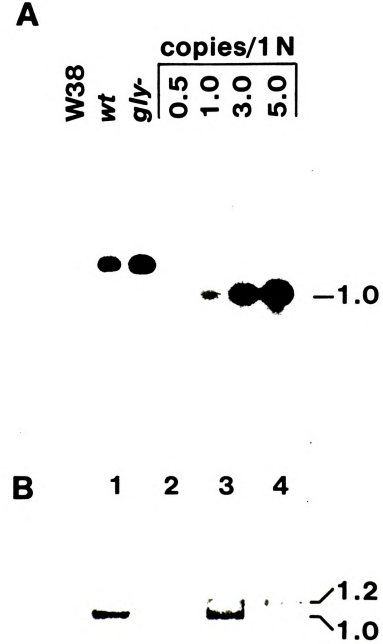
was subcloned into pGA643 and transformed into tobacco. Constructs and kanamycin-resistant tobacco plants containing the mutant barley lectin are designated as *gly* (Figure 1B).

***Detection of barley lectin cDNA and mRNA in transgenic tobacco.*** The structure and stable integration of *wt* and *gly* barley lectin cDNA into the tobacco genome was examined in independent transformants by Southern blot analysis. A radiolabeled restriction fragment containing a portion of the barley lectin cDNA and the T-DNA left border of pGA643 was used to probe tobacco genomic DNA restricted with HindIII. Three HindIII-restriction fragments (5 kbp to 9.0 kbp) and five fragments (18 kbp and 2.8 kbp to 4.0 kbp) were detected in tobacco genomic DNA isolated from *wt* and *gly* transformants, respectively (data not shown). Gene reconstruction experiments (Figure 2A) were performed with EcoRI-restricted tobacco DNA and purified BLc3 insert titered at 0.5-, 1.0-, 3.0-, and 5-copy equivalents per tobacco genome. Hybridization of gene reconstruction experiments with radiolabeled BLc3 indicate the presence of 3-copies of *wt* and 5-copies of *gly* barley lectin cDNA integrated into the tobacco genome of the individual transformants presented in Figure 2A. No hybridization was observed between barley lectin and tobacco DNA in untransformed plants (W38, Figure 2A) or in transgenic plants containing only the vector pGA643 (data not shown).

**Figure 2. Gene reconstruction analysis and accumulation of steady-state RNA levels of barley lectin in transgenic tobacco.**

**(A)** Southern blot containing 12 ug of tobacco genomic DNA restricted with EcoRI and probed with radiolabeled barley lectin pBLc3 cDNA insert (Lerner and Raikhel, 1989). Reconstruction lanes represent 0.5-, 1.0-, 3.0-, and 5.0-gene copy equivalents of barley lectin pBLc3 cDNA insert per haploid genome of tobacco. Tobacco DNA was isolated from untransformed tobacco (cv. W38) and transgenic tobacco plants containing cDNAs encoding wild-type (*wt*) or mutant (*gly*) barley lectin preproproteins. Approximate size of fragments in kbp are positioned to the right.

**(B)** Northern blot contiaining 25 ug of total RNA isolated from developing barley embryos (**lane 1**), untransformed tobacco (cv. W38) (**lane 2**) and transgenic tobacco plants containing *wt* (**lane 3**) or *gly* (**lane 4**) barley lectin cDNA constructs. The size of barley lectin mRNA species is indicated in kb to the right of the blot.

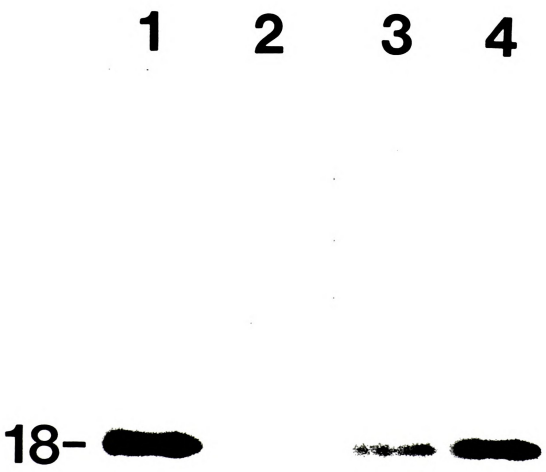

A
copies/1 N
W38
wt
gly-
0.5
1.0
3.0
5.0
—1.0
B
1
2
3
4
1.2
1.0

The relative levels of mRNA encoding *wt* or *gly* barley lectin in transgenic tobacco was investigated by Northern blot analysis. The Northern blot in Figure 2B represents the accumulation of *wt* and *gly* barley lectin steady-state mRNA in total RNA isolated from transgenic tobacco leaves detected by $^{32}$P-labeled barley lectin cDNA (BLc3). Two mRNA species of 1.2 kb and 1.0 kb were observed in tobacco transformants containing either the *wt* or *gly* barley lectin (Lanes 3 and 4, respectively, Figure 2B). The 1.0 kb barley lectin mRNA in tobacco transformants (Lanes 3 and 4, Figure 2B) corresponds in length to the 1.0 kb barley lectin mRNA in developing barley embryos (Lane 1, Figure 2B; Lerner and Raikhel, 1989). The 1.2 kb mRNA species is unique to transgenic tobacco plants and presumably represents utilization of an alternate polyadenylation site contained within the termination sequences of the plant expression vector pGA643 (An *et al.*, 1988). Examination of individual transformants revealed the differential accumulation of the 1.2 kb and 1.0 kb lectin mRNAs in both *wt* and *gly* plants (Lanes 3 and 4, Figure 2B; data not shown). However, densitometer scanning of the autoradiograph indicated that the overall accumulation of steady-state *wt* and *gly* barley lectin mRNAs are very similar. No hybridization was observed in total RNA isolated from transgenic plants containing only the vector pGA643 (data not shown) or in untransformed tobacco (Lane 2, Figure 2B) probed with barley lectin cDNA.

***Expression and assembly of active barley lectin in tobacco.*** Gramineae lectins possess the ability to specifically bind oligomers of the carbohydrate N-acetylglucosamine (GlcNAc). Since the carbohydrate binding site of WGA is comprised of amino acids contributed by both monomeric subunits (Wright, 1980), the assembly of active WGA is therefore contingent upon the formation of the dimer. Barley lectin shares 95% amino acid homology with WGA, including conservation of amino acids involved in carbohydrate binding (Lerner and Raikhel, 1989). This conservation is exemplified by the ability to form active heterodimers *in vitro* from monomeric subunits of WGA and barley lectin (Peumans, *et al.*, 1982b). Hence, the mechanism of dimerization and carbohydrate binding of WGA and barley lectin are presumably identical.

In order to determine if barley lectin is synthesized and assembled into an active lectin in transgenic tobacco plants, crude protein extracts prepared from *wt* or *gly* tobacco transformants were fractionated on an immobilized GlcNAc affinity matrix. The affinity-purified fractions were separated by SDS-PAGE and analyzed by immunoblotting (Figure 3). Since barley lectin and WGA are antigenically indistinguishable (Stinissen, *et al.*, 1983), polyclonal anti-WGA antiserum was used to detect barley lectin on immunoblots. The 18 kd mature subunit of barley lectin was readily discernible in wt or *gly* transgenic tobacco leaves (Lanes 3 and 4, respectively, Figure 3). Detection of mature 18 kd polypeptides on immunoblots following affinity chromatography (Figure 3) indicates that barley lectin is synthesized and assembled as an active

**Figure 3. Immunoblot detection of mature barley lectin in *wt* and *gly* tobacco transformants.** Acid soluble protein extracts from *wt* (**lane 3**) and *gly* (**lane 4**) transformed and untransformed (**lane 2**) tobacco leaves were concentrated by ammonium sulfate precipitation. Barley lectin was affinity purified as described in Materials and Methods, separated on SDS-PAGE, and electroblotted onto nitrocellulose. Immunodetection of barley lectin was performed with polyclonal anti-WGA antiserum and protein A-conjugated alkaline phosphatase. **Lane 1** is a control lane containing 1 ug of purified WGA. The molecular mass of mature WGA and barley lectin subunits in kd is shown on the left.


1
2
3
4
18-

GlcNAc-binding lectin in both *wt* and *gly* tobacco transformants. Anti-WGA antiserum does not cross-react with any polypeptide in untransformed tobacco (Lane 2, Figure 3). Similar results were obtained on immunoblots prepared from roots of *wt* and *gly* transgenic tobacco plants (results not shown).

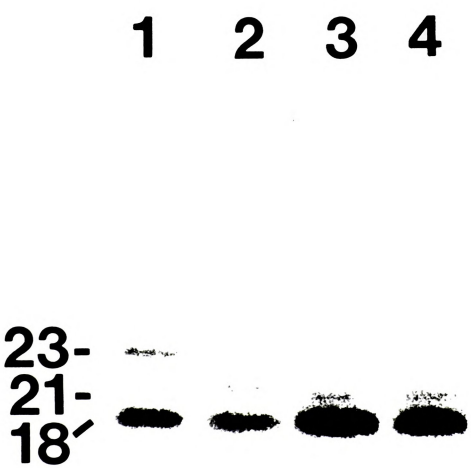
The accumulation of barley lectin in *wt* and *gly* tobacco plants was quantitated in total acid soluble protein extracts from transgenic tobacco leaves using double-bind ELISA. A range of 800 ng to 2 ug of affinity-purified barley lectin per 1 g/fw leaf tissue can be recovered from *wt* and *gly* tobacco transformants. The accumulation of barley lectin in tobacco leaves corresponds to 0.2% to 0.5% of total acid-soluble leaf proteins.

***Synthesis of wild-type (wt) and mutant (gly) barley lectin proproteins in tobacco protoplasts.*** In barley embryos, barley lectin is initially synthesized as a 23 kd glycosylated proprotein (Stinissen, *et al.*, 1985; Lerner and Raikhel, 1989). To ensure that barley lectin is synthesized and processed by similar mechanisms in tobacco, the post-translational modifications of radiolabeled barley lectin precursors in transgenic tobacco were examined. Tobacco protoplasts were prepared from axenic cultures and pulse-labeled for 12 hr in the presence of $^{35}$S-Trans label. Radiolabeled barley lectin was recovered from tobacco protoplasts by affinity chromotography on immobilized GlcNAc columns. Following affinity chromatography, eluant fractions were treated with Endo H, an enzyme which specifically cleaves high mannose oligosaccharide side-chains

between the GlcNAc residues of the glycan core. Radiolabeled proteins incubated in the presence or absence of Endo H were analyzed following separation by SDS-PAGE and fluorography (Figure 4). In addition to the mature 18 kd subunit, a 23 kd polypeptide was also evident in pulse-labeled *wt* tobacco protoplasts (Lane 1, Figure 4). The majority of the 23 kd polypeptide is converted to a 21 kd protein following treatment with Endo H (Lane 2, Figure 4), indicating that the 23 kd polypeptide contains a 2 kd high mannose glycan. These results infer that the signal sequence has been cleaved and that the synthesis and processing of barley lectin precursors in tobacco is analogous to processing mechanisms in barley. As expected, the *gly* barley lectin is synthesized as a 21 kd proprotein (Lane 3, Figure 4) which is resistant to Endo H (Lane 4, Figure 4). Comparison of the *wt* and *gly* 21 kd polypeptides (Lanes 2 and 4, respectively, Figure 4) shows a slight disparity in migration of these two proproteins. The slower migration of the *wt* 21 kd polypeptide is due to the presence of a GlcNAc residue (M, 221.2), which remains attached to $Asn_{180}$ of the propeptide after enzymatic deglycosylation with Endo H. Thus, both the *wt* and *gly* barley lectins are synthesized as the predicted glycosylated 23 kd and nonglycosylated 21 kd proproteins, respectively, and processed to 18 kd mature polypeptides similarly in transgenic tobacco and barley.

***Subcellular localization of wt and gly barley lectin in vacuoles.*** Barley lectin is localized in vacuoles in the peripheral cell-layers of embryonic and adult root

**Figure 4. Endo H digestion of radiolabeled barley lectin isolated from transgenic tobacco.** Radiolabeled barley lectin was affinity purified from *wt* (**lanes 1 and 2**) and *gly* (**lanes 3 and 4**) tobacco protoplasts pulse-labeled for 12 hr. Duplicate samples were incubated at 37 C for 23 hr in the absence (**lanes 1 and 3**) or presence (**lanes 2 and 4**) of Endo H. Samples were lyophilized and separated by SDS-PAGE. The position and molecular mass (kd) of barley lectin *wt* and *gly* proproteins (23 and 21 kd, respectively) and mature barley lectin (18 kd) are indicated to the left.


1
2
3
4
23-
21-
18-

caps of barley (Mishkind, *et al.*, 1983; Lerner and Raikhel, 1989). The subcellular location of *wt* and *gly* barley lectin in transgenic tobacco was ascertained by a combination of organelle fractionation, immunoblot analysis and EM immunocytochemistry. Protoplasts were prepared from both *wt* and *gly* transgenic tobacco plants (Figure 5A). Vacuoles were released from protoplasts (Figure 5B) and purified by centrifugation on a discontinuous ficoll gradient system. The purity of the vacuole preparation was evaluated by determining the enzymatic activity of two vacuolar-specific enzymes (acid phosphatase and $\alpha$-mannosidase) and a peroxisomal enzyme (catalase) in vacuoles and protoplasts. Catalase was employed as an extravacuolar enzyme marker for two reasons. One, peroxisomes are very fragile and consequently lyse during preparation of vacuoles, thereby liberating catalase into the cell lysate. Secondly, the high specific activity of catalase is readily detectable at very low concentrations in cell lysates. As shown in Table 1, the relative enzymatic activity of the vacuolar enzyme markers in vacuoles isolated from *wt* or *gly* protoplasts approaches 100%. Less than 2% of catalase activity is associated with the vacuoles, indicating that the vacuoles (Figure 5B) are essentially free of contaminating cytosol and unbroken protoplasts.

Protoplast and vacuole fractions from *wt* and *gly* tobacco plants were examined for the presence of barley lectin by immunoblot analysis. Barley lectin was purified by affinity chromatography from a protein lysate representing an equivalent number of *wt* or *gly* protoplasts and vacuoles and analyzed on

**Figure 5. Isolation of vacuoles from tobacco protoplasts expressing *wt* or *gly* barley lectin.**

**(A)** Protoplasts were prepared by enzymatic digestion of tobacco leaves collected from axenically cultured transgenic plants. Bar=10 um.

**(B)** Vacuoles stained with neutral red were isolated from tobacco protoplasts by centrifugation on a discontinuous 5%/10% Ficoll step gradient. Stained vacuoles were collected from the 0%/5% Ficoll interface and purified on a second 5%/10% Ficoll step gradient. Bar=10 um.

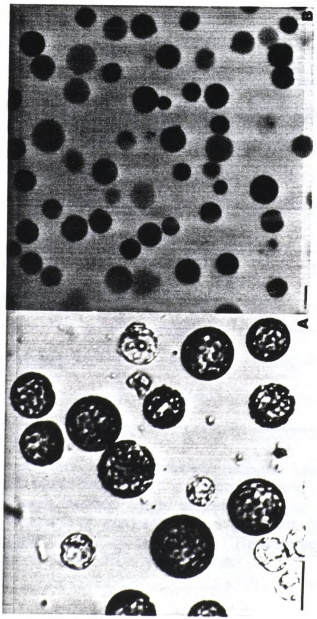


Table 1. Relative enzyme activity (%) in vacuoles prepared from transgenic tobacco protoplasts

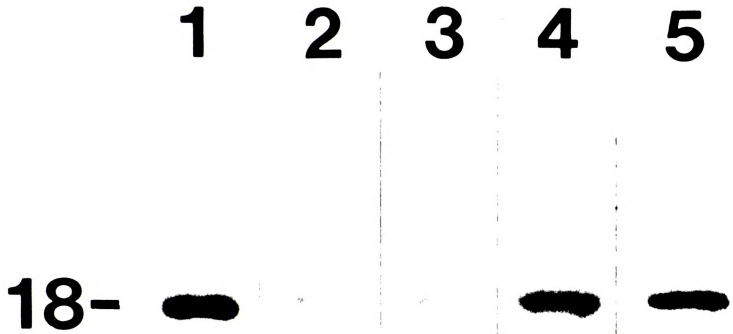
| | *wt* | *gly*$^-$ |
|---|---|---|
| Vacuole-specific enzymes | | |
| α-mannosidase | 106.8 ± 8.1 | 102.5 ± 1.5 |
| acid phosphatase | 85.6 ± 6.8 | 98.6 ± 10.3 |
| Extravacuolar enzyme | | |
| catalase | < 2.0 | < 2.0 |

Enzyme activities of two vacuole-specific enzyme markers and an extravacuolar enzyme were determined in protoplast and vacuole fractions prepared from transgenic tobacco plants expressing *wt* or *gly* barley lectin. Enzyme activity in vacuoles is expressed as a per cent (%) of the activity determined in the same number of protoplasts. Results represent the mean ± SD calculated from three individual experiments.

immunoblots (Figure 6) with polyclonal anti-WGA antiserum. The 18 kd mature subunit of barley lectin is readily discernible in protoplasts isolated from *wt* or *gly* tobacco plants (Lanes 2 and 4, Figure 6). Immunoblot analysis also revealed the presence of mature barley lectin in both *wt* and *gly* vacuoles (Lanes 3 and 5, Figure 6). These results indicate that barley lectin is correctly targeted to vacuoles in tobacco. Moreover, the absence of the propeptide glycan does not apparently preclude the targeting of barley lectin to tobacco vacuoles. The vacuolar distribution of barley lectin in *wt* and *gly* transgenic tobacco leaves was also confirmed by EM immunocytochemistry (results not shown). No immunoreactive component was observed in the cytoplasm of transgenic tobacco plants (data not shown).

***Kinetics of intracellular processing of wt and gly barley lectin in transgenic tobacco.*** Pulse-chase experiments were performed to assess the influence of the high-mannose glycan contained within the propeptide of the barley lectin proprotein on the rate of post-translational processing and accumulation of mature barley lectin in tobacco vacuoles. Both *wt* and *gly* tobacco protoplasts were pulse-labeled for 10 hr in the presence of $^{35}$S-Trans label and chased with unlabeled methionine and cysteine for an additional 10 hrs. At specified intervals during the chase period, radiolabeled barley lectin was recovered from lysed protoplasts by affinity chromatography and analyzed by SDS-PAGE and fluorography (Figure 7). The 23 kd *wt* proprotein and the 21 kd *gly* proprotein

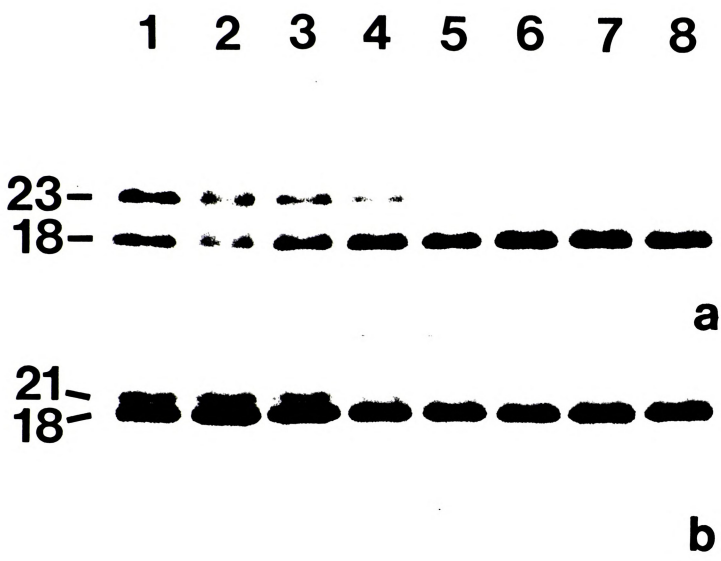
**Figure 6. Immunodetection of mature barley lectin in protoplasts and vacuoles isolated from *wt* and *gly* transgenic tobacco plants.** Mature barley lectin detected in protoplasts and vacuoles prepared from a tobacco plant expressing *wt* barley lectin (**lanes 2 and 3**, respectively) or *gly* barley lectin (**lanes 3 and 5,** respectively). Lane 1 is affinity purified mature WGA. The molecular mass of mature WGA and barley lectin is denoted on the left in kd.


1
2
3
4
5
18-

as well as the 18 kd mature barley lectin polypeptide were present in pulse-labeled protoplasts (Lane 1, Figure 7A and 7B). During the chase period, both the *wt* and *gly* radiolabeled proproteins gradually disappear over time (Figure 7A and 7B, respectively). The disappearance of the barley lectin proproteins is accompanied by a corresponding increase in the level of the 18 kd mature protein. The kinetics at which either the *wt* or *gly* proproteins are processed to mature polypeptides were determined by scanning densitometry. Half-life ($t_{1/2}$) determinations of the *wt* or *gly* barley lectin proproteins indicates that the *gly* 21 kd proprotein ($t_{1/2}$ = 1.0 hr) is processed to the mature protein at least 2-fold faster than the *wt* 23 kd proprotein ($t_{1/2}$ = 2.0 hr). These results indicate that *wt* and *gly* barley lectin proproteins are processed at different kinetics during transport through the endomembrane system of the secretory pathway.

***Effect of monensin on the post-translational processing of barley lectin in transgenic tobacco.*** Processing of the proprotein to mature barley lectin involves the selective removal of the COOH-terminal glycopeptide from the proprotein. To address the events involved in the post-translational processing of the proprotein of barley lectin, *wt* and *gly* tobacco protoplasts were pulse-labeled in the presence of the inhibitor monensin. Monensin is an ionophore that primarily disrupts transport vesicles and protein sorting from the *trans*-cisternae of the Golgi complex (Tartakoff, 1983; Chrispeels, 1983). Following a

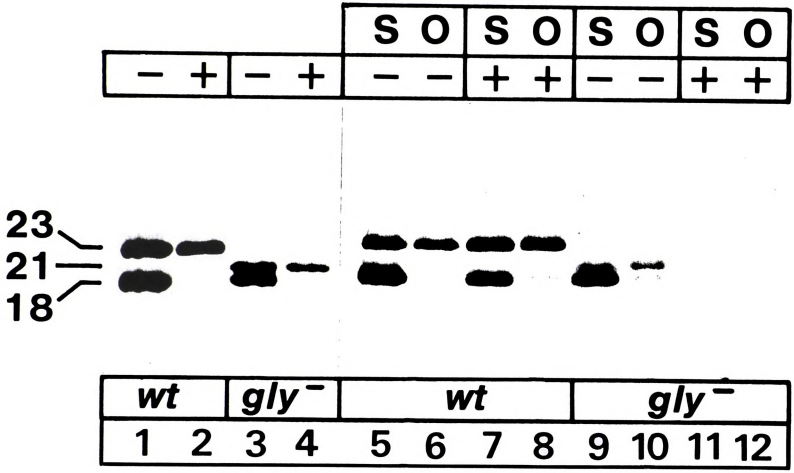
**Figure 7. Pulse-chase labeling experiments of tobacco protoplasts expressing *wt* or *gly* barley lectin.** Tobacco protoplasts expressing *wt* (**Panel A**) or *gly* (**Panel B**) barley lectin were pulse-labeled for 10 hr and chased for 0, 1, 2, 3, 4, 6, 8, and 10 hr (**lanes 1-8**). Radiolabeled barley lectin was affinity purified from lysed protoplasts and the eluants subjected to SDS-PAGE and fluorography. The position and molecular mass of the *wt* (23 kd) and *gly* (21 kd) barley lectin proproteins and the 18 kd mature polypeptide are denoted on the left.


1 2 3 4 5 6 7 8
23–
18–
a
21–
18–
b

1 hr preincubation in the presence of monensin, both *wt* and *gly* tobacco protoplasts were subsequently pulse-labeled for 12 hr. Radiolabeled barley lectin was affinity purified from lysed protoplasts and analyzed by SDS-PAGE and fluorography. The effect of monensin on the post-translational processing of *wt* and *gly* barley lectin proproteins in tobacco is presented in Figure 8. In the absence of monensin, both the 18 kd mature protein and the *wt* or *gly* proproteins are evident in pulse-labeled protoplasts (Lanes 1 and 3, respectively, Figure 8). However, the preponderance of barley lectin radiolabeled in the presence of monensin are the 23 kd *wt* or 21 kd *gly* proproteins (Lanes 2 and 4, respectively, Figure 8) indicating that monensin effectively inhibits processing of the proproteins to the mature polypeptide. Densitometer scanning of trace levels of 18 kd mature protein observed in both *wt* and *gly* protoplasts (Lanes 2 and 4, Figure 8) established that less than 4% of the proproteins are converted to the mature protein in the presence of monensin.

To establish that monensin disrupts processing of *wt* or *gly* proproteins within the Golgi complex, protoplasts pulse-labeled in the presence or absence of monensin were gently lysed and separated into soluble (cytosol + vacuolar contents) and organelle (enriched ER/Golgi) fractions. The molecular forms of radiolabeled barley lectin affinity-purified from soluble (S) or organelle (O) fractions isolated from *wt* or *gly* protoplasts pulse-labeled in the presence (+) or absence (-) of monensin are presented in Figure 8. Both proproteins and mature barley lectins are present in the soluble fraction of *wt* or *gly* protoplasts

**Figure 8. Inhibition of proteolytic processing of barley lectin proproteins in the presence of monensin.** Tobacco protoplasts expressing *wt* or *gly* barley lectin were pulse-labeled for 12 hr in 0.1% ethanol (-) or 50 uM monensin, 0.1% ethanol (+). Radiolabeled barley lectin was affinity purified from a portion of the protoplasts and analyzed by SDS-PAGE and fluorography. Soluble (**S**) and organelle (**O**) fraction from the remaining protoplasts were separated by Sepharose-4B chromatography. Radiolabeled barley lectin in subcellular fractions of pulse-labeled tobacco protoplasts were affinity purified and fractionated by SDS-PAGE and treated for fluorography. The position and molecular mass (kd) of barley lectin proproteins and mature polypeptide are indicated to the left.


S O S O S O S O
– + – + – – + + – – + +
23
21
18
wt gly⁻ wt gly⁻
1 2 3 4 5 6 7 8 9 10 11 12

incubated in the absence or presence of monensin (Lanes 5,7 and 9,11, respectively, Figure 8). However, only the proproteins are readily discernible in organelle fractions isolated from *wt* or *gly* protoplasts treated with or without monensin (Lanes 6,8 and 10,12, respectively, Figure 8). These results demonstrate that *wt* or *gly* proproteins are associated with ER/Golgi compartments. The lower levels of *gly* proprotein evident in soluble and particularly organelle fractions is congruent with a shorter half-life for *gly* proproteins (Figure 7). Although monensin is efficacious in preventing the processing of proproteins to mature barley lectin (see above), significant levels of mature 18 kd barley lectin can be observed in soluble fractions isolated from protoplasts (especially *wt*) treated with monensin (Lane 7, Figure 8). The appearance of mature barley lectin in monensin treated soluble fractions is presumably due to *in vitro* processing of proproteins by proteases during isolation of subcellular fractions.

Levels of radiolabeled barley lectin recovered from soluble and organelle fractions isolated from protoplasts incubated with monensin differ significantly in *wt* and *gly* tobacco transformants (Figure 8). Densitometer scanning demonstrated that the overall levels of the *wt* 23 kd proprotein in soluble and organelle fractions isolated from protoplasts incubated with (Lanes 5 and 6, Figure 8) or without (Lanes 7 and 8, Figure 8) monensin are very similar, although there is a discernible increase of the proprotein in the presence of monensin. However, the *gly* 21 kd proprotein is barely dectectable in

subcellular fractions of *gly* protoplasts treated with monensin (Lanes 11 and 12, Figure 8), especially in the organelle fraction (Lane 12, Figure 8). The monensin results suggest that the glycosylated 23 kd *wt* proprotein and the unglycosylated 21 kd *gly* proprotein are differentially transported from the Golgi complex.

Monensin primarily disrupts intracellular vesicular transport and consequently results in extracellular secretion of lysosomal proteins (Tartakoff, 1983). Pea vicilin (Craig and Goodchild, 1984) and ConA (Bowles, *et al.*, 1986) accumulate at the cell surface and in the periplasmic space between the cell wall and the plasma membrane in cotyledons treated with monensin. Thus, the presence and relative abundance of radiolabeled barley lectin was examined in the culture media of pulse-labeled *wt* and *gly* tobacco protoplasts incubated in the presence or absence of monensin. Radiolabeled barley lectin was isolated from the culture media by affinity chromatography and subsequently analyzed by SDS-PAGE and fluorography. Radiolabeled barley lectin is not discernible in the culture media of either *wt* or *gly* protoplasts pulse-labeled in the presence or absence of monensin (data not shown).

## DISCUSSION

Barley lectin is a member of a class of vacuolar proteins which are initially synthesized as glycosylated precursors and subsequently processed to mature nonglycosylated proteins by the post-translational cleavage of a COOH-terminal glycopeptide. This class of vacuolar proteins includes the Gramineae lectins

and a plant defense-related $\beta$-1,3-glucanase of tobacco (Shinshi, *et al.*, 1988). The transient association of an N-linked oligosaccharide side-chain with the proprotein provides a unique opportunity to investigate the functional significance of the N-linked glycan in the post-translational processing and transport of these vacuolar proteins.

***Barley lectin is correctly assembled and targeted to vacuoles in transgenic tobacco.*** The feasibility of expressing a monocot vacuolar protein in a heterologous dicot system was examined by introducing cDNAs encoding the *wt* barley lectin preproprotein under the transcriptional control of the constitutive CaMV 35S promoter into tobacco by *Agrobacteria*-mediated transformation. Analysis of transgenic plants established that the *wt* barley lectin is synthesized as the appropriate 23 kd proprotein in tobacco. The 23 kd *wt* proprotein is correctly modified by the covalent attachment of a 2 kd high mannose oligosaccharide side-chain, post-translationally processed to the mature 18 kd subunit and transported to vacuoles in tobacco analogous to barley embryos (Lerner and Raikhel, 1989). Synthesis of the correct glycosylated barley lectin proprotein in transgenic tobacco plants is indicative that the signal sequence of this monocot protein is recognized and cleaved by an ER signal peptidase in dicots. Correct utilization of $NH_2$-terminal signal sequences in heterologous systems have been documented for the vacuolar protein PHA (Sturm, *et al.*, 1988) and a chimeric construct employing the signal sequence of the vacuolar

storage protein patatin (Iturriaga, *et al.*, 1989). Predicated on the ability to isolate mature barley lectin by affinity chromatography on immobilized GlcNAc, the wt proproteins are assembled into the correct dimeric conformation required of an active lectin. In summary, the correct synthesis, assembly, processing and transport of barley lectin to vacuoles in tobacco indicates the existence of a common mechanism for post-translational processing and targeting of proteins to vacuoles in monocots and dicots. A number of storage proteins and lectins have been correctly expressed in seeds of heterologous systems (Beachy, *et al.*, 1985; Sengupta-Gopalan, *et al.*, 1985; Okamuro, *et al.*, 1986; Hoffman, *et al.*, 1987; Sturm, *et al.*, 1988). However, only patatin has been shown to be correctly processed in vegetative tissues of tobacco (Sonnewald, *et al.*, 1989). The present study is the first report to demonstrate the correct processing and stable accumulation of a embryo-specific monocot vacuolar protein in tobacco leaves and roots.

***Propeptide glycan is not required for correct assembly and transport of barley lectin in transgenic tobacco.*** The myriad of functions associated with the N-linked oligosaccharides of many mammalian glycoproteins (Olden, *et al.*, 1985) indicate that there is no universal role for N-linked glycans. The influence of the barley lectin proprotein glycan on assembly, processing and transport of this protein was investigated by examining the expression of a mutant gly$^-$ barley lectin in transgenic tobacco. The 21 kd nonglycosylated proprotein is correctly

synthesized, assembled as an active lectin, transported to vacuoles, and processed to the mature polypeptide analogous to *wt* barley lectin in transgenic tobacco and barley embryos. Although the absence of the propeptide glycan in tobacco plants expressing the *gly*⁻ proprotein of barley lectin apparently does not impede the formation of active lectin dimers, it is unknown whether the presence of the glycan or the glycopeptide may influence the rate of assembly of active lectin dimers. Active dimers can actually be assembled from mature nonglycosylated subunits *in vitro* (Peumans, *et al.*, 1982b).

Localization of mature barley lectin derived from the gly⁻ proprotein in vacuoles of tobacco also demonstrates that the high mannose glycan covalently attached to the COOH-terminal propeptide is not an absolute requirement for the targeting of barley lectin to vacuoles. Similar results have been observed for the glycoprotein PHA (Bollini, *et al.*, 1985; Voelker, *et al.*, 1989) even though barley lectin is only glycosylated as a precursor and unlike PHA, it is not a glycoprotein in its mature form. The glycans of the barley lectin proprotein and PHA are not essential for processing and targeting of these proteins to vacuoles. Conversely, the glycan of proConA apparently plays a direct role in processing and transport of ConA to vacuoles (Faye and Chrispeels, 1987).

***Propeptide glycan affects rate of post-translational processing and transport of barley lectin in transgenic tobacco.*** To assess the possibility that the N-linked glycan plays an indirect role in intracellular processing and transport of barley

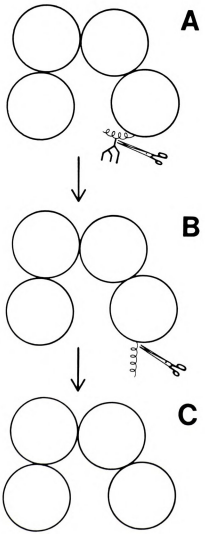
lectin, pulse-chase labeling and monensin experiments were performed with tobacco protoplasts expressing the *wt* or *gly* barley lectin proproteins. Pulse-chase experiments demonstrated that the glycosylated and unglycosylated proproteins are differentially processed to the mature protein with characteristic kinetics. The nonglycosylated (*gly*) 21 kd proprotein is processed to the mature 18 kd protein at a rate at least 2-fold faster than the glycosylated (*wt*) 23 kd proprotein. Fractionation of subcellular components established that the proproteins are associated with the Golgi compartment and indicates that the *gly* proprotein is also transported from the Golgi complex faster than the *wt* proprotein.

Monensin effectively inhibits the post-translational processing of both the *wt* and *gly* barley lectin proproteins to the mature subunit in tobacco protoplasts. However, monensin exerts a differential effect on the transport of these proproteins from the Golgi complex. In the presence of monensin, the *wt* glycosylated proprotein remains lodged within the Golgi complex or is transported from this compartment very slowly. The protracted rate of processing and transport of the *wt* proprotein relative to the *gly* proprotein implies that the N-linked glycan of the propeptide modulates the processing and transport of barley lectin to vacuoles. We propose that the deglycosylation of the propeptide precedes processing and transport and is the rate-limiting step of these processes. The post-translational removal of an internal glycopeptide from proConA is also believed to commence with a deglycosylation step

(Bowles, *et al.*, 1986). In contrast to the present study, monensin purportedly has limited effect on the processing of the rice lectin proprotein to the mature protein in developing embryos (Stinissen, *et al.*, 1985). However, similar inhibitory effects by monensin have been observed on the processing of proConA (Bowles, *et al.*, 1986) and pea vicilin proproteins (Craig and Goodchild, 1984).

***A model for the role of the glycan in the post-translational processing of barley lectin.*** The pulse-chase and monensin experiments indicate that the N-linked high mannose glycan of the barley lectin propeptide modulates the rate of processing and transport of the barley lectin proprotein from the Golgi complex to the vacuoles. Moreover, the modulation of these processes by the glycan is rate-limiting. The glycan therefore presumably plays an indirect or negative role in the regulation of processing and transport of barley lectin to vacuoles. We propose that the molecular mechanism by which the glycan regulates these processes relys upon a sequential two-step processing of the proprotein COOH-terminal glycopeptide (Figure 9). Concomitant with the formation of an active lectin dimer, the proprotein assumes a conformation in which the high mannose glycan sequesters the propeptide from the aqueous environment, thereby masking the availability of the propeptide for processing (Figure 9A). This predicted protein configuration is predicated on the conformation of the protein

**Figure 9. Proposed chain of events involved in the post-translational processing of barley lectin.** The processing model schematically depicts one subunit of a barley lectin dimer adapted from the structure of WGA (Wright, 1987). Each of the highly homologous domains of barley lectin is represented by a circle. A high mannose glycan tree is attached to the sole N-linked glycosylation site (Asn-Ser-Thr) residing within the COOH-terminal propeptide of barley lectin.


A
B
C

(Wright, 1987), the amphipathic characteristic of the propeptide and the hydrophilic nature of the glycan. In the *trans*-cisternae of the Golgi complex, the glycan is removed post-translationally in a regulated manner from the proprotein. As a consequence, deglycosylation exposes the propeptide to proteases and thereby facilitates further processing and transport of the proprotein (Figure 9B). The contribution of the glycan in the processing and transport of this plant vacuolar protein is congruous with the involvement of N-linked glycans in the proteolytic processing and stabilization of many mammalian glycoproteins (Olden, *et al.*, 1985).

## REFERENCES

Aebi, H. (1974) Catalase. In Bergermeyer, H.U. (ed.), *Methods of Enzymatic Analysis* **Vol. II.** Academic Press, NY, pp. 673-678.

Akazawa, T. and Hara-Nishimura, I. (1985) Topographic aspects of biosynthesis, extracellular, and intracellular storage of proteins in plant cells. *Ann. Rev. Plant Physiol.* **36,** 441-472.

An, G., Ebert, P.R., Mitra, A. and Ha, S.B. (1988) Binary vectors. *Plant Molec. Biol. Manual* **A3,** 1-19.

Beachy, R.N., Chen, Z.-L., Horsch, R.B., Rogers, S.G., Hoffman, N.J. and Fraley, R.T. (1985) Accumulation and assembly of soybean $\beta$-conglycinin in seeds of transformed petunia plants. *EMBO J.* **4,** 3047-3053.

Blake, M.S., Johnston, K.H., Russel-Jones, G.J. and Gotschlich, E.C. (1984) A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on Western blots. *Anal. Biochem.* **136,** 175-179.

Boller, T. and Kende, H. (1979) Hydrolytic enzymes in the central vacuole of plant cells. *Plant Physiol.* **63,** 1123-1132.

Bollini, R., Ceriotti, A., Daminati, M.G. and Vitale, A. (1985) Glycosylation is not needed for the intracellular transport of phytohemagglutinin in developing *Phaseolus vulgaris* cotyledons and for the maintenance of its biological activities. *Physiol. Plant.* **65,** 15-22.

Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R. and Burgess, J. (1986) Posttranslational processing of Concanavalin A precursors in jackbean cotyledons. *J. Cell Biol.* **102,** 1284-1297.

Bradford, M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **72,** 248-254.

Chrispeels, M.J. (1983) The Golgi apparatus mediates the transport of phytohemagglutinin to the protein bodies in bean cotyledons. *Planta* **158,** 140-151.

Craig, S. and Goodchild, D.J. (1984) Golgi-mediated vicilin accumulation in pea cotyledon cells is re-directed by monensin and nigericin. *Protoplasma* **122,** 91-97.

Faye, L. and Chrispeels, M.J. (1987) Transport and processing of the glycosylated precursor of Concanavalin A in jack-bean. *Planta* **170,** 217-224.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132,** 6-13.

Guy, M., Reinhold, L. and Michaeli, D. (1979) Direct evidence for a sugar transport mechanism in isolated vacuoles. *Plant Physiol.* **64,** 61-64.

Herman, E.M., Shannon, L.M. and Chrispeels, M.J. (1985) Concanavalin A is synthesized as a glycosylated precursor. *Planta* **165,** 23-29.

Hoffman, L.M., Donaldson, D.D., Bookland, R., Rashka, K. and Herman, E.M. (1987) Synthesis and protein body deposition of maize 15-kd zein in transgenic tobacco seeds. *EMBO J.* **6,** 3213-3221.

Horsch, R.B., Fry, J., Hoffman, N., Neidermeyer, J., Rogers, S.G. and Fraley, R.T. (1988) Leaf disc transformation. *Plant Molec. Biol. Manual* **A5,** 1-9.

Hooykaas, P.J.J. (1988) *Agrobacterium* molecular genetics. *Plant Molec. Biol. Manual* **A4,** 1-13.

Iturriaga, G., Jefferson, R.A. and Bevan, M.W. (1989) Endoplasmic reticulum targeting and glycosylation of hybrid proteins in transgenic tobacco. *The Plant Cell* **1,** 381-390.

Kobata, A. (1984) The carbohydrates of glycoproteins. In Ginsburg, V. and Robbins, P.W. (eds.), *Biology of Carbohydrates* **Vol. 2.** John Wiley & Sons, NY, pp. 87-161.

Kornfeld, S. (1986) Trafficking of lysosomal enzymes in normal and disease states. *J. Clin. Invest.* **77,** 1-6.

Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154,** 367-382.

Lerner, D.R. and Raikhel, N.V. (1989) Cloning and characterization of root-specific barley lectin. *Plant Physiol.* **91,** 124-129.

Lord, J.M. (1983) in Hall, J.L. and Moore, A.L. (eds.), *Isolation of Membranes and Organelles from Plant Cells.* Academic Press, NY, pp. 119-134.

Machamer, C.E. and Rose, J.K. (1988) Vesicular Stomatitis virus G proteins with altered glycosylation sites display temperature-sensitive intracellular transport and are subject to aberrant intermolecular disulfide bonding. *J. Biol. Chem.* **263,** 5955-5960.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mansfield, M.A., Peumans, W.J. and Raikhel, N.V. (1988) Wheat-germ agglutinin is synthesized as a glycosylated precursor. *Planta* **173,** 482-489.

Matzuk, M.M. and Boime, I. (1988) Site-specific mutagenesis defines the intracellular role of the asparagine-linked oligosaccharides of chorionic gonadotropin $\beta$ subunit. *J. Biol. Chem.* **263,** 17106-17111.

Mishkind, M.L., Palevitz, B.A., Raikhel, N.V. and Keegstra, K. (1983) Localization of wheat germ agglutinin-like lectins in various species of the Gramineae. *Science* **220,** 1290-1292.

Okamuro, J.K., Jofuku, K.D. and Goldberg, R.B. (1986) Soybean seed lectin gene and flanking nonseed protein genes are developmentally regulated in transformed tobacco plants. *Proc. Natl. Acad. Sci. USA* **83,** 8240-8244.

Olden, K., Bernard, B.A., Humphries, M.J., Yeo, T.-K., Yeo, K.-T., White, S.L., Newton, S.A., Bauer, H.C. and Parent, J.B. (1985) Function of glycoprotein glycans. *Trends Biochem Sci.* **10,** 78-82.

Peumans, W.J., Stinissen, H.M. and Carlier, A.R. (1982a) Isolation and partial characterization of wheat-germ-agglutinin-like lectins from rye (*Secale cereale*) and barley (*Hordeum vulgare*) embryos. *Biochem J.* **203,** 239-243.

Peumans, W.J., Stinissen, H.M. and Carlier, A.R. (1982b) A genetic basis for the origin of six different isolectins in hexaploid wheat. *Planta* **154,** 568-572.

Peumans, W.J., Stinissen, H.M. and Carlier, A.R. (1983) The rice lectin and its relationship to cereal lectins. *Biochem. Physiol. Pflanzen.* **178,** 423-431.

Raikhel, N.V. and Wilkins, T.A. (1987) Isolation and characterization of a cDNA clone encoding wheat germ agglutinin. *Proc. Natl. Acad. Sci. USA* **84,** 6745-6749.

Raikhel, N.V., Mishkind, M.L. and Palevitz, B.A. (1984) Characterization of a wheat germ agglutinin-like lectin from adult wheat plants. *Planta* **162,** 55-61.

Raikhel, N.V., Bednarek, S.Y. and Wilkins, T.A. (1988) Cell-type specific expression of a wheat-germ aggutinin gene in embryos and young seedlings of *Triticum aestivum*. *Planta* **126,** 406-414.

Rice, R.H. and Etzler, M.E. (1974) Chemical modification and hybridization of wheat germ agglutinins. *Biochem. Biophys. Res. Commun.* **59,** 414-419.

Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **56,** 5463-5467.

Sengupta-Gopalan, C., Reichert, N.A., Barker, R.F., Hall, T.C. and Kemp, J.D. (1985) Developmentally regulated expression of the bean $\beta$-phaseolin gene in tobacco seed. *Proc. Natl. Acad. Sci. USA* **82,** 3320-3324.

Shimomura, S., Inohara, N., Fukui, T. and Futai, M. (1988) Different properties of two types of auxin-binding sites in membranes from maize coleoptiles. *Planta* **175,** 558-566.

Shinshi, H., Wenzler, H., Neuhaus, J.-M., Felix, G., Hofsteenge, J. and Meins, Jr., F. (1988) Evidence for N- and C-terminal processing of a plant defense-

related enzyme: Primary structure of tobacco prepro-$\beta$-1, 3-glucanase. *Proc. Natl. Acad. Sci. USA* **85,** 5541-5545.

Shure, M., Wessler, S. and Federoff, N. (1983) Molecular identification and isolation of the *waxy* locus in maize. *Cell* **35,** 225-233.

Smith, J.J. and Raikhel, N.V. (1989) Production of an antibody specific for the propeptide of wheat germ agglutinin. *Plant Physiol.* **91,** 473-476.

Sonnewald, U., Sturm, A., Chrispeels, M.J. and Willmitzer, L. (1989) Targeting and glycosylation of patatin the major potato tuber protein in leaves of transgenic tobacco. *Planta* **179,** 171-180.

Stinissen, H.M., Peumans, W.J. and Carlier, A.R. (1983) Occurrence and immunological relationships of lectins in gramineous species. *Planta* **159,** 105-111.

Stinissen, H.M. Peumans, W.J. and Chrispeels, M.J. (1984) Subcellular site of lectin synthesis in developing rice embryos. *EMBO J.* **3,** 1979-1985.

Stinissen, H., Peumans, W.J. and Chrispeels, M.J. (1985) Posttranslational processing of proteins in vacuoles and protein bodies is inhibited by monensin. *Plant Physiol.* **77,** 495-498.

Struhl, K. (1985) A rapid method for creating recombinant DNA molecules. *Biotechniques* **3,** 452-453.

Sturm, A., Voelker, T.A., Herman, E.M. and Chrispeels, M.J. (1988) Correct glycosylation, Golgi-processing, and targeting to protein bodies of the vacuolar protein phytohemagglutinin. *Planta* **175,** 170-183.

Tartakoff, A.M. (1983) Perturbation of vesicular traffic with the carboxylic ionophore monensin. *Cell* **32,** 1026-1028.

Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76,** 4350-4354.

Vieira, J. and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods Enzymol.* **153,** 3-11.

Voelker, T.A., Herman, E.M. and Chrispeels, M.J. (1989) *In vitro* mutated phytohemagglutinin genes expressed in tobacco seeds: Role of glycans in protein targeting and stability. *The Plant Cell* **1,** 95-104.

Wilkins, T.A. and Raikhel, N.V. (1989) Expression of rice lectin is governed by two temporally and spatially regulated mRNAs in developing embryos. *The Plant Cell* **1,** 541-549.

Wright, C.S. (1980) Crystallographic elucidation of the saccharide binding mode in wheat germ agglutinin and its biological significance. *J. Mol. Biol.* **141,** 267-291.

Wright, C.S. (1987) Refinement of the crystal structure of wheat germ agglutinin isolectin 2 at 1.8 A resolution. *J. Mol. Biol.* **194,** 501-529.

Zimmerman, J.L. and Goldberg, R.B. (1977) DNA sequence organization in the genome of *Nicotiana tabacum*. *Chromosoma* **59,** 227-252.

## Chapter 5

## Summary

The isolation and characterization of cDNA clones encoding the lectins from wheat and rice has contributed significantly to our studies to elucidate the molecular mechanisms regulating the post-transcriptional and post-translational processing of the Gramineae lectins. The major conclusions obtained from the research described in this dissertation are summarized below.

### A. *Isolation and Characterization of cDNA Clones Encoding Wheat Germ Agglutinin (WGA) and Rice Lectin:*

1. A truncated cDNA clone encoding the proprotein of WGA (isolectin B) was isolated from a cDNA library constructed from developing wheat embryos using synthetic oligonucleotides.
2. Two full-length cDNA clones (cRL852 and cRL1035) encoding the preproprotein of rice lectin were isolated from a cDNA library constructed from developing rice spikelets using the WGA-B cDNA clone as a heterologous probe.

a. The two cDNA clones differ solely in the length of their 3'-untranslated region and represent two mRNA species encoding identical polypeptides.

b. The cDNA clone cRL1035 contains two putative polyadenylation sites; one of the sites is common to both cDNA clones.

3. The mature proteins of WGA-B and rice lectin exhibit 73% identity at the amino acid level of the mature subunit.

4. Rice lectin is more closely related to isolectin A of WGA (Appendix A).

5. The cDNA clones encoding WGA-B and rice lectin were discovered to encode COOH-terminal propeptides of 15 amino acids and 26 amino acids, respectively. The propeptide domains of WGA-B and rice lectin are less than 47% homologous at the amino acid level, thereby exhibiting a significantly lower degree of conservation than in the mature protein.

6. The $NH_2$-terminal domain of chitinase, a plant defense-related protein from *Phaseolus vulgaris*, exhibits high homology (48% to 60%) to all four domains of WGA, rice and barley lectin.

B. ***Expression of WGA and Rice Lectin in Developing Embryos:***

1. Levels of steady-state mRNA encoding WGA-B in developing embryos is enhanced several-fold by the addition of exogenous abscisic acid.
2. Rice lectin is encoded by two mRNA species derived from a single gene gene present at 1 to 2 copies per haploid genome.
3. The two mRNA species of rice lectin are presumably derived from alternative polyadenylation site selection during the post-transcriptional processing of the pre-mRNA.
4. The mRNA of rice lectin accumulates in root caps, peripheral cell-layers of the radicle, coleorhiza, scutellum and in all cell-layers of the coleoptile.
5. The expression of rice lectin is regulated at two molecular levels in developing embryos.
    a. The temporal expression of the rice lectin mRNAs is presumably regulated at the transcriptional and/or post-transcriptional levels.
    b. The differential accumulation of the two rice lectin mRNAs is controlled at the post-transcriptional level.

C. ***The Molecular Mechanisms of Post-Translational Processing of Rice and Barley Lectin:***

1. Properties of COOH-terminal propeptides of Gramineae lectins:
   a. The sole N-linked glycosylation site within the COOH-terminal propeptide is modified by the addition of a high mannose oligosaccharide side-chain.
   b. The primary amino acid sequence of Gramineae lectin COOH-terminal propeptides is not conserved (Appendix B).
   c. The COOH-terminal glycopeptides are hydrophobic, acidic domains (Appendix B).
   d. Secondary structure predictions indicate that the COOH-terminal propeptide domains of the Gramineae lectins form amphipathic $\alpha$-helices (Appendix B).
   e. A COOH-terminal propeptide of tobacco $\beta$-1,3-glucanase shares properties described above (a to c) with the Gramineae lectins.
2. The maturation of rice lectin involves the following series of post-translational processing events (Appendix C):
   a. Rice lectin is initially synthesized as a 23 kd preproprotein on the rough endoplasmic reticulum.

b. Concomitant with the cleavage of the signal sequence, the COOH-terminal propeptide is glycosylated with a 2 kd high mannose glycan to generate a 25 kd glycosylated proprotein.

c. The COOH-terminal glycopeptide is post-translationally processed to yield the mature 18 kd subunit of rice lectin. The post-translational removal of the COOH-terminal glycopeptide is hypothesized to occur by a two-step process, commencing with the deglycosylation of the propeptide.

d. The 18 kd subunit is processed further by an endoproteolytic cleavage to produce 10 kd and 8 kd polypeptides. Several amino acids are processed from the COOH-terminus of the 10 kd polypeptide following cleavage from the 18 kd subunit.

2. Barley lectin is correctly synthesized, assembled, post-translationally processed and transported to vacuoles in transgenic tobacco. These results demonstrate the first correct processing and stable accumulation of an embryo-specific monocot protein in vegetative tissues of a dicot.

3. The molecular mechanisms involved in the post-translational processing of barley lectin are similar in monocots and dicots.

4. The high mannose N-linked glycan of the COOH-terminal propeptide is not required for targeting of barley lectin to vacuoles in transgenic tobacco.

5. The glycan of the barley lectin proprotein modulates the rate of post-translational processing and transport through the endoplasmic reticulum and Golgi complex to the vacuole.

6. The post-translational processing of the COOH-terminal glycopeptide is proposed to occur via a two-step process, commencing with the deglycosylation of the propeptide.

Appendix A

## Appendix A

**Table 1.** Amino acid positions which distinguish isolectins A, B, and D of wheat germ agglutinin (WGA) and rice lectin

| Amino Acid | WGA-A | WGA-B | WGA-D | Rice |
|---|---|---|---|---|
| 9 | N | G | N | G |
| 37 | N | N | N | S |
| 53 | A | K | A | A |
| 56 | T | P | P | S |
| 59 | Q | H | H | Q |
| 66 | Y | H | H | Y |
| 93 | A | A | S | A |
| 109 | F | Y | F | Y |
| 119 | G | E | G | N |
| 123 | S | N | S | S |
| 171 | A | G | A | G |

Relatedness of rice lectin to the isolectins of wheat:
WGA-A 77.8% WGA-B 44.4% WGA-D 33.3%

Appendix B

Figure B.1 Acidic N-linked Glycosylated COOH-terminal Propeptide Domains of Gramineae Lectins and $\beta$-1,3-glucanase of Tobacco

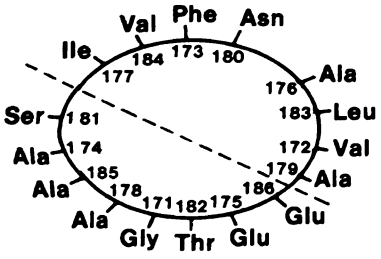
## COOH-terminal N-linked Glycopeptides

| | |
|---|---|
| **WGA** | V F A [E] A I A T N S T L L A [E] |
| **Barley lectin** | V F A [E] A I A A N S T L V A [E] |
| **Rice lectin** | [D] G M A A I L A N N G S V S F [E] G I I [E] S V A [E] L V |
| **β-1,3-Glucanase** | V S G G V W [D] S S V [E] T N A T A S L V S [E] M |

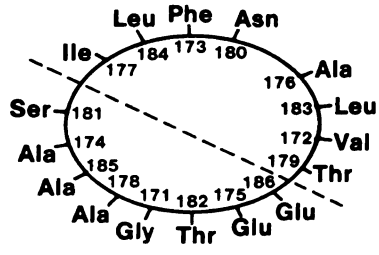
Figure B.2 Amphipathic $\alpha$-helices of Gramineae Lectin COOH-terminal Propeptide Domains

# Appendix B

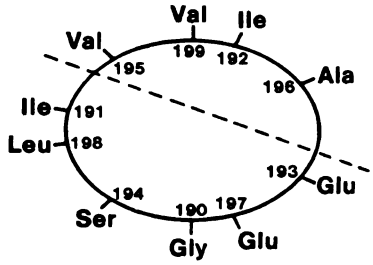
## AMPHIPATHIC CHARACTER OF GRAMINEAE LECTIN PROPEPTIDES




Barley Lectin



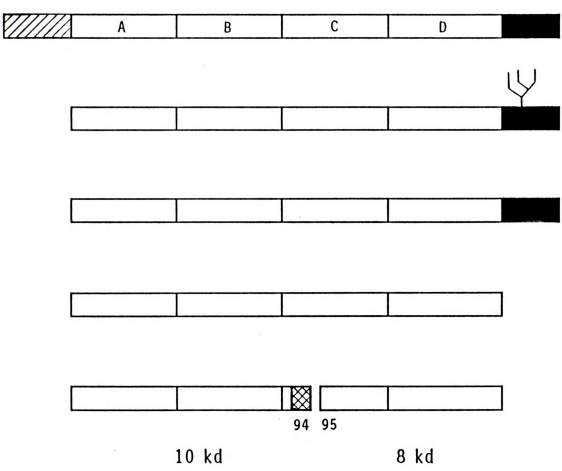

WGA-B




Rice Lectin

Appendix C

Figure C.1 Post-translational Processing of Rice Lectin in Developing Embryos


A
B
C
D
preproprotein
glycosylated
proprotein
proprotein
mature
subunit
processed
subunit
94 95
10 kd
8 kd