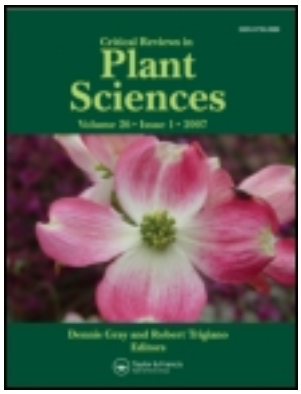

Critical Reviews in Plant Sciences
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/bpts20

Plant Lectins: A Composite of Several Distinct Families of Structurally and Evolutionary Related Proteins with Diverse Biological Roles

Els J. M. Van Damme [a], Willy J. Peumans [a], Annick Barre [b] & Pierre Rougé [b]

[a] Laboratory for Phytopathology and Plant Protection, Katholieke Universiteit Leuven, Willem de Croylaan 42, 3001 Leuven, Belgium

[b] Institut de Pharmacologie et Biologie Structurale, UPR CNRS 9062, 205 Route de Narbonne, 31077 Toulouse Cedex, France

Published online: 24 Jun 2010.

To cite this article: Els J. M. Van Damme , Willy J. Peumans , Annick Barre & Pierre Rougé (1998): Plant Lectins: A Composite of Several Distinct Families of Structurally and Evolutionary Related Proteins with Diverse Biological Roles, Critical Reviews in Plant Sciences, 17:6, 575-692

To link to this article: http://dx.doi.org/10.1080/07352689891304276

Critical Reviews in Plant Sciences, 17(6):575–692 (1998)

Plant Lectins: A Composite of Several Distinct Families of Structurally and Evolutionary Related Proteins with Diverse Biological Roles

Els J. M. Van Damme,[1] Willy J. Peumans,[1] Annick Barre,[2] and Pierre Rougé[2]*

[1]Laboratory for Phytopathology and Plant Protection, Katholieke Universiteit Leuven, Willem de Croylaan 42, 3001 Leuven, Belgium; [2]Institut de Pharmacologie et Biologie Structurale, UPR CNRS 9062, 205 Route de Narbonne, 31077 Toulouse Cedex, France

* Address all correspondence to: Els J. M. Van Damme.

Referee: Dr. Arpad Pusztai, The Rowett Research Institute, Division of Nutritional Sciences, Backsburn, AB2 9SB, Aberdeen, Scotland, United Kingdom

TABLE OF CONTENTS

ABSTRACT: Many plants contain carbohydrate-binding proteins that are commonly designated as lectins, agglutinins, or hemagglutinins. Due to the obvious differences in molecular structure, biochemical properties, and carbohydrate-binding specificity, plant lectins are usually considered a complex and heterogeneous group of proteins. Recent advances in the structural analysis of lectins and molecular cloning of lectin genes enable subdividision of plant lectins in a limited number of subgroups of structurally and evolutionary related proteins. Four major lectin families, namely, the legume lectins, the chitin-binding lectins composed of hevein domains, the type 2 ribosome-inactivating proteins, and the monocot mannose-binding lectins comprise the majority of all currently known plant lectins. In addition to these four large families the jacalin-related lectins, the amaranthin family, and the Cucurbitaceae phloem lectins are now recognized as separate subgroups. Each of the above-mentioned lectin families is discussed in detail. The description of the individual lectin families includes (1) a brief historical note, (2) an overview of the occurrence, molecular structure, and primary structure of the lectins, (3) a detailed discussion of the structure of the gene(s) and the biosynthesis and posttranslational processing of the primary translation products, (4) a summary of carbohydrate-binding specificity, (5) if relevant a note on the occurrence of lectin-related proteins, (6) a description of the three-dimensional structure of the lectins and the protomers, (7) a detailed discussion of the molecular evolution, and (8) a critical assessment of the physiological role of each group of lectins. Lectins that cannot be classified into one of the seven groups are discussed separately. General conclusions about the structure, evolution, and function of plant lectins are summarized in the concluding remarks.

KEY WORDS: amaranthin, Concanavalin A, Cucurbitaceae phloem lectin, defense, *Galanthus nivalis* agglutinin, gene structure, hevein, jacalin, processing, storage, type 2 RIP.

I. INTRODUCTION

A. Lectins: from Stillmark's 'Ricin' to Carbohydrate-Binding Proteins

Lectins, also called agglutinins or phytohemagglutinins, have the longest scientific history of all plant proteins. The first description of a lectin dates back to 1888, indeed, when Stillmark published his dissertation entitled 'Über Ricin ein giftiges Ferment aus den Samen von *Ricinus communis* L. und einigen anderen Euphorbiaceen' (About Ricin A Toxic Ferment from the Seeds of *Ricinus communis* L. and Some Other Euphorbiaceae Species) (Stillmark, 1888). Although it is now evident that Stillmark's 'ricin' was a complex mixture of true (toxic) ricin molecules and nontoxic agglutinins, his pioneering work was a milestone in biology because he was the first to link the toxicity of castor beans to the occurrence of an hemagglutinating proteinaceous factor. Moreover, his discovery was also a milestone in plant biochemistry because 'ricin' was the first plant protein to which a well-defined biological activity could be ascribed. Soon after the discovery of ricin similar toxic substances were identified in the seeds of *Croton tiglium* (crotin) and *Abrus precatorius* (abrin) and the bark of *Robinia pseudoacacia* (robin). In 1898 Elfstrand introduced for the first time the term 'Blutkörperchenagglutinin' (hemagglutinin) as a common name for all plant proteins that cause clumping of cells (Elfstrand, 1898). This novel term was clearly inspired by the striking similarity between the macroscopically visible activity of the plant proteins and that of human and animal serum agglutinins (which were first described by

Landois in 1875). The idea that toxicity is an intrinsic property of lectins was abandoned in the beginning of the century after Landsteiner and Raubitschek (1907) reported for the first time the presence of nontoxic lectins in the legumes *Phaseolus vulgaris* (bean), *Pisum sativum* (pea), *Lens culinaris* (lentil), and *Vicia sativa* (vetch). Following the work of Landsteiner and Raubitschek many more nontoxic plant hemagglutinins were discovered. Eventually, it became evident that lectins are widespread in the plant kingdom and that toxicity is the exception rather than the rule.

The next milestone in the history of plant lectins was the finding that some hemagglutinins exhibit a clear preference toward erythrocytes of a particular human blood group within the ABO system (Renkonen, 1948; Boyd and Reguera, 1949). This discovery of blood group specificity was the direct motive to the introduction of the novel term 'lectin' (from the Latin verb 'legere', which means 'to select'). Unfortunately, although the term lectin was originally introduced to emphasize the selective agglutination behavior of some hemagglutinins, it was later applied to all proteins with agglutinating activity. 'Hemagglutinin' is certainly a more appropriate term than 'lectin' because it refers to the capability of a protein to agglutinate erythrocytes but does not take into account that most lectins can also agglutinate other cells. Hence, the term 'agglutinin' should be preferred. In the absence of a clear consensus, the term lectin is actually most commonly used, but agglutinin and hemagglutinin still persist as synonyms.

The current confusion in lectin terminology to a great degree is due to the fact that different names have been introduced before the mechanism causing the macroscopically visible agglutination activity was understood in molecular terms. Although Summer and Howell (1936) already in 1936 observed that cane sugar inhibited the agglutination activity of Concanavalin A (ConA), it was only demonstrated in 1952 that the agglutination properties of lectins are based on a specific sugar-binding activity (Watkins and Morgan, 1952). As soon as lectins were recognized as carbohydrate-binding proteins they could be distinguished from other proteins on the basis of a well-defined functional criterion. For this reason lectins are now considered initially as carbohydrate-binding proteins rather than as (hem)agglutinins.

B. Definition and Subdivision of Plant Lectins

The use of carbohydrate-binding activity rather than (hem)agglutination activity as a functional criterion of lectins has some profound consequences on the definition of lectins. According to a first proper definition, which was based primarily on the sugar specificity and inhibition of the agglutination reaction, lectins are carbohydrate-binding proteins (or glycoproteins) of nonimmune origin that agglutinate cells and/or precipitate glycoconjugates (Goldstein et al., 1980). Although this definition was approved (with minor modifications) by the Nomenclature Committee of the International Union of Biochemistry (Dixon, 1981), it had some shortcomings because it was confined to multivalent carbohydrate-binding proteins. Therefore, Kocourek and Horejsi (1983) proposed extending the definition of lectins to also cover the poorly agglutinating toxins with genuine lectin subunits such as ricin and abrin (which at that time were erroneously considered as monovalent, but are in fact divalent because their respective lectin subunits possess two carbohydrate-binding sites). However, after it was observed that some lectins contain a second type of binding site that interacts with noncarbohydrate ligands, the definition proposed by Kocourek and Horejsi was also considered too restric-

tive. Hence, lectins were redefined as carbohydrate-binding proteins other than antibodies or enzymes (Barondes, 1988). Novel insights in the structure of lectins and lectin-related proteins as well as their corresponding genes soon argued for an update of Barondes' definition. For example, the finding that some plant enzymes (like the type 2 ribosome-inactivating proteins [RIP] and the class I chitinases) are fusion proteins built up of a carbohydrate-binding domain tandemly arrayed with a catalytic domain implies that the definition of lectins cannot exclude all enzymes. Similarly, the discovery of proteins that are related to lectins but lack active carbohydrate-binding domains necessitates including functionality as a criterion. Therefore, the presence of at least one noncatalytic domain that binds reversibly to a specific carbohydrate is now considered as the only criterion for a protein to be called a lectin. Based on these considerations, plant lectins were defined recently as 'all plant proteins possessing at least one non-catalytic domain, which binds reversibly to a specific mono- or oligosaccharide' (Peumans and Van Damme, 1995b). This definition, which is far less restrictive than all previous definitions, comprises a broad range of proteins with different agglutination and/or glycoconjugate precipitation properties. A further subdivision of lectins in 'merolectins', 'hololectins', 'chimerolectins', and 'superlectins' has been proposed on the basis of the overall structure of the mature lectins (Figure 1). Merolectins consist of a single carbohydrate-binding domain. By definition they are monovalent and hence cannot precipitate glycoconjugates or agglutinate cells. Hevein, the small chitin-binding protein from the latex of the rubber tree (*Hevea brasiliensis*) (Van Parijs et al., 1991) is a typical merolectin. Hololectins also are exclusively built up of carbohydrate-binding domains but contain at least two such domains that are either identical or very homologous and bind either the same or structurally similar sugar(s). Hololectins are by definition di- or multivalent and hence agglutinate cells and/or precipitate glyco-conjugates. Most plant lectins belong to the subgroup of hololectins. Chimerolectins are fusion proteins consisting of one or more carbohydrate-binding domain(s) tandemly arrayed to an unrelated domain. The latter domain may have a well-defined enzymatic

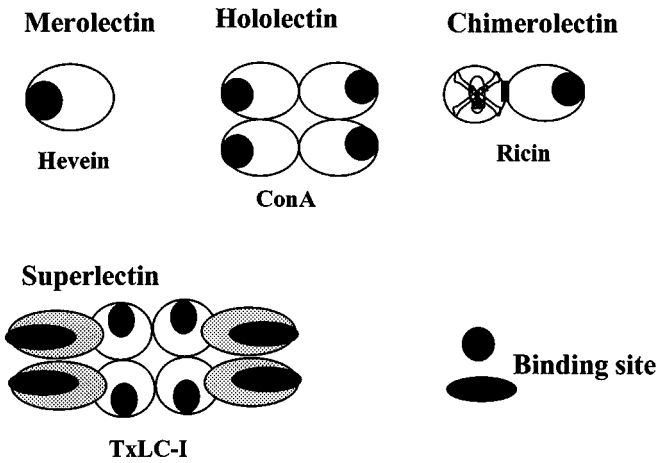

FIGURE 1. Schematic representation of merolectins, hololectins, chimerolectins, and superlectins.

or another biological activity but must act independently of the carbohydrate-binding domain. Depending on the number of carbohydrate-binding sites, chimerolectins behave as merolectins or as hololectins. For example, type 2 RIP with two (or possibly three) carbohydrate-binding sites per protomer behave as genuine hololectins, whereas class I plant chitinases with a single chitin-binding domain per molecule do not precipitate glycoconjugates or agglutinate cells. Like hololectins, superlectins consist exclusively of at least two carbohydrate-binding domains. However, unlike the hololectins the carbohydrate-binding domains of the superlectins recognize structurally unrelated sugars. Therefore, superlectins can also be considered a special group of chimerolectins composed of two tandemly arrayed structurally and functionally different carbohydrate-binding domains. This chimeric nature is illustrated by the protomer of the tulip lectin TxLC-I that consists of an *N*-terminal mannose-binding domain tandemly arrayed with an unrelated GalNAc-binding domain (Van Damme et al., 1996f).

II. RECENT ADVANCES IN THE ANALYSIS OF THE STRUCTURE OF PLANT LECTINS AND THEIR GENES

During the last 2 decades, the introduction of modern biochemistry in lectin research and especially the elaboration of powerful affinity-purification techniques greatly facilitated the isolation and characterization of numerous plant lectins. At present, roughly 300 different plant lectins have been purified and characterized in some detail for what concerns their molecular structure, biochemical properties, and carbohydrate specificity (Van Damme et al., 1998). A vast amount of work has also been devoted to in-depth studies of the biological activities of lectins. Unfortunately, because most of these studies have concentrated on a few lectins only, a small group of plant lectins is well characterized in terms of biological activities. This does not preclude, however, that plant lectins are widely used as tools and as bioactive proteins in biological and biomedical research, and that intensive efforts are being undertaken to exploit their unique carbohydrate-binding activity for therapeutical purposes. The unraveling of the carbohydrate specificity and biological activities also made a critical contribution to the elucidation of the role of lectins in the plant's defense. Compelling evidence has been obtained indeed that the specificity of most plant lectins is not directed against plant glycans but against foreign glycoconjugates (Peumans and Van Damme, 1995a,b). It has been proposed that dietary lectins bind to glycan-receptors exposed on the surface of the gastrointestinal tract of either herbivorous animals or phytophagous invertebrates that eventually results in adverse effects. *In vivo* feeding experiments confirmed the detrimental effects of orally administered lectins on insects and rodents. These experiments not only supported the idea that lectins play a role in the plant's defense but also paved the way for the use of lectin genes in transgenic (crop) plants as resistance factors against insects and/or mammalian pests (Gatehouse et al., 1995; Pusztai and Bardocz, 1996; Peumans and Van Damme, 1998).

All uses and applications of plant lectins eventually rely on their highly specific carbohydrate-binding activity and the biological activities based thereon. To better understand and exploit the sugar-binding activity of lectins, the lectin-carbohydrate interaction(s) have to be studied at the molecular level (Weis and Drickamer, 1996). Therefore, intensive efforts are being undertaken to determine the three-dimensional structure of the carbohydrate-binding sites of lectins and lectin-carbohydrate complexes using

X-ray diffraction and NMR analysis. During the last few years important progress has been made in the structural analysis of plant lectins. An increasing number of lectin structures has been determined and the resolution of the structures is steadily increasing (Rini, 1995). Moreover, several novel structures have been resolved that differ from all previously determined protein structures (Wright, 1997). In addition, structural analyses of lectin-carbohydrate complexes yielded detailed information about molecular interactions between the amino acids of the carbohydrate-binding sites and the sugar ligand(s). Although most plant lectins are not suited for NMR analysis because of their large size, this technique has been applied successfully for the resolution of the structure of a few small lectins and detailed studies of the interactions of these lectins with their carbohydrate ligands.

The rapid progress in the structural analysis of plant lectins is intimately linked to the recent advances in the biochemistry and molecular biology of these proteins. Because even in the early days of plant biochemistry lectins could easily be purified in reasonable quantities, they were the favorite proteins of many research groups. As a result, amino acid sequencing (which is a prerequisite for X-ray crystallography) of lectins developed much faster than that of any other group of plant proteins. At present, about 40 different lectins have been completely sequenced at the protein level and there is still some progress in this area. The wealth of lectin sequence information also stimulated the search for lectin genes in the early days of plant molecular biology. In the beginning most efforts were concentrated on the molecular analysis of the lectin genes in legumes, but within a few years lectin genes had also been isolated from several nonlegume species. At present, about 80 different lectins from species of diverse taxonomic groups have been cloned. Molecular analysis of these genes not only revealed the primary structure of the corresponding lectins but also yielded essential information about the biosynthesis and posttranslational modifications of the primary translation products of the lectin genes. In addition, molecular cloning allowed to corroborate the simultaneous occurrence of two or more lectins at the molecular level and led to the discovery of extended lectin gene families in some plant species. Finally, the sequence information obtained through the molecular cloning of lectin genes from numerous plant species gives a fairly good idea about the evolutionary relationships between the currently known plant lectins. Based on these relationships, the whole group of plant lectins can now be subdivided in a limited number of structurally and evolutionary related proteins.

III. DELINEATION OF SEVEN FAMILIES OF STRUCTURALLY AND EVOLUTIONARY RELATED PLANT LECTINS

Plant lectins are usually considered a very heterogeneous group of proteins because comparative biochemical studies clearly indicate that they differ from each other with respect to their biochemical/physicochemical properties, molecular structure, carbohydrate-binding specificity, and biological activities. A careful analysis of the available data indicates that the concept of heterogeneity, although partly supported by the biological reality, should not be overemphasized. For example, the lack of uniformity in the description of lectins gives the wrong impression that they are all different. Similarly, the incomplete or poor characterization of lectins often creates an apparent heterogeneity. The same holds true whenever impure lectin preparations have been used. Finally, descriptions of novel lectins

usually emphasize the differences rather than the similarities with existing lectins.

In the past several attempts have been made to subdivide plant lectins. For example, lectins have been subdivided according to their carbohydrate-binding specificity in so-called specificity groups. Based on this criterion mannose-, mannose/glucose-, mannose/maltose-, Gal/GalNAc-, GlcNAc/$(GlcNAc)_n$-, fucose-, and sialic acid-binding lectins have been distinguished (Goldstein and Poretz, 1986; Van Damme et al., 1998). This subdivision is certainly helpful for the use of lectins as tools but is probably artificial and irrelevant with respect to possible evolutionary relationships. Subdivisions according to evolutionary relationships have also been proposed on the basis of either serological relationships or sequence similarities (or both). For example, it is generally accepted that the legume lectins are a family of closely related proteins. Similarly, evolutionary relationships between chitin-binding lectins from species of different taxonomic groups have been inferred from their sequence similarities. Until a few years ago, a more general classification of plant lectins was hampered by the lack of relevant sequence information. Fortunately, the rapid progress in the structural analysis of lectins and the molecular cloning of lectin genes has provided detailed sequence information of close to 100 plant lectins. Analysis of the available sequences distinguishes seven families of evolutionary related proteins. Four of these families, namely, the legume lectins, the monocot mannose-binding lectins, the chitin-binding lectins composed of hevein domains, and the type 2 RIP, comprise numerous members. In contrast, the jacalin-related lectins, the amaranthin lectin family, and the Cucurbitaceae phloem lectins are at this moment only small protein families. Most of the currently known plant lectins can be classified in one of the seven lectin families. Some lectins, however, do not fit the classification system or cannot be classified because there is no sequence information available. In the next part of this contribution, the seven lectin families are critically reviewed. The description of each lectin family starts with a brief historical note followed by a general overview of the occurrence, molecular structure, and amino acid sequences of the members of each lectin family. Subsequently, the state of the art of the lectin gene structure and the conversion of the primary translation products into the mature lectin molecules is given. Furthermore, structural and functional aspects of the lectins are highlighted in a few headings describing the carbohydrate-binding specificity and the three-dimensional structure, respectively, of the lectins. If relevant, a section dealing with lectin-related proteins with nonfunctional carbohydrate-binding sites is included. After a general discussion of the molecular evolution of each lectin family, there is a brief critical note on the physiological role(s) of the different members belonging to the lectin family.

IV. LEGUME LECTINS

The term ‘legume lectin’ refers to a particular type of plant lectin that is found exclusively in the Leguminoseae. It should be emphasized that not all lectins found in legume species belong to the legume lectins. For example, the class I chitinase from bean (*Phaseolus vulgaris*) and the seed lectins from *Abrus precatorius* belong to the chitin-binding lectins and type 2 RIP, respectively. Legume lectins are not only the most famous lectin family but also played a determining role in the development of lectinology as a scientific discipline. In addition, several legume lectins have become indispensable tools in biological and biomedical research and are widely used as bioactive proteins in

research and medicine (Van Damme et al., 1998).

A. Historical Note

The history of legume lectins starts in 1890 with the report of Power and Cambier on the isolation of a presumed toxic lectin called 'robin' from the bark of the legume tree black locust (*Robinia pseudoacacia*) (Kocourek, 1986). In a subsequent paper, Landsteiner and Raubitschek (1907) described the presence of nontoxic lectins in the legumes *Phaseolus vulgaris* (bean), *Pisum sativum* (pea), *Lens culinaris* (lentil), and *Vicia sativa* (vetch) and demonstrated for the first time that lectins are not necessarily toxic proteins. The subsequent discovery of lectins in various legume seeds had a profound influence on further development of lectinology since from then on most of lectin research was focused on legume seed lectins. As a result, all major breakthroughs in plant lectin research have been achieved with legume lectins. For example, the first evidence that lectins are carbohydrate-binding proteins follows the observation that cane sugar inhibited the agglutination activity of ConA (Summer and Howell, 1936). Similarly, the discovery of blood group specificity (and the subsequent introduction of the term lectin) is based on the observation that extracts from some legume seeds specifically agglutinate blood cells from some blood groups within the ABO system without affecting cells from other groups (Renkonen, 1948; Boyd and Reguera, 1949). The concept of lectins as bioactive proteins also resulted from the observation that *Phaseolus vulgaris* seed lectin triggered quiescent, nondividing lymphocytes into a derepressed state of active growth and proliferation (Nowell, 1960). It should be emphasized that most of the pioneering work in the field of biochemistry, physiology, and molecular biology of plant lectins has been achieved with legume lectins. For example, ConA was the first plant lectin to be purified and crystallized (Summer and Howell, 1936) and was also the first lectin whose primary structure and three-dimensional structure were resolved (Edelman et al., 1972; Hardman and Ainsworth, 1972). Similarly, the first plant lectin gene was isolated from seeds of soybean (Vodkin et al., 1983).

B. Occurrence, Molecular Structure, and Amino Acid Sequences

Legume lectins are a large family of closely related proteins that heretofore have been found exclusively in representatives of the Leguminoseae (Fabaceae) (Sharon and Lis, 1990). At present about 100 legume lectins have been isolated from over 70 different species belonging to various taxonomic groups (Table 1). Most legume lectins have been isolated from mature seeds. Typical legume lectins usually account for 1 to 10% of the total soluble seed protein. In some species higher lectin concentrations (up to 50%) have been observed, whereas in others the lectin represents less than 0.1% of the total seed protein. Several legumes contain two or more different seed lectins. Taking into consideration that some seed lectins are very rare proteins, the occurrence of multiple lectins in legume seeds is probably the rule rather than the exception. Legume seed lectins are predominantly located in the storage parenchyma cells of the cotyledons and to a lesser extent in the primary axes. Localization studies further indicated that legume seed lectins are typically sequestered in storage protein vacuoles (also known as protein bodies) (Etzler, 1986).

Legume lectins have also been found in different vegetative tissues. Although the lectin concentration is usually (very) low in leaves, stems, roots and root nodules, lectins

TABLE 1
Legume Lectins: Occurrence, Molecular Structure, and Specificity

Species	Tissue	Structure[a]	Specificity	Sequence available[b]
Amphicarpaea bracteata	Seed	$[P32,30,28,27]_4$	GalNAc	
Arachis hypogaea	Seed	$[P27]_4$	Gal	Pr, Nu (PNA)
	Nodule	$[P27]_4$	Gal	Nu (NGL)
	Nodule	$[P28]_2$	Man/Glc	Nu (NML)
	Root	I $[P33]_4$	Gal	
	Root	II $[P33]_2$	Man/Glc	
Bauhinia purpurea	Seed	$[P32]_4$	GalNAc	Pr, Nu (BPA)
Bowringia milbraedii	Seed	$[P(13 + 12)]_2$	Man/Glc	Pr (BMA)
Butea monosperma	Seed	$[P32, 34]_2$	GalNAc	
Canavalia brasiliensis	Seed	$[P26]_4$	Man/Glc	Nu
Canavalia ensiformis	Seed	$[P26]_4$	Man/Glc	Pr, Nu (ConA)
Canavalia gladiata	Seed	$[P30]_4$	Man/Glc	Nu
Caragana arborescens	Seed	I $[P30]_4$	GalNAc	
	Seed	II $[P30]_2$	Gal	
Cicer arietinum	Seed	$[P26]_2$	Complex	
Cladrastis lutea	Bark	I $[P(17 + 15)]_4$	Man/Glc	Nu (CLA)
	Bark	II $[P30]_4$	Man/Glc	Nu
Crotalaria juncea	Seed	$[P30]_4$	Gal>GalNAc	
Crotalaria striata	Seed	$[P30]_4$	GalNAc>Gal	
Cytisus multiflorus	Seed	I $[P42]_2$	(GlcNAc)n	
	Seed	II $[P42]_2$	Gal	
Cytisus scoparius	Seed	I$[P30]_4$	Gal/GalNAc	
	Seed	II $[P30]_4$	GalNAc>Gal	Pr (CS-II)
Cytisus sessilifolius	Seed	I $[P27]_4$	(GlcNAc)n	Pr (CSA-I)
	Seed	II $[P34]_2$	Gal	Pr (CSA-II)
Dioclea grandiflora	Seed	$[P25]_4$	Man/Glc	Pr (DGL)
Dolichos biflorus	Seed	$[P28]_4$	GalNAc	Nu (DbL)
	Plant	$[P30]_2$	Complex	Nu
Erythrina corallodendron	Seed	$[P30]_2$	GalNAc>Gal	Pr, Nu (EcorL)
Erythrina cristagalli	Seed	$[P26, 28]_2$	GalNAc	
Falcata japonica	Seed	$[P34]_4$	GalNAc	
Galactia tashiroi	Seed	$[P24]_4$	GalNAc	
Glycine max	Seed	$[P30]_4$	GalNAc>Gal	Nu (SBA)
Griffonia simplicifolia	Seed	I-A4 $[P32]_4$[c]	GalNAc	
	Seed	I-B4 $[P33]_4$[c]	Gal	
	Seed	II $[P30]_4$	GlcNAc	
	Seed	IV [P29 + P27]	Fucose	
	Leaf	I-A4 $[P32]_4$[c]	GalNAc	
	Leaf	I-B4 $[P33]_4$[c]	Gal	
	Leaf	II $[P30]_4$	GlcNAc	Nu (GS-II)
	Leaf	IV [P29 + P27]	Fucose	
Lablab purpureus	Seed	$[P(15 + 12)]_2$	Man/Glc	Pr (DLA)
Laburnum alpinum	Seed	I $[P32]_4$	GlcNAc	Pr (LAA)
	Seed	II $[P36]_2$	Gal	
Lathyrus ochrus	Seed	$[P(6 + 20)]_2$	Man/Glc	Pr (LOLI)
Lathyrus nissolia	Seed	$[P26]_2$	Man/Glc	Pr
Lens culinaris	Seed	$[P(6 + 18)]_2$	Man/Glc	Pr (LCA)
Lotononis bainesii	Seed	$[P32 + P35]_2$	Gal	
	Root	$[P32 + P35]_2$	Gal	
Lotus tetragonolobus	Seed	$[P27]_{2–4}$	Fucose	Pr (LTA)

TABLE 1 (continued)
Legume Lectins: Occurrence, Molecular Structure, and Specificity

Species	Tissue	Structure[a]	Specificity	Sequence available[b]
Maackia amurensis	Seed	$[P33]_4$	NANA	Pr, Nu (MAHs)
	Seed	$[P37]_4$	NANA	Nu (MALs)
	Bark	$[P32]_4$[c]	NANA	Nu (MAHb)
	Bark	$[P37]_4$[c]	NANA	Nu (MALb)
Macrotyloma axillare	Seed	$[P27]_4$	GalNAc	
Onobrychis viciifolia	Seed	$[P26]_2$	Man/Glc	Pr (OVL)
Ononis spinosa	Root	$[P31]_4$	GalNAc>Gal	
Phaseolus acutifolius	Seed	$[P29]_4$	Complex	
Phaseolus coccineus	Seed	$[P35]_4$	Complex	
Phaseolus lunatus	Seed	$[P30]_{2–4}$	GalNAc	Nu (LBL)
Phaseolus vulgaris	Seed	E $[P31]_4$[c]	Complex	Nu (PHA-E)
	Seed	L $[P31]_4$[c]	Complex	Nu (PHA-L)
	Seed	Pinto $[P28]_2$	Complex	Nu (RPbA1-B)
Pisum sativum	Seed	$[P(17 + 6)]_2$	Man/Glc	Pr, Nu (PSA)
Psophocarpus tetragonolobus	Seed	I $[P29]_2$	GalNAc	Pr, Nu (WBA)
	Seed	II $[P29]_2$	GalNAc	
	Tuber	$[P29]_2$	Gal/GalNAc	
Robinia pseudoacacia	Seed	I $[P34]_4$	Complex	Nu (RPsAI)
	Seed	II $[P29]_4$[c]	Complex	Nu (RPsAII)
	Bark	I-A $[P31]_4$[c]	Complex	Nu (RPbA1-A)
	Bark	I-B $[P29]_4$	Complex	Nu (RPbA1-B)
	Bark	II $[P26]_4$	Complex	Nu (RPbAII)
Sophora japonica	Seed	$[P30]_4$	GalNAc>Gal	Nu (SJAsg)
	Bark	$[P30]_4$	GalNAc>Gal	Nu (SJAbg)
	Bark	$[P(18 + 13]_4$	Man/Glc	Nu (SJAbmII)
	Leaf	$[P32]_4$	GalNAc>Gal	
	Leaf	$[P34]_4$	GalNAc>Gal	
Ulex europaeus	Seed	I $[P29, 31]_4$	Fucose	Pr (UEA-I)
	Seed	II $[P25]_4$	GlcNAc	Pr (UEA-II)
	Seed	III $[P34]_4$	Gal	
Vicia cracca	Seed	I $[P33]_4$	GalNAc	
	Seed	II $[P(17 + 6)]_2$	Man/Glc	
Vicia faba	Seed	$[P(6 + 21)]_2$	Man/Glc	Pr (VFA)
Vicia graminea	Seed	$[P25]_4$	GalNAc	
Vicia villosa	Seed	A $[P34]_4$[c]	GalNAc	
	Seed	B $[P36]_4$[c]	GalNAc	
Wistaria floribunda	Seed	$[P28]_4$	GalNAc	
	Seed	$[P32]_2$	Man/Glc	

[a] [PX] stands for protomer with a molecular mass of X kDa. [P(Y + Z)] indicates that the protomer is cleaved in two polypeptides of Y and Z kDa.
[b] Pr, protein sequence; Nu, nucleotide sequence. The abbreviation in brackets refers to the sequence name used in the dendrogram (Figure 8).
[c] The lectin is a mixture of isolectins composed of two different subunits that associate randomly.

are the most abundant proteins (representing 20 to 50% of the total soluble protein) in the bark of several legume trees (e.g., *Robinia pseudoacacia*, *Maackia amurensis*, and *Sophora japonica*). However, it should be noted that the lectin concentration in these tissues is developmentally regulated (Nsimba-Lubaki and Peumans, 1986; Baba et al., 1991). Vegetative tissues can also contain two or more different lectins. For example, the bark of *Robinia pseudoacacia* (Van Damme et al., 1995c) and *Sophora japonica* (Van Damme et al., 1997a) contains three different lectins that are encoded by different genes. It is noteworthy that in all cases studied thus far (i.e., *Robinia pseudoacacia*, *Maackia amurensis*, and *Sophora japonica)*, the bark and seed lectins are encoded by a different set of genes (Van Damme et al., 1995b). Localization studies have demonstrated that the abundant bark lectins are like their seed counterparts located predominantly in the storage parenchyma cells and, in addition, are also sequestered in subcellular organelles comparable to the seed storage protein vacuoles.

1. Molecular Structure

All legume lectins are built up of protomers* of approximately 30 kDa. Most protomers consist of a single polypeptide chain of about 250 amino acid residues and give rise to the so-called 'one-chain' legume lectins. In some instances the protomers are (partly) cleaved into two smaller polypeptides that, depending on the position of the cleavage site, are either similar or dissimilar in size. The legume lectins composed of such cleaved protomers are usually referred to as 'two chain' legume lectins.

Legume lectins are a unique group of lectins in that they contain divalent cations at specific metal-binding sites. Each subunit possesses Mn^{2+} and Ca^{2+} ions, which are essential for the carbohydrate-binding activity of the lectin. Furthermore, the amino acids involved in the binding of the Mn^{2+} and Ca^{2+} ions are highly conserved in all legume lectins.

Many, but certainly not all, legume lectins are glycosylated. Glycosylated lectin protomers usually possess one or two glycan chains. Some glycan chains are of the high-mannose type (e.g., the oligosaccharide side chain of the soybean lectin [Dorland et al., 1981; Nagai and Yamaguchi, 1993]), whereas others are of the complex type (Figure 2). Both high-mannose and complex type glycans may be present on a single lectin protomer (e.g., *Phaseolus vulgaris* hemagglutinin [PHA, Sturm and Chrispeels, 1986; Sturm et al., 1992]). Differences in glycosylation often result in the formation of glycoforms, which can mistakenly be considered as isolectins.

Native legume lectins are composed of two or four protomers held together by noncovalent interactions. The formation of di- and tetramers combined with the occurrence of both glycosylated and unglycosylated one- and two-chain protomers implies that legume lectins can occur in eight different molecular forms (Figure 3). In addition, heterotetrameric legume lectins are quite common. In their cells some legumes express two different lectin polypeptides that can associate with each other into heterotetrameric lectins. For example, the E and L subunits synthesized in the cotyledons of bean (*Phaseolus vulgaris*) seeds associate randomly, giving rise to the homotetrameric E_4 and L_4 isoforms and the heterotetrameric E_1L_3, E_2L_2, and E_3L_1 isolectins (Feldsted et al., 1977). Similar mixtures of isoforms are also found in the seeds of *Griffonia simplicifolia* and *Vicia villosa* as well as in the bark of *Sophora japonica*, *Maackia amurensis*, and *Robinia pseudoacacia* (Table 1). The simultaneous expression of two similar lectin polypeptides is not sufficient for the formation of heterotetrameric

* In this article a protomer is defined as the mature processing product(s) of the primary translation product of the lectin mRNA.

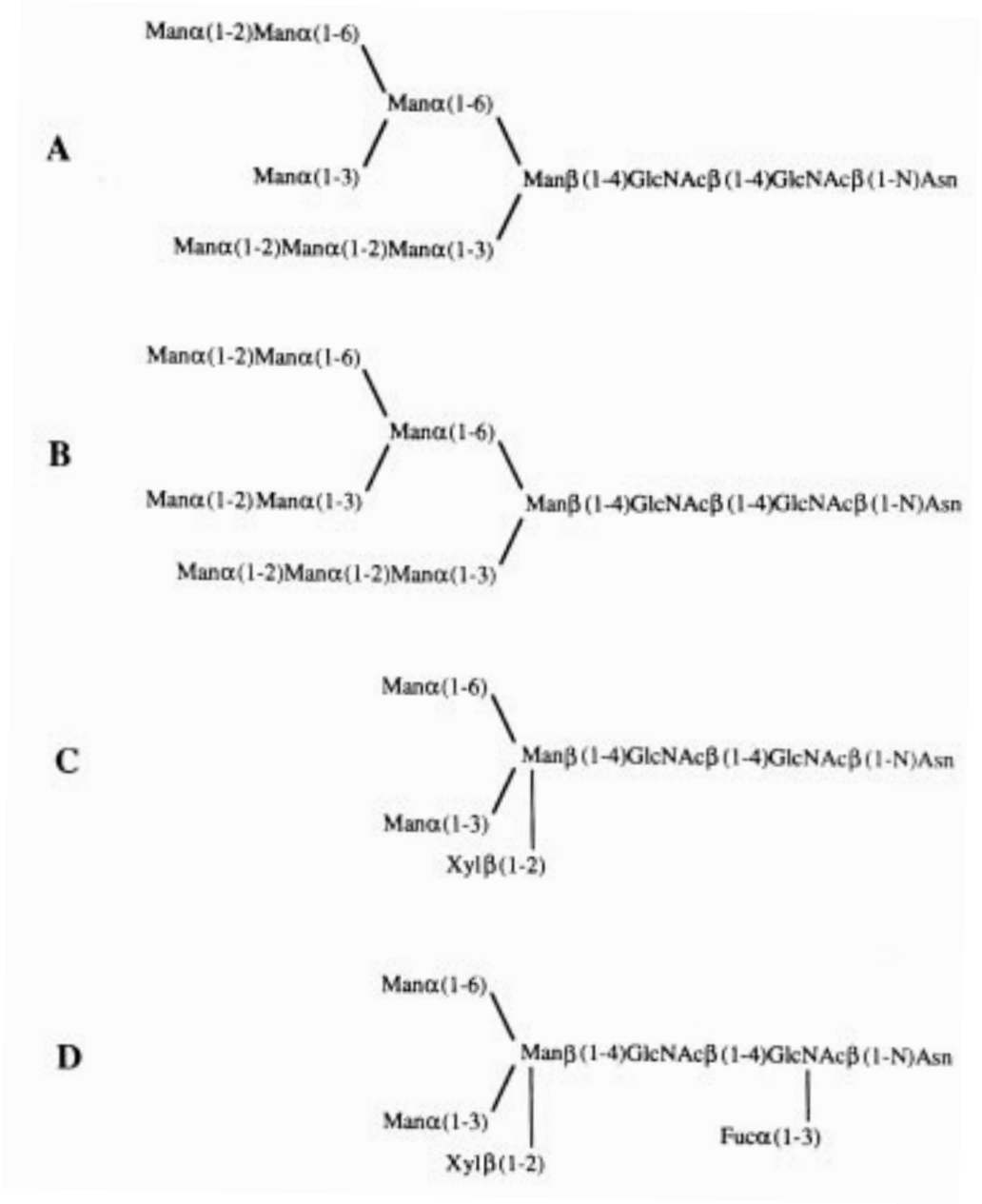

FIGURE 2. Schematic representation of the structure of *N*-glycan chains from plant lectins. Examples shown are (A) the high-mannose type *N*-glycans from kidney bean agglutinin (Sturm et al., 1992) and (B) soybean lectin (Dorland et al., 1981), (C) a complex type glycan from the kidney bean α-amylase inhibitor (Yamaguchi et al., 1992), and (D) a complex type glycan found on the lectins from mistletoe (ML-I) (Debray et al., 1992), *Robinia pseudoacacia* (Wantyghem et al., 1992), *Wisteria floribunda* (Ramirez-Soto et al., 1991), and on the kidney bean α-amylase inhibitor. (Yamaguchi et al., 1992).

lectins. This is clearly illustrated by the seed and bark lectins of *Maackia amurensis* and *Robinia pseudoacacia* (Van Damme et al., 1995b,c, 1997d; Yamamoto et al., 1994, 1997). Both legume trees express two highly homologous genes in their seeds as well as in their bark. Although the sequence identity between the respective seed and bark lectin

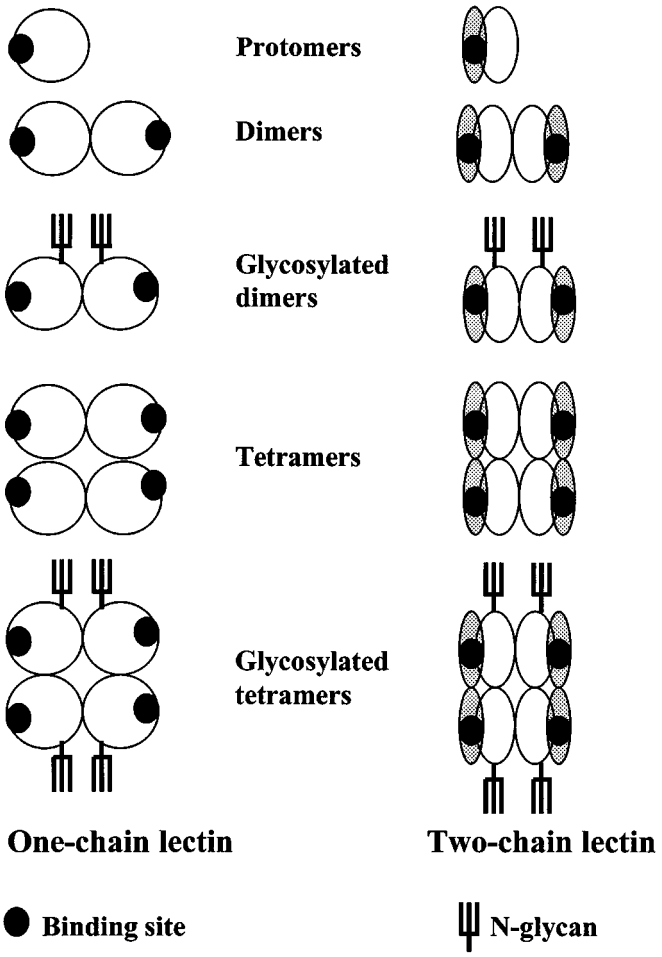

FIGURE 3. Schematic representation of the different molecular structures of native legume lectins.

polypeptides is virtually the same, the seeds contain exclusively homotetrameric lectins, whereas the bark contains both homotetrameric and heterotetrameric isoforms. Up to now no evidence has been presented for the occurrence of heterodimeric legume lectins (although they can at least in principle be formed through the same mechanism as the heterotetrameric lectins).

A few reports claim the presence of an interchain disulfide bridge in some legume lectins. Although there is little conclusive evidence at present, it cannot be precluded that some legumes contain lectins built up from disulfide bridge-linked subunits. For example, jequirity beans (*Abrus precatorius*) contain two lectins belonging to the type 2 RIP family, which consists of disulfide bridge-linked chains of 30 to 35 kDa possibly. Type 2 RIP also occur in some other legumes and have erroneously been identified as genuine legume lectins.

2. Primary Structure

Numerous legume lectins have been sequenced completely by chemical methods (Table 1). In addition, the amino acid sequences of many other lectins have been deduced from the nucleotide sequence of cDNAs or genomic DNA fragments. Table 1

gives an overview of the legume lectins that at present have been studied in some detail for what concerns their molecular structure and carbohydrate-binding properties. Although no sequence information is given, the table also indicates whether a lectin sequence is available. Detailed analyses and comparisons of the different lectin sequences have shown that despite the differences in, for example, carbohydrate-binding specificities among the legume lectins, they all share a high degree of sequence similarity over the entire length of the lectin protomer and contain highly conserved regions. This issue is discussed further below in the sections dealing with the three-dimensional structure and molecular evolution of the legume lectins.

C. Structure, Biosynthesis, and Posttranslational Modifications

The introduction of molecular biology greatly improved our basic knowledge of different aspects of legume lectins and their genes. At present, approximately 30 legume lectins have been cloned (Table 1). Sequence analysis of these genes not only yielded a wealth of information about the primary structure of many legume lectins but also helped to determine how the primary translation products of the lectin mRNAs are converted into the mature lectin polypeptides.

Studies of the biosynthesis, processing, and topogenesis of several legume lectins demonstrated that these proteins are synthesized on the endoplasmic reticulum (ER) as pre(pro)proteins and subsequently enter the secretory pathway. Cotranslational processing involves the removal of the signal peptide and in many cases also *N*-glycosylation. During or after intracellular transport from the ER via the Golgi apparatus to the storage protein vacuoles, further processing of both the polypeptide backbone and *N*-linked glycan chains takes place (Van Driessche, 1988).

Molecular analysis of legume lectin genes not only allowed refining the results of the biochemical and cellular biological experiments, but also solved some unanswered questions and led to some unexpected findings. The coding region of all legume lectin genes comprises an *N*-terminal signal sequence followed by a peptide of about 250 amino acid residues. These molecular data confirm that legume lectins are secretory proteins and provide unambiguous evidence that the protomers of both one- and two-chain legume lectins are derived from similar precursors. Comparisons of the amino acid sequences of the mature lectin polypeptide(s) and the deduced sequences of the cDNAs allowed unraveling the posttranslational processing of many legume lectins. A summary of the available data shows that the conversion of the primary translation products of the lectin mRNAs into the mature lectin polypeptide(s) can be achieved in different ways (Figure 4) (Van Damme and Peumans, 1996). In the simplest processing scheme the lectin mRNA encodes an unglycosylated prelectin that in a single step is converted into the mature lectin polypeptide by the cotranslational removal of the signal peptide and subsequently transported via the Golgi complex to the vacuoles. *N*-glycosylation of the prelectin makes the processing somewhat more complicated because then the cotranslational removal of the signal peptide is followed by trimming and modifications of the glycan chain(s). Many legume lectins are synthesized as preprolectins and require a more complex processing. Cotranslational cleavage of the signal peptide converts the preprolectin in the corresponding prolectin. This prolectin is subsequently transported via the Golgi to the vacuoles where it is converted into the mature lectin polypeptide by the removal of a *C*-terminal propeptide (Rini et al., 1987;

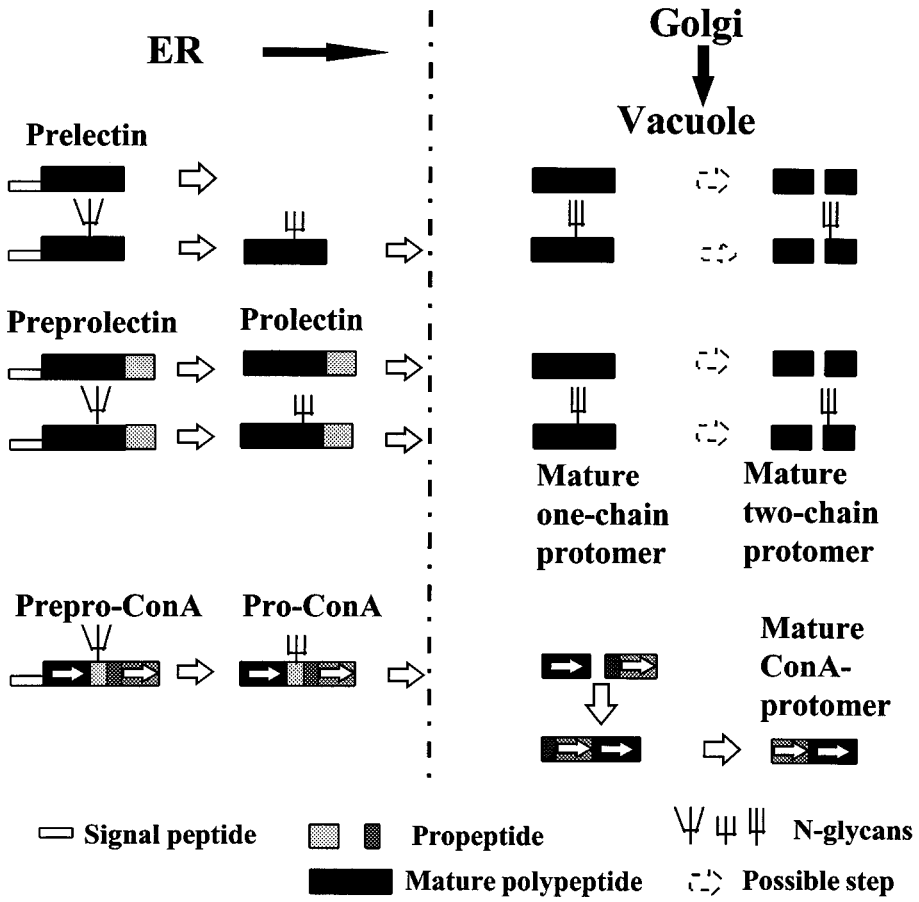

FIGURE 4. Schematic representation of the biosynthesis, co- and posttranslational modifications, and topogenesis of legume lectins.

Mandal et al., 1994). Evidently, this processing becomes even more complicated when the preprolectin is glycosylated. The above-described processing schemes summarize the biosynthesis and posttranslational modifications of one-chain lectins. Two-chain lectins are synthesized and processed similarly except that they undergo an additional proteolytic processing. Most members of the subclass of mannose/glucose binding legume lectins (e.g., *Pisum sativum, Lathyrus* sp., *Lens culinaris, Vicia* sp., *Cladrastis lutea,* and *Sophora japonica*) are cleaved into two smaller polypeptides by a specific vacuolar endoprotease. This proteolytic cleavage is in some cases accompanied by the removal of a short internal peptide (e.g., in the pea lectin [Higgins et al., 1983]). Depending on the position of the cleavage site, the resulting lectin polypeptides are either similar (e.g., the *Cladrastis lutea* lectin [Van Damme et al., 1995a]) or dissimilar (e.g., the *Pisum sativum* and *Vicia faba* lectins) in size. At present, it is not clear why two-chain lectins occur exclusively in the subclass of mannose/glucose-specific legume lectins. However, the simultaneous occurrence of one- and two-chain lectins in the cells of some legumes (e.g., in the bark of *Cladrastis lutea* and *Sophora japonica*) indicates that the cleavage of the lectin protomers is not determined by the presence or absence of a vacuolar protease, but depends on the proteins themselves (Van Damme et al., 1995a;

1997a). Most probably, 'cleavable' lectin protomers contain a sequence that is specifically recognized and cleaved by the endoprotease. It should be mentioned that some members of the subclass of mannose/glucose-specific legume lectins are single chain lectins, for example, *Lathyrus sphaericus* lectin (Richardson et al., 1987). On the other hand, the bark of *Cladrastis lutea* contains both a one- and a two-chain mannose/glucose-binding lectin that have a highly similar sequence but differ at the position where the two-chain lectin is cleaved (Van Damme et al., 1995a).

The biosynthesis and processing described above holds true for all legume lectins except for species belonging to the Diocleae tribe (comprising the genera *Canavalia* and *Dioclea*). Molecular cloning and biosynthesis studies of the latter lectins have revealed an unexpected and unique mode of protein processing that was first described for ConA (Carrington et al., 1985; Bowles et al., 1986). ConA is synthesized as an inactive glycosylated preprolectin, which is cotranslationally converted into a (inactive) glycosylated prolectin (proConA) by the removal of the signal peptide (Figure 4). In the next processing step an internal glycosylated peptide is excised from proConA. The resulting peptides are transposed and religated, and the processing is terminated by the removal of a tetrapeptide from the *N*-terminus of the ligation product. At present, the biological significance of this unusual processing is not fully understood. Possibly, this mechanism of temporary inactivation of the carbohydrate-binding activity of the lectin evolved to prevent unwanted interaction of the lectin with mannosylated glycoproteins in the ER.

Molecular analyses showed that the expression of most legume lectins is controlled by one or a few genes. However, as many legumes contain two or more lectins in different tissues and some lectins are encoded by different genes, extended lectin gene families are common in legumes. A further discussion of the legume lectin gene families necessitates making a clear distinction between 'identical' and 'different' lectins. In practice, there is no consensus about the use of the terms 'isolectins' and 'different lectins'. When a legume contains multiple lectins, those with the same or a similar specificity are usually considered 'isolectins', whereas lectins with a different specificity are regarded as truly different. However, in many cases the term isolectin is used in a broader sense and refers to lectins with a different specificity. For example, the E_4 and L_4 isoforms of PHA have a different specificity and exhibit different biological activities. Strictly speaking, lectins can be considered true isoforms only if they are derived from protomers with identical amino acid sequences. Differences in glycosylation and/or posttranslational modifications of the polypeptide backbone can generate so-called glycoforms or isoforms, respectively. All other lectins, irrespective of their sequence similarities, should be regarded as truly different proteins.

As mentioned above, the complexity of the lectin gene families varies according to the species. Pea seed lectin, for instance, is encoded by a small family of related genes, of which only one is functional (Kaminski et al., 1987). This gene is expressed in the seeds (where different posttranslational processing gives rise to two different isoforms of the pea seed lectin) as well as in the roots (Hoedemaeker et al., 1994). In soybean each genetic line contains two lectin genes, one of which encodes the seed lectin, whereas the function of the second gene is unknown (Goldberg et al., 1983). It was shown that in lectin-minus lines an insertion element blocks the expression of the lectin gene (Okamuro and Goldberg, 1992). Some legumes (e.g., *Griffonia simplicifolia*, *Ulex europaeus*) contain two or more seed lectins with different

specificities and amino acid sequences that are encoded by different genes or small families of genes. Many legumes also express one or more lectins in (a) specific vegetative tissue(s). There are some well-documented examples of legumes in which the seed and vegetative lectins are clearly encoded by different genes. For example, *Maackia amurensis* trees accumulate a hemagglutinating and a leukoagglutinating Neu5Acα(2,3)GalNAc-binding lectin in the seeds (Yamamoto et al., 1994, 1997) as well as in the bark (Van Damme et al., 1997d). Both the seed and bark lectins are controlled by two different genes, which implies that the *Maackia amurensis* lectin gene family comprises at least four members. Similarly, *Robinia pseudoacacia* contains two distinct seed lectins and three different bark lectins that are encoded by different genes (Van Damme et al., 1995b,c). These examples clearly illustrate the occurrence of complex lectin gene families in legumes but do not allow drawing general conclusions because only a limited number of lectins have been studied in detail. However, considering the fact that many legumes contain two or more seed lectins, and evidence is accumulating that numerous legumes express lectins in vegetative tissues, one can reasonably expect that extended lectin gene families are common in legumes. Until now, no introns have been reported in legume lectin genes.

D. Carbohydrate-Binding Specificity

Legume lectins strongly differ from each other with respect to their carbohydrate-binding specificity. A brief overview of the sugar-binding specificities of the currently known plant lectins indicates that the legume lectins cover a much broader range of specificities than any other lectin family (Table 1). Most of the so-called specificity groups, which are distinguished on the basis of the preferential binding of lectins to monosaccharides, are represented in the legume lectin family. Only mannose-binding and mannose/maltose-specific lectins have not been identified (yet) in legumes. It is also worth noting that fucose-specific and mannose/glucose-binding plant lectins have only been found in legumes. Moreover, almost all lectins with a so-called 'complex' specificity are typical legume lectins. The broad specificity range is at first sight difficult to reconcile with the high sequence similarities and the conserved structure of the legume lectin monomers. Recent studies have demonstrated, however, that substitutions of a few amino acids (which are involved in sugar binding activity) and variations in the length of a particular loop profoundly change the structure of the binding site without affecting the overall three-dimensional structure of the protomer (Young and Oomen, 1992; Sharma and Surolia, 1997). Most probably, the legume lectin family is the result of an evolution toward a diversity of carbohydrate-binding specificities. The question remains what was and possibly still is the driving force behind this evolution.

E. Occurrence of Lectin-Related Proteins

Molecular studies of legume lectins also revealed the occurrence of proteins, which are clearly related to the genuine lectins but are devoid of carbohydrate-binding activity. A first set of lectin-related proteins have been identified in the common bean (Mirkov et al., 1994). Studies of the *Phaseolus vulgaris* hemagglutinin family have shown that bean seeds express in addition to the genes encoding the E and L subunits of PHA (Hoffman and Donaldson, 1985) genes of at least two related proteins. One of these proteins has been identified as the bean α-amylase inhibitor (Hoffman et al., 1982) and the other as the so-called arcelin (Osborn et al., 1988a). According to more detailed reports,

several types of arcelins exist. Both the α-amylase inhibitor and the arcelins are dimeric proteins that share a high sequence similarity with the PHA polypeptides but are devoid of sugar-binding activity. Sequence comparisons indicated that the lack of carbohydrate-binding activity results most probably from deletions in the genes encoding the α-amylase inhibitor and arcelins (Mirkov et al., 1994). The apparent absence of active sugar-binding sites does not imply that the α-amylase inhibitor and arcelins have no biological activity. On the contrary, the bean α-amylase inhibitor is a potent insecticidal protein (Huesing et al., 1991c; Shade et al., 1994), and there are also indications that the arcelins have a protective effect against seed predating insects (Osborn et al., 1988a). It was shown that the genes encoding kidney bean lectins and lectin-like proteins are closely linked on the same chromosome (Moreno and Chrispeels, 1989; Chrispeels and Raikhel, 1991; Nodari et al., 1993).

Another lectin-related protein has been identified in *Cladrastis lutea* (Van Damme et al., 1995a). Protein purification and cDNA cloning have shown that one of the predominant bark proteins is a tetramer composed of subunits that show 50% sequence identity with the mannose/glucose-binding bark lectin but are devoid of carbohydrate-binding activity. The lack of sugar-binding activity is most probably due to the disruption of the normal binding site by the insertion of three extra amino acid residues. It is not known whether the *Cladrastis lutea* lectin-related protein has any biological activity.

F. Three-Dimensional Structure

1. Overall Three-Dimensional Structure

ConA was the first lectin to be crystallized (Summer and Howell, 1936) and analyzed by X-ray diffraction (Edelman et al., 1972; Hardman and Ainsworth, 1972). Subsequently, other seed lectins were successfully crystallized and their three-dimensional conformations solved at atomic resolution. They are in chronological order: PSA from *Pisum sativum* (Einspahr et al., 1986), favin from *Vicia faba* (Reeke and Becker, 1986), LoLI from *Lathyrus ochrus* (Bourne et al., 1990a), GS-IV from *Griffonia simplicifolia* (Delbaere et al., 1990), EcoL from *Erythrina corallodendron* (Shaanan et al., 1991), LCA from *Lens culinaris* (Loris et al., 1993), PNA from *Arachis hypogaea* (Banerjee et al., 1994), SBA from *Glycine max* (Dessen et al., 1995), PHA-L from *Phaseolus vulgaris* (Hamelryck et al., 1996a), ConBr from *Canavalia brasiliensis* (Sanz-Aparicio et al., 1997), the isolectin VVL-B4 from *Vicia villosa* (Osinaga et al., 1997), and WBA-I from *Psophocarpus tetragonolobus* (Pabru et al., 1998). In addition, molecular modeling was performed to predict the three-dimensional structure of the vegetative lectin Blec1 from *Pisum sativum* (Pak et al., 1992), the seed lectin DbL from *Dolichos biflorus* (Imberty et al., 1994), the bark and seed lectins from *Robinia pseudoacacia* (Van Damme et al., 1995b,c) and *Sophora japonica* (Van Damme et al., 1997a), and the bark lectin from *Cladrastis lutea* (Van Damme et al., 1995a).

Basically, all the legume lectin protomers, consisting either of a single chain (ConA, GS-IV, EcoL, PNA, SBA, PHA-L, VVL-B4) or two chains (lectins from the tribe Vicieae, except for isolectin VVL-B4 from *Vicia villosa*), show a very similar three-dimensional conformation. They are built from a curved seven-stranded β-sheet (front face) and a flat six-stranded β-sheet (back face) interconnected by turns and loops to form a flattened dome-shaped structure (Plate 1).* The orientation of the antiparallel strands of β-sheet changes from one face to the other, thus creating a very rigid and strong struc-

* Plate 1 appears following page 602.

ture that can partly explain the extreme resistance of legume lectins toward many proteolytic enzymes. Four loops located at the upper part of the dome form the monosaccharide-binding site of the lectin monomer that is responsible for the binding of simple sugars. As there are almost no helical structures, all legume lectin protomers belong to the class of beta-proteins (Richardson, 1981). Interestingly, a *cis*-peptide bond involving an aspartic residue (Asp^{81} in LoLI) occurs along the peptide chain that might orient the side chain of this residue, allowing it to interact with a Ca^{2+} ion. Another divalent cation, Mn^{2+} , is also linked to the polypeptide. Because demetallization of the lectins (e.g., ConA) causes loss of carbohydrate-binding activity, it is evident that both divalent cations are involved in the functional conformation of the monosaccharide binding site (Agrawal and Goldstein, 1968). Legume lectin protomers also contain a hydrophobic cavity, which was first described for ConA (Becker et al., 1975). The cavity consists mainly of conserved hydrophobic residues and may be responsible for the noncovalent binding of the plant hormone auxin (Edelman and Wang, 1978).

As mentioned above, some legume lectins are glycosylated. In some cases the glycan chains stabilize the noncovalent association of the lectin protomers. For example, the dimeric assembly of the bean α-amylase inhibitor is stabilized by direct and water-mediated hydrogen bonds between the GlcNAc-residues on Asn^{12} of both protomers. In other cases the glycan chains prevent the noncovalent association of the protomers. For instance, the glycan chains exposed on the surface of the protomers of the bean arcelins should prevent the dimers from associating into tetramers.

Most legume lectins consist of two protomers that interact with each other in a twofold symmetry plane. This most common mode of dimerization creates a 12-stranded β-sandwich in which the two facing monomers associate by their flat bottoms (Figure 5A). Accordingly, the loops forming the monosaccharide-binding sites are located at both ends of the dimer. This very rigid structure exhibits a front wall of 14 antiparallel strands of β-sheet associated to a back wall of 12 antiparallel strands of β-sheet. Glycan chains can prevent the formation of this canonical divalent structure. For example, in EcoL the two lectin monomers are associated differently (Shaanan et al., 1991). In addition, the dimeric association in GS-IV (Delbaere et al., 1990) and PNA (Banerjee et al., 1994) is similar but distinct from that of other legume lectins, even though PNA is not a glycoprotein and therefore interactions involving covalently linked glycan chains are apparently not always responsible for the different dimeric arrangements of the lectin monomers.

Whereas most of the lectins from the Vicieae tribe (PSA, LCA, LoLI) are bivalent dimers, many other legume lectins form tetrameric structures with four monosaccharide-binding sites. The noncovalent association of two dimers can be achieved in different ways. Like most other homotetrameric proteins, tetrameric lectins possess a 222 (D2) symmetry except for PNA, which displays an unusual quaternary arrangement exhibiting neither a 222 (D2) nor a fourfold (C4) symmetry (Banerjee et al., 1994). In PHA-L (Hamelryck et al., 1996a), two canonical dimers become associated through their back walls in such a way that two monosaccharide-binding sites occur at both sides of the tetramer (Figure 6A). The central cleft delineated by this type of association probably plays a major role in the binding of adenine and adenine-related plant hormones by both PHA-L and PHA-E (Hamelryck et al., 1996a). In this case, the glycan chains occurring at the surface of both monomers cannot prevent the association of the dimers because they are directed to the solvent.

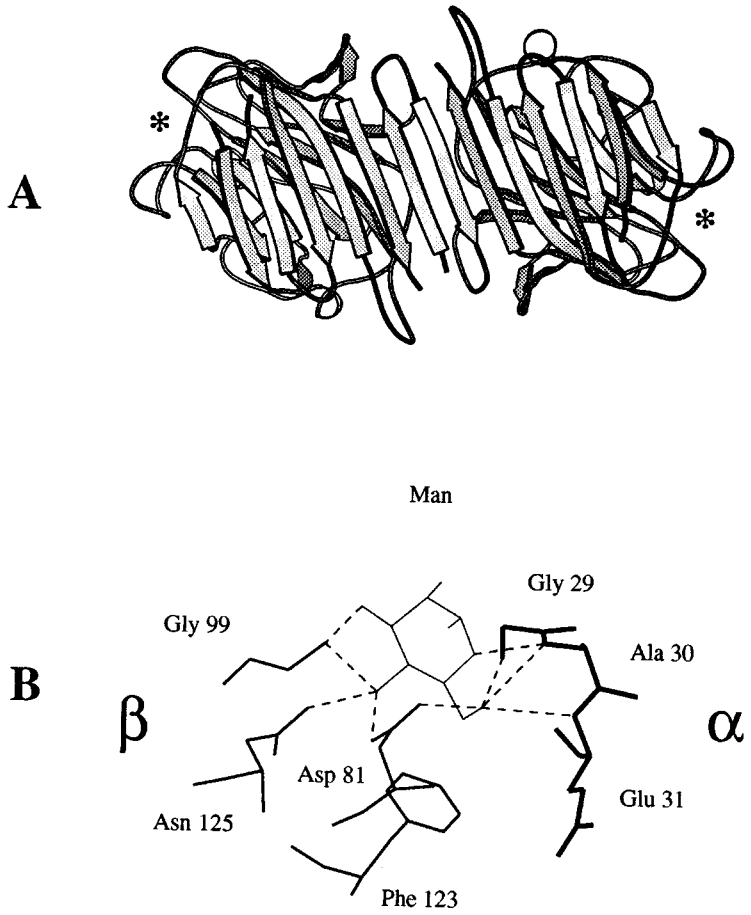

FIGURE 5. (A) Dimer of LoLI (code 1loe, Brookhaven PDB) showing the canonical twelve-stranded β-sandwich structure that consists of 12 strands of β-sheet forming the flattened back wall of the dimer (foreground) connected by turns and loops to the 14 strands of β-sheet forming the curved front wall (background) of the dimer. Asterisks indicate the location of the monosaccharide-binding sites located at both ends of the LoLI dimer. This cartoon was generated from the X-ray coordinates of LoLI using Molscript (Kraulis, 1991). (B) Monosaccharide-binding site of LoLI showing a mannose residue (thin line) connected to the amino acid residues forming the binding site by a network of nine hydrogen bonds (dot lines). Amino acid residues belonging to the α-chain (heavy line) and β-chain (thick line) of LoLI are indicated. Phe[123] interacts with mannose via hydrophobic contacts.

2. Structural Bases for the Recognition of Simple or Complex Sugars

Legume lectins can interact with both simple and complex sugars (Drickamer, 1995; Weis and Drickamer, 1996). The interaction with simple sugars and sugar derivatives, which is responsible for the broad carbohydrate-binding specificity of the lectins, for example, the mannose/glucose-binding specificity of ConA or LoLI, depends on the presence of a monosaccharide-binding site at the surface of the lectin monomer. Co-crystallization of ConA with D-manno-pyranoside (Derewenda et al., 1989) and of LoLI with D-mannose and D-glucose (Bourne et al., 1990b) has shown that a few amino acid residues belonging to four loops located at the top of the dome-shaped lectin monomer form a monosaccharide-binding site responsible for this specific

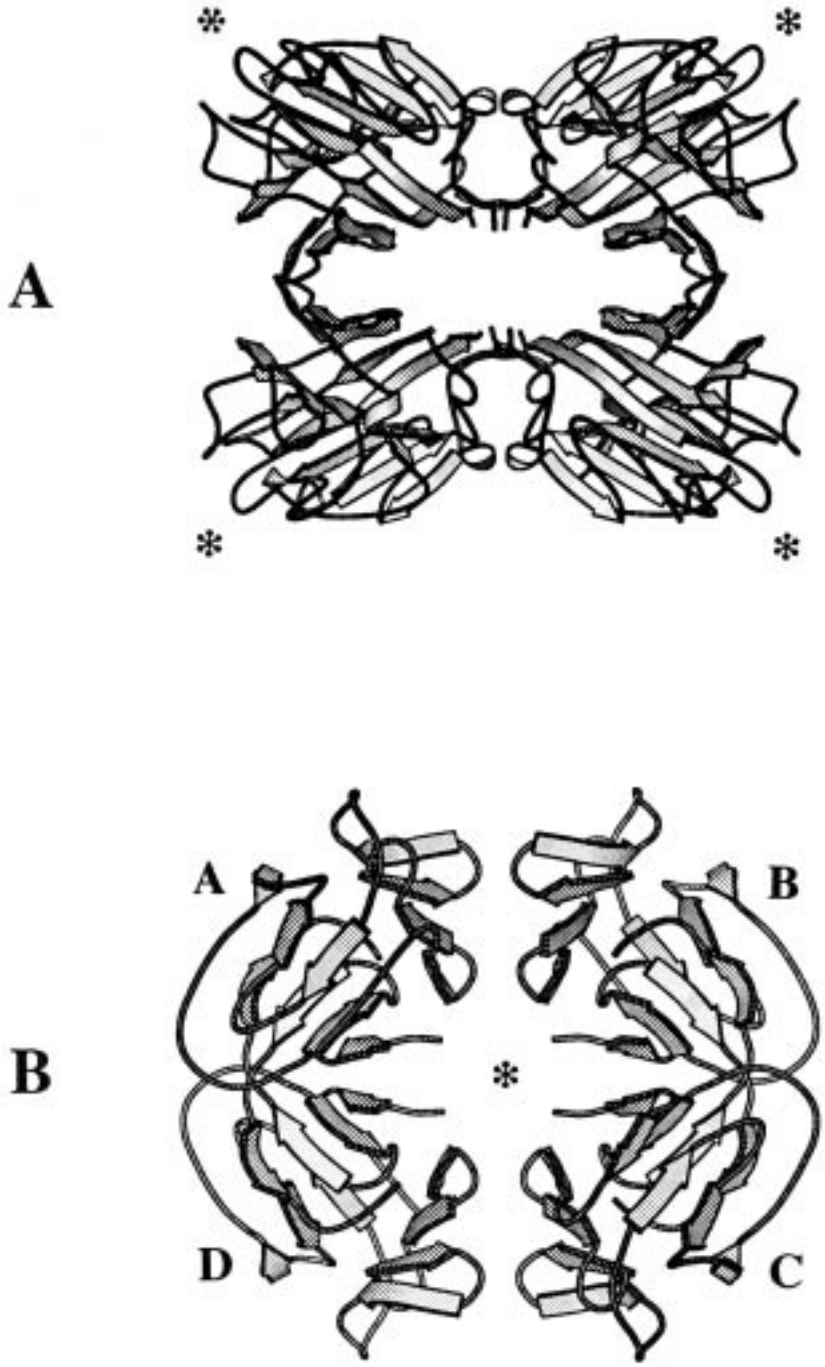

FIGURE 6. (A) Tetrameric structure of PHA-L (code 1fat, Brookhaven PDB). The noncovalent association of two dimers (upper and lower structures) creates a central cleft with exposed hydrophobic residues. Both dimers are associated by their back faces allowing exposition of the binding-sites (asterisks) toward the solvent. (B) Tetrameric structure of GNA (code 1msa, Brookhaven PDB) showing the association by hydrophobic contacts of two dimers A-D (left) and B-C (right) into a flattened crown-shaped structure with a central solvent channel (asterisk). In both dimers A-D and B-C, the noncovalent associations of monomers result from mainly hydrogen bonds and exchange of the *C*-termini of the monomers. Both cartoons were made using Molscript. (From Kraulis, 1991.)

recognition. These residues (Asp^{81}, Gly^{99}, Asn^{125}, Gly^{29}, Ala^{30}, and Glu^{31} in LoLI) create a network of seven hydrogen bonds with O3, O4, O5, and O6 of the simple sugar (Figure 5B). In addition, an aromatic residue (Phe^{123} in LoLI) creates a hydrophobic interaction with the pyranose ring of the sugar. A very similar binding was shown to mediate the interaction of PSA with a trimannoside (Rini et al., 1993). The specific binding of the galactosyl unit of lactose to EcoL (Shaanan et al., 1991) follows a similar mechanism, although the orientation of the pyranose ring into the binding site and the binding to O4 are different. This can explain the different monosaccharide-binding specificity of EcoL (Gal/GalNAc-specific) compared with ConA, PSA, or LoLI (Man/Glc-specific). Site-directed mutagenesis experiments performed on PSA (Van Eijsden et al., 1992), PHA (Mirkov and Chrispeels, 1993), and EcoL (Adar and Sharon, 1996), have shown that two of the residues forming the monosaccharide-binding site Asp^{81} and Asn^{125} in PSA are extremely important, as their replacement by another residue prevents the binding to any simple sugar. However, even though PHA-E contains all the amino acid residues forming the monosaccharide-binding site, moderate amounts of simple sugars cannot inhibit the hemagglutination of PHA-E with erythrocytes (Goldstein and Poretz, 1986).

In addition to the interaction with simple monosaccharides, legume lectins also interact with more complex glycans, such as the oligosaccharide side chains of glycoproteins. This binding, which is responsible for the fine carbohydrate-binding specificities of legume lectins, is much more specific and discriminates lectins with similar monosaccharide-binding specificities (Debray et al., 1981; Kornfeld et al., 1981). For example, biantennary glycans of the *N*-acetyllactosaminic type carrying a fucose residue α(1,6)-linked to the first GlcNAc attached to the Asn residue of the polypeptide chain, are the best inhibitors of the Man/Glc-specific lectins from the Vicieae tribe, e.g., LCA, PSA, LoLI, or LoLII (Debray and Rougé, 1984). In contrast, ConA, which is also a Man/Glc-specific lectin, does not discriminate between fucosylated and nonfucosylated glycans. X-ray analysis of LoLII complexed to a glycopeptide N2 isolated from human lactotransferrin revealed that the fucosylated biantennary dodecasaccharide of the *N*-acetyllactosaminic type binds to an extended carbohydrate-binding site (Bourne et al., 1994). This extended site comprises in addition to a central monosaccharide-binding site, a number of adjacent residues at the surface of the lectin that interact with several sugar units of both antennae of the complex glycan. The α(1,3)Man residue from the inner trimannoside core of the glycan binds to the monosaccharide-binding site, while other sugar units interact via hydrogen bonds with other amino acid residues located in the vicinity of this site. Many of these hydrogen bonds are mediated by ordered water molecules present at the surface of the lectin. In addition, several hydrophobic interactions involving surface-exposed aromatic residues play a key role in the binding. The α(1,6)-linked fucose residue is important because it allows the best presentation of the glycan moiety to the extended binding site of the lectin.

3. Three-Dimensional Structure of Lectin-Like Proteins from Kidney Bean

Molecular modeling of arcelin-1 and α-amylase inhibitor-1 using the coordinates of typical legume lectins indicated that these proteins lack one and two loops, respectively, that are required for the formation of the monosaccharide-binding site of legume

lectins (Rougé et al., 1993). The predicted lack of active carbohydrate binding sites is in good agreement with previous observations showing that arcelins are only weakly active and α-amylase inhibitor is completely inactive in agglutination tests with pronase-treated erythrocytes (Osborn et al., 1988b).

The three-dimensional structure of arcelin-5 from *Phaseolus vulgaris* was solved recently (Hamelryck et al., 1996b). Although arcelin-5 has been described as a dimer, the structure resolved by X-ray crystallography corresponds to a monomer (Goossens et al., 1994). In contrast, the three-dimensional structure of arcelin-1 from *Phaseolus vulgaris* (Mourey et al., 1998) shows a dimeric organization very similar to the canonical β-sandwich of legume lectins (Figure 7A). However, one of the loops forming the monosaccharide-binding site of lectins is lacking and other loops located in its vicinity are rearranged. Thus, the Asn residue ho-

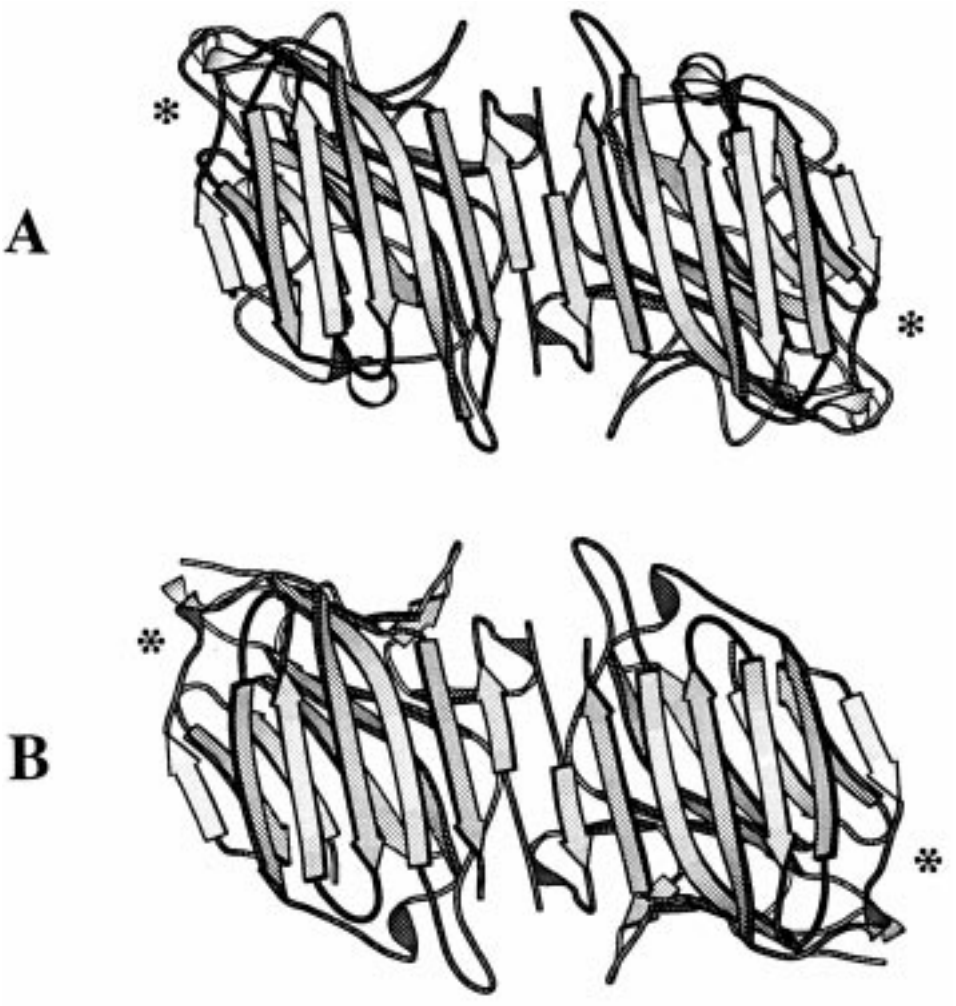

FIGURE 7. Dimeric structures of arcelin-1 (A) (code 1avb, Brookhaven PDB) and α-amylase inhibitor-1 (B) (code 1dhk, Brookhaven PDB) from kidney bean seeds. Both lectin-like proteins exhibit the canonical 12-stranded β-sandwich structure of legume lectins, but the dimers are truncated because they lack one (arcelin-1) or two (α-amylase inhibitor-1) loops that are located at both ends of the dimer and form the monosaccharide-binding site of typical lectins (asterisks). As a result, the monosaccharide-binding sites of these proteins are incomplete and not functional. Both cartoons were generated using Molscript. (From Kraulis, 1991.)

mologous to Asn[125] of LoLI or PSA, which was identified as a very important residue for the monosaccharide-binding activity, is deleted and the resulting incomplete monosaccharide-binding site is no longer reactive. Accordingly, none of the simple sugars and sugar derivatives used in hapten inhibition experiments can inhibit the weak hemagglutinating activity of arcelin-1 (Fabre et al., 1998). However, the weak binding activity of arcelin-1 toward various glycoproteins, for example, fetuin, asialofetuin, or thyroglobulin, suggests that arcelin-1 still contains a weakly active extended carbohydrate-binding site even though its monosaccharide-binding site is nonfunctional. The three-dimensional structure of arcelin-1 shows a disulfide bond between Cys[144] and Cys[180]. In addition, each arcelin-1 monomer possesses three *N*-glycosylation sites (Asn[12], Asn[68], Asn[107]) that are occupied by glycan chains of the high-mannose and complex types (Fabre et al., 1998; Mourey et al., 1998). One of them, linked at Asn[68], protrudes on the back face of the arcelin-1 monomer and could prevent the noncovalent association of two dimers into a tetramer by creating a steric hindrance between both dimers.

The three-dimensional structure of α-amylase inhibitor-1 in a complex with porcine pancreatic α-amylase has been solved by X-ray analysis at 1.9 Å (Bompard-Gilles et al., 1996). The inhibitor is a dimer built from the noncovalent association of two monomers that exhibit the canonical β-sandwich structure of typical legume lectins (Figure 7B). Like arcelin-1, the dimer is truncated at both ends because two loops belonging to the monosaccharide-binding site are lacking. As a result, α-amylase inhibitor-1 lacks two out of seven amino acid residues forming the carbohydrate-binding site. Because no hemagglutinating activity can be detected for this protein, the extended carbohydrate-binding site is nonfunctional. During the posttranslational processing of α-amylase inhibitor-1, a proteolytic cleavage occurs at Asn[77] resulting in two polypeptides of 10.8 and 15 kDa, respectively, which are associated by noncovalent bonds in the α-amylase inhibitor-1 monomer. This cleavage could be responsible for some of the conformational changes that occur in the vicinity of what remains of the monosaccharide-binding site and are required for the interaction with α-amylase (Pueyo et al., 1993). Like arcelin-1, α-amylase inhibitor-1 contains three *N*-glycosylation sites (Asn[12], Asn[65], Asn[140]) with glycans belonging to the high-mannose and complex types (Bompard-Gilles et al., 1996). The branched glycans projecting out from Asn[12] residues are located at the monomer-monomer interface and enhance the stability of the dimer. In contrast, the glycan chains linked to Asn[65] and Asn[140] residues protrude in the solvent at the back face of the α-amylase inhibitor-1 monomer and, like in arcelin-1, may prevent the formation of tetrameric structures by steric hindrance between two dimers.

4. Lectin-Kinase Receptors

Recently, a single chain plasma membrane receptor consisting of an extracellular lectin-like domain connected to an intracellular kinase domain through a hydrophobic transmembrane helix was identified in *Arabidopsis thaliana* (Hervé et al., 1996). Structural predictions and molecular modeling of the lectin-like domain from the X-ray coordinates of LoLI and PSA suggested an overall folding very similar to that of typical legume lectins, even though there is an extra loop of 17 residues in the amino acid sequence of the receptor (Hervé et al., unpublished results). However, most of the residues homologous to those involved in the binding of monosaccharides are completely different and this is especially true for Asp[81] and Asn[125], which play a key role in the

sugar-binding properties of legume lectins. Accordingly, one can reasonably expect that the monosaccharide-binding site of this lectin-like domain is inactive. In contrast, almost all the residues forming the hydrophobic cavity in legume lectins are well conserved, suggesting that these lectin-kinase receptors may be involved in the recognition of some plant hormones (Edelman and Wang, 1978).

G. Molecular Evolution

To construct a phylogenetic tree of the legume lectin family, a distance matrix has been calculated from the complete amino acid sequences of 42 lectins and 3 lectin-related proteins (Figure 8). It should be mentioned that for practical reasons not all sequences have been included. For instance, only one of the *Lathyrus* lectins is shown. Similarly, the arcelins and α-amylase inhibitors are confined to a single example. Due to the high sequence identity within the *Lathyrus* lectins and arcelins/α-amylase inhibitors, the omission of some members of these two subgroups has no effect on the overall structure of the phylogenetic tree.

The phylogenetic tree of the legume lectins consists of four major branches (indicated by the capitals S, P, V, and G) (Figure 8). Branch S clusters the GalNAc-specific *Sophora japonica* seed and bark lectins, the Neu5Acα(2,3)Gal/GalNAc-specific *Maackia amurensis* seed and bark lectins, and the lectin-related protein from *Cladrastis lutea*. The clustering of these lectins is in agreement with the classification of the genera *Sophora*, *Maackia*, and *Cladrastis* in the Sophoreae tribe. A further subdivision of the S branch in an Ss and Sm side branch separates the *Maackia amurensis* lectins from the *Sophora japonica* and *Cladrastis lutea* lectins. Both side branches comprise bark and seed lectins. The closely related *Sophora japonica* seed and bark lectins apparently arose by gene duplication from a common ancestor. In the case of the *Maackia* lectins two gene duplications must have taken place. A first duplication of an ancestral lectin gene gave rise to *Maackia amurensis* hemagglutinin (MAH) and leukoagglutinin (MAL) genes that then evolved further into the seed- and bark-specific MAH and MAL genes. The S branch does not comprise all Sophoreae lectins. As discussed below, the mannose/glucose-specific bark lectins from *Sophora japonica* and *Cladrastis lutea* and the *Bowringia milbraedii* seed lectin belong to the P branch.

Branch P comprises an extended group of lectins and lectin-related proteins from three different tribes. Side branch Pp clusters the lectins and lectin-related proteins from the Phaseoleae tribe, indicating that the *Phaseolus vulgaris*, *Phaseolus lunatus*, *Dolichos biflorus*, and *Glycine max* lectins and lectin-related proteins probably have evolved from a common ancestor. The topology of the Pp side branch is in good agreement with the taxonomy of the Phaseoleae tribe. *Glycine max* and *Dolichos biflorus* contain only one and two sets of lectin genes, respectively. In contrast, *Phaseolus vulgaris* possesses three distinct sets of genes encoding lectins, arcelins, and α-amylase inhibitors. Both PHA genes as well as the genes encoding the arcelins and α-amylase inhibitors probably evolved from a single ancestral gene through gene duplications. This assumption is in good agreement with the fact that PHA, arcelin, and α-amylase inhibitor occur at a single locus of the *Phaseolus vulgaris* genome (Nodari et al., 1993). According to the dendrogram, duplication of the ancestral *Phaseolus vulgaris* lectin gene yielded the direct ancestors of the PHA-genes and the arcelin/α-amylase-ancestor genes. Duplication of the PHA ancestor gene gave rise to the actual PHA-E and PHA-L genes, whereas duplication of

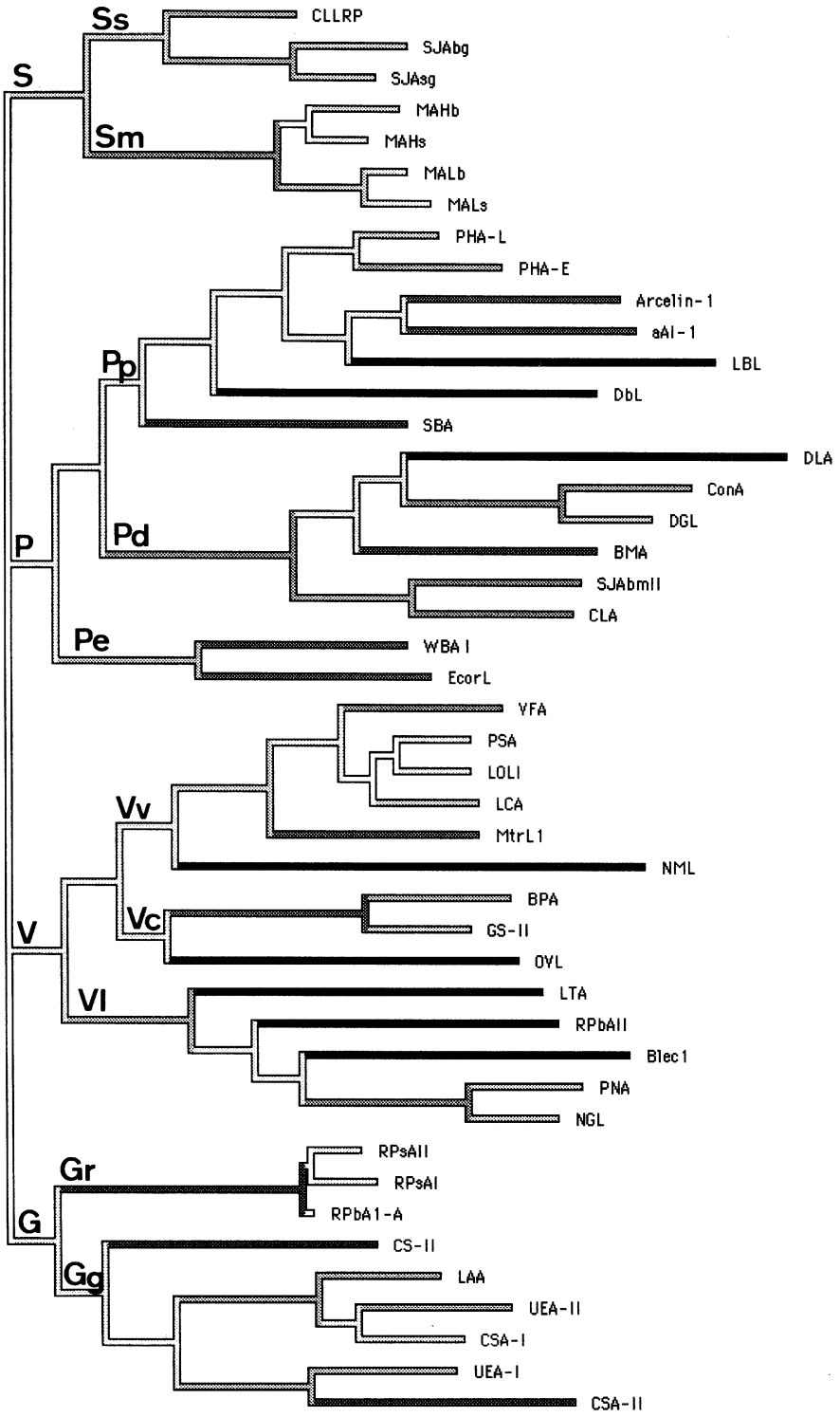

FIGURE 8. Phylogenetic tree of sequences encoding legume lectins and related proteins. The dendrogram was made on a Macintosh LC 630 using the program MacClade (Maddison and Maddison, 1992). The abbreviations of the lectins are listed in Table 1. In addition to lectin sequences, the dendrogram also includes the sequences of the putative lectins from *Medicago truncatula* seeds (Mtrl1) and *Pisum sativum* buds (Blec1) as well as the lectin-related proteins from *Cladrastis lutea* (CLLRP) and *Phaseolus vulgaris* (arcelin-1 and α-amylase inhibitor-1 (aAI-1)).

the arcelin/α-amylase inhibitor ancestor genes yielded an arcelin gene and an α-amylase-inhibitor gene. Interestingly, the arcelins and α-amylase-inhibitors have three and one gaps, respectively, in the aligned sequences when compared with the lectin sequences. Because these gaps correspond to deletions of one and two loops forming the monosaccharide binding site, this can explain why arcelins have a low carbohydrate-binding activity, whereas α-amylase inhibitors are completely inactive (Rougé et al., 1993; Mirkov et al., 1994). The dendrogram shows only one arcelin gene and one α-amylase-inhibitor gene. However, several bean arcelins and α-amylase-inhibitors have been identified. Therefore, it is evident, that both the arcelin and α-amylase-inhibitor genes have evolved further. More details about the evolutionary relationships among the phytohemagglutinin-arcelin-α-amylase inhibitor family of the common bean and its relatives have been described by Mirkov et al. (1994).

Side branch Pd clusters the mannose/glucose-specific lectins from the unrelated Diocleae and Sophoreae tribes. It is surprising, indeed, that the *Dioclea grandiflora* and *Canavalia ensiformis* lectins are closely related to the mannose/glucose-binding lectins from *Bowringia milbraedii*, *Sophora japonica*, and *Cladrastis lutea*. Because there is still some uncertainty about the exact position of the *Bowringia milbraedii* lectin in the dendrogram (because only the mature lectin polypeptides have been sequenced and not the cDNA) the side branch Pd can probably be subdivided in a Diocleae group and a Sophoreae group. The side branch Pe corresponds to a small cluster of Gal-specific lectins from *Psophocarpus tetragonolobus* and *Erythrina corallodendron*.

Branch V also comprises an extended group of lectins from different tribes. A first side branch Vv clusters the mannose/glucose-specific lectins from the Vicieae tribe (comprising the genera *Lens*, *Lathyrus*, *Medicago*, *Pisum*, and *Vicia*) and the mannose/glucose-binding nodule lectin from *Arachis hypogaea* (tribe Hedysareae). The second small side branch Vc groups the lectins from *Bauhinia purpurea* and *Griffonia simplicifolia*, which are the only documented lectin-containing species of the Caesalpinoideae subfamily. Because the Caesalpinoideae are considered as more primitive than the Papilionaceae, the *Bauhinia purpurea* and *Griffonia simplicifolia* lectins may be the closest relatives of the common ancestor of the legume lectins. Side branch Vc also includes the lectin from *Onobrychis viciifolia* (Hedysareae tribe). Side branch Vl comprises a heterogeneous cluster of lectins from different taxonomic groups. Besides the fucose-specific lectin from *Lotus tetragonolobus* (Lotaceae) and the closely related GalNAc-binding seed and nodule lectins from *Arachis hypogaea* (Hedysareae) (which have evolved most probably by gene duplication), the Vl cluster also contains a nonagglutinating *Robinia pseudoacacia* bark lectin RPbAII and a presumed bud-specific *Pisum sativum* (Vicieae) lectin called Blec1.

Branch G clusters lectins from a relatively narrow taxonomic group comprising species of the tribes Galegeae (*Robinia pseudoacacia* and *Laburnum alpinum*) and Genisteae (*Cytisus scoparius*, *Cytisus sessilifolius*, and *Ulex europaeus*). The *Robinia pseudoacacia* lectins form a separate side branch Gr comprising two seed-specific lectins and two bark-specific lectins (only one of which is shown in the dendrogram). All four lectin genes have a common ancestor and most probably arose by two consecutive gene duplication events. The side branch Gg consists of a GalNAc-specific lectin from *Cytisus scoparius*, a small cluster of three (GlcNAc)n-specific lectins from *Laburnum alpinum*, *Ulex europaeus*, and *Cytisus sessilifolius*, and two fucose-specific

lectins from *Ulex europaeus* and *Cytisus sessilifolius*. According to the dendrogram the (GlcNAc)n-specific lectins from *Laburnum alpinum*, *Ulex europaeus*, and *Cytisus sessilifolius* are more closely related to each other than to their conspecific fucose-specific homologues. This indicates that the common ancestor of the Genisteae lectins already had given rise (through gene duplication) to the direct ancestors of the (GlcNAc)n-specific and fucose-specific lectins before the Genisteae species diverged from each other.

The simultaneous occurrence in a single species of lectins belonging to two different branches of the dendrogram enables to reconstruct the outlines of the molecular evolution of the legume lectins (Figure 9). For example, the expression of S and P branch lectins in *Sophora japonica* and *Cladrastis lutea* indicates that these two branches evolved from a common ancestral SP branch. Similarly, the occurrence of V and G branch lectins in *Robinia pseudoacacia* suggest a common ancestral VG branch. Finally, because all legume lectins definitely have a common origin, the ancestral SP and VG branches must have evolved from a single common stem, namely, the legume lectin domain LLD.

Recently, a putative receptor kinase has been discovered in *Arabidopsis thaliana* that

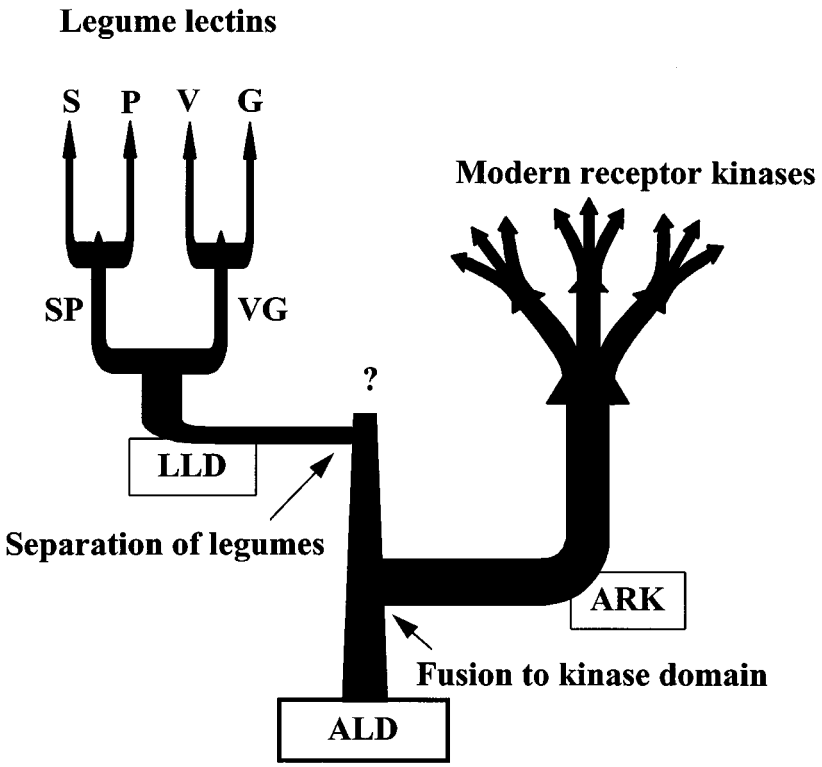

FIGURE 9. Hypothetical model of the molecular evolution of legume lectins. The ancestral legume lectin domain (ALD) followed two independent evolutionary routes. First, ALD evolved as a carbohydrate-binding protein but became extinct or silenced except in the direct ancestor of the modern legumes where it gave rise to the legume lectin domain (LLD) and further evolved into the legume lectins as indicated in detail in Figure 8. Second, early in the evolution of higher plants ALD fused to a kinase domain giving rise to an ancestral receptor kinase (ARK) that further evolved into the modern receptor kinases similar to the one discovered in *Arabidopsis thaliana*.

contains an extracellular domain with 39% sequence similarity to the *Medicago truncatula* lectin, and according to molecular modeling studies has the same overall three-dimensional structure as the legume lectin protomers (Hervé et al., 1996). Although it remains to be demonstrated that the presumed lectin domain possesses carbohydrate-binding activity, the discovery of the *Arabidopsis thaliana* gene indicates that the legume lectin(-like) domain also occurs outside the legume family. One can speculate that receptor kinases with a legume lectin-like domain may play an important role in the transduction of foreign and endogenous oligosaccharide signal molecules. If so, the question of the evolution of the legume lectins has to be readdressed. A possible scenario is that a legume lectin-like domain arose during early evolution of higher plants as a specific receptor or binder of either foreign or endogenous carbohydrates. In the next evolutionary step, the legume lectin-like domain fused with a protein kinase domain to form a receptor kinase for the transduction of carbohydrate signal molecules. Due to the much higher specificity and activity of the receptor kinases, the single legume lectin-like domain became superfluous and the corresponding gene was silenced or even eliminated from the genome. Later, during evolution of the higher plants (more precisely around the ascent of the ancestors of the modern legumes), the legume lectin-like domain was 'rediscovered' for defense purposes. The silenced gene was reactivated or alternatively the legume lectin-like domain was excised from a receptor kinase gene and used as a kind of template for the development of carbohydrate-binding defense-related proteins.

H. Physiological Role

Although the physiological role of legume lectins has been an important issue in lectin research for decades, the possible function of legume lectins is still unclear. Many efforts have been undertaken to prove the hypothesis that legume lectins are responsible for the specific interaction between legumes and their symbiotic nitrogen-fixing *Rhizobia* (Diaz et al., 1989; Brewin and Kardailsky, 1997). However, until now there was no conclusive evidence that legume lectins play a determining role in this process. On the contrary, the discovery that the host specificity of *Rhizobium* strains depends on the structure of the nod factors (also called lipochitin oligosaccharides) they secrete almost excludes an involvement of the legume lectins, as they simply do not have the right specificity. Moreover, the presumed role of the legume lectins in *Rhizobium*-legume symbiosis is even more questioned by the recent discovery of a novel type of carbohydrate-binding protein with a high affinity for Nod factors in the roots of *Dolichos biflorus* (Etzler et al., 1997).

Two major observations argue against the involvement of legume lectins in specific recognition processes within the plant. First, many legume lectins are so abundant that the plant cannot contain a correspondingly high level of specific receptors. Second, most legume lectins have a preferential specificity for carbohydrates that are absent in plants but are important components of animal glycoconjugates (e.g., sialic acid, GalNAc, some complex-type *N*-glycans) (Peumans and Van Damme, 1995a,b). The obvious abundance taken together with the fact that the lectins from seeds and vegetative storage tissues behave as typical storage proteins for what concerns their cellular and intracellular location, as well as the spatial and temporal regulation of their expression, strongly suggest that legume lectins have a transient storage function. To explain the preferential specificity toward typical animal glycans, the hypothesis has been put forward that legume lectins play a role in the plant's defense against insects and/or pre-

dating animals (Chrispeels and Raikhel, 1991; Peumans and Van Damme, 1995b). According to the currently accepted ideas, dietary lectins, which are usually highly resistant to gut proteases, bind to glycan receptors exposed on the surface of the epithelial cells along the gastrointestinal tract of the plant-eating organisms. After binding and possible endocytosis, the lectins elicit specific reactions in the target cells that eventually may result in a noxious effect (Pusztai and Bardocz, 1996). For example, ingestion of high doses of PHA causes acute nausea followed by vomiting and diarrhoea (Pusztai, 1991; Pusztai and Bardocz, 1996). The eventual discomfort is so severe that experimental animals refuse to continue eating a PHA-containing diet. Although some older reports also claim that PHA is toxic to insects, Murdock et al. (1990) have clearly shown that PHA is not toxic to cowpea weevil (*Callosobruchus maculatus*), and the previous toxic effects attributed to PHA are due to the contamination of the lectin with the α-amylase inhibitor. Some other legume lectins are toxic to insects or interfere with insect development when tested *in vitro*. For example, the lectin from the seeds of peanut has an inhibitory effect on the development of cowpea weevil larvae (Murdock et al., 1990). Similarly, the *Bauhinia purpurea* lectin is lethal to neonate larvae of *Ostrinia nubilalis* and inhibits the growth of *Diabrotica undecimpunctata* larvae (Czapla and Lang, 1990). Tests with transgenic plants have shown that tobacco plants expressing the pea lectin gene have increased resistance to *Heliothis virescens* (Boulter et al., 1990). It should be emphasized, however, that the lectin-related proteins are more potent anti-insect proteins than the legume lectins themselves. For example, low levels of dietary *Phaseolus vulgaris* α-amylase inhibitor effectively inhibit larval growth of typical seed predating insects such as the cowpea weevil (*Callosobruchus maculatus*) and the Azuki bean weevil (*Callosobruchus chinensis*) (Ishimoto and Kitamura, 1989). Moreover, transgenic pea seeds expressing the bean α-amylase inhibitor acquire resistance against the cowpea and Azuki bean weevil (Shade et al., 1994).

Summarizing, it can be concluded that legume lectins have a dual role. Under normal conditions they are genuine storage proteins. However, as soon as the plant is eaten the lectins end up in the gastrointestinal tract of the predator and act as defense proteins.

V. CHITIN-BINDING LECTINS

By definition the family of chitin-binding lectins comprises all proteins containing at least one hevein domain. The term 'hevein domain' refers to hevein, a small 43 amino acid protein from the latex of the rubber tree (*Hevea brasiliensis*) (Waljuno et al., 1975). Despite its small size, hevein possesses a fully active chitin-binding site. All structural units of about 40 amino acid residues that exhibit sequence similarity to hevein and possess chitin-binding activity are considered hevein domains. It should be noted that there are also chitin-binding lectins without hevein domain(s). For example, the chitin-binding legume lectins and Cucurbitaceae phloem lectins show no sequence similarity to the hevein domains and are not clearly related to the chitin-binding lectins discussed here.

A. Historical Note

The documented history of the chitin-binding lectins dates back to the work of Marcusson-Begun in 1926, who investigated the agglutinating principle of potato tubers (Kocourek, 1986). Another momentous step was the study of Aub and collaborators (Aub et al., 1963) who observed that a lipase-

containing extract from wheat germ agglutinated several transformed cell lines but did not cause agglutination of their untransformed counterparts. A few years later, Burger and Goldberg (1967) separated the agglutinating activity from the lipase activity and identified 'wheat germ agglutinin' (WGA) as a lectin. From then on the history of the chitin-binding lectins is closely linked to the pioneering work done with WGA, which was the first chitin-binding lectin to be isolated and characterized (Nagata and Burger, 1972; LeVine et al., 1972). The first three-dimensional structure of a chitin-binding lectin was obtained by X-ray crystallographic analysis of WGA. It should be noted that the WGA structure was already completed at the 2.2 Å level (Wright, 1977) before the complete amino acid sequence of the protein became available. Finally, WGA was also the first chitin-binding lectin to be cloned (Raikhel and Wilkins, 1987). The concept of hevein as a chitin-binding domain is relatively recent. Although already isolated and sequenced in 1975 (Waljuno et al., 1975), the chitin-binding activity of hevein was not recognized until 1991 (Van Parijs et al., 1991). Since then, sequencing and/or molecular cloning have demonstrated that the chitin-binding activity of many proteins relies on the presence of one or more domains that are structurally and functionally related to hevein.

B. Occurrence, Molecular Structure, and Amino Acid Sequences

The family of chitin-binding lectins is widespread in the plant kingdom. According to the novel definition of lectins, class I chitinases are (chimero)lectins because of the presence of an *N*-terminal chitin-binding hevein domain (Linthorst, 1991; Collinge et al, 1993). Because these chitinases probably occur in all plants, chitin-binding lectins can be considered a ubiquitous group of plant proteins. Class I chitinases have been found in various vegetative tissues and in seeds (e.g., in wheat embryos) and are usually present in (very) low concentrations. Chitin-binding lectins other than the class I chitinases have been identified in many species of different taxonomic groups. Hevein and hevein-like proteins have been isolated from the latex of *Hevea brasiliensis* (rubber tree, Euphorbiaceae), seeds of *Amaranthus caudatus* (grain amaranth, Amaranthaceae), and fruits and leaves of elderberry (*Sambucus nigra*, Caprifoliaceae). In contrast to hevein, which is an abundant latex protein, the elderberry hevein-like protein is present in very small quantities. The seed-specific *Amaranthus* chitin-binding lectin represents about 1% of the soluble seed protein.

Chitin-binding lectins are widespread in the Gramineae family. Up to now, Gramineae lectins have been found in the embryos of all species of the tribe Triticeae (comprising the genera *Aegilops*, *Agropyrum*, *Eremopyrum*, *Elymus*, *Haynaldia*, *Heteranthelium*, *Hordeum*, *Psatyrostachys*, *Secale*, *Triticosecale*, *Triticum*) and the genera *Brachypodium* and *Oryza* (Peumans and Stinissen, 1983). All Gramineae seed lectins are strictly confined to the embryo where they are located in the peripheral tissues (i.e., the embryonic root cap, the coleorhiza, and in the outer cell layers of the coleoptile, the radicle, and the scutellum). Gramineae lectins account for 0.2 to 0.5% of the soluble embryo protein (corresponding to less than 0.02% of the total seed protein). Lectins, which are indistinguishable from the Gramineae seed lectins but are definitely not derived from the embryos, are also present in small quantities in root caps and leaves of some cereals (Raikhel et al., 1993). Leaf-specific Gramineae lectins that differ slightly from the embryo lectins have been isolated from *Agropyrum repens* (couch grass) and *Phragmites australis* (com-

mon reed). Both leaf-specific lectins are also minor proteins.

Chitin-binding lectins are common in the Solanaceae family. Typical so-called Solanaceae lectins have been characterized from *Solanum tuberosum* (potato), *Lycopersicum esculentum* (tomato), *Cyphomandra betacea* (tamarillo), and *Datura stramonium* (thorn apple). The potato and tomato lectin occur as minor proteins in all vegetative tissues, whereas the tamarillo lectin has been described as a minor fruit-specific protein. In contrast, the *Datura* lectin is strictly confined to the seeds and is an abundant protein (representing up to 50% of the total seed protein).

Besides those Gramineae and Solanaceae species, agglutinating chitin-binding lectins are also found in roots and leaves of *Phytolacca* sp. (pokeweed, Phytolaccaceae), seeds and rhizomes of *Urtica dioica* (stinging nettle, Urticaceae), seeds of *Chelidonium majus* (greater celandine, Papaveraceae), and in green tissues of *Viscum album* (mistletoe, Viscaceae). All these lectins are minor proteins in the respective tissues. In summary, it can be concluded that chitin-binding lectins are present in monocots and dicots and occur in seeds as well as in different vegetative tissues.

1. Molecular Structure of the Native Lectins

The family of chitin-binding lectins comprises merolectins, hololectins, as well as different types of chimerolectins and is rather heterogeneous in the structure of the native proteins (Figure 10). In this section their structure is discussed in order of increasing complexity.

The group of merolectins composed of a single hevein domain are the simplest chitin-binding lectins of which two types have been described. The first types are hevein and the hevein-like proteins. Hevein, which is an abundant latex protein of the rubber tree, consists of a single polypeptide chain of 43 amino acids (Waljuno et al., 1975). It is rich in glycine and contains eight cysteine residues that are all involved in intrachain disulfide bridges. Recently, a hevein-like protein with sequence similarity to hevein has been isolated from the fruits of *Sambucus nigra* (elderberry; W. Peumans, unpublished results). A second type of single domain chitin-binding lectins are the chitin-binding antimicrobial peptides from *Amaranthus caudatus* (AcAMP) (Broekaert et al., 1992). AcAMP1 and AcAMP2 are single chain polypeptides of 29 and 30 amino acid residues, respectively. Both proteins have an identical amino acid sequence, except that AcAMP1 is one residue shorter at the *C*-terminus than AcAMP2. Sequence alignment with hevein indicates that the AcAMPs consist of a single truncated hevein domain, as they contain only six cysteine residues (which form three intrachain disulfide bridges). Most probably, the AcAMPs correspond to the minimal structures required for chitin-binding activity.

The family of chitin-binding lectins comprises hololectins composed of polypeptides containing two, three, four, or seven tandemly arrayed hevein domains. At present, several hololectins built up of polypeptide chains with two hevein repeats have been described. The *Urtica dioica* (stinging nettle) agglutinin (UDA) is a monomer of a single polypeptide chain of 89 amino acid residues that consists of two hevein repeats separated by a four amino acid hinge region (Peumans et al., 1984a; Beintema and Peumans, 1992). UDA contains two hevein domains and hence is a hololectin capable of agglutinating cells. The chitin-binding lectin from *Viscum album* (called *Viscum album* chitin-binding protein or VisalbCBP) is a homodimer of two identical subunits of 10.8 kDa (Peumans et al., 1996). VisalbCBP shares a high

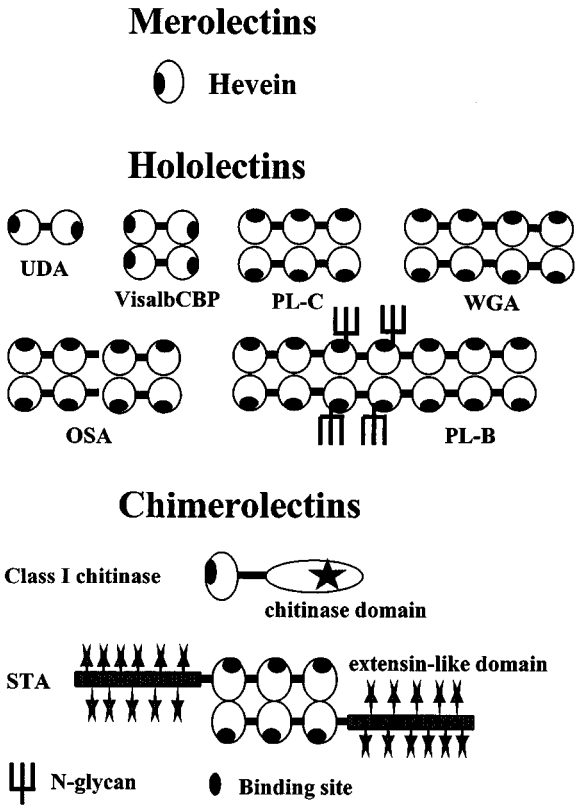

FIGURE 10. Schematic representation of the different molecular structures of native chitin-binding lectins. Examples shown are: hevein, UDA (*Urtica dioica* agglutinin), *Viscum album* chitin-binding protein (VisalbCBP), *Phytolacca americana* lectins PL-C and PL-B, wheat germ agglutinin (WGA), *Oryza sativa* agglutinin (OSA), class I chitinase and *Solanum tuberosum* agglutinin (STA).

sequence similarity at its *N*-terminus with the hevein domain and has an amino acid composition similar to that of other chitin-binding hololectins. Therefore, one can reasonably assume that the subunits of VisalbCBP consist of two hevein repeats and that the native lectin contains four chitin-binding domains. A cysteine/glycine-rich chitin-binding lectin was also found in seeds of *Chelidonium majus* (Peumans et al., 1985a). The lectin, called *Chelidonium majus* agglutinin or CMA, is a dimer composed of two different subunits of 9.5 and 11.5 kDa. At present, no sequence information is available for CMA. However, because CMA closely resembles the chitin-binding hololectins in amino acid composition, we assume that both subunits consist of two hevein repeats and that the native lectin contains four chitin-binding sites. Roots of *Phytolacca americana* contain a complex mixture of chitin binding lectins differing from each other in molecular size. Amino acid sequencing of the mature proteins has shown that the *Phytolacca americana* lectins Pa-5/PL-D, Pa-4/PL-C, and PL-B are built up of subunits consisting of two, three, and seven tandemly arrayed hevein domains, respectively (Yamaguchi et al., 1995, 1996, 1997). PL-B contains two *N*-linked glycans per subunit and is the only documented example of mature *N*-glycosylated chitin binding proteins. WGA and all other Gramineae lectins characterized thus far (i.e., the lectins from *Hordeum* sp., *Secale* sp., *Brachypodium* sp., *Agropyrum repens, Phragmites australis,* and

Oryza sp.) are built up of protomers consisting of four hevein repeats, and thus contain eight chitin-binding domains in total (Raikhel et al., 1993). Biochemical and molecular analysis of WGA demonstrated, indeed, that the three composing isolectins are dimers (36 kDa) of 171–residue polypeptides that contain four repetitive hevein domains (Raikhel and Wilkins, 1987; Smith and Raikhel, 1989). Most Gramineae lectins have essentially the same structure as WGA (Peumans and Stinissen, 1983; Stinissen et al., 1983). However, there are two exceptions. First, the leaf-specific lectin from *Agropyrum repens* (ARLL, Cammue et al., 1985) and also one of the leaf lectins from *Phragmites australis* (PALL-I, Peumans and Cammue, 1986) consist of subunits that are 15 to 20 amino acid residues longer than the WGA polypeptides. Second, the protomers of the lectins from *Oryza sativa* and some other wild rice species are for 50 to 90% posttranslationally cleaved into two smaller polypeptides of 8 and 10 kDa, respectively. As a result, they contain polypeptides of 18 kDa as well as 8 and 10 kDa.

The family of chitin-binding lectins comprises at least two different types of chimerolectins. Class I chitinases contain a single *N*-terminal hevein domain of about 40 residues linked through a small variable glycine/proline-rich hinge domain to the catalytically active chitinase domain (Collinge et al., 1993; Beintema, 1994). Because native class I chitinases are monomeric proteins of 25 to 30 kDa, they possess only one carbohydrate-binding site and hence cannot agglutinate cells. Protomers of the Solanaceae lectins are also considered chimeric polypeptides. Although no complete amino acid sequences are available yet, there is firm evidence that at least the 65-kDa subunits of the potato lectin (called *Solanum tuberosum* agglutinin or STA) consist of an *N*-terminal chitin-binding domain of three hevein repeats linked to an O-glycosylated serine-hydroxyproline-rich domain resembling the cell wall protein extensin (Kieliszewski et al., 1994; Allen et al., 1996). Mass spectrometric analyses of native and deglycosylated STA polypeptides indicated that the lectin contains 52.3% covalently bound carbohydrate consisting mainly of *O*-linked arabinoside-hydroxyproline and to a lesser extent galactoside-serine (Allen et al., 1996). Because native STA is a dimer of two identical or slightly different subunits of 65 kDa, the lectin contains six hevein domains per molecule and is capable of agglutinating cells. Several other Solanaceae lectins have been studied in some detail. Although there are still discrepancies between the molecular mass of the native lectins and the lectin polypeptides, it seems likely that all Solanaceae lectins are dimers composed of subunits of 30 to 70 kDa. For example, the tomato fruit lectin has been described as a dimer of two 70-kDa subunits. The seed lectin from *Datura stramonium* is a mixture of isolectins composed of two identical or slightly different subunits of 28 and 32 kDa (Broekaert et al., 1987).

2. Primary Structure of Chitin-Binding Lectins

Several chitin-binding lectins (e.g., AcAMP, hevein, UDA, WGA, and three *Phytolacca* lectins) have been completely sequenced by chemical methods (Table 2). Partial sequence information has also been obtained for a number of other chitin-binding lectins by sequencing of their *N*-terminus (e.g., VisalbCBP) or tryptic fragments (e.g., potato lectin). In addition, the primary structure of many chitin-binding lectins has been deduced from the nucleotide sequence of cDNAs. Sequence comparisons indicate that all chitin-binding lectins share sequence similarity in their hevein repeat(s). However, their non-hevein domains (i.e., class I chitinases and Solanaceae lectins) are definitely unrelated.

TABLE 2
Chitin-Binding Lectins: Occurrence, Molecular Structure, and Specificity

Species	Tissue	Structure[a]	Specificity	Sequence available[b]
Merolectins				
Amaranthus caudatus	Seed	[P4]	(GlcNAc)n	Pr, Nu (Ac-AMP2)
Hevea brasiliensis	Latex	[P5]	(GlcNAc)n	Pr, Nu (hevein)
Hololectins				
Agropyrum repens	Embryo	$[P18]_2$	GlcNAc	
	Leaf	$[P20]_2$	GlcNAc	
Brachypodium sylvaticum	Embryo	$[P18]_2$	GlcNAc	
Chelidonium majus	Seed	[P10 + P12]	(GlcNAc)n	
Cyphomandra betacea	Fruit	$[P25]_2$	(GlcNAc)n	
Datura stramonium	Seed	$[P28]_2$, $[P32]_2$, [P28 + P32]	(GlcNAc)n	
Hordeum vulgare	Embryo	$[P18]_2$	GlcNAc	Nu (HVA)
Oryza sativa	Embryo	$[P18]_2$,$[P(8 + 10)]_2$	GlcNAc	Nu (OSA)
Oryza alta	Embryo	$[P18]_2$	GlcNAc	
Oryza minuta	Embryo	$[P18]_2$,$[P(8 + 10)]_2$	GlcNAc	
Phragmites australis	Leaf	I $[P20]_2$	GlcNAc	
	Leaf	II $[P18]_2$	GlcNAc	
Phytolacca americana	Root	Pa-1 $[P14]_?$		
	Root	Pa-2/A $[P22]_2$	(GlcNAc)n	
	Root	Pa-3$[P22]_?$	(GlcNAc)n	
	Root	Pa-4/C $[P14]_2$	(GlcNAc)n	Pr (PL-C)
	Root	Pa-5/D $[P9]_2$	(GlcNAc)n	Pr (PL-D)
	Root	Pa-B $[P40]_2$	(GlcNAc)n	Pr (PL-B)
Secale cereale	Embryo	$[P18]_2$	GlcNAc	
Triticum aestivum	Embryo	$[P18]_2$	GlcNAc	Pr, Nu (WGA)
Triticum durum	Embryo	$[P18]_2$	GlcNAc	Nu
Urtica dioica	Rhizome	[P9]	(GlcNAc)n	Pr, Nu (UDA)
Viscum album	Plant	$[P11]_2$	(GlcNAc)n	
Chimerolectins				
Lycopersicon esculentum	Fruit	$[P70]_2$	(GlcNAc)n	
Solanum tuberosum	Tuber	$[P65]_2$	(GlcNAc)n	
Class I Chitinases	Plant	[P25–30]	(GlcNAc)n	Nu

[a] [PX] stands for protomer with a molecular mass of X kDa. [P(Y + Z)] indicates that the protomer is cleaved in two polypeptides of Y and Z kDa.

[b] Pr, protein sequence; Nu, nucleotide sequence. The abbreviation in brackets refers to the sequence name used in the dendrogram (Figure 13).

C. Structure, Biosynthesis, and Posttranslational Modifications

Molecular cloning and analysis of cDNAs and/or genomic DNA fragments made an important contribution to the elucidation of the primary structure of chitin-binding lectins. More important, however, the molecular approach unraveled the biosynthesis and posttranslational processing of the primary translation products of the lectin mRNAs and revealed unexpected relationships between chitin-binding lectins and some other plant proteins.

In addition to the numerous class I chitinases that have been cloned from different species, the list of cloned chitin-binding lectins is confined to AcAMP, hevein, and a few hevein-like proteins, UDA, the WGA-like lectins from *Triticum aestivum, T. durum,* and *Hordeum vulgare,* and the *Oryza sativa* lectin (Table 2). Molecular analysis combined with the data from biochemical and cellular biological studies reconstructed the conversion of the primary translation products of the genes into the mature proteins (Figure 11). Basically, the chitin-binding lectins can be divided in two subgroups on the basis of their overall processing. AcAMP, Gramineae lectins, and class I chitinases are synthesized as preproproteins containing a small *C*-terminal propeptide. Hevein and UDA are also synthesized as preproproteins, but *C*-terminal processing involves the removal of a large peptide.

The cDNA encoding AcAMP consists of a putative signal sequence and a 30–residue sequence corresponding to the mature protein followed by an extra 31–residue propeptide at the *C*-terminus (De Bolle et al., 1993). These data indicate that AcAMP is synthesized as a preproprotein that undergoes co-translational removal of a signal peptide and posttranslational cleavage of the

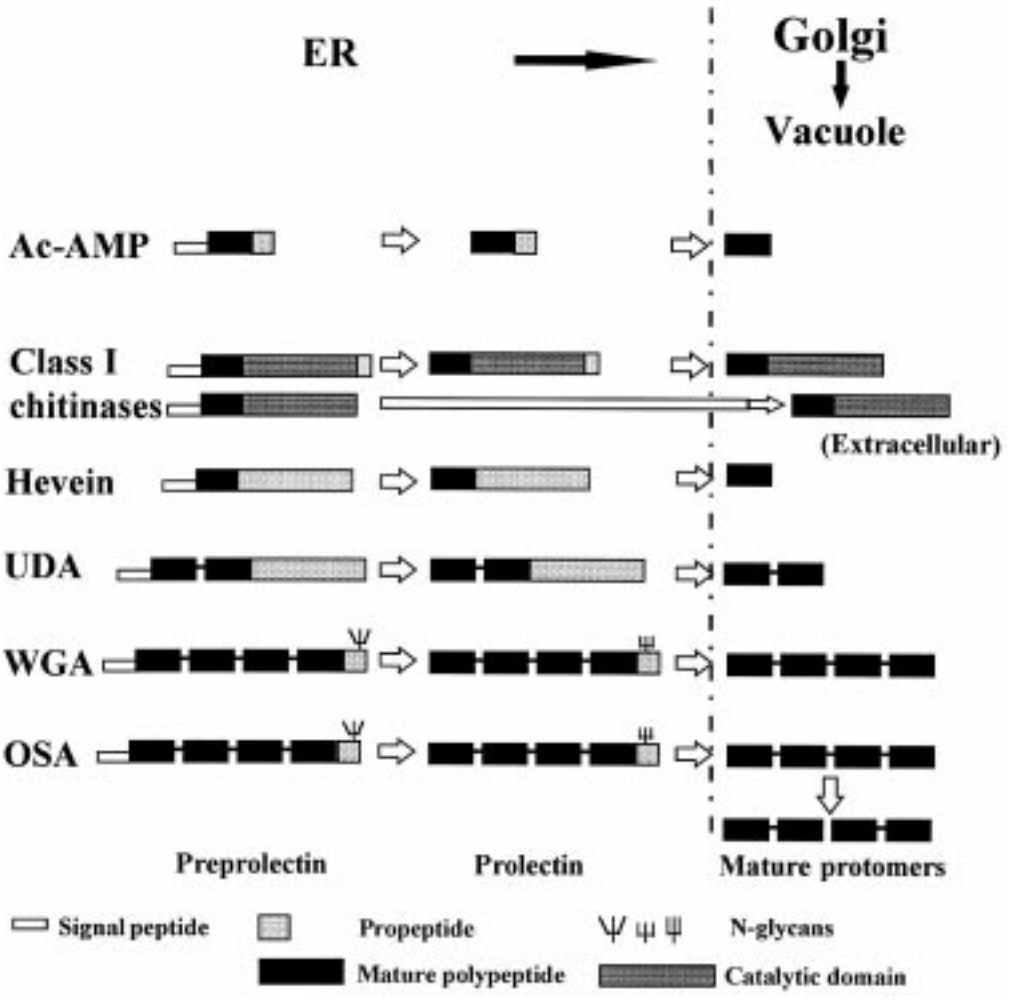

FIGURE 11. Schematic representation of the biosynthesis, co- and posttranslational modifications, and topogenesis of the chitin-binding lectins. Examples shown are *Amaranthus caudatus* antimicrobial protein (Ac-AMP), class I chitinase, hevein, UDA (*Urtica dioica* agglutinin), wheat germ agglutinin (WGA), and *Oryza sativa* agglutinin (OSA).

C-terminal propeptide to give rise to a truncated 30–residue hevein domain. The overall structure of the preproprotein suggests that AcAMP follows the secretory pathway.

Three Gramineae lectins, namely, WGA and the barley and rice lectins, have been studied in detail for what concerns their biosynthesis, processing, and topogenesis. According to the cDNA sequences, the wheat and barley lectins are synthesized as preproproteins consisting of a 26–residue *N*-terminal signal peptide, a sequence of 171 amino acids with four tandemly arrayed hevein domains corresponding to the mature lectin polypeptide, and a 15–residue *C*-terminal propeptide with a putative glycosylation site (Raikhel and Wilkins, 1987; Lerner and Raikhel, 1989). Biosynthesis studies confirmed the co-translational removal of the signal peptide and demonstrated that the primary translation product of the lectin mRNA is co-translationally glycosylated in the *C*-terminal propeptide. Further processing of the glycosylated prolectin takes place during transport from the ER via the Golgi to the vacuoles or after arrival in the vacuoles, and involves the proteolytic removal of the glycosylated *C*-terminal propeptide. The role of the barley *C*-terminal propeptide has been studied in detail using heterologous expression systems (Bednarek et al., 1990; Bednarek and Raikhel, 1991). Inactivation of the glycosylation site does not affect the folding, processing, or targeting of the mature protein, but deletion of the entire propeptide leads to missorting and secretion of the lectin. Fusion of the barley *C*-terminal propeptide to secreted proteins redirects them to the vacuole, indicating that the barley propeptide is both necessary and sufficient for vacuolar targeting in plants. The biosynthesis and processing of the rice lectin are very similar to that of WGA and barley lectin, but include an additional proteolytic cleavage. According to the cDNA sequence, rice lectin is synthesized as a preproprotein consisting of a 28–residue signal peptide, a sequence of 173 amino acids with four tandemly arrayed hevein domains, and a 26–residue *C*-terminal propeptide with a putative glycosylation site (Wilkins and Raikhel, 1989). Biosynthesis studies confirmed the co-translational removal of the signal peptide and glycosylation of the *C*-terminal propeptide. Further processing of the glycosylated prolectin occurs during or after transport from the ER via the Golgi to the vacuoles. After removal of the glycosylated *C*-terminal propeptide, the 173–residue lectin protomer is proteolytically cleaved into two smaller polypeptides. It should be emphasized, however, that the second processing leaves about 10% of the protomers unaffected and occurs only in some *Oryza* species.

cDNA clones of class I chitinases encode primary translation products composed of an *N*-terminal signal peptide, a sequence corresponding to the mature enzyme either followed by a short *C*-terminal propeptide or not (Collinge et al., 1993). The biosynthesis of vacuolar class I chitinases apparently follows the secretory pathway. These class I chitinases are synthesized as preproproteins and co-translationally processed by the removal of the signal peptide. The resulting proprotein is subsequently transported from the ER to the vacuoles and during or after transport converted into the mature enzyme by the cleavage of the *C*-terminal propeptide (Neuhaus et al., 1991). Class I chitinases lacking the *C*-terminal propeptide eventually end up in the apoplastic space.

In contrast to the above-described examples where the major part of the primary translation products eventually ends up in mature proteins, hevein and UDA are only small remnants of much larger precursors. According to the cDNA sequence, the primary translation product of the hevein mRNA is a 204 amino acid polypeptide consisting of a signal peptide, a 43–residue sequence corresponding to mature hevein, and a large (144 residues) *C*-terminal peptide (Broekaert

et al., 1990). Interestingly, this *C*-terminal peptide itself consists of a glycine/proline-rich hinge region followed by a sequence with high sequence similarity to the wound-inducible genes of potato and the acidic pathogenesis-related protein PR4 from tobacco and P2 from tomato. Biosynthesis studies revealed that the signal peptide of preprohevein is co-translationally processed and that the resulting prohevein is incompletely cleaved in a subsequent posttranslational processing step (Lee et al., 1991). As a result, rubber tree latex contains not only mature hevein and a 15-kDa protein corresponding to the *C*-terminal domain, but also the 20-kDa prohevein. The location of hevein in lutoid bodies of the latex that are vacuole-derived organelles confirms that the synthesis and transport of hevein follow the secretory pathway.

The biosynthesis and processing of UDA are very similar to that of hevein. According to the cDNA sequence, the primary translation product contains a signal peptide, a sequence of 89 amino acids corresponding to mature UDA, and a *C*-terminal extension of 171 amino acid residues (Lerner and Raikhel, 1992). No data have been reported on the biosynthesis and processing of UDA, but one can reasonably expect that the signal peptide of preproUDA is co-translationally removed and that the resulting proUDA is processed further by the proteolytic cleavage of a large *C*-terminal peptide. UDA preparations do not contain detectable amounts of uncleaved proUDA, suggesting that most precursor molecules are processed. It is not known whether the cleaved *C*-terminal peptide occurs as a stable protein in the cells of the stinging nettle. Interestingly, the *C*-terminal part has a high sequence similarity with the catalytic domains of both class I and class II chitinases. At present, it is not known whether proUDA (or its possible processing product) has chitinase activity *in vivo*.

At present there is no information available about the genes encoding other chitin-binding lectins. This is rather unfortunate because some important questions remain unanswered. For example, the structure and biosynthesis of the Solanaceae lectins cannot be established without cDNA or genomic DNA sequences. Similarly, the relationships between the different *Phytolacca* lectins can only be solved by a molecular approach.

Molecular cloning yielded valuable information about the chitin-binding lectin genes. Most, but not all, chitin-binding lectins are encoded by a single gene or small gene families. AcAMP is probably under the control of a single gene. Hevein and a putative hevein-like protein from *Arabidopsis* (Potter et al., 1993) are also encoded by a single gene. Class I chitinases are often encoded by small gene families. No data have been reported on the UDA gene(s). However, taking into consideration that UDA is a complex mixture of at least 10 different isoforms (Van Damme and Peumans, 1987), it seems likely that the lectin is encoded by a gene family. The number of lectin genes in Gramineae species primarily depends on the ploidy level. Biochemical studies demonstrated that in both diploid and allopolyploid species each individual genome directs the synthesis of its own lectin. As a result, diploids such as rye, barley, and rice contain a single lectin, whereas tetraploid and hexaploid wheats contain two and three lectins, respectively (Peumans and Stinissen, 1983). Molecular analyses of the genes encoding the Gramineae lectins are in good agreement with the biochemical data. Rye, barley, and rice express a single lectin gene, whereas hexaploid wheat possesses a separate gene for the A, B, and D isoform of WGA (Raikhel et al., 1993).

Heretofore, only a few genomic clones encoding chitin-binding lectins have been analyzed and sequenced. According to the published data, introns are common in class I chitinase genes. Intron sequences have also been found in the gene encoding a hevein-like protein from elderberry (E. Van Damme,

unpublished results). Interestingly, in both cases the introns are located in the second domain of the protein. Although no general conclusions can be drawn as yet about the occurrence of introns in genes encoding different types of chitin-binding lectins, these examples clearly demonstrate that introns do occur in this lectin family.

D. Carbohydrate-Binding Specificity

The sugar-binding activity and specificity of the chitin-binding lectins are determined exclusively by their hevein domains. All chitin-binding lectins, including class I chitinases, are retained on a chitin column and hence are believed to bind native chitin. Only WGA, UDA, and the potato and thorn apple lectins have been studied in detail for their carbohydrate-binding activity. In addition, the specificity of the lectins from *Chelidonium majus*, *Viscum album*, *Agropyrum repens*, and *Phragmites australis* has been determined partly by hapten inhibition assays of the agglutination of red blood cells. According to the results of these studies, the carbohydrate-binding sites of the chitin-binding lectins have a complex structure that is most complementary to the trimer or tetramer of GlcNAc. This explains why the inhibitory potency of GlcNAc-oligomers increases with chain length up to four residues. For example, in the case of WGA, the relative inhibitory potency of the di-, tri-, and tetramers of GlcNAc is 131-, 3700-, and 4500-fold higher, respectively, than that of the monomer (Goldstein and Poretz, 1986). The occurrence of complex sugar-binding sites also explains why most chitin-binding lectins are not inhibited by GlcNAc but require at least the dimer. It should be emphasized, indeed, that the Gramineae lectins are the only chitin-binding lectins that are inhibited by GlcNAc and bind to immobilized GlcNAc. Most other members require at least a dimer for inhibition and binding. Moreover, class I chitinases, hevein, and AcAMP are only slightly retained even on a column of immobilized GlcNAc-oligomers (although they bind quantitatively to crude chitin). Specificity studies also demonstrated that the multiple binding sites of some chitin-binding lectins are not identical. For example, WGA possesses both high- and low-affinity sites. Similarly, the specificity and affinity toward GlcNAc-oligomers of the two binding sites of UDA are different (Shibuya et al., 1986). A final remark concerns some peculiarities concerning the specificity of the Gramineae lectins. First, WGA (and probably all other Gramineae lectins) in addition binds GlcNAc and sialic acid. Second, the leaf-specific *Agropyrum repens* lectin and one of the leaf lectins from *Phragmites australis* are better inhibited by GalNAc than GlcNAc and are also quantitatively retained on a column of immobilized GalNAc (Peumans, unpublished results). At present, the reason for this unusual specificity is not understood. However, it may be correlated with the fact that both lectins are composed of subunits that are definitely larger than those of the other Gramineae lectins.

In summary, one can conclude that the overall specificity of the chitin-binding lectins has been conserved during evolution. This apparent conservation of specificity indicates that the binding to chitin- or (GlcNAc)-containing glycoconjugates is probably essential for their function.

E. Three-Dimensional Structure

1. Three-Dimensional Structure of Hevein and Structurally Related Lectins

The three-dimensional structure of hevein was solved by ^{1}H-NMR spectrometry (Andersen et al., 1993). Hevein contains three

strands of β-sheet and two short α-helices (Plate 1). The most striking conformational feature of hevein consists of a stretch of amino acid residues located at the *N*-terminal end of the polypeptide chain that forms two strands of antiparallel β-sheet followed by an α-helix. Three of the four intrachain disulfide bonds of hevein (Cys^3-Cys^{18}, Cys^{12}-Cys^{24}, Cys^{17}-Cys^{31}, and Cys^{37}-Cys^{41}) stabilize the folding of the *N*-terminal region. The two strands of antiparallel β-sheet contain several residues (Ser^{19}, Trp^{21}, Trp^{23}) involved in the binding of GlcNAc-containing oligosaccharides (Asensio et al., 1995). Another residue, Tyr^{30}, belongs to the first α-helix and also participates in the binding of GlcNAc-oligomers.

A three-dimensional model of UDA based on the NMR coordinates of hevein is quite similar to that of hevein. However, due to the deletion of two amino acid residues, it apparently lacks the second α-helix of hevein. Like hevein, the two strands of antiparallel β-sheet of the UDA monomer (residues 16 to 19 and 23 to 26) are probably involved in the binding of *N*,*N*′,*N*″-triacetylchitotriose because ^{1}H-NMR studies showed that a number of these β-sheet residues are perturbed in the presence of the trisaccharide (Hom et al., 1995). These binding properties are probably responsible for the antiviral (Balzarini et al., 1992), antifungal (Broekaert et al., 1989), and insecticidal properties (Huesing et al., 1991b) of UDA and may explain the potent superantigen activity of the lectin (Musette et al., 1996).

The short cysteine-rich antimicrobial peptide Ac-AMP2 from *Amaranthus caudatus* seeds (Broekaert et al., 1992) exhibits a three-dimensional conformation similar to that of hevein and UDA (Martins et al., 1996). The two strands of the antiparallel β-sheet are present, but the polypeptide chain is considerably shortened, mainly at the *C*-terminal end. Hence, Ac-AMP2 lacks the α-helix and the strand of β-sheet located at the *C*-terminus of hevein. However, most of the residues homologous to those involved in the binding of GlcNAc-containing oligosaccharides to hevein (Ser^{16}, Phe^{18}, Tyr^{20}, Tyr^{27}) are well conserved in Ac-AMP2.

2. Overall Three-Dimensional Structure of Wheat Germ Agglutinin

The three-dimensional structure of wheat germ agglutinin is rather complex because the native lectin corresponds to a dimer built from the noncovalent association of two monomers, I and II, each of which consists of four structurally similar hevein domains of 43 residues (A1, B1, C1, and D1 for monomer I and A2, B2, C2, and D2 for monomer II) (Wright, 1977; Wright, 1987) (Figure 12A). Except for a very short α-helix of only five residues, each hevein domain is devoid of regular secondary structures and consists mainly of coil and turn structures. Each domain also contains four disulfide bridges and the resulting 32 disulfide bridges stabilize the overall folding of the dimer. The two dimers associate in a "head-to-tail" fashion in such a way that the facing domains of both monomers correspond to the pairs A1-D2, B1-C2, C1-B2, and D1-A2. Each of the four domains that constitute a monomer possesses a carbohydrate-binding site.

3. Structural Basis for the Recognition of Simple or Complex Sugars by Wheat Germ Agglutinin

The binding-specificity of WGA is not restricted to GlcNAc but extends to various complex GlcNAc- and sialic acid-containing glycoconjugates. Each monosaccharide-binding site consists of four residues belonging to one domain of monomer I, namely, an invariant Ser and three (mainly) aromatic residues, and two other (mainly) polar resi-

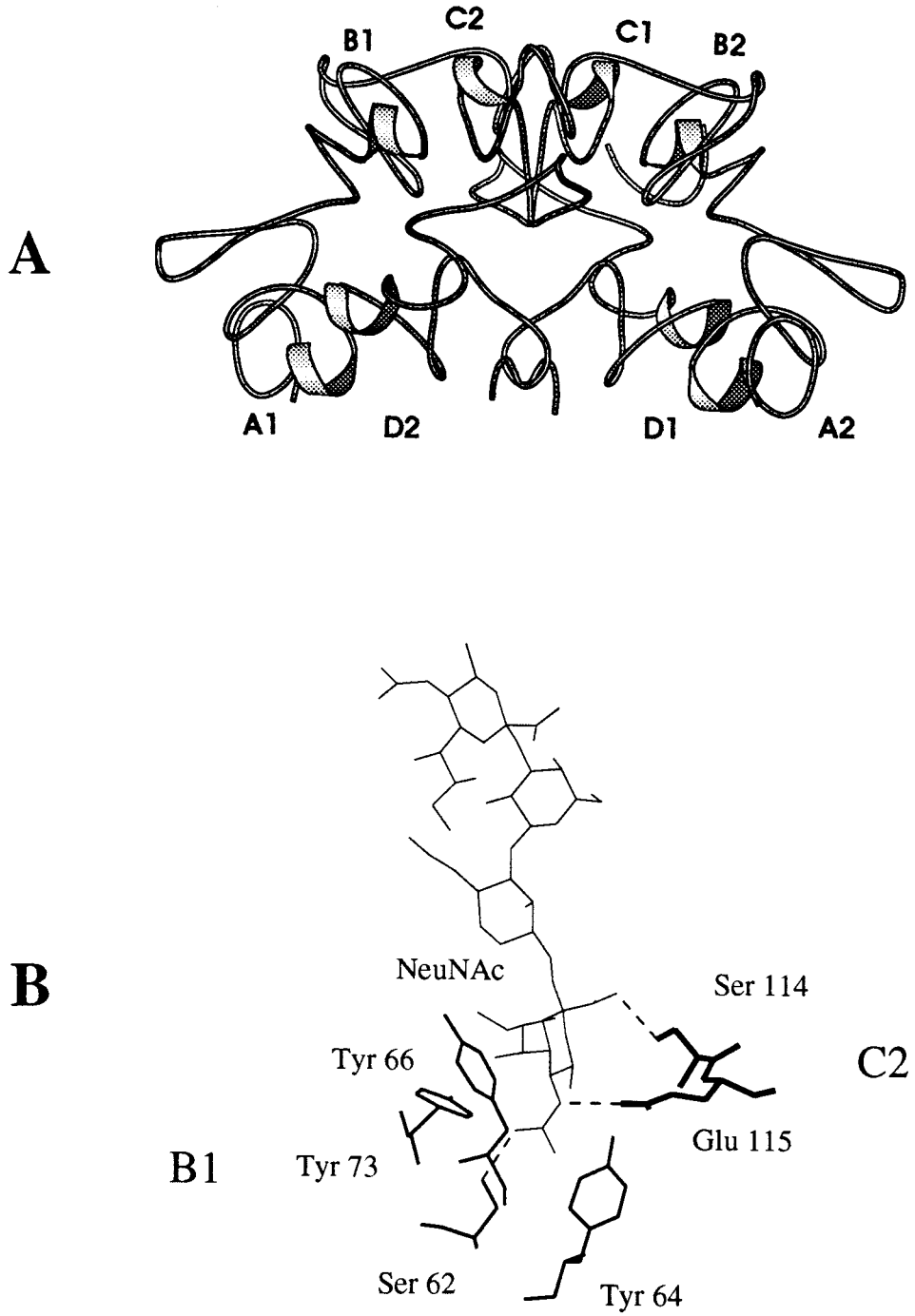

FIGURE 12. (A) Three-dimensional structure of the WGA dimer (codes 7wga and 9wga, Brookhaven PDB). Both monomers consist of four hevein-like domains (A1-B1-C1-D1 and A2-B2-C2-D2) and are noncovalently associated in a head-to-tail fashion to form four binding sites located at the interface of two domains A1-D2, B1-C2, C1-B2, and D1-A2. Each hevein-like domain contains a short α-helix. The cartoon was generated with the Molscript programme (Kraulis, 1991). (B) Binding of the tetrasaccharide of glycophorin-A to the B1-C2 binding-site of WGA. The NeuAc residue (thin line) interacts with the amino residues belonging to the domains B1 (thick lines) and C2 (heavy lines) via a network of three hydrogen bonds (dotted lines). Tyr^{64}, Tyr^{66}, and Tyr^{73} interact with NeuAc via hydrophobic contacts.

dues belonging to the adjacent domain of monomer II. For example, Ser^{62}, Tyr^{64}, Tyr^{66}, Tyr^{73}, and Ser^{114}, Glu^{115} form the carbohydrate binding site of the pair B1-C2. The binding domain B1 is referred as the principal (pr) binding domain, while C2 corresponds to the helper (hl) binding domain. Three hydrogen bonds connect the nonreducing NeuAc of the T5 sialoglycopeptide of glycophorin A (Wright, 1992; Wright and Jaeger, 1993) to the polar residues of the site (Ser^{62}, Ser^{114}, and Glu^{115}), while the three aromatic residues (Tyr^{64}, Tyr^{66}, and Tyr^{73}) interact with NeuAc by hydrophobic contacts (Figure 12B). B1(pr)/C2(hl) and C1(pr)/B2(hl) are the highest-affinity carbohydrate-binding sites, whereas the two other sites, and especially A1(pr)/D2(hl), which contains two helper residues (Pro^{157} and Gly^{158}) and are unable to create hydrogen bonds, correspond to low-affinity binding sites. The same amino acid residues are involved in the binding of *N*,*N*′-diacetylchitobiose, but the binding scheme is slightly different with four hydrogen bonds instead of three (Wright and Kellog, 1996). The high- (B1(pr)/C2(hl)) and low-affinity (D1(pr)/A2(hl)) binding sites for this disaccharide are clearly different.

F. Molecular Evolution

The molecular evolution of chitin-binding lectins involves two different mechanisms. First, there has been a 'classic' evolution in the sequence of the hevein domain by amino acid substitutions and short deletions or insertions. Second, the hevein domain has been used as a kind of mobile element that has been inserted into other genes leading to either multi-hevein-domain proteins or chimeric proteins consisting of (a) hevein domain(s) fused to an unrelated structural or enzymatic domain. Because the non-lectin domains of the different types of chimeric chitin-binding proteins are unrelated (e.g., chitinase versus extensin-like domain), it is not possible to construct a phylogenetic tree for the complete sequences of these proteins. Therefore, this tree has been built for the hevein domains only. A distance matrix has been calculated from the amino acid sequences of 30 different hevein domains (Figure 13). It should be mentioned that the hevein domains of the class I chitinases have not been included because these enzymes clearly form a separate branch of the chitin-binding lectin family.

According to the dendrogram in Figure 13, the phylogenetic tree of the sequenced hevein domains consists of three major branches P, H, and G2, and a minor branch G1. Branch P clusters the hevein domains of the *Phytolacca americana* root (holo)lectins PL-B, PL-C, and PL-D, whereas branch H comprises a heterogeneous set of hevein domains from species of various taxonomic groups. The remaining branches G1 and G2 consist exclusively of the hevein domains of the Gramineae lectins.

According to the dendrogram, the hevein domains of the Gramineae lectins from *Oryza sativa*, *Hordeum vulgare*, and *Triticum aestivum* cluster in two separate branches. Branch G1 exclusively comprises the *C*-terminal domains of the respective lectins, whereas branch G2 contains the domains 1, 2, and 3. The topology of the branches G1 and G2 is indicative for the molecular evolution of the Gramineae lectins. First, the close clustering of the corresponding domains of each lectin strongly suggests that the Gramineae lectins evolved from a common four-domain ancestor. Second, the relative distances between the individual domains of each lectin draw important conclusions about the evolution of the four-domain ancestor of the Gramineae lectins starting from a single hevein domain. It has been suggested previously that WGA arose from two consecutive duplications of an ancestral hevein domain (Wright et al., 1991).

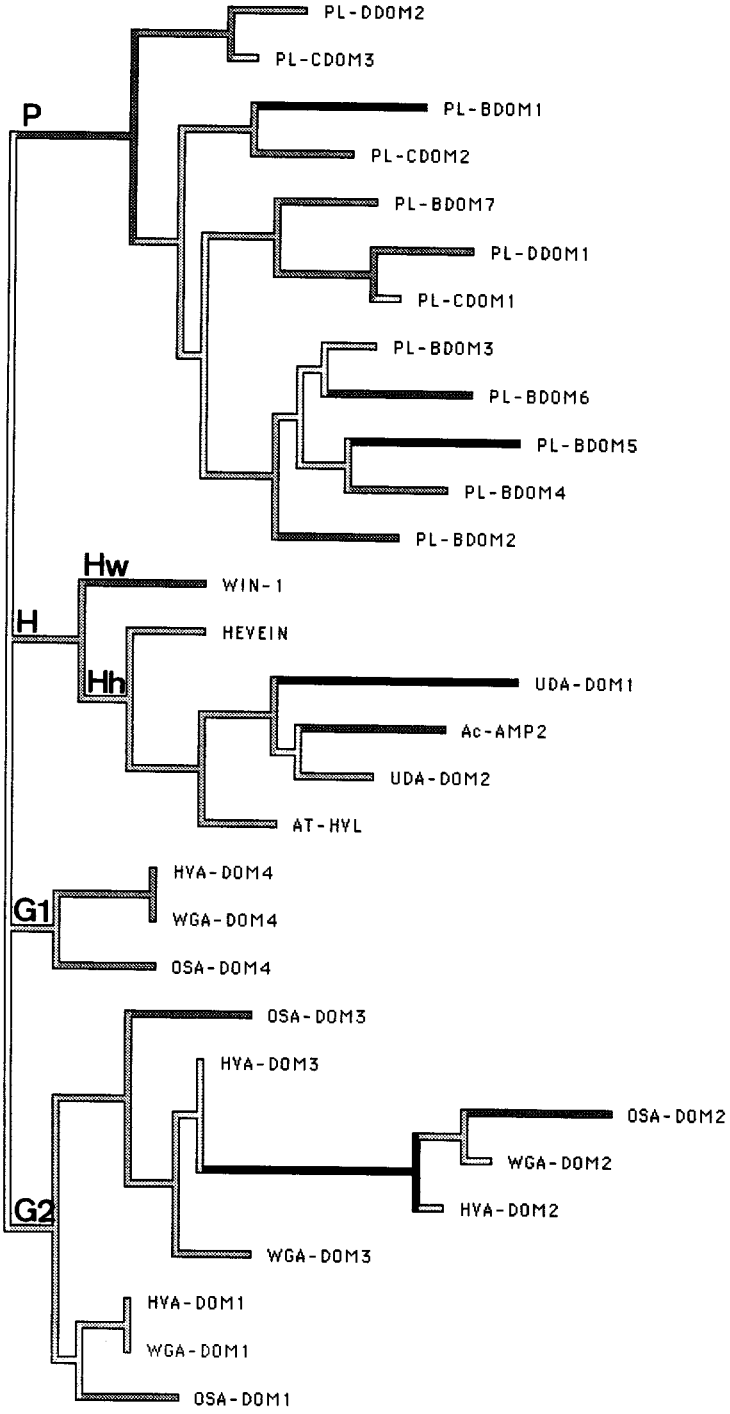

FIGURE 13. Phylogenetic tree of the hevein domains composing chitin-binding lectins and related proteins. The dendrogram was made on a Macintosh LC 630 using the MacClade program (Maddison and Maddison, 1992). The abbreviations of the lectins are listed in Table 2. Besides lectins the dendrogram also includes the sequences of the wound-inducible protein from potato (win) and the hevein-like protein from *Arabidopsis thaliana* (AT-HVL).

According to this theory, a first duplication yielded a two-domain intermediate (comparable to UDA) that, through a second duplication event, gave rise to a four-domain lectin. Unfortunately, the structure of the dendrogram does not support this model because duplication of a two-domain intermediate implies that domain 1 and 3 and domain 2 and 4 form two separate clusters (which is clearly not the case). Based on this observation and the recent sequencing of a three-domain chitin-binding lectin from *Phytolacca americana* (see below), it is suggested that the four-domain Gramineae lectins evolved by three consecutive duplications of a single hevein domain (and hence involve a two-domain and a three-domain intermediate). A putative model of this theory is presented in Figure14A. According to this model, the evolution of WGA started with a one-domain hevein-like protein [W]. Most probably, the original hevein domain corresponds to the far ancestor of the actual domain 4 (because the branch G1 is apparently related closer to the common ancestor of the monocot and dicot hevein domains than the branch G2). A self-duplication of [W] followed by an in tandem insertion yielded a two-domain lectin [W′-W] that evolved into a lectin [X-W]. Duplication and in tandem insertion of the [X] domain then gave rise to a three-domain lectin [X-X′-W] that evolved further into a lectin [X-Y-W]. In the next step, duplication and in tandem insertion of [Y] gave rise to a four-domain lectin [X-Y′-Y-W] that eventually evolved toward the ancestor [X-Z-Y-W] of the Gramineae lectins. Because our model is in agreement with the general structure of the dendrogram, one can reasonably assume that the evolution of the four-domain Gramineae lectins involved both a two-domain and a three-domain intermediate. As discussed below, the proposed model of consecutive duplications and in tandem insertions of a single hevein domain also applies to the molecular evolution of the two, three-, and seven-domain chitin-binding lectins from *Phytolacca americana*.

Branch P of the dendrogram clusters all the hevein domains of the three sequenced *Phytolacca americana* root lectins PL-B, PL-C, and PL-D with subunits consisting of seven, three, and two hevein domains, respectively. The topology of the cluster suggests a complex molecular evolution of the *Phytolacca americana* lectin family. To unravel this apparently complex evolution, the same model of consecutive duplications and in tandem insertions of a single hevein domain as described above for the Gramineae lectins was used. As shown in Figure 14B, self-duplication and in tandem insertion of the ancestral *Phytolacca americana* hevein domain [X] yielded a two-domain intermediate [X-X′] that evolved into a lectin [X-Y]. [X-Y] further evolved giving rise to the polypeptide [D1-D2] of PL-D. In addition, duplication and in tandem insertion of the [X] domain resulted in a three-domain intermediate [X-X″-Y] that eventually yielded the direct ancestor [X-Z-Y] of the polypeptide [C1-C2-C3] of PL-C. In an independent duplication and insertion event, an intermediate [X‴-X] was formed that further evolved into the two-domain ancestor [Q-R] of PL-B. Five consecutive duplications and in tandem insertions of a single hevein domain converted the ancestor [Q-R] into a seven-domain polypeptide [Q-S-T-V-W-U-R] that served as the direct ancestor of the PL-B subunit. The model of the molecular evolution of PL-B shown in Figure 14B does not exactly fit the topology of the P branch, and therefore should be considered a preliminary attempt to explain the origin of the seven-domain lectin. Most probably this preliminary model can be refined as soon as the nucleotide sequences of the *Phytolacca americana* lectins become available.

Branch H comprises a cluster of hevein domains from species of various taxonomic

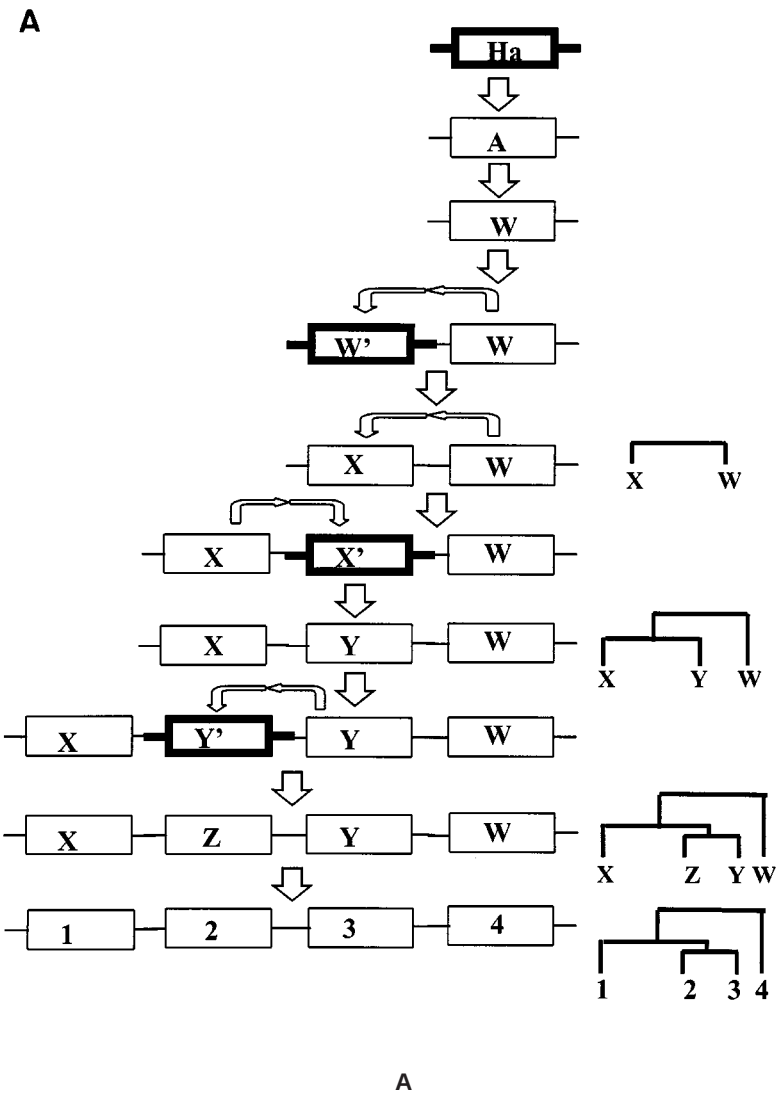

FIGURE 14. (A) Model of the evolution of the four-hevein domain protomers of wheat germ agglutinin. According to this model, the WGA polypeptide is the result of three subsequent hevein domain duplication and in tandem insertion events. (B) Model of the evolution of the multihevein domain protomers of the *Phytolacca americana* lectins PL-B, PL-C, and PL-D. Ha stands for ancestral hevein domain. 'Transposed' hevein domains are shown in bold.

groups. Except for Ac-AMP, all hevein domains of branch H are encoded by chimeric genes. Side branch Hw, for example, clusters the so-called win proteins. Side branch Hh contains hevein, a hevein-like protein from *Arabidopsis thaliana*, the two hevein domains of UDA and Ac-AMP. The *Arabidopsis* hevein-like protein is apparently more closely related to UDA and Ac-AMP than to hevein from *Hevea brasiliensis*. Although both hevein domains of UDA are closely related, it is striking that the second domain of UDA is more closely related to the truncated hevein domain of Ac-AMP than to the first hevein domain of UDA. Most probably UDA arose from a self-duplicated in tandem insertion of the hevein domain of a class I chitinase-like protein.

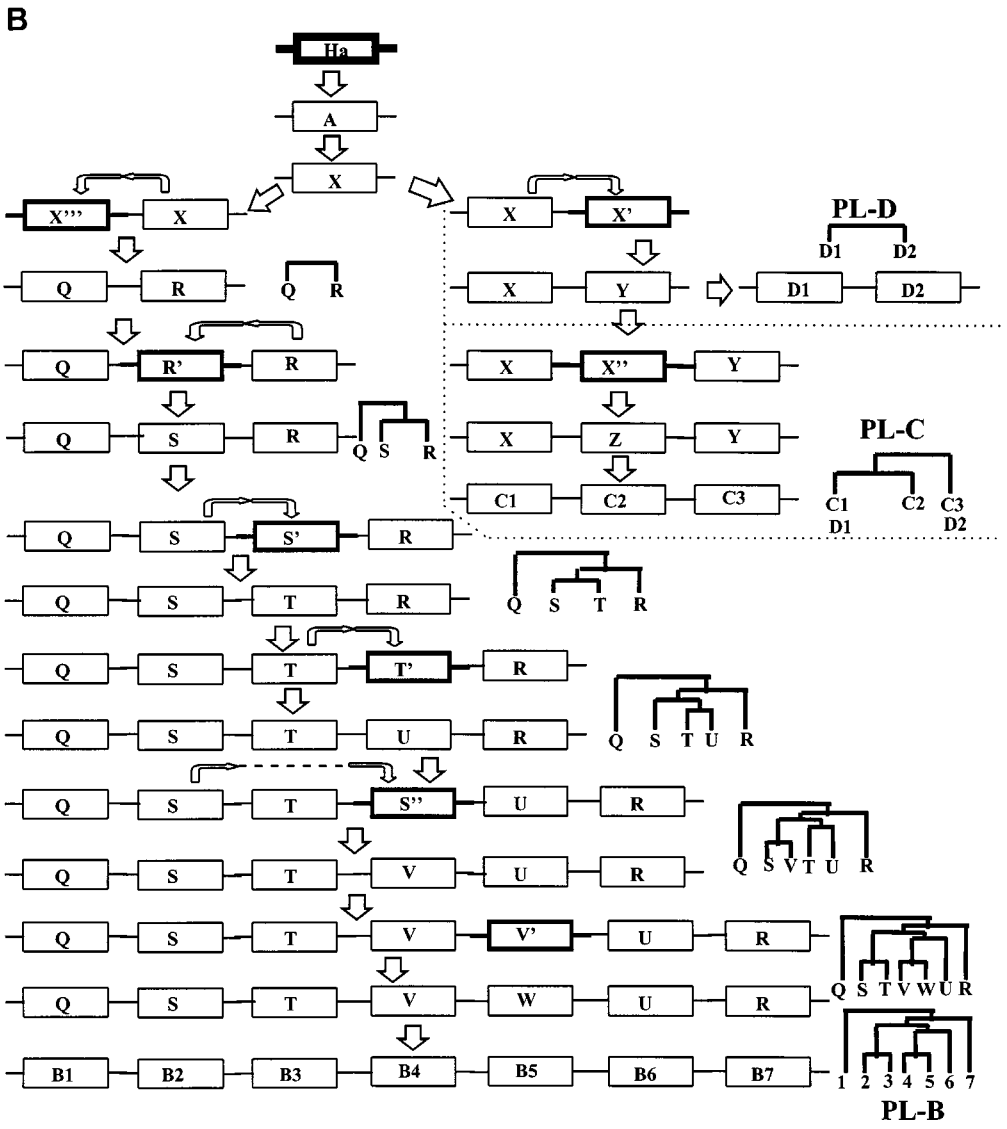

FIGURE 14B

An overview of the structure of the lectin protomers suggests that duplications and insertions of hevein domains played an important role in the molecular evolution of the chitin-binding lectin family. The occurrence of multiple-domain chitin-binding lectins in different unrelated families (e.g., Gramineae, Papaveraceae, Phytolaccaceae, Solanaceae, Viscaceae, and Urticaceae) also addresses the question whether all these lectins evolved from a common ancestor or arose by independent amplification events in several taxonomic groups (Figure 15). Although the concept of a common ancestor cannot be excluded *a priori*, several observations argue in favor of independent hevein domain amplification events. First, the multidomain lectins are scattered over a wide taxonomic range covering both dicots and monocots without a clustering of the two- and three-domain lectins in distinct taxonomic groups. Second, the occurrence of multidomain

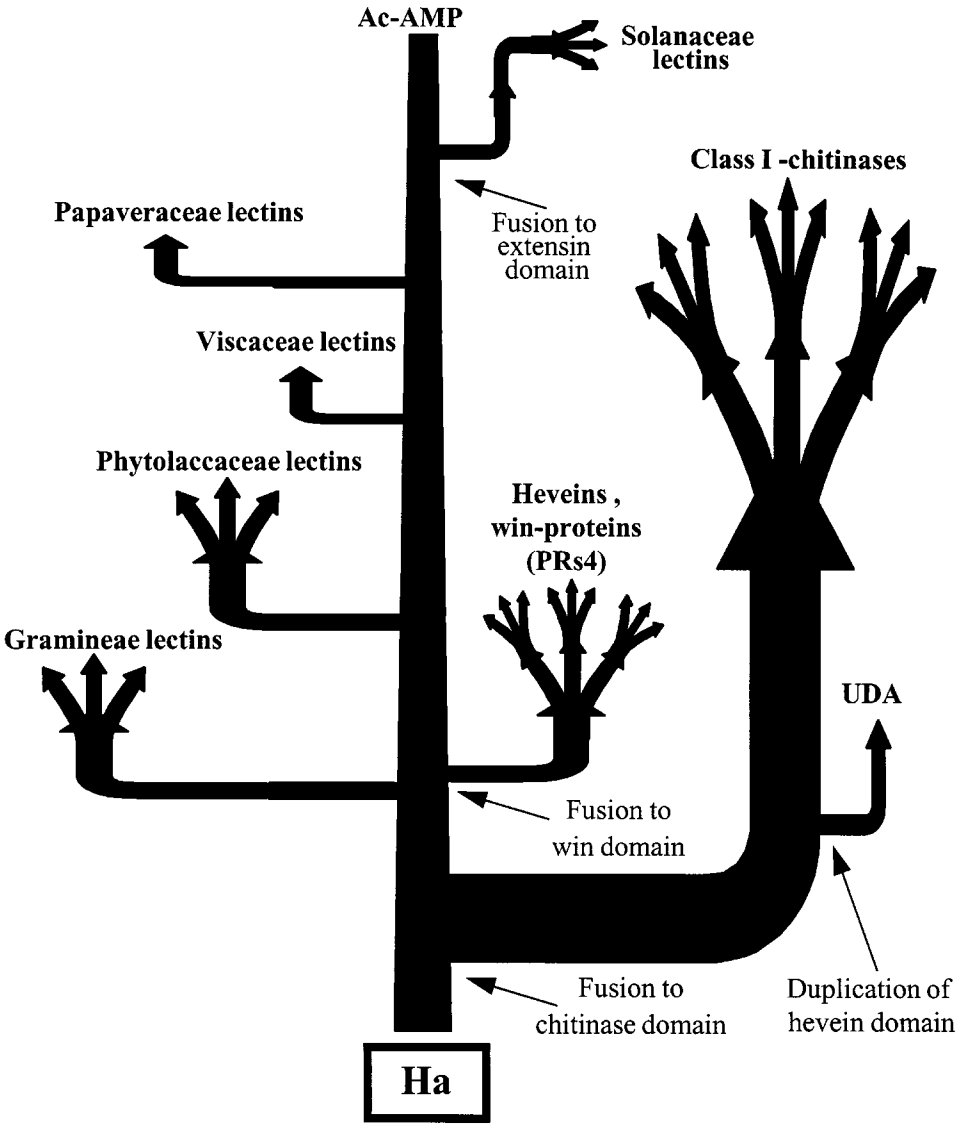

FIGURE 15. Hypothetical model of the molecular evolution of chitin-binding lectins. The ancestral hevein domain (Ha) followed different evolutionary routes. One route leads from Ha to the actual one-hevein domain lectins like the *Amaranthus caudatus* antimicrobial protein (Ac-AMP). Another route starts with the fusion of Ha to a chitinase domain. During evolution of this chimeric lectin into the modern class I chitinases, the hevein domain was duplicated whereby the genes encoding the *Urtica dioica* agglutinin (UDA) evolved. In an independent evolutionary event, Ha fused to a domain corresponding to the C-terminal domain of the actual PRs4. Further evolution of the ancestral chimeric protein eventually yielded the actual heveins, win-proteins, and other PRs4. Ha (or possibly a three-hevein domain derivative) also fused to an extensin-like domain that resulted in the formation of the chimeric Solanaceae lectins. Finally, the hevein domain was amplified at several locations by gene duplication and in tandem insertion events. This mechanism led independently to the origin of the Gramineae, Phytolaccaceae, Viscaceae, and Papaveraceae lectins. It should be noted that the order in which the different lectin groups separated from each other was chosen arbitrarily. Because the genes encoding the Phytolaccaceae, Viscaceae, and Papaveraceae lectins have not been cloned yet, it cannot be excluded that these lectins are derived from chimeric precursors.

lectins is usually confined to a few closely related genera of the relevant families, which indicates that amplification of the hevein domain took place after the species diverged from a common ancestor of the family. Third, the occurrence in *Phytolacca americana* of chitin-binding lectins with two, three, and seven hevein domains demonstrates that some species are specialized in the amplification of the hevein domain. Based on these observations, it is suggested that the currently known multidomain chitin-binding lectins evolved independently in different taxonomic groups through the amplification of a ubiquitous highly conserved hevein domain.

Insertion of hevein domain(s) in unrelated genes is most probably the evolutionary mechanism that has been used by plants to develop the chimeric chitin-binding lectins. Shinshi et al. (1990) suggested that class I chitinases arose through the insertion, by a kind of transposition event, of a hevein domain into the coding region of an ancestral chitinase gene. Because class I chitinases are ubiquitous in higher plants, their common ancestor probably arose very early in the evolution of the plant kingdom. Hevein domains also occur in the proform of UDA, which consists of two hevein domains tandemly arrayed with a domain that has sequence similarity with the catalytic domain of class I chitinases. Possibly, the UDA gene evolved from a 'normal' class I chitinase through a duplication/in tandem insertion of the hevein domain. Heveins as well as the win-proteins and all other group 4 pathogenesis-related proteins (PRs4) probably arose in a similar way to the class I chitinases through the insertion of a hevein domain in the coding region of a gene encoding an ancestral protein corresponding to the *C*-terminal domain of the modern PRs4.

In addition in class I chitinases and PRs4, and structurally similar (pro)proteins, hevein domains occur also in the chimeric Solanaceae lectins, which consist of an *N*-terminal chitin-binding domain of three hevein repeats linked to a *C*-terminal extensin-like domain. At present it is not clear whether a bloc of three hevein domains was inserted into an extensin-like gene or amplification took place after a single hevein domain fused to an extensin-like domain. Because the catalytic domain of the class I chitinases is undoubtedly unrelated to the extensin-like domain of the Solanaceae lectins, one can reasonably assume that the formation of these two different groups of chimeric chitin-binding lectins results from at least two independent evolutionary events involving the insertion of hevein domains into foreign genes.

The origin of the ancestral hevein domain is still enigmatic. Possibly, it arose during early evolution of plants as a specific receptor for bacterial or microbial cell-wall polysaccharides like chitin, chitosan, or muraminic acid, which was an important part in the signaling process by potentially dangerous organisms. Once this function of the hevein domain was superseded by more efficient signaling molecules (like receptor kinases), the hevein domain lost its original role. Fusion of the hevein domain to the catalytic domain of chitinases generated an enzyme with improved defensive capabilities. Therefore, this evolutionary line was maintained throughout the evolution of higher plants. At several occasions this line gave rise to side branches of catalytically less active or inactive homologues. Possibly the hevein domain(s) of these proteins acquired a specific function. The acquisition of a specific role may also be the driving force behind the evolution of the hololectins that arose in several stages through duplication and in tandem insertion of the hevein domain. Although speculative, the same explanation can be given for the fusion of hevein domains to an extensin-like domain.

G. Physiological Role

In the following discussion on the physiological role of the chitin-binding lectins a clear distinction is made between class I chitinases and other members of this lectin family. By virtue of the presence of a catalytic domain, which acts independently from the chitin-binding domain, class I chitinases possess a biological activity that is unrelated to their carbohydrate-binding activity. Because functions based on the enzymatic activity fall beyond the scope of this review, the physiological role of class I chitinases is not discussed in detail. According to the currently accepted ideas, class I chitinases are (together with other defense-related proteins) involved in the plant's defense against fungi (Collinge et al., 1993). Although there is no doubt that the antifungal activity of class I chitinases is based on their chitinolytic activity, it is not clear whether the *N*-terminal hevein domain also takes part in or reinforces the defense properties of these enzymes.

The physiological role of the nonenzymatic chitin-binding lectins is still unclear. In the past, the specificity of these proteins toward chitin has been interpreted as an argument in favor of a defensive role against either fungi or insects (or possibly both). Many efforts have been made to prove the hypothesis that chitin-binding lectins act as antifungal proteins. However, experiments with highly purified chitinase-free preparations have demonstrated unambiguously that most chitin-binding lectins (e.g., all Gramineae and Solanaceae lectins) have no antifungal activity and the previously observed antifungal effects were due to contaminating chitinases. Only a few small chitin-binding lectins definitely exhibit antifungal properties different from those of chitinases. The small Ac-AMPs consisting of a single truncated hevein domain are the most potent antifungal chitin-binding proteins isolated to date (Broekaert et al., 1992). Hevein also has antifungal activity, but it is less potent than Ac-AMPs (Van Parijs et al., 1991). Besides hevein and Ac-AMP, the two-domain chitin-binding lectin UDA also inhibits the growth and development of fungi (Broekaert et al., 1989). UDA does not affect the metabolism of the fungi, but disturbs the synthesis and/or deposition of chitin in the cell wall and thoroughly changes the morphology of the mycelium (Van Parijs et al., 1992). However, it is questionable whether UDA is involved in the nettle's defense against pathogenic microbes because the lectin has no fungicidal activity. Based on the location of UDA, namely, in rhizomes/roots and seeds, and its dramatic effect on the hyphal structure, it has been proposed that the nettle lectin helps to control the colonization of rhizomes and roots by endomycorrhiza (Peumans and Van Damme, 1995b). The observation that the nettle lectin is at least in part responsible for the inability of nettle plants to form arbuscular mycorrhizal symbiosis with the fungus *Glomus mosseae* supports this hypothesis (Vierheilig et al., 1996). According to the available data, both the nettle lectin and the hevein-type chitin-binding lectins are fungistatic rather than fungicidal agents. Therefore, it is still questionable, whether these chitin-binding proteins alone can protect the plant against (pathogenic) fungi. Most probably, the antifungal chitin-binding proteins are only part of a more complex defense system in which they act synergistically with other antifungal proteins.

Some chitin-binding lectins may be involved in the plant's defense against bacteria. It has been observed, for example, that the seed lectin from *Datura stramonium* blocks the movements of normally motile bacteria at the air-water interface. Because the thorn apple lectin is abundantly present in the seed coat and the epidermis and rapidly elutes after imbibition of the seeds, it

has been proposed that this lectin plays a role in the seed's defense against potentially harmful microbes (Broekaert and Peumans, 1986).

Chitin-binding lectins have been associated with the plant's defense against insects. Feeding trials with artificial diets indicated that several chitin-binding lectins interfere with the growth and development of insects. For example, WGA, rice lectin, UDA, and the potato and thorn apple lectins have an inhibitory effect on the development of the larvae of the cowpea weevil (*Callosobruchus maculatus*) (Murdock et al., 1990; Huesing et al., 1991a). It should be mentioned, however, that this typical seed predator is only moderately affected by these chitin-binding lectins. WGA inhibits the larval growth of the Southern corn rootworm (*Diabrotica undecimpunctata*) and kills the neonate *Ostrinia nubilalis* (European corn borer) larvae at fairly low concentrations (Czapla and Lang, 1990). Similarly, the *Phytolacca americana* lectin is lethal to larvae of the Southern corn rootworm (Czapla and Lang, 1990). The results of feeding trials leave no doubt that some chitin-binding lectins in artificial diets have deleterious effects on chewing insects. However, there is no conclusive evidence to date that any of the anti-insect chitin-binding lectins protect the plant against insects in a natural environment. In addition, it still remains to be demonstrated that the expression of a chitin-binding lectin in a transgenic plant offers an increased protection against chewing insects. Tests with artificial diets further indicated that WGA negatively affects the growth and reproduction of the phloem feeding rice brown planthopper (*Nilaparvata lugens*) (Powell et al., 1995a). Again the question remains whether the results of the feeding trials can be extrapolated to field conditions.

The physiological role of the nonenzymatic chitin-binding lectins is not necessarily directly related to plant defense. Hevein, for example, is apparently involved in the coagulation of latex by bringing together rubber particles (Girdol et al., 1994). It is not clear whether this activity of hevein is due to its carbohydrate-binding activity or relies on other properties. Moreover, the latex-coagulating activity can certainly not be extrapolated to hevein-like proteins in plant species without laticifers.

No evidence has been presented yet that chitin-binding lectins are involved in recognition processes within the plant. Such a role can certainly not be excluded *a priori* because the natural target carbohydrates of chitin-binding lectins (i.e., GlcNAc-oligomers) are important signaling molecules in plants. It is difficult to envisage that the currently known chitin-binding lectins can fulfill a signaling role because they possess no signal transduction domain. However, there are possibly chimeric lectins consisting of a hevein domain linked to, for example, a transmembrane domain and a protein kinase domain. Another possibility is that chitin-binding lectins 'capture' endogenous GlcNAc-oligomers and store these potential signaling molecules in inactive or inaccessible forms. A similar mechanism can be used to prevent unwanted stimulation of, for example, root tips cells by exogenous signaling molecules with an oligo-GlcNAc backbone (like the rhizobial Nod factors). Such an 'exogenous-signal-capturing' mechanism could explain why the Gramineae lectins are located exclusively in the organs that come in contact with the soil during germination.

VI. TYPE 2 RIP AND RELATED LECTINS

Ribosome-inactivating proteins (RIP) are commonly known as proteins that catalytically inactivate eukaryotic ribosomes (Barbieri et al., 1993). Their working mecha-

nism is based on the enzymatic removal of a specific adenine residue from a highly conserved loop (A4324 of rat rRNA) in the large subunit of the ribosomal RNA. Cleavage of this specific adenine residue causes a conformational change in the affected loop and prevents binding of elongation factor EF2 to the ribosome. As a result, protein synthesis is arrested and the cell dies. During recent years convincing evidence has been presented that the enzymatic activity of RIP is not restricted to the cleavage of a single adenine residue from eukaryotic ribosomes or rRNA. Most RIP are indeed, capable, of releasing multiple adenine residues from various polynucleotides such as RNA, DNA, and poly(A). Hence, the term polynucleotide:adenosine glycosidase (PAG) would be more appropriate (Barbieri et al., 1996).

RIP are subdivided in two groups. Type 1 RIP consist of a single polypeptide of about 30 kDa with PAG activity (except in a few cases where the 30-kDa protomer is cleaved into smaller polypeptides). Type 2 RIP contain in addition to an enzymatically active A chain and B chain that harbors one or more carbohydrate binding sites and hence possesses lectin activity (unless, as has been demonstrated for one of the *Sambucus nigra* type 2 RIP, all carbohydrate-binding sites lost their sugar-binding activity). Because the A chain of type 2 RIP shares a high sequence similarity with type 1 RIP, type 2 RIP are considered as chimeric proteins composed of a RIP subunit and a lectin subunit.

Both type 1 and type 2 RIP strongly inhibit protein synthesis in a cell-free system. However, most RIP are not toxic to intact cells because they have no access to (cytoplasmic) ribosomes. In general, type 1 RIP are not cytotoxic except in a few cases where the cells (e.g., macrophages and trophoblasts) actively import the RIP by endocytosis. Many, but not all, type 2 RIP can enter intact cells by endocytosis after their B chain is bound to glycan receptors on the cell surface. Once endocytosed, the A chain may be delivered to the cytosol and inactivate the ribosomes, resulting in cell death. Type 2 RIP show marked differences in cytotoxicity ranging from virtually nontoxic to extremely toxic (causing 50% cell death at concentrations below 1 ng/ml). The same holds true for the oral toxicity of type 2 RIP. Some are virtually harmless indeed, whereas others (e.g., ricin) are lethal at a relatively low dose.

Type 2 RIP is a family of typical chimerolectins. Recent work revealed the occurrence of hololectins composed of polypeptides corresponding to the carbohydrate-binding domain of the chimeric type 2 RIP. Some of these hololectins are derived from type 2 RIP (Van Damme et al., 1996c), whereas others are encoded by truncated type 2 RIP genes (Van Damme et al., 1997c). In addition, type 2 RIP with inactive carbohydrate-binding domains have also been isolated (Van Damme et al., 1996d). Because all these proteins are definitely related at the molecular level, they are classified in the group of 'type 2 RIP and related lectins'.

A. Historical Note

The history of type 2 RIP begins with the description of 'ricin' by Stillmark (1888), who isolated the toxic principle of castor bean seeds and attributed the toxicity of the protein to its agglutinating activity. Soon after the discovery of ricin, similar proteins were also isolated from seeds of *Abrus precatorius* (abrin) and *Croton tiglium* (crotin), and in the early 1920s from roots of *Adenia* (Syn. *Modecca*) *digitata* (modeccin). Following the discovery of many nontoxic lectins at the beginning of this century, plant lectin research rapidly focused on legume lectins. As a result, type 2 RIP played a minor role in plant lectin research until Lin

et al. (1970) discovered that these toxins were more toxic to Ehrlich ascites than to normal cells. This important observation caused a revival in type 2 RIP research, which soon resulted in the elucidation of the two-chain structure of the toxins and their inhibitory activity on protein synthesis. The mode of action of type 2 RIP was eventually elucidated by Endo et al. (1987), who discovered the RNA *N*-glycosidase activity of the A chain of the toxins.

For a long time research on type 2 RIP has been dominated by the potent toxins ricin and abrin. Ricin was the first type 2 RIP whose primary structure was determined by chemical methods and whose three-dimensional structure was resolved by X-ray crystallographic analysis (Montfort et al., 1987). Furthermore, ricin was also the first type 2 RIP to be cloned (Lamb et al., 1985). At present, several type 2 RIP and their corresponding genes have been isolated and characterized. In addition, a number of proteins have also been identified that are either derived from or related to type 2 RIP.

B. Occurrence, Molecular Structure, and Amino Acid Sequences

Presently, only a limited number of type 2 RIP and related lectins have been isolated and characterized. As shown in Table 3, genuine type 2 RIP have been found in the family Euphorbiaceae (*Ricinus communis*, *Croton* sp.), Leguminoseae (*Abrus pre-catorius*), Viscaceae (*Viscum album* and *Phoradendron californicum*), Passifloraceae (*Adenia digitata* and *A. volkensii*), Ranunculaceae (*Eranthis hyemalis*), Lauraceae (*Cinnamonum camphora*), Sambucaceae (*Sambucus* sp.), Cucurbitaceae (*Momordica charantia*), and Iridaceae (*Iris* sp.). Type 2 RIP-derived or related lectins have only been reported in *Sambucus nigra*.

Type 2 RIP have been found in seeds and different vegetative tissues. In some species (e.g., castor bean) the type 2 RIP are confined to the seeds. Other species express type 2 RIP mainly (or possibly exclusively) in vegetative tissue(s) (e.g., modeccin in the roots of *Adenia digitata*). Finally, there are also some plant species that contain type 2 RIP in seeds as well as in various vegetative tissues. For example, elderberry expresses tissue-specific type 2 RIP and related lectins in seeds, bark, leaves, and fruits.

Many plants express several type 2 RIP in seeds and/or vegetative tissues. Castor beans, for instance, contain a complex mixture of type 2 RIP consisting of several ricins and so-called *Ricinus communis* agglutinins (which are both genuine type 2 RIP) (Lin and Li, 1980). Similarly, jequirity beans contain mixtures of abrins and *Abrus precatorius* agglutinins (Olsnes, 1978). The occurrence of multiple type 2 RIP has also been documented in vegetative tissues. A classic example is the toxic mistletoe type 2 RIP ML-I, ML-II, and ML-III (Franz et al., 1981). An even more complex mixture of type 2 RIP occurs in elderberry (*Sambucus nigra*) bark, which in addition three genuine type 2 RIP contains, also two type 2 RIP with an inactive B chain, and a hololectin derived from a type 2 RIP (Van Damme et al., 1996b–d, 1997b,c).

Expression levels for individual type 2 RIP vary strongly in different species and tissue(s). Seed-specific type RIP like the castor bean and jequirity bean toxins and agglutinins represent 1 to 5% of the soluble seed protein. Some isoforms, however, may be present at much lower concentrations. In other species the type 2 RIP are minor proteins accounting for less than 0.1% of the proteins. Some vegetative type 2 RIP (and related proteins) are predominant proteins, whereas others are barely detectable. In some instances abundant and rare type 2 RIP occur in the same tissue. For example, elderberry bark contains three very abundant type 2 RIP/lectins (SNA-II, SNLRP1, SNLRP2, each representing >20% of the

TABLE 3
Type 2 Ribosome-Inactivating Proteins and Related Lectins: Occurrence, Molecular Structure, and Specificity

Species	Tissue	Structure[a]	Specificity	Sequence available[b]
Merolectins				
Sambucus nigra	Bark	[P22]	NANA	Nu
	Fruit	[P22]	NANA	Nu
Hololectins				
Sambucus nigra	Bark	II $[P30]_2$	GalNAc>Gal	Nu
	Seed	III $[P30]_2$	GalNAc>Gal	
	Fruit	IVf $[P32]_2$	Gal/GalNAc	Nu (SNA-IV)
	Leaf	IVl $[P32]_2$	Gal/GalNAc	Nu
	Leaf	IV4l $[P32]_4$	Gal/GalNAc	
Chimerolectins				
Abrus precatorius	Seed	[P(34 + 32)]	Gal>GalNAc	Pr, Nu (Abrin)
	Seed	$[P(33 + 29)]_2$	Gal	Pr (APA)
Adenia digitata	Root	[P(28 + 38)]	Gal>GalNAc	
Adenia volkensii	Root	[P(29 + 36)]	Gal	
Cinnamonum camphora	Seed	$[P(30 + 33)]_?$	Unknown	
Eranthis hyemalis	Tuber	[P(30 + 32)]	GalNAc	
Iris hybrid	Bulb	[P(27 + 34)]	GalNAc	
Momordica charantia	Seed	$[P(28 + 30)]_2$	Gal>GalNAc	
Phoradendron californicum	Plant	[P(31 + 38)]	Gal	
Ricinus communis	Seed	[P(32 + 34)]	Gal>GalNAc	Pr, Nu (Ricin)
	Seed	$[P(32 + 36)]_2$	Gal>>GalNAc	Pr, Nu (RCA)
Sambucus canadensis	Bark	I $[P(32 + 35)]_4$	NANA	
Sambucus ebulus	Bark	I $[P(32 + 37)]_4$	NANA	
	Leaf	$[P(26 + 30)]_2$	GalNAc	
Sambucus nigra	Seed	Vs $[P(26 + 32)]_2$	GalNAc>Gal	
	Bark	I $[P(32 + 35)]_4$	NANA	Nu (SNA-I)
	Bark	I′ $[P(32 + 35)]_2$	NANA	Nu (SNA-I′)
	Bark	V $[P(26 + 32)]_2$	GalNAc>Gal	Nu (SNA-V)
	Fruit	If $[P(32 + 35)]_2$	NANA	Nu
	Fruit	Vf $[P(26 + 32)]_2$	GalNAc>Gal	Nu
Sambucus racemosa	Bark	I $[P(30 + 36)]_4$	NANA	
Sambucus sieboldiana	Bark	I $[P(31 + 37)]_4$	NANA	Nu (SSA-I)
	Bark	[P(27 + 32)]	GalNAc>Gal	Nu (Sieboldin)
Viscum album	Plant	I $[P(29 + 34)]_{1–2}$	Gal	
	Plant	II [P(29 + 34)]	Gal/GalNAc	
	Plant	III [P(25 + 30)]	GalNAc>Gal	
Type 2 RIP with inactive B chain				
Sambucus nigra	Bark	[P(32 + 32)]	—	Nu (LRPSN)

[a] [PX] stands for protomer with a molecular mass of X kDa. [P(Y + Z)] indicates that the protomer is cleaved in two polypeptides of Y and Z kDa.

[b] Pr, protein sequence; Nu, nucleotide sequence. The abbreviation in brackets refers to the sequence name used in the dendrogram (Figure 20).

total protein), one abundant type 2 RIP (SNA-I, representing about 5% of the total protein), one type 2 RIP (SNA-V) that accounts for about 1% of the total protein, and a very rare type 2 RIP (SNA-I′), which represents only 0.01% of the total protein.

In summary, type 2 RIP and related proteins occur in diverse taxonomic groups of both monocots and dicots, where they are found in seeds and different types of vegetative tissues. Many plants contain two or more different type 2 RIP. Expression levels of the individual type 2 RIP vary from very high to very low.

1. Molecular Structure of the Native Lectins

The family of type 2 RIP/lectins comprises proteins with an apparently different molecular structure. Besides the chimeric type 2 RIP, several hololectins and at least one merolectin have been identified, which are derived from either genuine type 2 RIP or truncated type 2 RIP genes (Figure 16).

All type 2 RIP are built up of similar protomers consisting of disulfide bridge-linked A and B chains. The A chain (25 to 30 kDa) possesses *N*-glycosidase activity, whereas the B chain (30 to 35 kDa) has the carbohydrate-binding activity. Both A and B chains are derived from a single precursor of about 60 to 65 kDa. This precursor is posttranslationally processed through the excision of a linker between the *N*-terminal A domain and the *C*-terminal B domain. After processing, the A and B chains are still held

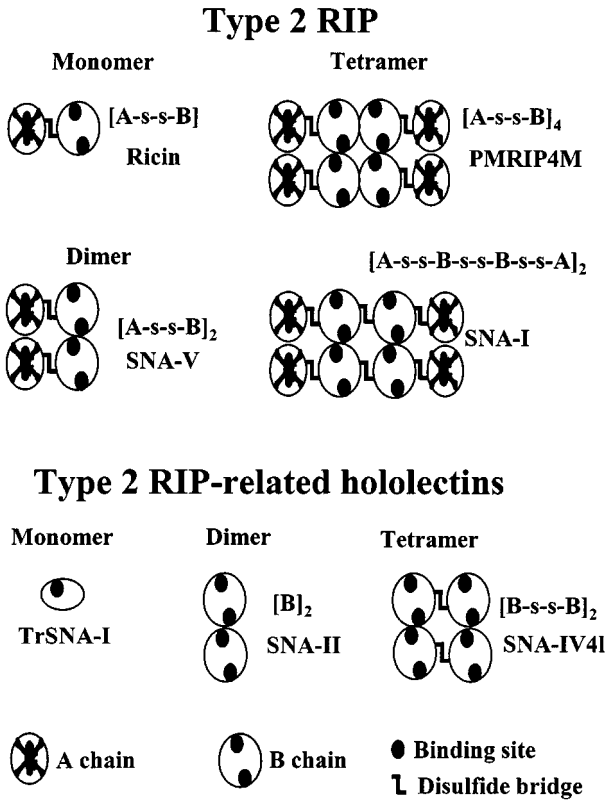

FIGURE 16. Schematic representation of the molecular structure of native type 2 RIP and related lectins. Examples shown are ricin, *Sambucus nigra* agglutinin V (SNA-V), *Polygonatum multiflorum* tetrameric type 2 RIP (PMRIP4M) (W. Peumans, unpublished results), *Sambucus nigra* agglutinin I (SNA-I), *C*-terminal fragment of the B chain of SNA-I (TrSNA-I), *Sambucus nigra* agglutinin II (SNA-II), and *Sambucus nigra* agglutinin IV4l (SNA-IV4l).

together by a disulfide bond between cysteine residues at the *C*-terminus of the A domain and the *N*-terminus of the B chain. The resulting protomer (also called [A-s-s-B]-pair) has a typical chimeric structure and exhibits enzymatic as well as carbohydrate-binding activity (Barbieri et al., 1993). All currently known type 2 RIP consist of glycosylated protomers. According to the overall carbohydrate content, most type 2 RIP contain two to four glycan chains. Usually, both the A and B chain are glycosylated. For example, ricin contains four *N*-glycan chains.

Native type 2 RIP consist of one or two, and, in a few exceptional cases, four identical [A-s-s-B]-pairs. Because the [A-s-s-B]-pair is a single structural unit, type 2 RIP consisting of one, two, and four [A-s-s-B]-pairs are considered as monomeric, dimeric, and tetrameric proteins, respectively. Monomeric type 2 RIP include the classic toxins ricin, abrin, modeccin, and volkensin, as well as the cytotoxic lectins from *Eranthis hyemalis*, *Iris* sp., and *Phoradendron californicum*. Two of the *Viscum album* lectins (ML-II and ML-III) also occur as monomeric type 2 RIP. The molecular structure of the mistletoe lectin ML-I depends on its concentration. At low concentration, ML-I is a monomer, but above 20 μg/ml it forms dimers (Franz et al., 1981). The nontoxic *Sambucus nigra* lectin-related protein (SNLRP) is also a monomeric type 2 RIP. However, because SNLRP has no carbohydrate-binding activity, it cannot be regarded as a lectin (Van Damme et al., 1996d).

Dimeric type 2 RIP comprise the *Ricinus communis* (RCA) and *Abrus precatorius* (APA) agglutinins, the lectin from *Momordica charantia,* and several RIP from *Sambucus nigra* (e.g., SNA-I′, SNA-V) and other *Sambucus* sp. In the past it was generally thought that the protomers of the dimeric type 2 RIP are held together by noncovalent interactions. Recently, however, it has been proposed on the basis of X-ray crystallographic analysis that the two protomers of native RCA are linked through an intermolecular disulfide bond between the two A chains (Sweeney et al., 1997). Because the cysteine residue involved in this disulfide bond is absent from the A chain of ricin, the difference in molecular structure between the dimeric lectin and the monomeric toxin can be explained in terms of the presence or absence of this extra-cysteine residue. The formation of an intermolecular disulfide bond between the two A chains of dimeric type 2 RIP can certainly not be generalized. For example, the A chains of SNA-I′ and SNA-V of elderberry lack the extra-cysteine residue present in the A chain of RCA but still occur as dimers. Moreover, there is also evidence that dimeric type 2 RIP is stabilized by noncovalent interactions between the B chains of the protomers.

Besides monomers and dimers, there are also a few documented examples of tetrameric type 2 RIP. SNA-I, a Neu5Acα(2,6)Gal/GalNAc-binding type 2 RIP from elderberry bark consists of four protomers that are linked in pairs through an intermolecular disulfide bridge between the B chains of two adjacent [A-s-s-B]-pairs (Van Damme et al., 1996b). The formation of the intermolecular disulfide bond clearly depends on the presence of an extra-cysteine residue in the B chain of SNA-I. It has been demonstrated indeed that SNA-I′, which shares a high sequence identity with SNA-I but lacks the extra-cysteine residue in its B chain, is a dimeric type 2 RIP (Van Damme et al., 1997b). Tetrameric homologues of SNA-I have also been found in other elderberry species such as *S. sieboldiana* (Kaku et al., 1996), *S. ebulus*, *S. canadensis,* and *S. racemosa*.

The first evidence for the occurrence of a hololectin consisting of subunits homologous to the B chain of type 2 RIP was reported in a paper describing the biochemical

characterization of a GalNAc-specific lectin from elderberry bark (Kaku et al., 1990a). This lectin, called *Sambucus nigra* agglutinin II or SNA-II, is a homodimeric protein composed of 30-kDa subunits that share a marked sequence similarity at their *N*-terminus with the *N*-terminal sequence of the ricin B chain. Initially, the link between SNA-II and type 2 RIP remained unclear because at that time there was no indication for the occurrence of type 2 RIP in elderberry. However, after the identification of SNA-I as a genuine type 2 RIP, the possible relationship between SNA-II and type 2 RIP was investigated in detail. Molecular cloning of a GalNAc-specific type 2 RIP from elderberry (called SNA-V) eventually revealed that the subunits of SNA-II correspond to the B chain of SNA-V except that they lack the first eight residues (including the cysteine residue, which is involved in the disulfide bond linking the A and B chains) (Van Damme et al., 1996c). Inasmuch as the SNA-II polypeptide is derived from the same precursor as the genuine type 2 RIP SNA-V, SNA-II is considered as a type 2 RIP-derived hololectin.

Soon after the discovery of SNA-II, two other homodimeric GalNAc-specific hololectins consisting of 30-kDa subunits were isolated from seeds (SNA-III) and fruits (SNA-IV) of elderberry. *N*-terminal sequencing revealed both SNA-III and SNA-IV differ at their *N*-terminus from SNA-II and hence must be different proteins. Molecular cloning of SNA-IV eventually demonstrated that the subunits of this lectin are encoded by a type 2 RIP gene, which has a major deletion in the A chain (Van Damme et al., 1997c). Accordingly, SNA-IV is considered a hololectin encoded by a truncated type 2 RIP gene. A similar protein, called SNA-IVl has also been isolated from the leaves. Further analyses of the fruit and leaf lectins resulted in the isolation of molecular variants of SNA-IV and SNA-IVl. These variants, which are homotetramers consisting of four identical subunits of about 30 kDa, are linked in pairs through an interchain disulfide bond. *N*-terminal sequencing demonstrated that the tetrameric lectins have slightly different *N*-termini compared with the dimeric lectins SNA-IV and SNA-IVl (Peumans, unpublished results). Most probably, the subunits of the tetrameric lectins have an extra-cysteine residue in their sequence.

All the hololectins described above are glycoproteins. Based on to the total carbohydrate content, the respective lectin polypeptides contain two *N*-linked glycans. Presently, the occurrence of hololectins derived from type 2 RIP or encoded by truncated type 2 RIP genes has been reported only in *Sambucus nigra*. One can reasonably expect, however, that homologues of the *Sambucus nigra* lectins also occur in other *Sambucus* species.

Besides chimerolectins and hololectins, elderberry also contains merolectins derived from type 2 RIP. Detailed analyses of total preparations of fetuin-binding lectins from fruits and bark of *Sambucus nigra* revealed the occurrence of single chain proteins of about 22 kDa. *N*-terminal sequencing indicated that these small proteins correspond to the *C*-terminal part of the B chain of the NeuAcα(2,6)Gal/GalNAc-binding RIP from fruits (SNA-If) and bark (SNA-I). Accordingly, the small lectins have been called truncated SNA-If and truncated SNA-I (or TrSNA-If and TrSNA-I), respectively. Both proteins contain two glycan chains and are proteolytic processing/degradation products of the respective parent enzymes. At present, it is not clear whether the fragments occur *in vivo* or are generated during extraction and purification. TrSNA-If and TrSNA-I bind to immobilized fetuin equally well as their parent type 2 RIP but do not agglutinate red blood cells because they possess only one carbohydrate-binding site per molecule (Peumans et al., 1998).

2. Primary Structure of Type 2 RIP and Related Lectins

Up to now, only a few type 2 RIP have been sequenced completely by chemical methods (Table 3). However, the primary structure of several type 2 RIP has been deduced from the nucleotide sequence of cDNAs (Table 3). Although none of the type 2 RIP-related holo- and merolectins has been sequenced completely at the protein level, their primary structure can be deduced from the nucleotide sequences of cDNA or genomic clones. All type 2 RIP share a high sequence identity in both their A and B chains. Similarly, the type 2 RIP-related holo- and merolectins exhibit a high sequence identity with (part of) the B chains of the genuine type 2 RIP.

A closer examination of the primary structure of the type 2 RIP protomers clearly indicates that the A and B chains are two distinct domains. The A chain, which has a high sequence similarity with the type 1 RIP and harbors the RIP activity, apparently consists of a single domain. In contrast, the carbohydrate-binding B chain consists of six tandemly arrayed subdomains of about 40 residues with a high internal sequence similarity (Figure 17A).

C. Structure, Biosynthesis, and Posttranslational Modifications

A great deal of our current knowledge of type 2 RIP and related lectins has been obtained through molecular cloning and analy-

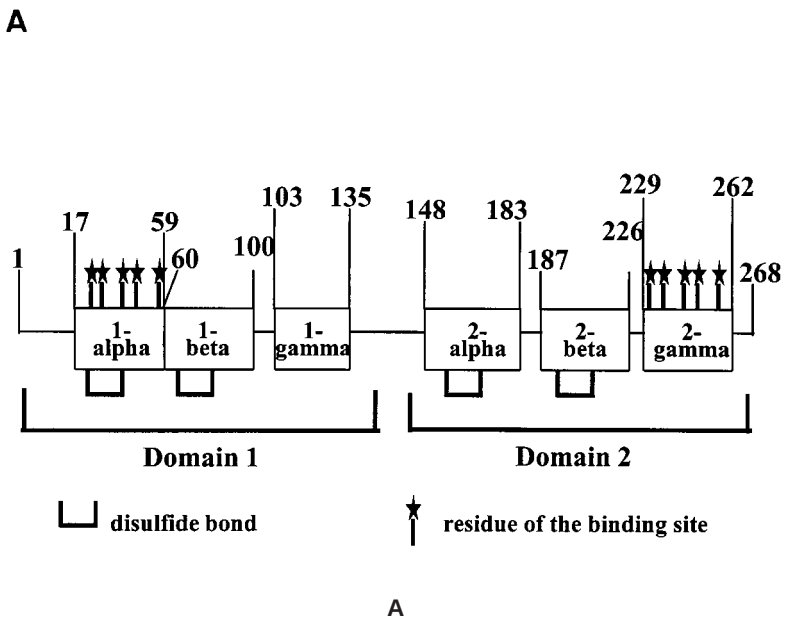

FIGURE 17. (A) Domain and subdomain structure of the B chain of ricin. Numbers refer the position of the amino acid residues along the ricin B chain. (B) Model of the molecular evolution of the ricin B chain. Two subsequent duplications and in tandem insertions of a single subdomain [X] resulted in the formation of an ancestral domain [1–2–3]. Duplication/in tandem insertion of the domain [1–2–3] yielded the typical two-domain B chain consisting of six subdomains. (C) Hypothetical model of the molecular evolution of type 2 and related lectins. The ancestral subdomain [Xa] evolved through a three-subdomain intermediate [Ia] into a six-subdomain ancestral lectin [Ba]. Fusion of [Ba] to an *N*-glycosidase domain yielded the direct ancestor of the type 2 RIP. Deletion of a large part of the A-domain of an elderberry type 2 RIP (SNA-V) gene generated a gene encoding a hololectin (SNA-IV). Another elderberry type 2 RIP evolved toward a lectin-related protein (LRPSN).

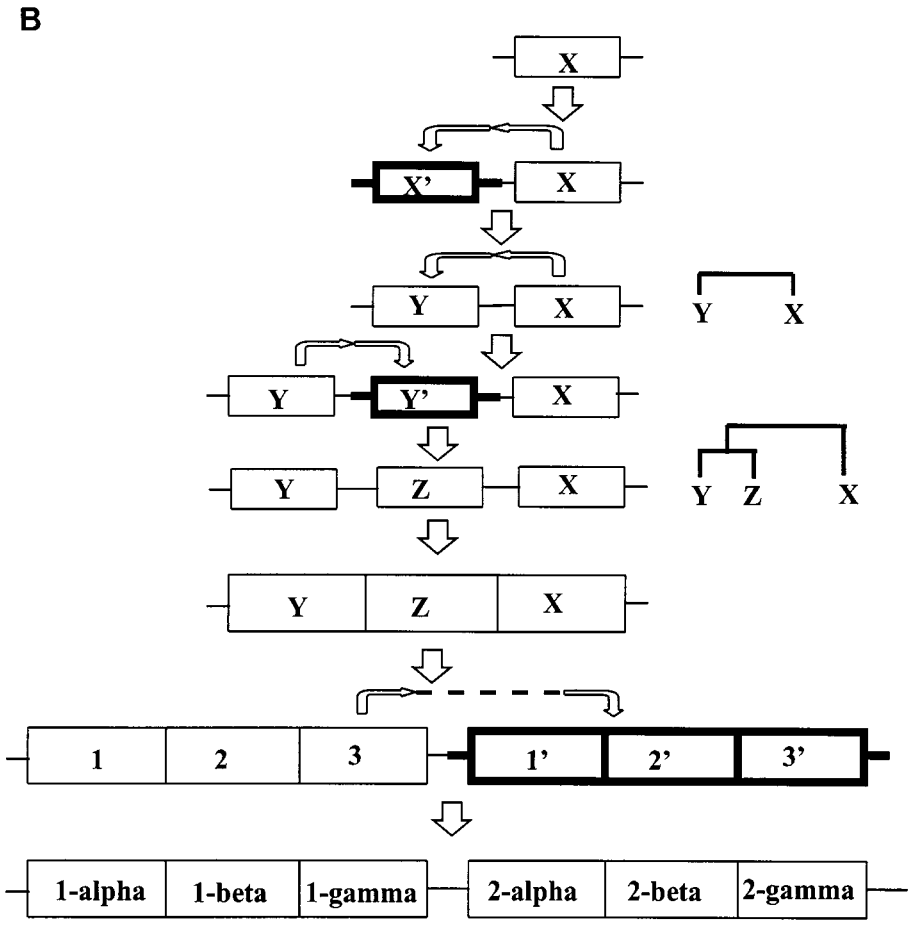

FIGURE 17B

sis of cDNAs and/or genomic DNA fragments. The determination of the structure of the genes encoding type 2 RIP as well as the unraveling of the biosynthesis and posttranslational processing of the primary translation products of these genes is mainly based on the pioneering molecular work on castor bean seed toxins and agglutinins. Similarly, most of the current insights into various structural aspects of type 2 RIP-related hololectins and merolectins are based on molecular cloning and analysis of the type 2 RIP/lectin family of *Sambucus nigra*. Because understanding of the molecular structure and biosynthesis/posttranslational processing of genuine type 2 RIP is a prerequisite to establish how the related holo- and merolectins are synthesized, the castor bean seed system is discussed first. Following the *Sambucus nigra* system is explained in detail to highlight the relationship between genuine type 2 RIP and related holo- and merolectins (Figure 18).

1. Genuine Type 2 RIP

Although the results of biochemical analyses and molecular cloning clearly demonstrated that all type 2 RIP consist of [A-s-s-B]-pairs, which are derived from similar precursor molecules, precise information

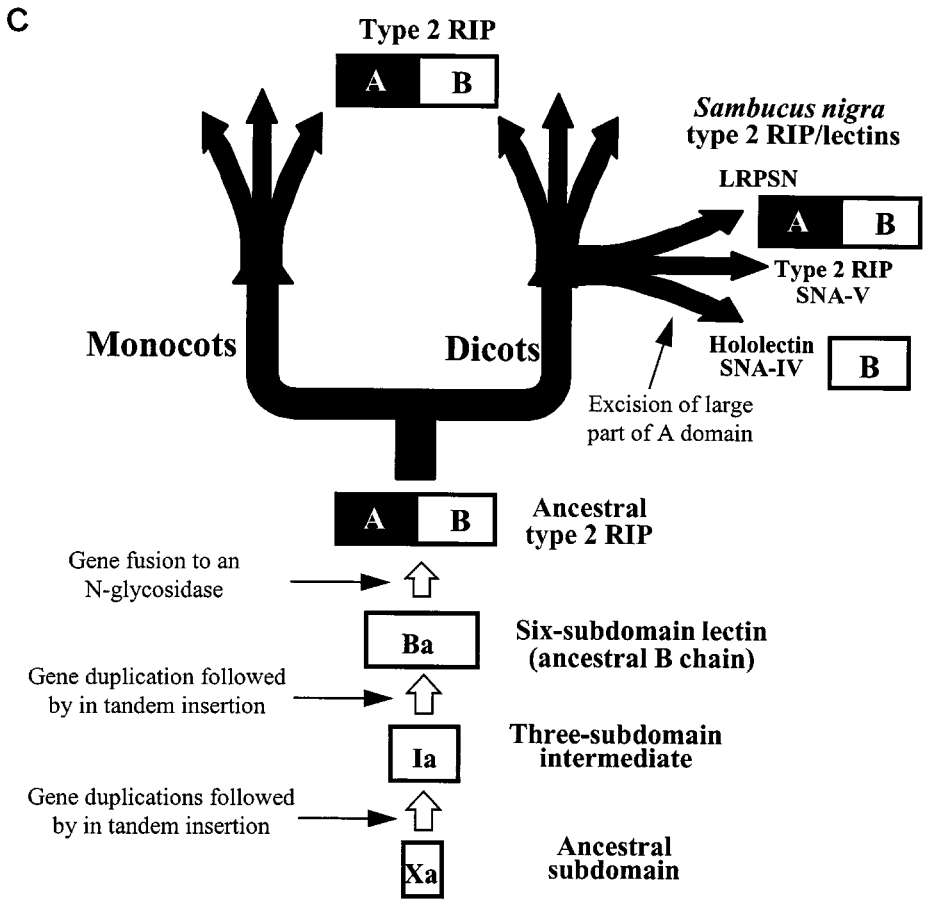

FIGURE 17C

about the conversion of the primary translation products of their mRNAs into the mature protomers has been obtained only for ricin. The combined results of biochemical, cellular biological, and molecular biological studies indicate that the synthesis of ricin follows the secretory pathway (Lord, 1985). Ricin is synthesized on the ER as a prepropotein that undergoes co-translational cleavage of a signal peptide and *N*-glycosylation at four sites, two in both A and B chains. The resulting glycosylated proricin is transported from the ER via the Golgi complex to protein storage vacuoles. During this transport some of the oligosaccharide side chains of proricin are modified into complex type *N*-glycans. Final modifications of the proricin polypeptide backbone take place in the protein bodies. A first modification consists of the excision of a 12–residue linker peptide between the A and B domain by an endopeptidase. As this excision occurs within a disulfide loop connecting the *C*-terminus of the A chain and the *N*-terminus of the B chain, both chains remain covalently linked. A second modification of the polypeptide backbone is the cleavage of a short peptide from the *N*-terminus of proricin. Proricin is fully capable of binding carbohydrate but definitely has no polynucleotide:adenosine glycosidase activity. The toxin is apparently transported out of the ER in a catalytically inactive form and becomes proteolytically activated only after

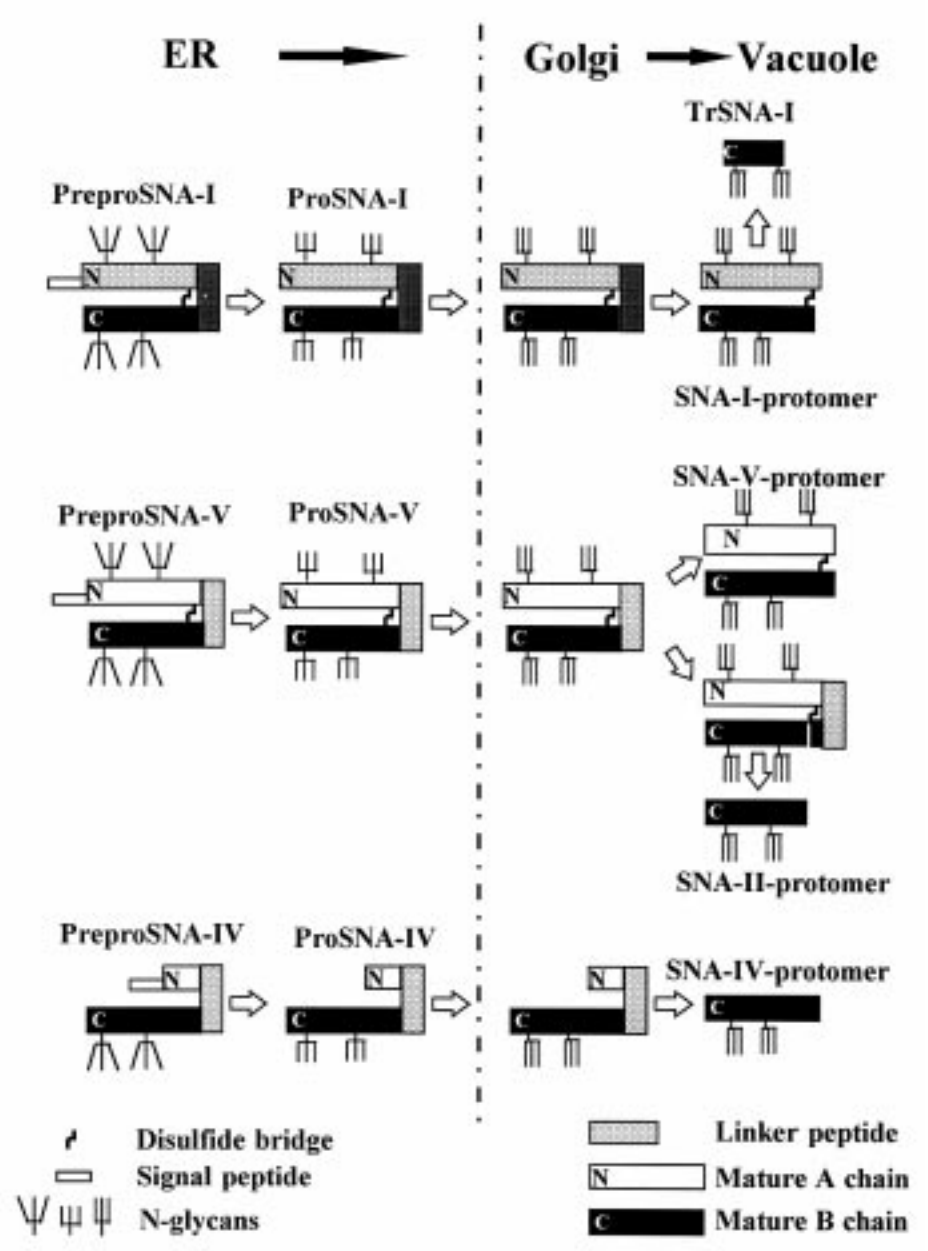

FIGURE 18. Schematic representation of the biosynthesis, co- and posttranslational modifications, and topogenesis of type 2 RIP and related lectins. Examples shown are the elderberry lectins SNA-I, SNA-V, and SNA-IV, respectively. Note that SNA-I yields also a C-terminal fragment (TrSNA-I) and that proSNA-V is converted into SNA-V and SNA-II. Inserted 'N' and 'C' indicate the *N*- and *C*-terminus of the protomer, respectively.

its arrival in the protein bodies (Lord et al., 1994). This mechanism probably prevents that by accidental miscompartimentalization into the cytosol ricin kills the cell.

The biosynthesis, processing and topogenesis, as described for ricin, probably applies to all genuine type 2 RIP. It is likely, however, that minor differences occur in the length of the linker sequence between the A and B chains, or the *N*-terminal propeptide. In addition, it has been demonstrated that the precursor of the type 2 RIP SNA-V from

elderberry bark undergoes not only the 'normal' but also an 'aberrant' processing that eventually gives rise to a genuine type 2 RIP SNA-V and a hololectin SNA-II composed of two slightly truncated B chains, respectively (Van Damme et al., 1996c).

Molecular cloning of ricin and related type 2 RIP also provided an explanation for the complex mixture of toxins and agglutinins in castor bean seeds. Ricin and the agglutinin are definitely encoded by different genes. It was shown by Tregaer and Roberts (1992) that ricin and RCA are encoded by a small gene family of approximately eight members, some of which are not functional. The expression of the genes is tissue-specific and developmentally regulated. Moreover, some castor bean varieties contain in addition to the classic ricin (ricin D), a variant toxin called ricin E, and express two distinct toxin genes. Interestingly, ricin E and ricin D are identical except that the *C*-terminal part of the B chain of ricin E is replaced by the corresponding domain of the castor bean agglutinin. Most probably, the gene encoding ricin E arose by recombination between the genes of ricin D and the agglutinin (Ladin et al., 1987).

Jequirity beans contain, like castor beans, a complex mixture of type 2 RIP, consisting of several isoforms of abrin and the *Abrus precatorius* agglutinin (APA). At present, no information is available on the biosynthesis, processing, and intracellular transport of the *Abrus* type 2 RIP. However, based on the sequence information from the cDNA clones (Hung et al., 1993), it can be predicted that abrin is synthesized, transported, and processed in the same way as ricin. In addition, molecular cloning demonstrated that the different isoforms of abrin are encoded by distinct although related genes (Wood et al., 1991).

Sequence analysis of genomic clones or PCR-amplified fragments encoding type 2 RIP did not yield evidence for the occurrence of introns, although they have been found in some type 1 RIP genes (Kataoka et al., 1993).

2. Type 2 RIP-Related Hololectins

Heretofore, type 2 RIP-related hololectins have been found exclusively in *Sambucus* species. Protein purification and molecular cloning have demonstrated that elderberry trees express complex mixtures of type 2 RIP and related proteins in bark, leaves, fruits, and seeds. At present, three types of hololectins and one type of merolectin were isolated. The first hololectin is the very abundant bark-specific lectin, SNA-II. This hololectin is a homodimer consisting of two identical subunits that have exactly the same *N*-terminal amino sequence as the B chain of the type 2 RIP SNA-V, except that they lack the first eight residues. Attempts to isolate cDNA clones encoding the SNA-II polypeptide were unsuccessful. Moreover, Northern blot analysis of elderberry bark RNAs exclusively yielded a very strong signal of about 2 kb (which is typical for type 2 RIP mRNAs) when hybridized with a specific probe for SNA-II. Because the apparent lack of an mRNA with a length corresponding to the SNA-II polypeptide (about 1 kb) could not be reconciled with the abundance of SNA-II in the bark, it was concluded that both SNA-V and SNA-II are encoded by the same gene. Hence, SNA-V and SNA-II are derived from the same precursor through differential processing (Figure 18). About 5% of the proSNA-V undergoes the 'normal' (i.e., ricin-like) processing, resulting in the formation of the type 2 RIP SNA-V. The bulk of the proSNA-V undergoes an alternative processing whereby a cleavage occurs after amino acid number eight of the B chain, resulting in a slightly truncated B chain. At present, the fate of the remainder of the precursor containing the A chain and the linker

sequence is not known. However, as no free A chains have been found in the elderberry bark (whereas the SNA-II polypeptides account for about 20% of the total bark protein) the *N*-terminal part of the aberrantly processed precursor is probably degraded. Hitherto, SNA-II has been found exclusively in the bark of *Sambucus nigra*. Because fruits, leaves, and seeds also contain type 2 RIP (but no SNA-II homologue) the aberrant processing of proSNA-V is probably a bark-specific event. An SNA-V-like type 2 RIP (called sieboldin) has also been isolated and cloned from the bark of *S. sieboldiana* (Rojo et al., 1997), but it is not clear whether this species contains an SNA-II homologue.

A second type of RIP-related hololectins are the dimeric GalNAc-specific fruit and leaf lectins SNA-IVf and SNA-IVl, respectively. Both lectins are homodimers consisting of subunits that show a high sequence similarity to the B chain of SNA-V except for the first 10 residues. Interestingly, the SNA-IV polypeptides lack the cysteine residue that is involved in the interchain disulfide bond of type 2 RIP. Molecular cloning of cDNA and genomic DNA fragments has shown that SNA-IVf and SNA-IVl are encoded by genes that strongly resemble type 2 RIP genes but have a major deletion in the A chain (Van Damme et al., 1997c). Although the biosynthesis and processing of SNA-IV has not yet been studied, it can be predicted on the basis of the deduced sequence of the primary translation product and the structure of the mature protein that SNA-IV is synthesized as a preproprotein that is co-translationally processed by the removal of the signal peptide and *N*-glycosylation at a single site (Figure 18). The resulting prolectin is subsequently transported from the ER to a vacuolar compartment and, during or after transport, further processed by proteolytic cleavage of the truncated A chain and linker sequence.

Sequencing of PCR-amplified fragments encoding SNA-IVl and SNA-IVf did not yield evidence for the occurrence of introns in the genes encoding these type 2 RIP-related hololectins. At present no sequence information is available for SNA-III or the tetrameric variant of SNA-IV.

3. Type 2 RIP-Derived Merolectins

Fruits and bark of elderberry contain small fetuin-binding proteins consisting of a single polypeptide with an *N*-terminal sequence identical to a sequence located in the middle of the B chain of the type 2 RIP SNA-If and SNA-I, respectively. Because no evidence could be obtained for the occurrence of small mRNAs encoding the *C*-terminal part of the B chain of SNA-If or SNA-I, it is postulated that the small lectins are processing or degradation products of the parent type 2 RIP (Peumans et al., 1998). Accordingly, the small lectins have been called truncated SNA-If and truncated SNA-I. Attempts to isolate similar fragments from other type 2 RIP like SNA-V or related hololectins (e.g., SNA-II or SNA-IV) have been unsuccessful.

D. Carbohydrate-Binding Specificity

The sugar-binding activity and specificity of type 2 RIP and related lectins are determined solely by the carbohydrate-binding site(s) of the B chain(s). Most type 2 RIP are effectively inhibited by Gal, GalNAc, or both Gal and GalNAc but clearly show a preference for di- or oligosaccharides (Table 3). A few type 2 RIP are neither Gal- nor GalNAc-specific but exclusively bind NeuAcα(2,6)Gal/GalNAc-containing glycans. Type 2 RIP-derived hololectins and merolectins exhibit the same specificity as the parent type 2 RIP.

Detailed specificity studies have only been performed for a limited number of type 2 RIP and related lectins, such as the toxins

and agglutinins from castor bean, the mistletoe lectins, and several *Sambucus* lectins. The results of these studies are usually less clear-cut than those of similar studies with, for example, legume lectins because the B chain of all type 2 RIP and the subunits of the related hololectins possess two (or possibly three) carbohydrate-binding sites with a (slightly) different specificity and/or affinity. Most probably, the overall specificity of the type 2 RIP is mainly determined by the specificity of the binding site with the highest affinity for the sugar.

Type 2 RIP usually occur as mixtures of two or more RIP and/or related lectins. In general, the individual type 2 RIP/lectins exhibit a similar although not identical specificity. For example, the castor bean toxin (ricin) and agglutinin preferentially bind Gal and GalNAc, respectively. Similarly, ML-I and ML-III are classified as Gal and GalNAc-specific lectins, respectively, whereas ML-II has an intermediate specificity. In contrast, *Sambucus* species contain complex mixtures of Gal-, GalNAc-, and NeuAcα(2,6)Gal/GalNAc-specific type 2 RIP and related lectins. The simultaneous occurrence of type 2 RIP with different specificities indicates that even within a single species the type 2 RIP family evolved toward a diversification of sugar specificity. It should be emphasized, however, that the overall specificity range of the type 2 RIP family is still restricted to Gal, GalNAc, and NeuAcα(2,6)Gal/GalNAc.

E. Occurrence of Lectin-Related Proteins

Two abundant proteins of elderberry bark exhibit ribosome-inactivating activity and show the same molecular structure as monomeric type 2 RIP but are devoid of carbohydrate-binding activity. Molecular cloning revealed that the proteins are encoded by cDNAs which have a high sequence similarity with the genes encoding genuine type 2 RIP/lectins of elderberry. Accordingly, the proteins have been called *Sambucus nigra* lectin-related protein (SNLRP). Molecular modeling indicated that the carbohydrate-binding sites of the SNLRP are not functional because some of the critical amino acid residues have been substituted. Because SNLRP exhibits RIP activity and has the same molecular structure as other type 2 RIP, it belongs to the family of type 2 RIP. However, SNLRP is definitely not a lectin because it has no carbohydrate-binding activity (Van Damme et al., 1996d).

F. Three-Dimensional Structure

1. Three-Dimensional Structure of Ricin and Related Proteins

The three-dimensional structure of ricin was determined at 2.5 Å resolution by X-ray crystallographic analysis (Rutenber et al., 1991; Katzin et al., 1991; Rutenber and Robertus, 1991) (Figure 19A). The B chain consists of two similar domains called domain 1 and domain 2. Both domains consist of four subdomains designated 1λ, 1α, 1β, and 1γ for domain 1, and 2λ, 2α, 2β, and 2γ for domain 2. Subdomains α, β, and γ are homologous (Figure 17). They correspond to the amino acid sequences 17 to 59 (1α), 60 to 100 (1β), 103 to 135 (1γ) for the first domain, and 148 to 183 (2α), 187 to 226 (2β) and 229 to 262 (2γ) for the second domain. The B chain contains no extended regular secondary structures such as α-helices or β-sheets but consists mainly of coil structures linked by turns and loops (Figure 19A). Only a few short strands of β-sheet occur along the B chain. Four disulfide bridges between cysteine residues 20 to 39, 63 to 80, 151 to 164, and 190 to 207 stabilize the folding of the B chain. Each domain has a tertiary folding characteristic of the β-trefoil family (Murzin et al., 1992). The B chain

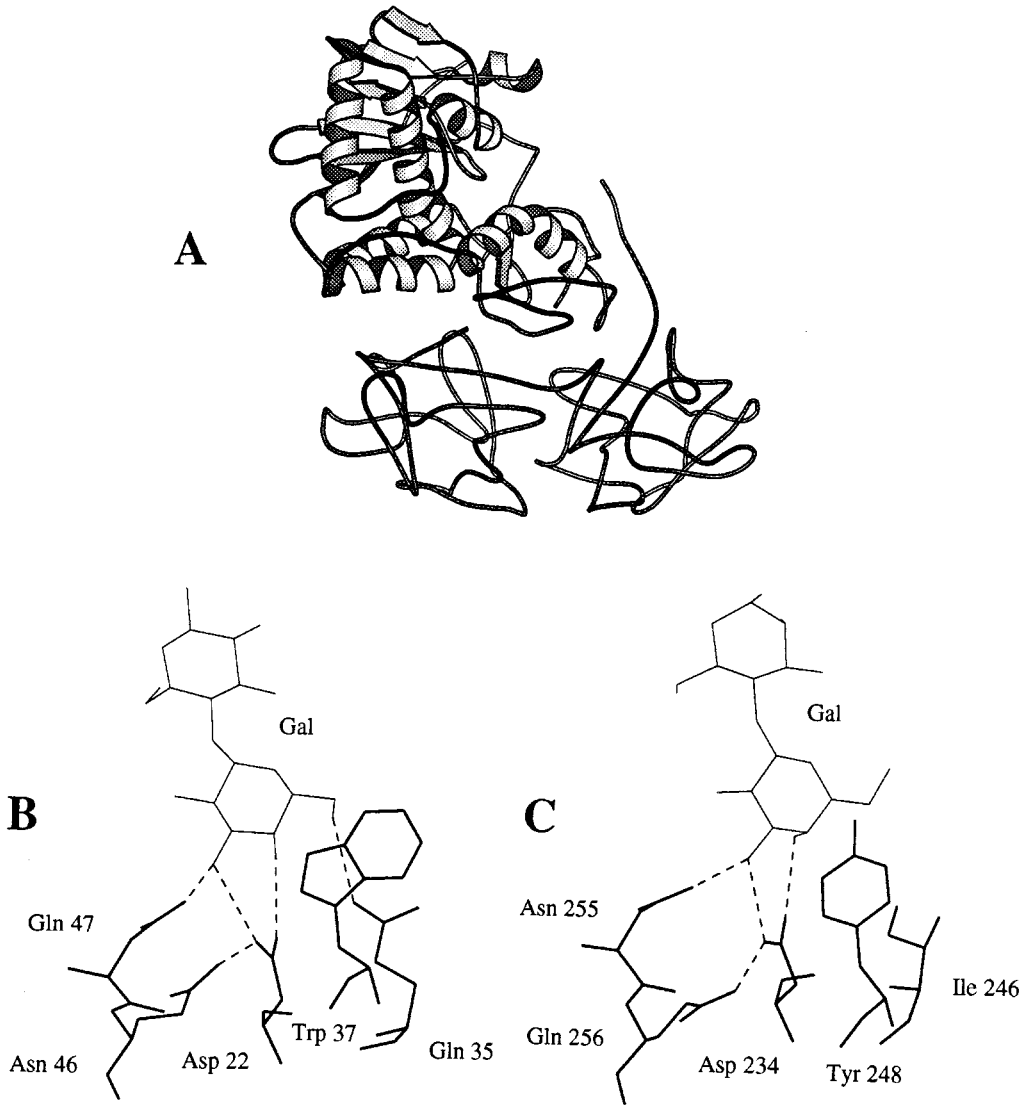

FIGURE 19. (A) Three-dimensional structure of ricin (code 2aai, Brookhaven PDB). The A chain (upper structure) contains eight α-helices and a left-handed twist of six strands of β-sheet, whereas the B chain (lower structure) is devoid of any regular secondary structure. A disulfide bridge covalently links the A and B chain. The B chain consists of two domains each of which possesses at least one saccharide-binding site. The cartoon was generated with Molscript (Kraulis, 1991). (B, C) Binding of the galactose residue of lactose to saccharide-binding sites I (B) and II (C), respectively, of the ricin B chain. The figures show the interaction between the sugar (thin line) with the amino acid residues forming the sites (thick lines) through a network of five (site I) or four (site II) hydrogen-bonds (dotted lines). Residues Trp^{37} of site I and Ile^{246} and Tyr^{248} of site II interact with galactose via hydrophobic contacts.

contains two *N*-glycosylation sites, Asn^{95}–Gly^{96}–Thr^{97} and Asn^{135}-Asn^{136}–Thr^{137}. Both sites are occupied by biantennary oligosaccharides, which extend into the solvent (Rutenber and Robertus, 1991).

The ricin A chain consists of three distinct domains and contains regular secondary structures, namely, eight α-helices numbered A-H, and six strands of β-sheet designated a to f that exhibit a left-handed

twist of about 110° when observed along the hydrogen bonds. As already mentioned, the toxicity of the A chain depends on its *N*-glycosidase activity toward an adenine residue occurring in an exposed loop of the 28S rRNA of the ribosomal large subunit (Endo et al., 1988). Although the active site of the A chain contains residues Tyr^{80}, Tyr^{123}, Glu^{177}, Arg^{180}, and Trp^{211} (Katzin et al., 1991; Kim and Robertus, 1992; Chaddock and Roberts, 1993) other residues located nearby this site, that is, Asn^{78}, Arg^{134}, Gln^{173}, Ala^{178}, Glu^{208}, and Asn^{209}, are probably also necessary to maintain its catalytic conformation. Most of these residues are conserved in all RIP sequenced to date. Two potential *N*-glycosylation sites, Asn^{10}–Phe^{11}–Thr^{12} and Asn^{236}–Gly^{237}–Ser^{238}, occur along the ricin A chain. At present it is not clear whether only one (Kimura et al., 1988) or both sites (Foxwell et al., 1985) are occupied by oligomannose type glycans.

X-ray crystallographic studies of abrin (*Abrus precatorius*) showed that this type 2 RIP exhibits the same overall three-dimensional conformation as ricin (Tahirov et al., 1995). Recently, molecular modeling of the deduced amino acid sequences of some type 2 RIP (SNA-I, SNA-I′, SNA-V) and RIP-related proteins (SNA-IV) from elderberry (*Sambucus nigra*) using the X-ray coordinates of ricin indicated that despite the differences in carbohydrate-binding specificity between the different proteins (Neu5Ac(α2–6)Gal/GalNAc versus Gal/GalNAc) and some structural discrepancies, all these proteins possess B chains with a three-dimensional conformation very similar to that of ricin and abrin.

2. Structural Basis for the Recognition of Simple or Complex Sugars by the B Chain of Chimerolectins

The ricin B chain possesses two carbohydrate-binding sites located at the *N*-terminal subdomain 1α and the *C*-terminal subdomain 2γ (Figure 17A). Each site comprises five critical amino acid residues, namely, Asp^{22}, Gln^{35}, Trp^{37}, Asn^{46}, Gln^{47} for site 1 and Asp^{234}, Ile^{246}, Tyr^{248}, Asn^{255}, and Gln^{256} for site 2. Although both sites bind the Gal residue of lactose *via* a network of four (domain 1) or three (domain 2) hydrogen bonds (Figure 19B,C), the site of domain 1 acts as a low-affinity site, whereas the site of domain 2 behaves as a high-affinity site for lactose (Yamasaki et al., 1985; Hatakeyama et al., 1986). The binding of lactose is completed by hydrophobic contacts occurring between the pyranose ring of Gal and the aromatic (Trp^{37}, Tyr^{248}) or hydrophobic (Ile^{246}) residues of the carbohydrate-binding sites. The exclusive binding of GalNAc to site 2 of the ricin B chain apparently depends on a rotation (of about 15°) of this sugar inside the site, which prevents a steric clash between the *N*-acetyl group of GalNAc and the side chain of the Asp^{250} residue located in the vicinity of the binding site (Rutenber and Robertus, 1991). Because there is a steric hindrance between the *N*-acetyl group and Asp^{44} of site 1 (homologous to Asp^{250} of site 2), the rotation of GalNAc cannot occur. Although ricin strongly interacts with oligosaccharides bearing terminal galactosides, no complex between ricin and biantennary glycans is yet available to corroborate the occurrence of an extended carbohydrate-binding site on the surface of the ricin B chain. Superposition studies using a biantennary glycan with two galactosides as terminal residues showed that both galactosides can bind to the carbohydrate-binding sites belonging to two different domains of two adjacent ricin molecules (Rutenber and Robertus, 1991). However, such an interaction is likely to be artificial because it depends on the packing of ricin molecules in the crystal. These carbohydrate-binding properties of the ricin B chain are of particular interest because they allow ricin to recognize complex sugars on the cell sur-

face, thus targeting the toxic A chain into the cells. Several recent results suggest the occurrence of a third carbohydrate-binding site in the ricin B chain (Fu et al., 1996; Frankel et al., 1996; Venkatesh and Lambert, 1997).

Docking experiments, in combination with hapten inhibition experiments with simple and complex sugars, revealed pronounced differences in specificity between the B chains of different type 2 RIP/lectins from elderberry (Shibuya et al., 1987; Kaku et al., 1990a). For example, SNA-I and SNA-I′ react with α(2,6)-sialylated Gal- or GalNAc-residues, whereas SNA-IV and SNA-V are readily inhibited by GalNAc and, to a lesser extent, by Gal. These differences in carbohydrate-binding activity between the different elderberry type 2 RIP/lectins apparently depend on specific amino acid substitutions in the respective carbohydrate-binding sites.

G. Molecular Evolution

All type 2 RIP sequences show a high sequence similarity in both their A and B chains. In addition, all cloned *Sambucus nigra* type 2 RIP-related hololectins exhibit a high sequence identity with the B chain of the genuine elderberry type 2 RIP. Because the type 2 RIP lectin family comprises both chimero- and hololectins, the phylogenetic tree is not for the complete protein sequences but for the sequences of the B chains only. The dendrogram shown in Figure 20 is based

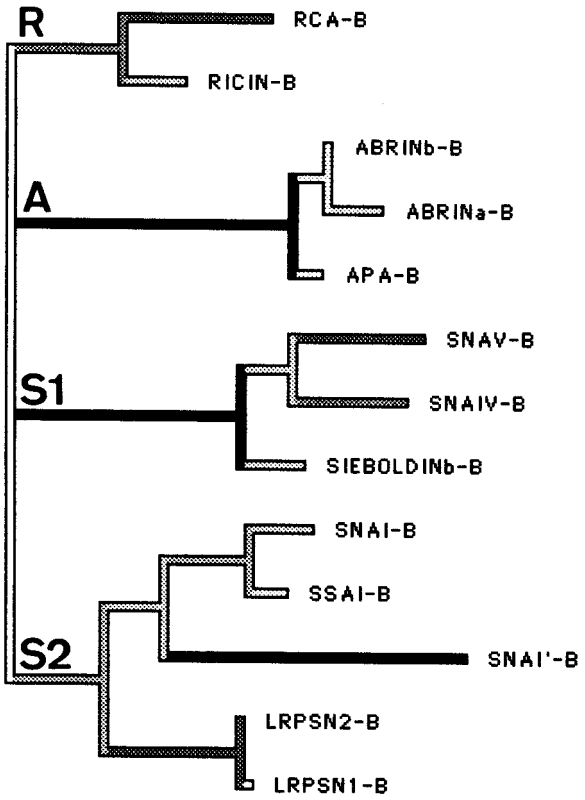

FIGURE 20. Phylogenetic tree of the B chains of type 2 RIP and related lectins. The dendrogram was made on a Macintosh LC 630 using the program MacClade (Maddison and Maddison, 1992). Abbreviations of the B chains are listed in Table 3.

on a distance matrix calculated from the amino acid sequences of 11 different carbohydrate-binding B chains and 2 inactive B chains.

The phylogenetic tree of the RIP sequences consists of four branches. Branches R and A comprise the B chains of *Ricinus communis* and *Abrus precatorius* type 2 RIP, respectively. Both branches further bifurcate into a toxin and an agglutinin side branch. The remainder of the dendrogram consists exclusively of the B chains of the *Sambucus* RIP/lectins, which apparently form two distinct branches. A closer examination indicates that branch S1 clusters the Gal/GalNAc-specific B chains from both *S. nigra* and *S. sieboldiana*, whereas the branch S2 comprises the Neu5Acα(2,6)Gal/GalNAc-specific lectins from both species as well as the inactive B chains of the *S. nigra* lectin-related proteins. The overall topology of the branches S1 and S2 reconstructs the molecular evolution of the *Sambucus* RIP/lectins. Most probably, the common ancestor of the B chains of all *Sambucus* RIP/lectins has been duplicated, giving rise to the direct ancestors of the Gal/GalNAc- and Neu5Acα(2,6)Gal/GalNAc-specific lectins. The very close relationship between SNA-V and sieboldin-B on the one hand, and SNA-I and SSA-I on the other hand, suggests that this duplication event has taken place before the different *Sambucus* species diverged from each other. The documented occurrence of three closely related Gal/GalNAc-binding lectins in *S. nigra* (i.e., SNA-III, SNA-IV, and SNA-V) implies that at least two other duplication events have taken place during the evolution of this lectin cluster. Because no sequences have been reported of the SNA-III, SNA-IV, and SNA-V homologues of *S. sieboldiana*, it is not possible to conclude whether the presumed duplications have taken place before or after the separation of the different *Sambucus* species. The topology of branch S2 indicates that the evolution of the cluster of the Neu5Acα(2,6)Gal/GalNAc-specific lectins also involved at least two gene duplications. A first duplication explains the divergence of the inactive B chains of the lectin-related protein LRPSN from the true lectins. The second duplication eventually resulted in the formation of two distinct Neu5Acα(2,6)Gal/GalNAc-specific type 2 RIP (i.e., SNA-I and SNA-I′). No sequences have been reported of the SNA-I′ and LRPSN homologues of *S. sieboldiana*. Therefore, it is impossible to predict when these duplications occurred. The bifurcation pattern of branch S2 indicates that the lectin-related proteins diverged from the Neu5Acα(2,6)Gal/GalNAc-specific type 2 RIP at an early evolutionary stage. At present it is not clear whether the lectin-related proteins fulfill a specific role (e.g., by virtue of their enzymatically active A chain) or act as bark-specific storage proteins. In the latter case one can conclude that elderberry used lectin genes to develop genes encoding storage proteins.

Sequence comparisons of all B chain sequences shown in the dendrogram (which are all of dicot origin) revealed a high degree of sequence similarity. In addition, partial sequencing of two type 2 RIP from *Polygonatum multiflorum* has shown that the B chains of these monocot lectins are also highly similar to their dicotyledonous homologues (Van Damme, unpublished results). Therefore, one can reasonably assume, that the B chains of all type 2 RIP and related lectins arose from a common ancestor that already existed before the monocots and dicots separated from each other during the evolution of higher plants. A closer analysis of the sequence of the B chains further revealed that they are composed of two similar domains, each consisting of three tandemly arrayed subdomains of 35 to 40 amino acid residues (Rutenber et al., 1987) (Figure 17A). At present, it is generally accepted that the B chain of the type 2 RIP and related hololectins

consist of six tandem repeats two (or possibly three) of which have a carbohydrate-binding site. The obvious similarities in the sequences and three-dimensional structures of the two domains indicate that the B chain arose from the duplication followed by an in tandem insertion of an ancestral three-subdomain lectin. This ancestral three-domain lectin itself resulted most probably from two consecutive duplications and in tandem insertions of a single ancestral subdomain. The residual sequence similarity between the three subdomains of, for example, SNA-I (Figure 17B) suggests that a first duplication and in tandem insertion of the ancestral subdomain [X] yielded a two-subdomain polypeptide [X′-X] that evolved further into a lectin [Y-X]. Duplication and in tandem insertion of the [Y] domain then gave rise to a three-subdomain polypeptide [Y-Y′-X] that evolved further into a lectin [Y-Z-X]. This presumed sequence of events explains why subdomains 1 and 2 are much closer related to each other than to subdomain 3. It is interesting to note here that the same model also explains the evolution of the three-subdomain subunits of the monocot mannose-binding lectins.

The origin of the ancestral subdomain of the B chain of type 2 RIP has not been traced yet. This ancient domain of about 40 amino acid residues possibly arose during early evolution of plants as a specific receptor/signaling molecule for microbial Gal/GalNAc-containing glycans. After more efficient signaling molecules had been developed, the small Gal/GalNAc-binding domain lost its original role. In a later stage of plant evolution, the Gal/GalNAc-binding domain was rescued to develop a defense mechanism against the novel threat presented by the newly evolved (predating) animals (Figure 17C). The development of this defense mechanism involved two major evolutionary events. First, amplification and in tandem insertion of the ancestral single domain yielded a carbohydrate-binding protein with a high affinity for glycoconjugates on the surface of the intestinal tract of animals. Second, fusion of this lectin to an *N*-glycosidase generated a chimeric toxin/lectin that further evolved into the modern type 2 RIP. As already mentioned above, the ancestor of the type 2 RIP probably existed before the angiosperms separated into the monocots and dicots. It is not clear whether type 2 RIP and related lectins are confined to some taxonomic groups or are widespread in higher plants. However, studies of *Sambucus nigra* demonstrated that this species is apparently specialized in the expression of type 2 RIP and related lectins. In addition, the same studies clearly indicated that type 2 RIP are still evolving and presented evidence for an event whereby the excision of a large part of the A domain of a genuine type 2 RIP gene results in a gene encoding a typical hololectin (Van Damme et al., 1997c).

H. Physiological Role

Because one can reasonably expect that the physiological role of the type 2 RIP/lectins is determined by the biological activities of the respective A and/or B chains, the enzymatic and nonenzymatic members of this lectin family probably fulfill different functions. Genuine type 2 RIP have both *N*-glycosidase and carbohydrate-binding activities. Due to the extreme cytotoxicity of the classic toxins ricin and abrin, type 2 RIP are believed to play a role in the plant's defense against plant-eating organisms. It should be emphasized, however, that the oral toxicity of the toxins is much lower than that of intraperitoneally or subcutaneously injected toxin (Ishiguro et al., 1983; Pusztai, 1991). However, taking into consideration the high concentration of ricin and abrin, these seed toxins can certainly offer some protection against mammalian predators (e.g., seed pre-

dating rodents). Moreover, it is possible that most animals in their natural habitat are much more sensitive to the toxins than laboratory animals. Finally, the oral toxicity of ricin may strongly increase in the presence of other castor bean seed proteins (e.g., the presumed nontoxic castor bean agglutinin). Insects react differently after feeding toxic type 2 RIP. For example, ricin was highly toxic to the coleoptera *Callosobruchus maculatus* and *Anthonomus grandis* but had no effect on the lepidoptera *Spodoptora litoralis* and *Heliothis virescens* (Gatehouse et al., 1990). The insensitivity of the lepidopteran species to ricin indicates that they can either inactivate the toxin or lack the appropriate glycan receptors for the B chain.

Most type 2 RIP are far less toxic after oral uptake than the classic toxins and are certainly not lethal to experimental animals. The biological significance of the relatively mild effects of these low-toxicity type 2 RIP is not fully understood yet, but is probably related to defense against herbivorous animals. According to recently proposed ideas, the presence of large quantities of moderately toxic proteins provokes unpleasant effects in the gut of predating animals, which cause a reaction of avoidance and protect the plant (or a population of a given species) from further damage (Peumans and Van Damme, 1995b). Such a defense mechanism can explain why type 2 RIP are particularly abundant in tissues that are most attractive to herbivorous animals (e.g., elderberry bark and evergreen mistletoe leaves and stems). The presumed protective effect against predating animals can also give an explanation for the simultaneous occurrence of two or more type 2 RIP with different sugar-binding specificities. For example, the presence of large quantities of a Neu5Acα(2,6)Gal/GalNAc-specific and a GalNAc-specific type 2 RIP in the bark of elderberry widens the range of potential glycan-receptor molecules in the gastrointestinal tract of predating animals and hence increases the chances of deleterious effects on the small intestine and/or other organs. Moreover, the two distinct type 2 RIP may act synergistically.

To date, little information has been reported on the possible toxic or antinutritive effects of low-toxicity type 2 RIP on insects. However, because the lectin from winter aconite (*Eranthis hyemalis*) tubers definitely has larvicidal activity against the southern corn rootworm (*Diabrotica undecimpunctata*) in an artificial feeding system (Kumar et al., 1993), the involvement of type 2 RIP in the plant's defense against insects cannot been excluded.

The observation that some type 2 RIP (e.g., ricin, abrin, modeccin, and the winter aconite lectin) exhibit antiviral activity *in vitro* against plant viruses raised the question of the possible involvement of these proteins in the plant's defense against viruses (Barbieri et al., 1993; Kumar et al., 1993). At present, the working mechanism of the antiviral activity of type 2 RIP is not known. It has been suggested that these lectins indirectly arrest viral multiplication *in vivo* because they selectively penetrate the virally infected cells and block viral replication through the inactivation of cellular ribosomes. A similar working mechanism has been proposed for the antiviral type 1 RIP, but recent experiments with transgenic plants expressing the *Phytolacca americana* antiviral protein (PAP) revealed a totally different mode of action. It has been demonstrated that the expression of PAP in transgenic tobacco root stocks induced virus resistance in a normally virus-sensitive untransformed graft, and that RNA *N*-glycosidase activity was a prerequisite for the induction process (Smirnov et al., 1997). The discovery of this PAP induced viral resistance in distant organs, which is apparently mediated by a transportable reaction product of the enzymatic activity of the type 1 RIP, raises the question whether type 2 RIP pos-

sibly act in a similar way. Although no answer can be given yet, this question actualizes the suggestion that type 2 RIP play a role in active defense systems based on a hypersensitivity reaction (Peumans and Van Damme, 1995b). According to the basic idea, type 2 RIP are a kind of suicidal proteins that temporarily become nonactive by their sequestration in the vacuole or the extracellular space. As soon as the cell is damaged or infected by microorganisms, the natural barrier between the cytoplasm and the vacuole or extracellular space is broken and the RIP get access to and inactivate their cytoplasmic ribosomes. As a result, protein synthesis is arrested and the cell dies. Such a scenario is reminiscent of the first part of the sequence of events leading to the hypersensitive response in plants to viral or microbial infection. If type 2 RIP effectively induce a hypersensitive response, their presumed (broad) protective action is indirect and involves an active defense system.

Type 2 RIP are not necessarily or exclusively defense-related proteins but may also transiently act as storage proteins. Several type 2 RIP from seeds and vegetative storage tissues are present in high concentrations and behave as typical storage proteins for the regulation of their expression. For example, the castor bean type 2 RIP behave as typical seed storage proteins (Lord et al., 1994). Similarly, the elderberry bark type 2 RIP are not only the most abundant bark proteins but are also developmentally regulated as typical vegetative storage proteins (Nsimba-Lubaki and Peumans, 1986). Most probably, the abundant type 2 RIP have a dual role. Under normal conditions they fulfill a transient storage function, but as soon as the plant is attacked by predating animals or phytophagous invertebrates they are used as nonspecific defense proteins.

Type 2 RIP-derived or -related hololectins have no enzymatic activity, which implies that their physiological role is determined exclusively by their carbohydrate-binding properties. Because their recent discovery, little information is available on the biological activities of the type 2 RIP-derived or -related hololectins. Orally administered SNA-II clearly had deleterious effects on the small intestine of rats (Pusztai et al., 1990). This observation not only demonstrates that the B chain on its own has toxic or antinutrient properties but also suggests that the B chain of type 2 RIP is capable of inducing specific reactions (other than the uptake of the intact RIP) in animal cells. Up to now, type 2 RIP-related/derived hololectins have been found exclusively in *Sambucus* species, where they occur together with genuine type 2 RIP. This apparent close relationship may be indicative for a synergism between enzymatic and nonenzymatic lectins.

Detailed studies of the elderberry type 2 RIP/lectins revealed that type 2 RIP-derived/related lectins are abundant proteins in bark, fruits, and seeds. Their high concentrations in bark and seeds can be reconciled with a storage role in these storage tissues. However, a transient storage function cannot be envisaged for the fruit lectin because the whole content of the fruits represents a net loss for the plant. Therefore, in the fruit a totally different role has been proposed. Elderberry depends for the dissemination of its seeds on fruit-eating birds (e.g., starlings, blackbirds). After ingestion of the fruits, the abundant GalNAc-specific hololectin causes a slight diarrhea, which speeds up the gut transit and prevents damage of the seeds (Van Damme et al., 1997c).

VII. MONOCOT MANNOSE-BINDING LECTINS

The term ‘monocot mannose-binding lectins’ refers to a superfamily of strictly mannose-specific lectins, which up to now

have been found exclusively in a subgroup of the monocotyledonous plants. All monocot mannose-binding lectins consist of subunits with a similar sequence and overall three-dimensional structure. Monocot mannose-binding lectins are structurally and evolutionary unrelated to the mannose/glucose-specific legume lectins and the mannose/maltose-specific Convolvulaceae lectins that are classified as jacalin-related lectins.

A. Historical Note

The monocot mannose-binding lectins are a relatively new family of plant lectins. Although the occurrence of lectin activity in garlic was already reported in 1980 (Nachbar and Oppenheim, 1980), the first description of a monocot mannose-binding lectin dates back to 1987 when a lectin with an exclusive specificity toward mannose was isolated from snowdrop bulbs (Van Damme et al., 1987). The importance of the discovery of the snowdrop lectin (called *Galanthus nivalis* agglutinin or GNA) was realized after detailed specificity studies revealed the unique carbohydrate-binding properties of this novel lectin (Shibuya et al., 1988), and molecular cloning demonstrated that GNA differed from all previously characterized plant lectins in its primary structure (Van Damme et al., 1991b). The isolation of closely related mannose-specific lectins from other Amaryllidaceae species (Van Damme et al., 1988) and from several Alliaceae species (Van Damme et al., 1991a; 1993a), and the subsequent molecular cloning of the corresponding genes provided the first unequivocal evidence for the occurrence of a family of GNA-related lectins in monocots. Once similar lectins had also been found in several other monocot families, the term 'monocot mannose-binding lectins' was introduced for this novel lectin family.

Despite their recent discovery, monocot mannose-binding lectins already received a lot of attention in different scientific disciplines. Initially, most of the interest was in biomedical research in an attempt to exploit the potent anti-(retro)viral activity of GNA and some other monocot mannose-binding lectins (Balzarini et al., 1991; 1992). Later, the attention shifted toward the exploitation of the anti-insect properties of GNA in plant protection (Gatehouse et al., 1995; Hilder et al., 1995).

B. Occurrence, Molecular Structure, and Amino Acid Sequences

Heretofore, monocot mannose-binding lectins have been found in six different monocot families, namely, Alliaceae, Amaryllidaceae, Araceae, Bromeliaceae, Liliaceae, and Orchidaceae (Table 4). Their occurrence is best documented for the Amaryllidaceae, Alliaceae, and Orchidaceae species (Van Damme et al., 1995e). Based on currently available data, it seems likely that most species of these three families contain lectins. Monocot mannose-binding lectins have also been isolated and characterized from several Araceae species (Van Damme et al., 1995d). Moreover, many other Araceae species contain agglutinins that have not yet been fully characterized but probably belong to the monocot mannose-binding lectins, indicating that these lectins are very common in the Araceae family. Within the (extended) Liliaceae family monocot mannose-binding lectins have been isolated from *Tulipa* sp., *Polygonatum multiflorum* (Solomon's seal), *Scilla campanulata* (blue bell), and *Aloe arborescens* (Koike et al., 1995; Van Damme et al., 1996a; 1996f; Van Damme, unpublished results). In addition, monocot mannose-binding lectins may also occur in closely related genera. Although these examples indicate these lectins occur in different unrelated genera of the Liliaceae, they are not ubiquitous in this family. At

TABLE 4
Monocot Mannose-Binding Lectins: Occurrence, Molecular Structure and Specificity

Species	Tissue	Structure[a]	Sequence available[b]
Merolectins			
Epipactis helleborine	Leaf	[P14]	Nu (EHMBP)
Listera ovata	Leaf	[P14]	Nu (LOMBP)
Hololectins			
Allium ascalonicum	Bulb	$[P12]_2$	Nu (AAA)
Allium cepa	Bulb	$[P12]_4$	Nu (ACA)
Allium porrum	Leaf	$[P13]_4$	Nu (APA)
Allium sativum	Bulb	I [P(11 + 12)]	Nu (ASAI)
	Bulb	II $[P12]_2$	Nu (ASAII)
	Leaf	$[P12]_2$	Nu (ASA-L)
	Root	$[P15]_2$	Nu (ASA-R)
Allium ursinum	Bulb	I [P11 + P12]	Nu (AUAG1, G2)
	Bulb	II $[P12]_2$	Nu (AUAG0)
	Leaf	$[P12]_2$	Nu (AUA-L)
	Root	Ir $[P11 + P12]_2$	
	Root	IIr $[P12]_2$	
Aloe arborescens	Leaf	$[P12]_4$	Pr (ALOE)
Arum maculatum	Tuber	$[P(12 + 12)]_2$	Nu (AMA)
Clivia miniata	Leaf	$[P13]_2$	Nu (CMA)
Colocasia esculenta	Tuber	$[P(12 + 12)]_2$	Nu (CEA)
Cymbidium hybrid	Leaf	$[P12]_2$	Nu (CHA)
Dieffenbachia sequina	Leaf	$[P(12 + 14)]_2$	
Epipactis helleborine	Leaf	$[P12]_2$	Nu (EPA)
Galanthus nivalis	Bulb	$[P12]_4$	Pr, Nu (GNA)
Hippeastrum hybrid	Bulb	$[P12]_4$	Nu (HHA)
Leucojum aestivum	Bulb	$[P12]_2$	
Leucojum vernum	Bulb	$[P12]_2$	
Listera ovata	Leaf	$[P12]_2$	Nu (LOA)
Lycoris aurea	Bulb	$[P12]_2$	
Narcissus pseudonarcissus	Bulb	$[P12]_2$, $[P12]_4$	Nu (NPA)
Neoregelia flandria	Leaf	$[P12]_2$	
Polygonatum multiflorum	Rhizome	$[P14]_4$	Nu (PMA)
Tulipa hybrid	Bulb	$[P12]_2$	Nu (TL-MII)
	Bulb	$[P(14 + 14)]_4$	Nu (TxLCI)
Xanthosoma sagittifolium	Tuber	$[P(12 + 12)]_4$	

[a] [PX] stands for protomer with a molecular mass of X kDa. [P(Y + Z)] indicates that the protomer is cleaved in two polypeptides of Y and Z kDa.

[b] Pr, protein sequence; Nu, nucleotide sequence. The abbreviation in brackets refers to the sequence name used in the dendrogram (Figure 24).

present, the lectin from *Neoregelia flandria* is the only example of a monocot mannose-binding lectin in a species from the family Bromeliaceae (Yagi et al., 1996).

Monocot mannose-binding lectins have been found in most vegetative tissues such as leaves, flowers, ovaries, bulbs, tubers, rhizomes, roots (Van Damme et al., 1995e), and even in nectar (Peumans et al., 1997a), but so far not in seeds except in *Clivia miniata*. It should be mentioned, however, that ripe seeds of this ornamental plant con-

tain miniature plants with fully differentiated green tissues.

In most species analyzed thus far, monocot mannose-binding lectins occur in almost all vegetative tissues. For instance, snowdrop and daffodil express very similar lectins in bulbs, roots, leaves, flower stalks, ovaries, and flowers. However, lectins from different tissues may differ in their isolectin composition. Furthermore, the lectin concentration in different tissues of snowdrop and daffodil is developmentally regulated (Van Damme and Peumans, 1990a). Garlic (*Allium sativum*) is the only documented example of a plant that accumulates different monocot mannose-binding lectins in different vegetative tissues. Thus, garlic contains two different bulb-specific lectins (Van Damme et al., 1992b), as well as one leaf-specific and one root-specific lectin (Smeets et al., 1997c).

A number of plants have been described that express two different monocot mannose-binding lectins in the same tissue(s). The simultaneous occurrence of two different hololectins is well documented for the bulbs of garlic and ramsons (Van Damme et al., 1992b, 1993b). A mixture of a hololectin and a merolectin has been found in green tissues from the orchids *Listera ovata* and *Epipactis helleborine* (Van Damme et al., 1994a,b). Tulip bulbs in addition to a hololectin also contain a superlectin (Van Damme et al., 1996f).

The relative abundance of the monocot mannose-binding lectins varies according to the species, the tissue, and the developmental stage. In general, the monocot mannose-binding lectins account for 0.1 to 5% of the total protein. However, some monocot mannose-binding lectins are highly abundant proteins representing up to 50% of the total protein (e.g., the bulb lectins of garlic and ramsons), whereas others are present in minute quantities and represent less than 0.01% of the total protein (e.g., in the leaves of some orchids).

1. Molecular Structure of the Native Lectins

The superfamily of monocot mannose-binding lectins is quite heterogeneous in the molecular structure of the native proteins (Figure 21). According to the size of the protomers, these lectins can be divided into one-domain protomers of 11 to 14 kDa and two-domain protomers of about 30 kDa. Lectins composed of 30-kDa protomers are subdivided further in a subgroup with protomers consisting of two identical or highly similar domains and a subgroup with protomers consisting of two dissimilar domains.

a. Monocot Mannose-Binding Lectins Composed of One-Domain Protomers

Most of the currently known monocot mannose-binding lectins are built up of one-domain protomers. This group comprises besides a few merolectins mainly hololectins.

The orchids *Listera ovata* and *Epipactis helleborine* contain mannose-binding proteins consisting of a single one-domain protomer of about 14 kDa (Van Damme et al., 1994a). Both proteins definitely bind mannose but have no agglutinating activity, and hence are considered merolectins.

The majority of monocot mannose-binding lectins consists of two or four identical one-domain protomers that are held together by noncovalent interactions. All Amaryllidaceae and Bromeliaceae lectins, most of the Alliaceae and Orchidaceae lectins, and some Liliaceae lectins are homodimers or homotetramers. It should be mentioned here that the daffodil lectin behaves as a trimer after analytical ultracentrifugation. Besides the homodimeric and homotetrameric monocot mannose-binding lectins, there is also one example of a heterodimeric lectin consisting of two slightly different one-domain

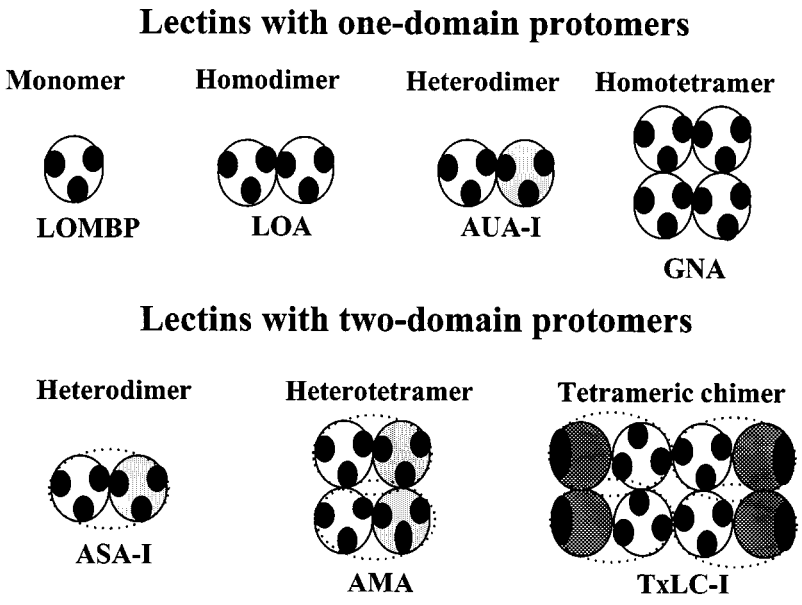

FIGURE 21. Schematic representation of the molecular structure of the native monocot mannose-binding lectins. Examples shown are: the *Listera ovata* monomeric mannose-binding protein (LOMBP), *Listera ovata* agglutinin (LOA), *Galanthus nivalis* agglutinin (GNA), *Allium ursinum* agglutinin I (AUA-I), *Allium sativum* agglutinin I (ASA-I), *Arum maculatum* agglutinin (AMA), and *Tulipa* lectin C-I (TxLC-I). Domains belonging to a single protomer are circled by a dotted line.

protomers in the bulbs of ramsons (Van Damme et al., 1993b; 1995e).

b. Monocot Mannose-Binding Lectins Composed of Two-Domain Protomers

Several monocot mannose-binding lectins are built up of two-domain protomers. The final structure of the mature lectins is not only determined by the number of protomers but also by the posttranslational processing of the protomers. Depending on the degree of sequence similarity between the two domains, the resulting lectins behave as hololectins or as superlectins.

Presently, three different types of monocot mannose-binding lectins composed of two-domain protomers have been described in detail. A first example is the so-called *Allium sativum* bulb lectin I (ASA-I) that behaves as a heterodimer composed of two different polypeptides of 11.5 and 12.5 kDa (Van Damme et al., 1992b). Molecular cloning of the gene encoding ASA-I revealed that both polypeptides are derived from a single precursor consisting of two highly similar tandemly arrayed lectin domains. Hence, ASA-I must be considered a lectin that is composed of a single two-domain protomer. As both domains are highly homologous (77% sequence identity, 82% sequence similarity) and exhibit specificity toward mannose, ASA-I behaves as a hololectin.

A second example are the Araceae lectins. According to biochemical analyses, Araceae lectins are heterotetramers consisting of two different polypeptides of either identical or slightly different size. Molecular cloning of the genes encoding the lectins from *Arum maculatum* (Van Damme et al., 1995d) and *Colocasia esculenta* (Hirai et al., 1993) unambiguously demonstrated that the two polypeptides of these lectins are derived from two-domain precursors consisting of

two tandemly arrayed lectin domains. Hence, the Araceae lectins are dimers of two identical two-domain protomers. Because the two domains of the lectin protomers clearly show sequence similarity (although less than in the case of ASA-I, namely, 41% sequence identity and 50% sequence similarity) and both exhibit mannose-binding activity, the Aracaeae lectins are considered hololectins.

A third type of a monocot mannose-binding lectin composed of two-domain protomers has been isolated from tulips (Van Damme et al., 1996f). Tulip bulbs in addition to a dimeric monocot mannose-binding lectin composed of one-domain protomers (called TxLM-II), contain a second mannose-binding lectin (called TxLC-I) with an apparently complex molecular structure and specificity. TxLC-I is built up of four identical subunits of 28-kDa, which are partly cleaved into two different polypeptides of 14 kDa. Molecular cloning revealed that the 28 kDa polypeptide consists of an *N*-terminal domain similar to the mannose-binding domain encoding the 12-kDa monocot mannose-binding lectins linked to an unrelated *C*-terminal domain that harbors a GalNAc-binding site. Because the two domains of the protomers are dissimilar and in addition recognize structurally different sugars (*in casu* mannose and GalNAc), TxLC-I is considered a superlectin. Native TxLC-I clearly consists of four two-domain protomers.

2. Primary Structure

Although only a few monocot mannose-binding lectins have been completely sequenced by chemical methods, detailed sequence information is available for many of these lectins because their genes have been cloned (Table 4). Sequence comparisons indicate that all monocot mannose-binding lectins share a high degree of sequence similarity. In addition, a close examination of their primary structure reveals that the lectin domain consists of three tandemly arrayed subdomains of about 40 residues with a high internal sequence similarity (Van Damme et al., 1991b). As discussed below, each subdomain corresponds to a mannose-binding site.

C. Structure, Biosynthesis, and Posttranslational Modifications

As GNA was discovered very recently, molecular cloning and analysis techniques could be used from the very beginning of the research on monocot mannose-binding lectins. cDNA cloning yielded a wealth of information about the primary structure of many of these lectins and made a critical contribution to the elucidation of their molecular structure. In addition, the availability of numerous deduced sequences helped to unravel the conversion of the primary translation products of the lectin mRNAs into the mature proteins (Figure 22).

Although the biosynthesis, processing, and topogenesis of only a few monocot mannose-binding lectins have been studied in detail, the sequence of events leading from the primary translation product(s) to the mature lectin polypeptides can be predicted for all monocot mannose-binding lectins. Basically, two different types of synthesis and processing are distinguished depending on whether the lectins consist of one- or two-domain protomers.

Like GNA, all one-domain protomers are synthesized on the ER as preproproteins (Van Damme and Peumans, 1988). After cotranslational removal of the signal peptide the proproteins undergo a proteolytic cleavage of the *C*-terminal peptide. This posttranslational processing probably occurs during transport from the ER to their final destination (which is still unknown but most probably is the vacuole or a vacuole-derived

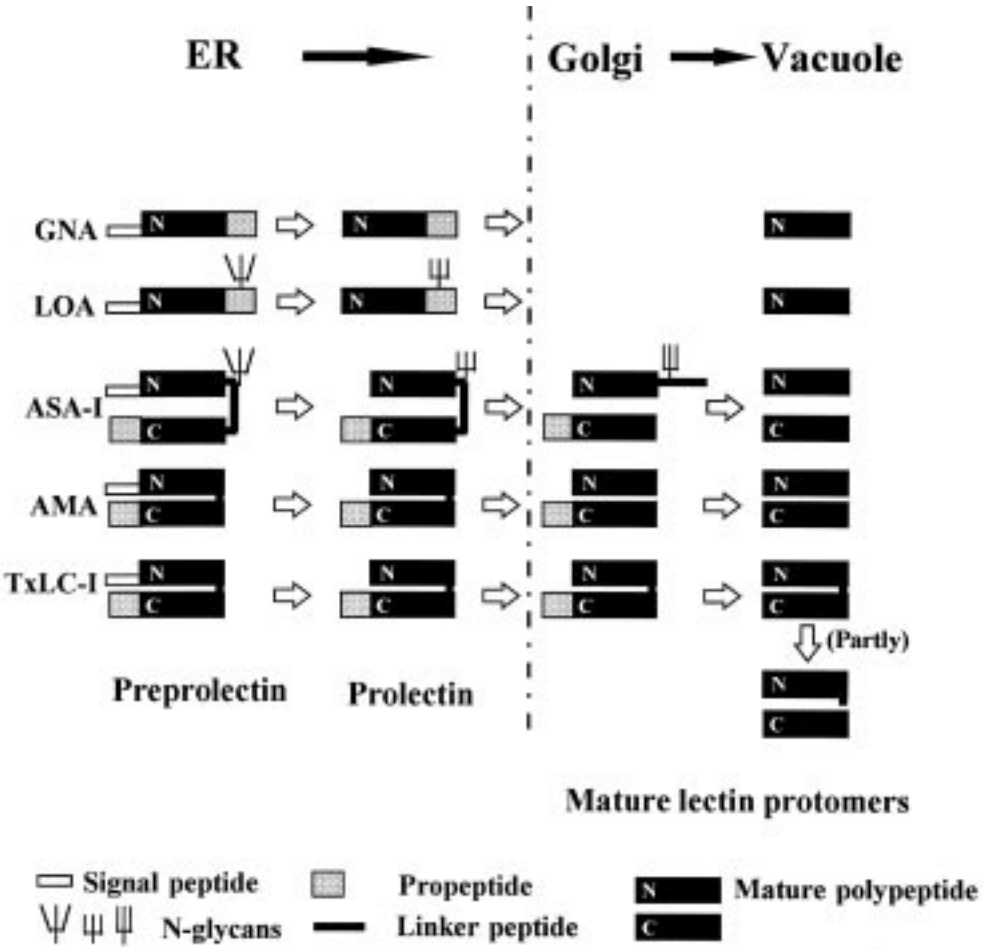

FIGURE 22. Schematic representation of the biosynthesis, co- and posttranslational modifications, and topogenesis of the monocot mannose-binding lectins. Examples shown are *Galanthus nivalis* agglutinin (GNA), *Listera ovata* agglutinin (LOA), *Allium sativum* agglutinin I (ASA-I), *Arum maculatum* agglutinin (AMA), and *Tulipa* lectin C-I (TxLC-I). Inserted 'N' and 'C' indicate the *N*- and *C*-terminus of the protomer, respectively.

organelle). For some monocot mannose-binding lectins, the proproteins are co-translationally glycosylated in the *C*-terminal peptide (e.g., orchid lectins). Two-domain protomers are also synthesized on the ER as preproproteins and are in some cases *N*-glycosylated (e.g., the garlic lectin ASA-I). After co-translational processing of the signal peptide, the resulting proproteins are subsequently transported from the ER into (most probably) the vacuolar compartment. During or after transport, the proprotein undergoes further proteolytic modifications. Most of the information about the final processing of the proproteins comes from *in vivo* biosynthesis studies with the garlic bulb lectin ASA-I (Smeets et al., 1994) (Figure 22). The precursor proASA-I consists of two lectin domains connected through a glycosylated linker sequence. The preprotein is first cleaved between the end of the linker sequence and the *N*-terminus of the second lectin domain, resulting in an *N*-terminal domain carrying a glycosylated linker at its *C*-terminus and a *C*-terminal domain with the original *C*-terminal propeptide. Both fragments undergo a final processing by the proteolytic removal of the glycosylated linker

and the *C*-terminal propeptide, respectively. It is not known whether the linker sequence is deglycosylated before cleavage. The biosynthesis and processing of protomers of the Araceae lectins (Van Damme et al., 1995d) and the tulip lectin TxLC-I (Van Damme et al., 1996f) is less complex than the processing of ASA-I because in *Arum* and tulip there is no evidence for the occurrence of a large linker sequence between the two lectin domains of the precursor. Furthermore, the lectin precursor possesses no glycosylation site. Cleavage of the proprotein between the two lectin domains results in two different mature lectin polypeptides. In the case of TxLC-I, the cleavage between the first and the second lectin domain is only partial. As a result, most of the mature lectin consists of protomers in which the two domains are still held together.

Molecular analysis of the genes encoding monocot mannose-binding lectins also yielded interesting information about the occurrence of lectin gene families. Sequencing of multiple cDNA clones and Southern blot analysis clearly indicated that all Amaryllidaceae lectins are encoded by extended gene families (Van Damme et al., 1992a). Estimates varying between 20 to 100 genes are probably realistic because they are in good agreement with the molecular data and the complex isolectin composition (up to 100 isoforms) of typical Amaryllidaceae lectins, like the snowdrop and the daffodil lectin (Van Damme and Peumans, 1990b). It should also be emphasized that all (sequenced) members of the lectin gene family of the Amaryllidaceae have a high sequence similarity in their respective coding regions and encode highly similar lectin polypeptides. Similar to the Amaryllidaceae, Alliaceae, Orchidaceae, Araceae, and Liliaceae species also express multiple lectin genes encoding monocot mannose-binding lectins. The situation is very complicated in the *Allium* species that contain two or more monocot mannose-binding lectins. Garlic, for instance, expresses at least four different sets of lectin genes. Two sets encode the bulb-specific lectins ASA-I and ASA-II (Van Damme et al., 1992b). A third set controls the expression of the leaf- and root-specific lectins and the fourth encodes a yet unidentified root-specific lectin-related protein (Smeets et al., 1997a). Similarly, ramsons express a complex family of lectin genes. Two subsets encode the bulb lectins AUA-I and AUA-II (Van Damme et al., 1993b), whereas the third subset controls the expression of a leaf-specific lectin (Smeets et al., 1997a).

Orchidaceae species express lectin gene families that encode the dimeric lectins and monomeric mannose-binding proteins (Van Damme et al., 1994a,b). Molecular cloning has shown that all Liliaceae species studied thus far express at least two different sets of lectin genes. Both tulip and *Scilla campanulata* contain two different types of monocot mannose-binding lectins that are clearly encoded by two distinct genes (Van Damme et al., 1996f; Van Damme, unpublished results). Similarly, two sets of genes control the expression of a lectin and a lectin-related protein in the rhizomes of *Polygonatum multiflorum* (Van Damme et al., 1996a).

No intron sequences have been found in the genes encoding the monocot mannose-binding lectins.

D. Carbohydrate-Binding Specificity

Hapten-inhibition assays with simple sugars have demonstrated that all currently known monocot mannose-binding lectins (except the superlectin TxLC-I from tulip and the two-domain lectin from *Scilla campanulata*) are exclusively inhibited by mannose (Van Damme et al., 1995e). In general, the mannose concentrations required

for an efficient inhibition are high (IC_{50} = 20 to 200 m*M*), suggesting that these lectins have a low affinity for the monosaccharide. Oligomannosides are more potent inhibitors of the monocot mannose-binding lectins, indicating that the carbohydrate-binding sites accommodate preferentially oligomannosyl residues. Further, detailed specificity studies revealed that the monocot mannose-binding lectins definitely differ from each other in fine specificity. For example, GNA has the highest affinity for terminal Manα(1,3)Man (Shibuya et al., 1988), whereas the daffodil lectin and *Listera ovata* agglutinin preferentially bind Manα(1,6)Man (Kaku et al., 1990b) and Manα(1,3)Manα(1,3)Man, respectively (Saito et al., 1993).

The superlectin TxLC-I from tulip bulbs is not inhibited by any simple sugar. However, a combination of mannose and GalNAc completely inhibits the agglutination activity of the lectin (Cammue et al., 1986; Van Damme et al., 1996f). Molecular cloning of the lectin revealed that the unusual behavior of TxLC-I is due to the presence of a mannose-binding and a GalNAc-binding domain in the protomers of this lectin (Van Damme et al., 1996f).

E. Occurrence of Lectin-Related Proteins

During the last few years evidence has been presented for the occurrence of proteins that are evolutionary related to the monocot mannose-binding lectins but are devoid of carbohydrate-binding activity. Fruits of *Curculigo latifolia* (Hypoxidaceae) contain a sweet protein called 'curculin' that consists of a single polypeptide with a high sequence similarity to the mannose-binding domain of the monocot mannose-binding lectins (Yamashita et al., 1990). Curculin has no carbohydrate-binding activity. According to the results of molecular modeling experiments, all three mannose-binding sites of curculin are nonfunctional because steric hindrances prevent the binding of the sugar (Barre et al., 1997). Because curculin has no carbohydrate-binding activity, it is not considered a lectin but a 'lectin-related protein'. The occurrence of a lectin-related protein has also been documented in *Polygonatum multiflorum* (Van Damme et al., 1996a) and *Allium sativum* (Smeets et al., 1997c). Screening of a cDNA library constructed with mRNA from rhizomes of Solomon's seal yielded in addition to clones encoding the mannose-binding lectin clones encoding a protein consisting of two tandemly arranged domains with a high sequence similarity to the mannose-binding lectin. Although the protein has not yet been isolated, molecular modeling clearly demonstrated that all the putative mannose-binding sites of both domains are nonfunctional because steric hindrances prevent the binding of the sugar (Van Damme et al., 1996a). Similarly, screening of a cDNA library from garlic roots yielded a cDNA clone encoding a putative protein consisting of two in tandem domains with a high sequence similarity to the lectin. Although this putative protein has not been purified, molecular modeling of the sequence leaves no doubt that both domains have no carbohydrate-binding activity (Smeets et al., 1997c).

F. Three-Dimensional Structure

1. Overall Three-Dimensional Structure

The *Galanthus nivalis* agglutinin (GNA) has been crystallized (Wright et al., 1990) and analyzed by X-ray diffraction at 2.3 Å resolution (Hester et al., 1995). GNA is a homotetramer composed of four identical noncovalently bound monomers of 109 residues (12 kDa). The GNA protomer consists

of three tandemly arrayed subdomains (I, II, and III) each of which corresponds to a four-stranded β-sheet. Completion of the four-stranded bundle forming subdomain I involves the folding of the extended *C*-terminal end of the adjacent polypeptide chain, which contains a strand of β-sheet. The three sequential subdomains are connected by loops and form a 12-stranded β-barrel exhibiting three mannose-binding sites located in the clefts formed by the three bundles of β-sheet (Plate 1). Two cysteine residues Cys^{29} and Cys^{52} are linked through a disulfide bridge. In the native lectin, four monomers (A, B, C, and D) are noncovalently associated into the GNA homotetramer, which resembles a flattened crown with a central 16-Å-wide solvent channel (Figure 6B). Two monomers (i.e., A and D and B and C) associate into tight dimers through hydrogen-bond contacts stabilized by *C*-terminal strand exchange. These A-D and B-C dimers further associate into tetramers mainly through hydrophobic interactions. Because each monomer possesses three mannose-binding sites, the GNA tetramer contains 12 apparently functional mannose-binding sites. Evidently, this structural organization is quite different from those found in other plant lectins.

Many other monocot lectins belonging to the families Alliaceae, Araceae, Orchidaceae, and Liliaceae have been modeled using the X-ray coordinates of GNA (Barre et al., 1996). Despite some discrepancies that consist of the deletion or insertion of a few residues along the amino acid sequences, all the modeled lectins exhibit an overall three-dimensional structure very similar to that of GNA. Interestingly, docking experiments with D-mannose showed that in addition to the fully reactive lectins with three functional mannose-binding sites (e.g., GNA), there are also weakly reactive lectins with only two (e.g., *Polygonatum multiflorum* agglutinin) or one (e.g., *Arum maculatum* agglutinin) functional site(s) and even an apparently inactive lectin with no mannose-binding sites (e.g., *Allium ursinum* lectin II). The mannose-binding activity of these lectins is destroyed by the replacement of some crucial amino acids in the carbohydrate-binding site by more extended or hydrophobic residues. The former create steric clashes that prevent the binding of D-mannose, whereas the latter cannot participate in the network of hydrogen bonds, allowing the binding of D-mannose into the site. It was also shown that there is a very good correlation between the number of active binding sites and the sugar-binding activity of this group of lectins.

2. Structural Bases for the Recognition of Simple or Complex Sugars

The three mannose-binding sites of GNA are made up of identical amino acid residues, namely, Gln, Asp, Asn, and Tyr (Gln^{89}, Asp^{91}, Asn^{93}, and Tyr^{97} for subdomain I; Gln^{57}, Asp^{59}, Asn^{61}, and Tyr^{65} for subdomain II; Gln^{26}, Asp^{28}, Asn^{30}, and Tyr^{34} for subdomain III) that bind O2 (Asp, Asn), O3 (Gln), and O4 (Tyr) of mannose through a network of four hydrogen bonds (Figure 23A). A hydrophobic residue, Val^{95} (subdomain I), Val^{63} (subdomain II) and Val^{32} (subdomain III), interacts with C3 and C4 of mannose through hydrophobic interactions. The interaction of Asp and Asn residues with O2, which has an axial position in D-mannose, explains the specificity of GNA toward D-mannose. The equatorial position of O2 in D-glucose prevents the binding of this sugar to GNA. Although more specific, the binding of mannose by GNA is weaker than that observed in legume lectins because of the smaller number of residues forming the binding site (four residues instead of seven) and the binding to only three oxygen residues (O2, O3, and O4) instead of four (O3, O4, O5, and O6) as in legume lectins.

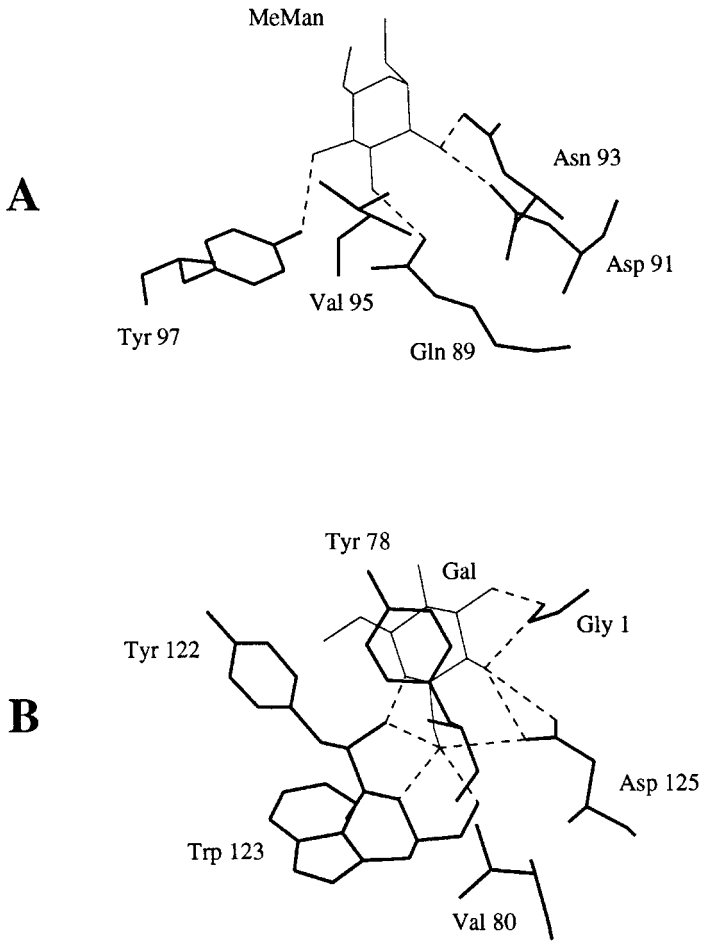

FIGURE 23. (A) Structure of the mannose-binding site of GNA showing the network of four hydrogen bonds (dotted lines) connecting a methyl-α-mannoside residue (thin line) with the amino acid residues forming the monosaccharide-binding site (thick lines). Val^{95} interacts with mannose via hydrophobic contacts. (B) Structure of the monosaccharide-binding site of jacalin showing the interaction of a galactose residue (thin line) with the amino acid residues of the binding site (thick lines) via a network of nine hydrogen bonds (dotted lines). Tyr^{78} and Val^{80} interact with galactose through hydrophobic contacts.

The recognition of mannose-containing glycoproteins by GNA and other related monocot lectins, which accounts for their inhibitory activity on the *in vitro* replication of retroviruses (Balzarini et al., 1991, 1992), suggested the occurrence of an extended carbohydrate-binding site. Analysis of GNA complexed with di- (Hester and Wright, 1996), tri-, and pentamannosides (Wright and Hester, 1996) not only confirmed the existence of an extended binding site but also revealed two types of interactions with more complex oligomannosides. The first extended site, which involves the binding sites of two adjacent subdomains III and the surrounding regions belonging to two different dimers (A and B or C and D), lets GNA to interact with the two D-mannose units of α(1,3)-linked oligomannosides. The second extended site is formed at the interface between two GNA tetramers and comprises the mannose-binding site of subdomain III of monomer (B) of a first GNA tetramer and the mannose-binding site II of monomer (A) of another GNA tetramer. Most probably the second extended site is not functional in solution as it depends

on the crystal packing of the GNA tetramers in the crystals.

3. Three-Dimensional Structure of Related Proteins

Curculin, the sweet-tasting and taste-modifying protein from *Curculigo latifolia* (Yamashita et al., 1990), shares a high sequence identity (41%) and similarity (65%) with GNA (Hester et al., 1995). Molecular modeling using the X-ray coordinates of GNA showed that the three-dimensional structure of the monomeric curculin closely resembles that of the protomers of the monocot mannose-binding lectins. However, docking experiments with D-mannose indicated that none of the three mannose-binding sites is functional in curculin because the substitution of some critical amino acid residues results in steric hindrances that prevent the binding of mannose (Barre et al., 1997).

The monocot mannose-binding lectins also share sequence similarity with an actin-binding and vesicle-associated protein called comitin. Comitin has first been identified in *Dictyostelium discoideum* (Stratford and Brown, 1985) but also occurs in mammalian cells (Weiner et al., 1993). Native *Dictyostelium discoideum* comitin is a dimeric protein composed of two identical subunits consisting of a 139-residue *N*-terminal mannose-binding core domain and a 41 *C*-terminal domain with a characteristic GYP(P)Q repeat. The core domain of comitin shares a high sequence identity (34%) and similarity (57%) with GNA and other related monocot mannose-binding lectins (Jung et al., 1996). Despite two extended extra sequences corresponding to loops inserted between the bundles of strands of antiparallel β-sheet, the three-dimensional model of comitin is superimposable on that of GNA and other monocot mannose-binding lectins. According to the results of the modeling experiments the mannose-binding sites of subdomains II and III remain unchanged (when compared with GNA), whereas three amino acids have been replaced in the binding site of subdomain I. Docking experiments further confirmed that the binding sites of subdomains II and III are fully active, whereas that of subdomain I is not functional.

G. Molecular Evolution

The monocot mannose-binding lectins are a homogeneous family of closely related proteins that apparently are confined to a taxonomic group spanning the families Amaryllidaceae, Alliaceae, Araceae, Bromeliaceae, and Orchidaceae. As most of the monocot mannose-binding proteins have been cloned and some sequenced, the evolutionary relationships within this lectin family can be traced fairly well. Due to the occurrence of several two-domain lectins, it is impossible to construct a phylogenetic tree using the complete sequences of the lectin precursors. Therefore, the dendrogram shown in Figure 24 is based on a distance matrix calculated from the (deduced) amino acid sequences of individual lectin domains. In addition to 30 lectin domains, the dendrogram also includes five domains of lectin-related proteins devoid of carbohydrate-binding activity.

The dendrogram of the monocot–mannose-binding lectins is composed of two major branches. Branch A consists of two side branches, Ao and Aa, which cluster the Orchidaceae and Amaryllidaceae lectins, respectively. According to the topology of the A branch, the Amaryllidaceae lectins are more closely related to each other than the Orchidaceae lectins. It should be noted that the monomeric mannose-binding proteins from *Epipactis helleborine* and *Listera ovata* form a separate subcluster. Branch L exhib-

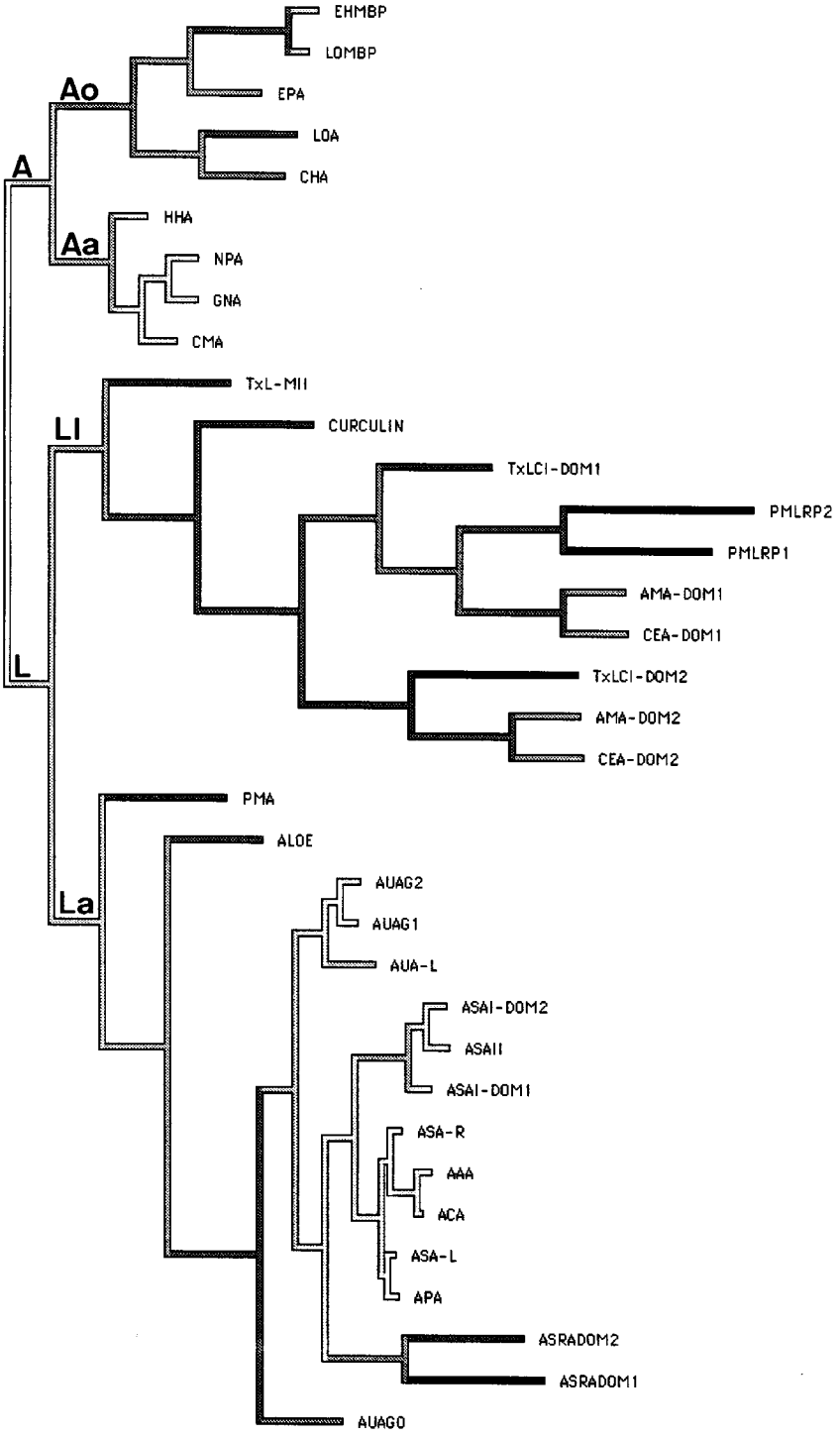

FIGURE 24. Phylogenetic tree of individual domains of the monocot mannose-binding lectins and related proteins. The dendrogram was made on a Macintosh LC 630 using the program MacClade (Maddison and Maddison, 1992). Abbreviations of the lectins are given in Table 4. The dendrogram also contains the sequences of the lectin-related proteins from *Curculigo latifolia* (curculin), *Polygonatum multiflorum* (PMLRP1 and PMLRP2), and *Allium sativum* (ASRADOM1 and ASRADOM2). Dom1 and Dom2 refer to the *N*- and *C*-terminal domains of the two-domain protomers.

its a more complex topology than branch A. The first side branch Ll comprises a cluster of lectins and lectin-related proteins from Liliaceae and Araceae species. As shown in Figure 24, the first two bifurcations of the Ll side branch separate the one-domain tulip lectin TL-MII and curculin, respectively, from a cluster of sequences encoding typical two-domain lectins. The topology of this cluster suggests that TxLC-I and the Araceae lectins AMA and CEA have a common two-domain ancestor because domains 1 and 2 of the respective lectins form two separate subclusters. Most probably, this presumed two-domain ancestor arose from the self-duplication and tandem insertion of an ancestral single-domain lectin. The close similarity between TxLC-I and the Araceae lectins suggests that this duplication event has taken place before the Liliaceae and Araceae families diverged from each other. It is difficult, however, to explain why lectins from the not so closely related Araceae and Liliaceae species belong to the same cluster. The two-domain lectin-related protein from *Polygonatum multiflorum* PMLRP apparently arose from an (in evolutionary terms) recent self-duplication and tandem insertion of a domain similar to the *N*-terminal domain of TxLC-I and the Araceae lectins AMA and CEA. Because *Polygonatum multiflorum* contains also a one-domain lectin PMA that clusters in the side branch La of the dendrogram (see below), it is possible that two sets of lectin genes evolved independently in this species.

Side branch La also has a complex topology. Two subsequent bifurcations separate the *Polygonatum multiflorum* and *Aloe arborescens* lectins from the extended group of *Allium* lectins that form a complex dendrogram on their own with several consecutive bifurcations. The first two bifurcations separate the ramsons lectins AUAG0 and [AUAG1, AUAG2, AUA-L], respectively, from the main line. In the next bifurcation, the two domains of the putative root lectin-related protein ASRA segregate from the rest of the *Allium* lectins. Because ASRADOM1 and ASRADOM2 form a separate cluster and have little similarity to the other garlic lectins, ASRA most probably arose by the duplication and tandem insertion of the direct ancestor domain of ASRADOM1 and ASRADOM2. A fourth bifurcation of the *Allium* branch separates the bulb-specific garlic lectins (ASA-I and ASA-II) from the so-called *Allium* leaf lectins. The bulb-specific garlic lectins follow a distinct evolutionary pathway. After the separation of the ancestral bulb lectin domain from the main line, a self-duplication and tandem insertion occurred, giving rise to the two-domain lectin ASA-I. In a subsequent independent evolutionary event, the second domain of ASA-I was duplicated and inserted elsewhere whereby a one-domain lectin gene ASAII was formed. The evolution of the leaf-specific *Allium* lectins did not involve duplication events. A common ancestor gave rise to four closely related lectins that according to the dendrogram form two subsets comprising the lectins from shallot and onion, and leek and garlic, respectively. The evolutionary tree of the *Allium* lectins is indicative of the physiological role of the different members of this lectin family. It is indeed striking that all the branches that diverged from the main line (leading to the leaf-specific lectins) consist of lectins with low or no carbohydrate-binding activity, whereas the leaf lectins are highly active agglutinins. In addition, most of the lectins of the side groups are abundant bulb-specific proteins, whereas the leaf lectins are only minor proteins. These observations suggest that conservation and/or enhancement of carbohydrate-binding activity was the most important criterion in the evolution of the main line (which is in perfect agreement with the presumed specific role of the leaf lectins in the plant's defense).

The evolution of the diverging groups clearly followed different criteria because the resulting lectins are either poorly active or inactive. Most probably, some *Allium* species used lectin genes to develop storage genes encoding tissue-specific storage proteins, which still possess some residual carbohydrate-binding activity (Smeets et al., 1997a).

Domain duplications played an important role in the evolution of the monocot mannose-binding lectins. An overview of the dendrogram indicates that two domain-lectins or lectin-related proteins have evolved by at least four independent duplication steps and in tandem insertion events (Figure 25A,B). In addition, a closer examination of the sequences reveals that the typical lectin domain itself consists of three similar subdomains of about 30 to 35 amino acid residues. Because each of these subdomains corresponds to one of the three mannose-binding sites of the lectin subunits, one can reasonably assume that the lectin domain has evolved from two subsequent duplications of a small ancestral mannose-binding polypeptide of about 35 residues. A comparison of the sequences of the three subdomains of GNA (Figure 25A) suggests that a first duplication of the ancestral subdomain [X] yielded a two-subdomain polypeptide [X′-X] that evolved further into a lectin [Y-X]. Duplication of the [Y] domain then gave rise to a three-subdomain polypeptide [Y-Y′-X] which further evolved into a lectin [Y-Z-X]. This presumed sequence of events explains why subdomains 1 and 2 are more closely related to each other than to subdomain 3.

The monocot mannose-binding lectins show an extensive sequence similarity with the mannose-binding domain of the actin-binding protein called comitin from *Dictyostelium discoideum* and mammalian cells. Comitin is a chimeric (cytoplasmic) protein consisting of an *N*-terminal mannose-binding domain and an unrelated *C*-terminal domain (Jung et al., 1996). Because of the simultaneous binding to mannose residues on the surface of vesicle membranes and to actin, comitin is capable of linking these membranes to the cytoskeleton. The mannose-binding domain of comitin not only exhibits a striking sequence similarity with GNA but also has a similar overall folding. Therefore, it is tempting to speculate, that the monocot mannose-binding lectins and the mannose-binding domain of comitin have a common three-subdomain ancestor. At present, it is not clear whether plants also contain comitin-like proteins. If so, the genes encoding the monocot mannose-binding lectins may have arisen from the *N*-terminal domain of comitin. Considering the rather narrow taxonomic distribution of the moconot mannose-binding lectins, such an evolution must have taken place after the common ancestor of the Amaryllidaceae, Alliaceae, Araceae, Bromeliaceae, Liliaceae, and Orchidaceae diverged from the other plants. If plants do not contain comitin-like proteins (which is possible because up to now no evidence has been obtained for the presence of comitin-like sequences in any plant species), the evolutionary origin of the monocot mannose-binding remains enigmatic unless one accepts the idea that the comitin domain has been acquired by the common ancestor of the aforementioned monocot families through a horizontal gene transfer (Figure 26).

H. Physiological Role

Similar to other lectin families, the physiological role of the monocot mannose-binding lectins is still poorly understood. Since the discovery of GNA in 1987 several functions have been proposed for the monocot mannose-binding lectins that are related to

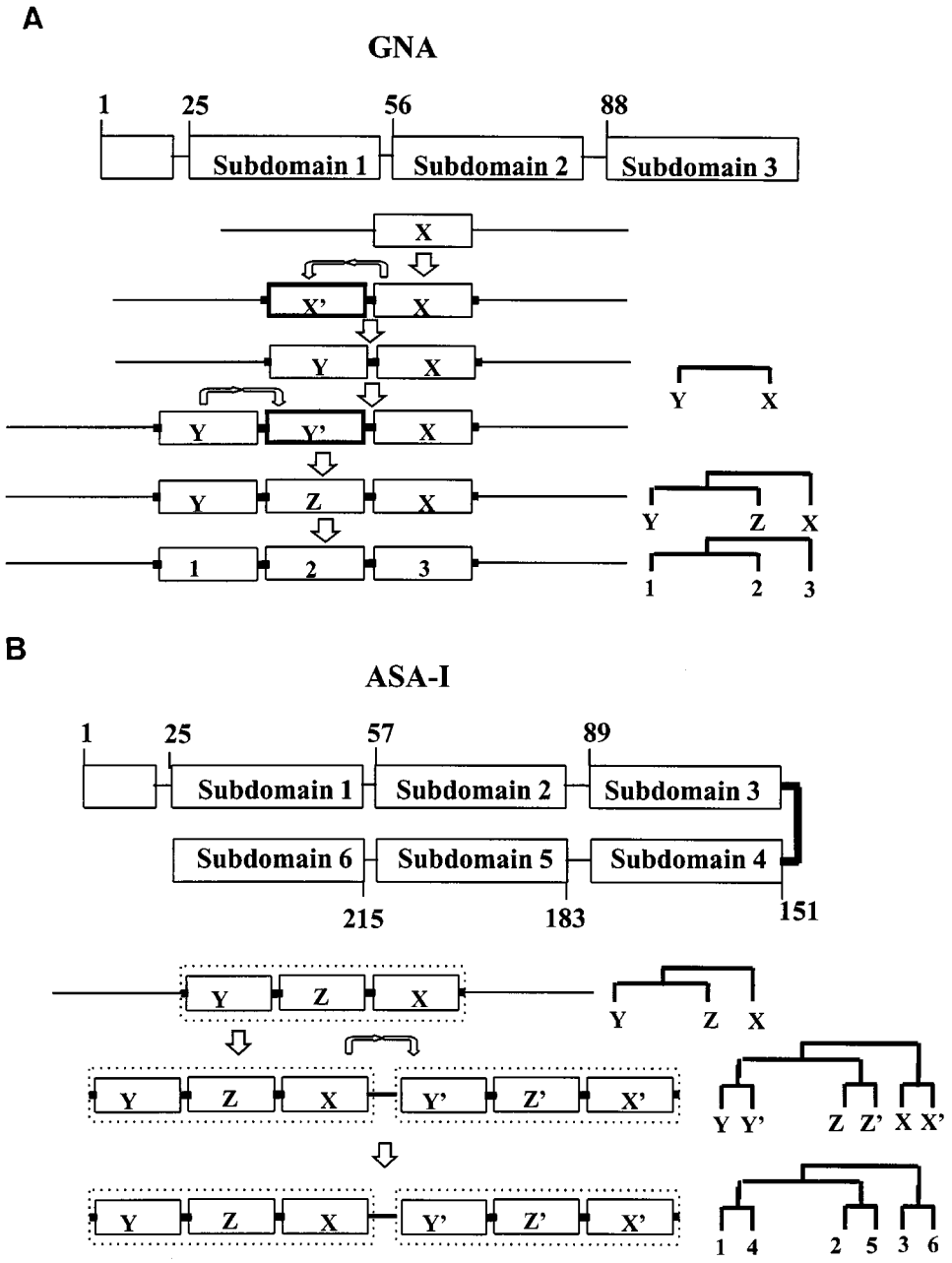

FIGURE 25. (A) Model of the molecular evolution of the three-subdomain protomer of GNA. Two subsequent duplications and in tandem insertions of a single subdomain [X] resulted in the formation of an ancestral [Y-Z-X] domain. (B) Model of the molecular evolution of the six-subdomain protomer of ASA-I. Two subsequent duplications and in tandem insertions of a single subdomain [X] resulted in the formation of an ancestral domain [Y-Z-X]. Duplication and in tandem insertion of the domain [Y-Z-X] yielded the typical two-domain ASA-I protomer consisting of six subdomains.

either nitrogen storage or defense. The presumed storage role is inferred from the observation that many monocot mannose-binding lectins are abundant proteins in vegetative storage tissues and behave as typical storage proteins for what concerns the regu-

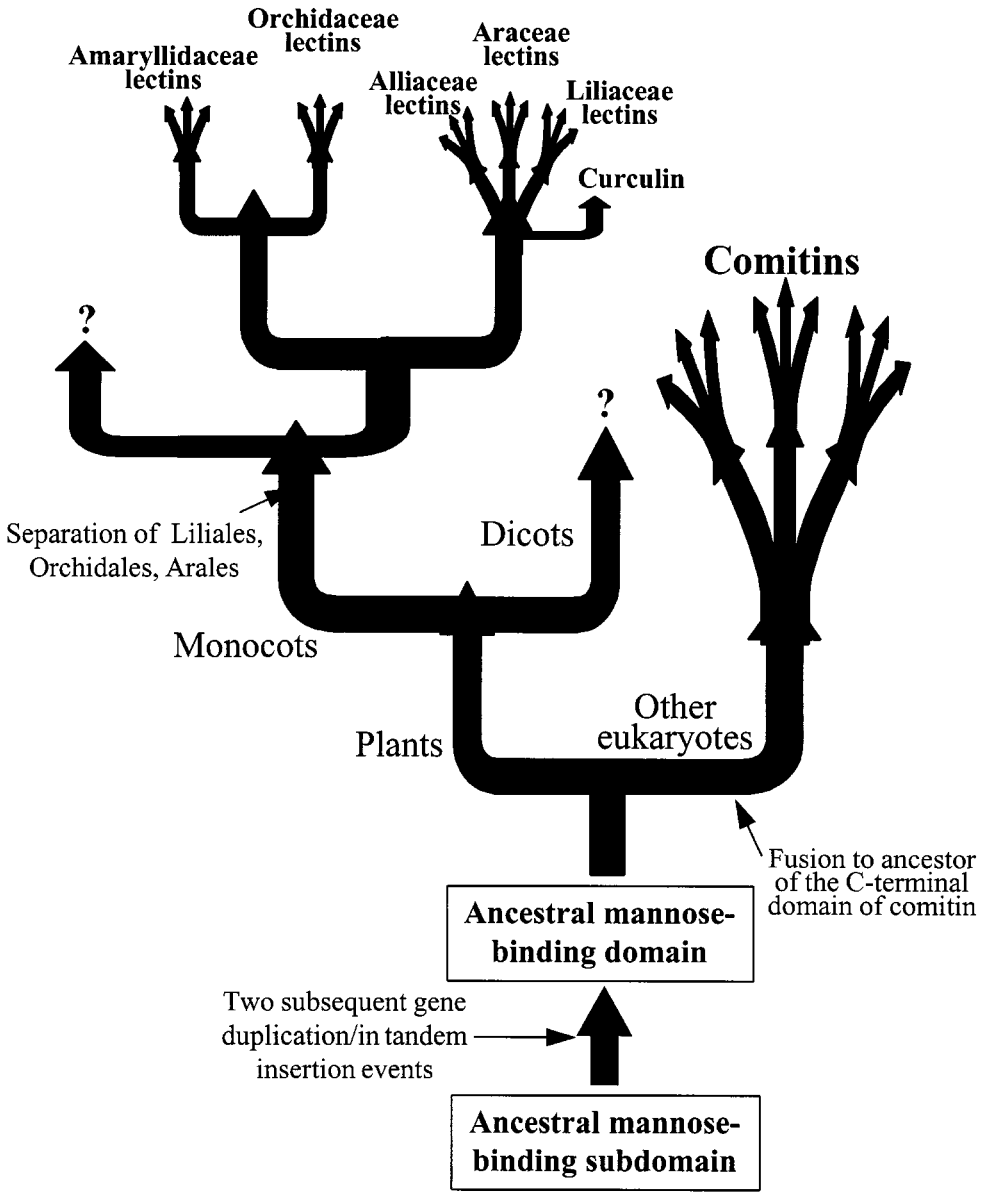

FIGURE 26. Hypothetical model of the molecular evolution of the monocot mannose-binding lectins. Two subsequent duplications and in tandem insertions of an ancestral mannose-binding subdomain yielded a mannose-binding domain consisting of three tandemly arrayed subdomains that followed two different evolutionary routes. One route started with a fusion to an unrelated domain and led to the comitin found in slime molds and mammalians. The second route eventually led to the current monocot mannose-binding lectins and the lectin-related sweet protein curculin.

lation of their expression. For example, the lectins of daffodil and snowdrop plants are located predominantly in the bulbs where they accumulate during late bulb formation and disappear when the tissue is consumed by the young shoot (Van Damme and Peumans, 1990a). Similarly, the bulb-specific garlic lectins, which are the most abundant bulb proteins, accumulate exclusively in the developing cloves and progressively disappear when the old clove is consumed by the young plant (Smeets et al., 1997a).

Monocot mannose-binding lectins are also believed to play a role in the plant's defense against phytophagous insects and other invertebrates. Experiments with artificial diets clearly demonstrated that the snowdrop lectin has detrimental effects on the development and reproduction of especially sucking insects (Rahbé et al., 1995). As transgenic plants expressing the snowdrop lectin also exhibit an increased resistance against sucking insects and nematodes, one can reasonably assume that at least this lectin is involved in the plant's defense against invertebrates (Hilder et al., 1995). The presumed defensive role of the monocot mannose-binding lectins is further supported by their preferential binding to high mannose-type glycan chains because these oligosaccharide structures are typical constituents of insect glycoproteins. A final argument in favor of the defensive role against sucking insects is the observation that some of these lectins accumulate in the phloem sap. The fact that a typical monocot mannose-binding lectin is the most abundant protein in the phloem exudate of flowering stalks of leek (*Allium porrum*) plants is in agreement with the presumed defensive role against phloem feeding insects (Peumans et al., 1997a). Although there is fairly good evidence that some monocot mannose-binding lectins protect the plant against invertebrates, such a defensive role can certainly not be generalized. Detailed studies of the agglutinating and antiviral activity have demonstrated that some monocot mannose-binding lectins have low activity compared with GNA. For example, the bulb-specific garlic lectins ASA-I and ASA-II are about 25- and 500-fold, respectively, less active than the leaf-specific lectin (which closely resembles GNA). Moreover, the mixture of ASA-I and ASA-II is virtually inactive against at least some of the insects when included in an artificial diet (Powell et al., 1995b). Based on these observations, it has been suggested that garlic expresses two types of monocot mannose-binding lectins with totally different physiological roles. The leaf lectin, which is present in small quantities, probably fulfills a specific defense-related role, whereas the abundant bulb-specific lectins are regarded as vegetative storage proteins (Smeets et al., 1997a). In this respect, garlic is much more specialized than, for example, the Amaryllidaceae species, which accumulate the same or at least very similar lectins in different tissues. For example, GNA and NPA preparations from leaves and bulbs are equally active. The high activity of Amaryllidaceae bulb lectins against insects does not preclude a storage role. Most probably, these bulb lectins are genuine storage proteins that can also be mobilized as defense proteins whenever the plant is attacked by phytophagous invertebrates.

VIII. JACALIN-RELATED LECTINS

Jacalin is the trivial name for the lectin from the seeds of jack fruit (*Artocarpus integrifolia*). At present, the term 'jacalin-related lectins' is used as a collective name for all lectins that are structurally and evolutionary related to the jack fruit lectin. The jacalin family comprises two subgroups of lectins. A first subgroup are the GalNAc-specific Moraceae seed lectins, which are very similar to the jack fruit lectin. The second subgroup are the Convolvulaceae lectins, which share sequence similarity with the Moraceae lectins but exhibit specificity toward mannose/maltose.

A. Historical Note

The first evidence for the occurrence of a lectin of the jacalin family dates back to the report of Jones et al. (1967) showing that seeds of *Maclura pomifera* contain a potent hemagglutinin. Purification of the *Maclura pomifera* seed lectin was achieved in 1977

(Bausch and Poretz, 1977) followed by that of jacalin in 1981 (Moreira and Ainouz, 1981). Detailed specificity studies of the *Maclura pomifera* lectin (Sarkar et al., 1981) and jacalin (Sastry et al., 1986) soon revealed that both lectins have a strong preference for Galβ(1,3)GalNAc residues. After the discovery of the specific IgA-binding activity and mitogenicity of jacalin (Roque-Barreira et al., 1985), the jack fruit lectin became an important tool in immunology. The subsequent discovery of the potent anti-HIV (Kabir and Daar, 1994) activity and insecticidal properties (Czapla and Lang, 1990) stimulated further research on jacalin and related lectins. Both jacalin and the *Maclura pomifera* lectin were completely sequenced by chemical methods (Young et al., 1991). Jacalin has also been cloned (Yang and Czapla, 1993) and its three-dimensional structure was resolved by X-ray crystallographic analysis (Sankaranarayanan et al., 1996).

Recently, molecular cloning of the rhizome lectin from *Calystegia sepium* (called Calsepa) led to the unexpected discovery that this mannose/maltose-specific lectin is evolutionarily related to the Moraceae lectins (Van Damme et al., 1996e). Therefore, Calsepa (and other closely related Convolvulaceae lectins) are now considered a subgroup of the jacalin family.

B. Occurrence and Molecular Structure

Previously, lectins of the jacalin family have been found in Moraceae and Convolvulaceae species. Within the Moraceae family, the occurrence of jacalins is documented for the seeds of jack fruit (*Artocarpus integrifolia*) and several other closely related *Artocarpus* species, as well as for the seeds of osage orange (*Maclura pomifera*). Most probably, seeds of other Moraceae species also contain similar lectins. Within the Convolvulaceae family lectins have been isolated from rhizomes of *Calystegia sepium* (hedge bindweed) and *Convolvulus arvensis* (bindweed). In addition, there is also evidence for the occurrence of a similar lectin in sweet potato (*Ipomea batatas*). Most probably, the occurrence of jacalins is not restricted to the Moraceae and Convolvulaceae families. For example, the lectin from tubers of the Jerusalem artichoke (*Helianthus tuberosus*, Asteraceae) resembles the Convolvulaceae lectins (Peumans, unpublished results). In addition, a lectin with sequence similarity to jack fruit lectin was reported in *Musa acuminata* (Musaceae) (Clendennen and May, 1997).

Moraceae lectins are usually considered typical seed proteins. Both the *Artocarpus* and *Maclura* lectins are present in large quantities in the seeds. According to some reports, jacalin is the most abundant seed protein representing over 50% of the total soluble protein (Kabir et al., 1993). However, Moraceae lectins are not strictly confined to the seeds. It has been reported, for instance, that the *Artocarpus lakoocha* lectin occurs in all plant tissues except the fruit flesh (Wongkham et al., 1995).

Convolvulaceae lectins typically occur in underground storage tissues. The lectins from *Calystegia sepium* and *Convolvulus arvensis*, called Calsepa and Conarva, respectively, are abundant rhizome proteins representing about 25 and 5%, respectively, of the total soluble protein. Both lectins are also present in small quantities in young shoots. The sweet potato lectin is only a minor tuber protein representing less than 0.1% of the soluble protein.

1. Molecular Structure

The jacalin family comprises two subgroups of lectins with different molecular structures. All Moraceae lectins characterized thus far are composed of four identical

protomers consisting of a large α-chain and a small β-chain. Both chains are derived from a single precursor and are not held together by disulfide bonds. Each protomer possesses a single carbohydrate-binding site. The mature lectin from *Artocarpus integrifolia* consists of four α- and four β-chains of 133 residues and 20 residues, respectively. Native jacalin is a complex mixture of isoforms. Part of the complexity is due to the simultaneous occurrence of (most probably) five different polypeptide chains. Differential posttranslational processing creates some additional microheterogeneity (Young et al., 1995). Only one of the jacalin isoforms contains an *N*-linked glycan in its α-chain. Hence, jacalin is a mixture of glycosylated and unglycosylated forms. Lectins very similar to jacalin have also been isolated from *A. altilis, A. champeden, A. integer, A. lakoocha,* and *A. tonkinensis* (Blasco et al., 1996). Although these *Artocarpus* lectins have not been studied in great detail, there is no doubt that they have the same overall structure. The seed lectin from *Maclura pomifera* also consists of four α- and four β-chains of 133 residues and 20 residues, respectively. Most probably, the *Maclura pomifera* lectin is a mixture of two (genetic) isoforms. Differential posttranslational processing of these two isoforms creates additional microheterogeneity. None of the *Maclura pomifera* isoforms is glycosylated.

The molecular structure of the native Convolvulaceae lectins is apparently less complex than that of the Moraceae lectins. Both the *Calystegia sepium* and *Convolvulus arvensis* rhizome lectins consist of two identical unglycosylated protomers of 153 residues. Molecular cloning revealed the occurrence of at least two genetic isoforms of Calsepa (Van Damme et al., 1996e).

2. Primary Structure

The lectins from *Artocarpus integrifolia* and *Maclura pomifera* have been completely sequenced by chemical methods and were shown to contain an internal repeat in their α-chains (Young et al., 1991). Furthermore, the lectin from jack fruit has been cloned (Yang and Czapla, 1993). The primary structure of the lectins from *Calystegia sepium* and *Convolvulus arvensis* can be deduced from the nucleotide sequences of the respective cDNAs (Van Damme et al., 1996e; Van Damme, unpublished results). Like the Moraceae lectins, both Convolvulaceae lectins also contain an internal repeat.

C. Structure, Biosynthesis, and Posttranslational Modifications

cDNA cloning of jacalin and the lectins from *Calystegia sepium* and *Convolvulus arvensis* determines the structure and the possible posttranslational modifications of the primary translation products of the respective lectin mRNAs. Although the mature proteins are similar, their biosynthesis and processing are different (Figure 27).

Jacalin is synthesized as a preproprotein consisting of a signal peptide followed by a propeptide of 39 residues, the β-peptide of 20 residues, a linker of four amino acids, and the α-peptide of 133 residues. Most probably, jacalin is synthesized on the ER and co-translationally processed by the removal of the signal peptide. The resulting propeptide is subsequently transported via the ER to the protein bodies. During or after transport, the propeptide is cleaved at three different sites, resulting in the removal of the *N*-terminal propeptide and the excision of the linker between the β-chain and α-chain. The β-chain and α-chain remain together by noncovalent bonds and form the protomer of jacalin. One of the isoforms of jacalin contains an *N*-linked glycan chain, and thus is co-translationally glycosylated. cDNA cloning revealed the occurrence of four different although highly similar molecular forms of jacalin. Southern blot analysis further con-

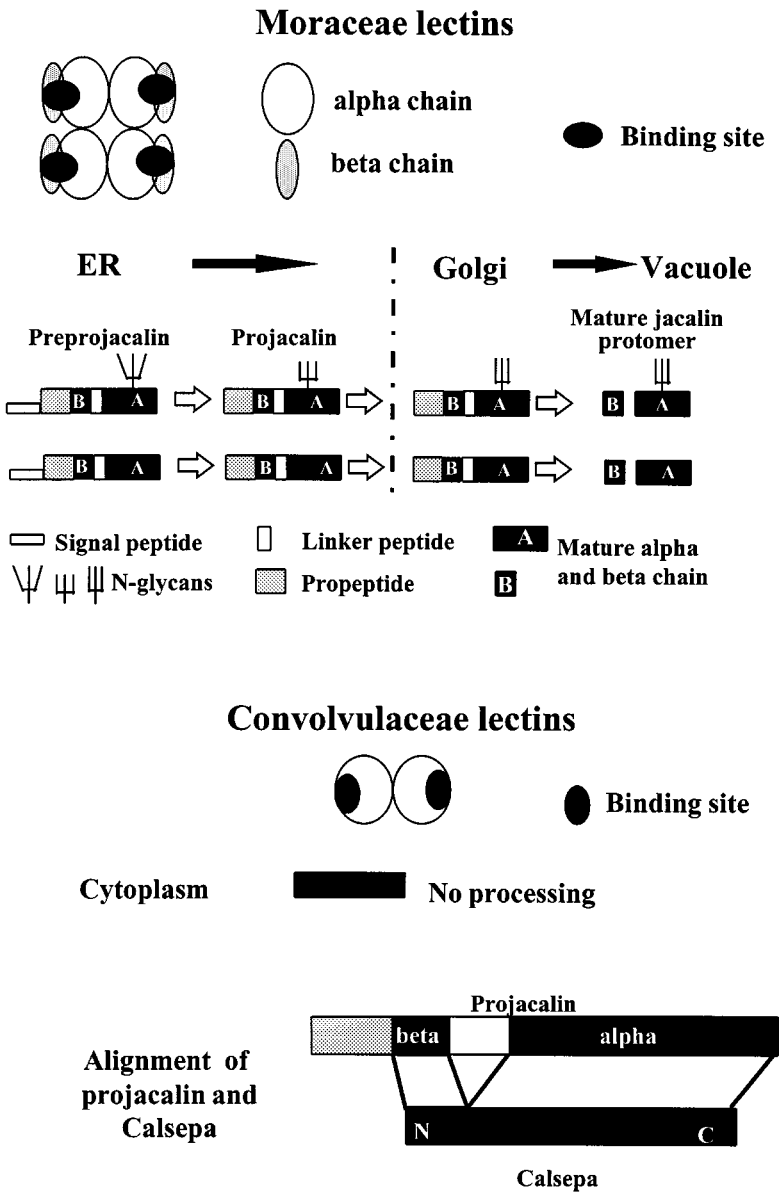

FIGURE 27. Schematic representation of the molecular structure, biosynthesis, and processing of jacalin-related lectins. Examples shown are jacalin (from *Artocarpus integrifolia*) and the *Calystegia sepium* agglutinin (Calsepa). A and B represent the α- and β-chains of jacalin, whereas N and C refer to the *N*- and *C*-terminus of Calsepa.

firmed that jacalin is encoded by a family of genes (Yang and Czapla, 1993). Most probably, the processing scheme described for jacalin holds true for all *Artocarpus* lectins as well as for the *Maclura pomifera* lectin.

cDNA cloning of Calsepa and Conarva revealed that the mature lectin polypeptides correspond to the entire open reading frame of the respective lectin genes (Van Damme et al., 1996e). The apparent absence of a signal peptide indicates that the Convolvulaceae lectins are synthesized in the cytoplasm and undergo no posttranslational modifications. Interestingly, the Calsepa sub-

unit corresponds to a polypeptide composed of the (peptide bond-linked) homologues of both the β-chain and α-chain of jacalin. Analysis of multiple cDNAs demonstrated the occurrence of two highly similar isoforms of Calsepa, suggesting that the lectin is encoded by a family of genes. In addition, sequence analysis of PCR-amplified genomic DNA fragments revealed the presence of a single intron in the Calsepa gene(s) (Van Damme, unpublished results).

D. Carbohydrate-Binding Specificity

The jacalin lectin family comprises two subgroups with a completely different carbohydrate-binding specificity. Detailed specificity studies have demonstrated that all *Artocarpus* lectins as well as the *Maclura pomifera* lectin strongly react with terminal nonreducing α-D-galactosyl residues and have a high affinity for Galβ(1,3)GalNAc (Sastry et al., 1986; Sarkar et al., 1981). At present, the sugar specificity of the Convolvulaceae lectins has not yet been studied in great detail. However, preliminary studies clearly indicated that Calsepa and Conarva recognize both mannose and maltose (Peumans et al., 1997b).

E. Three-Dimensional Structure

Recently, the three-dimensional structure of jacalin has been resolved by X-ray diffraction analysis (Sankaranarayanan et al., 1996). Each protomer consists of a threefold symmetric β-prism made of three four-stranded β-sheets (Plate 1). Eleven out of the twelve strands (numbered from 1 to 12) are formed by the α-chain whereas the β-chain forms the last strand (number 11). Four subunits, 1, 1′, 2, and 2′, are noncovalently associated into a tetrameric structure that probably results from the association of two dimers 1–2 and 1′-2′. The β-chains, which occupy the central region of the tetramer, play a crucial role in the association of the four subunits.

Each protomer possesses a single monosaccharide-binding site made of residues Gly^{1}, Tyr^{122}, Trp^{123}, and Asp^{125} of the α-chain that create a network of nine hydrogen bonds with O3, O4, O5, and O6 of methyl-α-D-galactose (Figure 23B). Because Gly^{1} of the α-chain is only liberated after proteolytic processing of projacalin into the α and β chains, the sugar-binding activity of the lectin requires a correct posttranslational processing of the lectin precursor.

Based on the high sequence homology between jacalin and the *Maclura pomifera* lectin (Young et al., 1991), it is predicted that both lectins have very similar three-dimensional conformations. One can also reasonably expect that the subunits of the mannose/maltose-specific lectin isolated from *Calystegia sepium* (Van Damme et al., 1996e), which shares 30% sequence identity and 49% sequence similarity with jacalin, have a similar overall structure.

F. Molecular Evolution

The family of the jacalin-related lectins comprises the two distinct subfamilies of Moraceae and Convolvulaceae lectins. Because only a few members of each subfamily have been sequenced and/or cloned, no dendrogram is presented here. During the last few months evidence has been presented that lectins similar to the Convolvulaceae lectins also occur in *Helianthus tuberosus* (Jerusalem artichoke) (Peumans, unpublished results) and *Musa acuminata* (banana) (Clendennen and May, 1997). The occurrence of similar mannose-specific lectins in the families Convolvulaceae, Asteraceae, and Musaceae not only demonstrates that the jacalin family is widespread within the plant kingdom but also suggests that the GalNAc-specific Moraceae lectins are most likely a

side branch of the main evolutionary line (Figure 28). Further sequencing and cloning of the *Helianthus tuberosus* and *Musa acuminata* agglutinins certainly will provide more information about the molecular evolution of the family of jacalin-related lectins.

The protomers of both Moraceae and Convolvulaceae lectins consist of two internal repeats with a relatively low sequence similarity. By analogy with some other lectin families, it is tempting to speculate that the two-domain protomers of the jacalin-related lectins also arose by duplication and tandem insertion of an ancestral domain. However, because jacalin possesses only one binding site consisting of amino acid residues located at the *N*-terminus and at the *C*-terminus of the protomer, the two presumed subdomains do not correspond to separate binding sites. Therefore, it remains to be demonstrated, that the ancestor domain of the jacalin-related lectins was a single-domain carbohydrate-binding protein.

Although all jacalin-related lectins are undoubtedly related, the Moraceae and Convolvulaceae lectins definitely have a different intracellular localization in the vacuoles and the cytoplasm, respectively. As mentioned previously, the Moraceae lectins most likely represent a side branch of the family of jacalin-related lectins that diverged from the main line of cytoplasmic mannose-

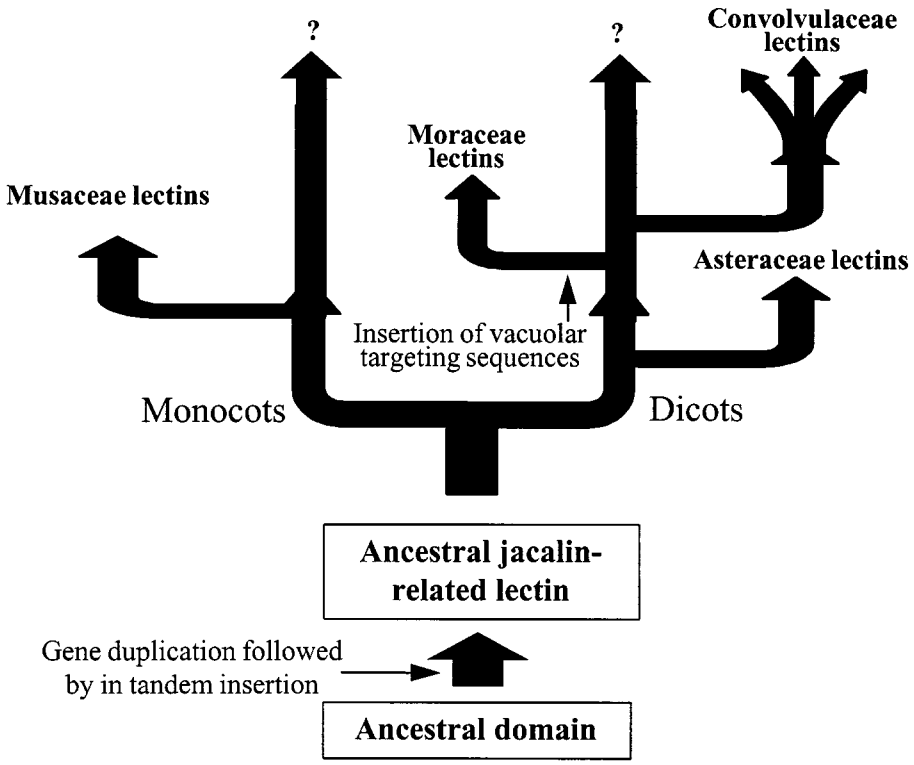

FIGURE 28. A hypothetical model of the molecular evolution of the jacalin-related lectins. The ancestor of the modern jacalin-related lectins arose — before the monocots and dicots separated from each other — by a duplication and in tandem insertion of an ancestral domain. Further evolution of the common ancestor gave rise to the Musaceae, Asteraceae, and Convolvulaceae (and probably other) cytoplasmic lectins. The Moraceae lectins arose after the insertion of vacuolar targeting sequences in the cytoplasmic lectins. This evolutionary event most probably took place after the Moraceae separated from the other dicots.

specific lectins. By analogy with, for example, the *Allium* lectins, it is tempting to speculate that some Moraceae species developed genes encoding storage proteins starting from the lectin genes. To make the original lectin genes suitable as storage protein genes, extra sequences (e.g., a signal peptide) had to be introduced to direct the protein product toward the secretory system. This explains why the primary translation products of the Moraceae lectins undergo a more complex processing than these of the Convolvulaceae lectins.

G. Physiological Role

Previously, no conclusive evidence has been reported for a well-defined role for the jacalin-related lectins. However, some speculation can be made on the basis of the currently available data on the physiology and biological activities of the lectins of both the Moraceae and Convolvulaceae lectins.

Due to their earlier discovery, the Moraceae lectins have been studied in more detail than the Convolvulaceae lectins. As mentioned above, Moraceae lectins occur in many tissues but are particularly abundant in the seeds. *In vitro* studies indicated that jacalin and the *Maclura pomifera* lectin interact with human and animal cells and are capable of inducing specific processes. No data have been reported on the effects of orally administered Moraceae lectins to higher animals. Feeding trials with artificial diets have demonstrated that the Moraceae lectins have anti-insect properties. For example, the *Maclura pomifera* lectin had a significant inhibitory effect on the larvae of the cowpea weevil (*Callosobruchus maculatus*) (Murdock et al., 1990). Similarly, jacalin and the *Maclura pomifera* lectin inhibited larval growth of the Southern corn rootworm (*Diabrotica undecimpunctata*) (Czapla and Lang, 1990) and jacalin affected the survival of the potato leafhopper (*Empoasca fabae*) (Habibi et al., 1993). Based on these observations, it seems likely that the Moraceae seed lectins are, like most other abundant seed lectins, storage proteins with an additional defensive function against potential seed predating animals and/or insects.

At present little information is available on the role of Convolvulaceae lectins. It was shown that these lectins are abundant rhizome-specific proteins located exclusively in the cytoplasm (Peumans, unpublished results) and have mitogenic activity (Peumans et al., 1997b). Although these data indicate that the Convolvulaceae lectins are (rhizome-specific) storage proteins with a possible protective activity against potential predating soil-borne vertebrates or invertebrates, further research on the possible toxic or antinutrient activity is required to support this hypothesis. It should be mentioned, however, that the rhizomes of for example, *Calystegia sepium* and *Convolvulus arvensis*, are not or rarely attacked by soil invertebrates or voles. This apparent resistance may be a reason for the persistency of these weeds, which create difficulties in agriculture worldwide.

IX. AMARANTHIN LECTIN FAMILY

The term ‘amaranthin’, which originally referred to the seed lectin of *Amaranthus caudatus*, is now used as a collective name for the closely related GalNAc-specific seed lectins from various *Amaranthus* species. The amaranthins are not related to any other lectin family.

A. Historical Note

The occurrence of amaranthins was reported initially by Bird (1954), who observed

that seed extracts from various *Amaranthus* species agglutinated red blood cells of different animals. However, it took until 1988 before amaranthins were purified and characterized from *A. leucocarpus* and *A. cruentus* (Koeppe and Rupnov, 1988; Zenteno and Ochoa, 1988). The first primary structure of an amaranthin has been reported only recently by Transue et al. (1997), who determined the three-dimensional structure of the *Amaranthus caudatus* lectin by X-ray crystallographic analysis.

B. Occurrence, Molecular Structure, and Amino Acid Sequences

Amaranthins have been found in seeds of *Amaranthus caudatus, A. spinosus, A. leucocarpus,* and *A. cruentus.* No related lectins have been identified in any other plant family. Amaranthins are typical seed proteins representing about 3 to 5% of the soluble protein (Rinderle et al., 1989). All amaranthins are homodimeric proteins composed of subunits of approximately 33 kDa, which are not glycosylated. Only the lectin from *Amaranthus caudatus* (amaranthin) has been sequenced completely by chemical methods (Transue et al., 1997). The lectin polypeptide is 299 residues long and consists of two homologous domains. Amaranthin shares a very high sequence similarity with a previously cloned seed-specific protein from *Amaranthus hypochondriacus* (Raina and Datta, 1992). Most probably, this unidentified protein corresponds to the *Amaranthus hypochondriacus* lectin.

C. Structure, Biosynthesis, and Posttranslational Modifications

At present no data have been reported on the biosynthesis and/or processing of the amaranthins. However, some predictions can be made on the basis of the sequence of the mature amaranthin and the deduced sequence of the presumed *Amaranthus hypochondriacus* lectin. Both sequences are almost identical except that the deduced sequence of the *Amaranthus hypochondriacus* lectin has four extra residues at its *C*-terminus. The obvious absence of a putative signal peptide indicates that the amaranthins are synthesized in the cytoplasm. Similarly, the 4–residue *C*-terminal extension in the deduced sequence of the *Amaranthus hypochondriacus* lectin suggests that a short *C*-terminal propeptide is cleaved from the primary translation product of the amaranthins (Figure 29). No sequence information is available on the number of genes encoding amaranthins and the possible occurrence of introns.

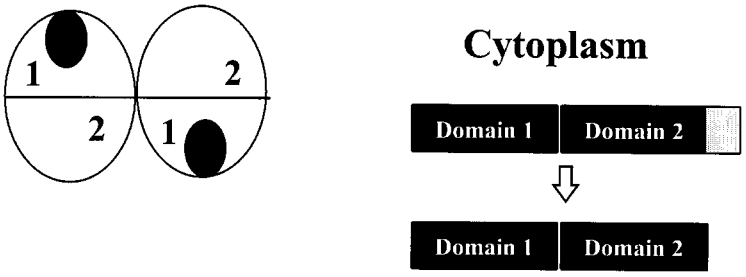

FIGURE 29. Schematic representation of the molecular structure, biosynthesis, and processing of Amaranthaceae lectins.

D. Carbohydrate-Binding Specificity

Detailed specificity studies have been performed only with the *Amaranthus caudatus* lectin. The lectin is inhibited by GalNAc but has a much higher affinity for the disaccharide Galβ(1,3)GalNAc (Rinderle et al., 1989).

E. Three-Dimensional Structure

Recently, the three-dimensional structure of the *Amaranthus caudatus* agglutinin was solved at 2.2 Å resolution (Transue et al., 1997). The protomers consist of two domains (called *N*- and *C*-domains) linked by a short helix. Each of these domains has a β-trefoil structure similar to that of the two domains of the ricin B chain. However, unlike the ricin B chain (which has no extended secondary structure), the domains of amaranthin consist of six strands of antiparallel β-sheet capped by three β-hairpins into a β-barrel. The two domains are linked by a 3_{10} helix and form an oval protomer with the capping β-hairpins of each domain located at both ends of the monomer. Two monomers associate head to tail into a dimer by extensive noncovalent contacts between both monomers. In this dimeric organization, the *N*-domain of one monomer faces the *C*-domain of the other monomer. The dimer exhibits two surface-exposed carbohydrate-binding sites that appear as two shallow depressions formed at the interface between the *N*- and *C*-domains of the two facing monomers.

The binding of benzylated T-antigen disaccharide (Gal-β1,3–GalNAc-α-*O*-benzyl) to amaranthin shows a complex hydrogen bonding pattern involving amino acid residues that belong mainly to surface-exposed hairpins and turns. The GalNAc-reducing end of the disaccharide interacts with the *N*-domain residues Asn^{74}–His^{75}–Tyr^{76}–Trp^{77} of hairpin 2, residues Tyr^{124}, Val^{126}, and Phe^{130} of hairpin 3 and residue Phe^{135} of a β-strand of hairpin 3. The nonreducing Gal of the disaccharide interacts through hydrogen bonds often mediated by water molecules with Met^{261}–Gln^{262}–Lys^{263}–Thr^{264} of a turn of the *C*-domain and the side chain of Asn^{74} of the *N*-domain. In this binding no hydrophobic contact occurs between the disaccharide and the lectin surface. In addition, the pyranose rings of the T-antigen are approximately parallel to the protein surface. Thus, this sugar binding pattern is quite different from that found in other proteins exhibiting a β-trefoil fold, such as ricin or fibroblast growth factors.

F. Molecular Evolution

The amaranthins are a very small lectin family with no apparent sequence similarity to any other protein. Sequencing of the *Amaranthus caudatus* agglutinin revealed that the protomers consist of two repeats (domains) of about 150 amino acid residues with a relatively high sequence similarity. Each of these repeats folds into a β-trefoil motif. By analogy with many other plant lectins, it is tempting to speculate that the two-domain protomers of the amaranthins result from a duplication and tandem insertion of an ancestral domain. Due to the location of the carbohydrate-binding sites at the interface between two domains, the question remains whether the putative ancestral domain of the amaranthins was a single-domain carbohydrate-binding protein. X-ray crystallographic analysis demonstrated that each domain of the *Amaranthus caudatus* agglutinin consists of three subdomains. Because there is some residual sequence similarity between the three subdomains, the ancestral domain of the amaranthins may have arisen from two subsequent duplications and in tandem insertions of a small ancestral

polypeptide of about 50 residues. According to this scenario the putative molecular evolution of the amaranthin protomers closely resembles that of the B chain of the type 2 RIP. In a first step, two subsequent duplications and in tandem insertions of an ancestral polypeptide of about 50 residues gave rise to a three-subdomain polypeptide. Duplication and in tandem insertion of this putative three-subdomain polypeptide then resulted in the formation of a two-domain lectin polypeptide.

G. Physiological Role

Amaranthins are typical seed proteins that interact at least *in vitro* with human and animal cells. As no data have been reported on the possible toxic or antinutrient properties of the amaranthins, their physiological role remains unclear. Taking into consideration the exclusive location of the amaranthins in the seeds and their high affinity for the T-antigen, these lectins could be involved in the plant's defense against seed predators.

X. CUCURBITACEAE PHLOEM LECTINS

The Cucurbitaceae phloem lectins are a small family of chitin-binding agglutinins found in the phloem exudate of Cucurbitaceae species. They are not related to other Cucurbitaceae lectins (e.g., the type 2 RIP from *Trichosanthes kirilowii*) and do not contain hevein domains.

A. Historical Note

The presence of an agglutinating factor in pumpkin (*Cucurbita maxima*) was observed in 1968 by Hossaini (Hossaini, 1968). One decade later, Sabnis and Hart (1978) demonstrated that lectins are major protein components of the phloem exudate of several Cucurbitaceae species, and reported for the first time the partial characterization of the pumpkin lectin. In subsequent years the carbohydrate-binding specificity and molecular structure of several phloem lectins were determined (Allen, 1979; Read and Northcote, 1983). Cloning of the pumpkin lectin in 1992 eventually yielded the full details of the sequence of the phloem lectins (Bostwick et al., 1992).

B. Occurrence, Molecular Structure, and Amino Acid Sequences

Cucurbitaceae phloem lectins have been identified in phloem exudates of *Cucurbita*, *Citrullus*, *Cucumis*, *Sechium*, *Luffa*, and *Coccinia* species. The phloem exudate lectins (also called protein PP2) are abundant proteins. According to immunocytochemical localization and *in situ* hybridization studies, the phloem lectin genes are specifically expressed in the companion cells of the phloem tissue (Bostwick et al., 1992).

All Cucurbitaceae phloem lectins consist of unglycosylated subunits of about 25 kDa. Exact values calculated from deduced sequences are 24,478 and 24,550 Da for the 228 residue protomers of the *Cucurbita maxima* and *Cucurbita pepo* lectins, respectively (Bostwick et al., 1992; Wang et al., 1994). In solution, the phloem lectins form dimers. According to the original report, the two subunits of the *Cucurbita maxima* lectin are covalently linked through two interchain disulfide bonds (Read and Northcote, 1983). However, the subunits of the *Luffa acutangula* lectin are not linked by disulfide bridges (Anatharam et al., 1986).

None of the Cucurbitaceae phloem lectins has been completely sequenced by chemical methods, but deduced sequences are avail-

able for the *Cucurbita maxima* and *Cucurbita pepo* phloem lectins (Bostwick et al., 1992; Wang et al., 1994).

C. Structure, Biosynthesis, and Posttranslational Modifications

No data are available on the biosynthesis and/or processing of the Cucurbitaceae phloem lectins. However, some predictions can be made from the deduced sequences of the *Cucurbita maxima* and *Cucurbita pepo* phloem lectins (Figure 30). Neither of the sequences contains a putative signal sequence. This, taken together with the observation that the *N*-terminal sequence of a cyanogen bromide cleavage fragment of the *Cucurbita pepo* lectin starts with amino acid residue 2 of the open reading frame of the cDNA, indicates that the phloem lectins are synthesized in the cytoplasm. It is not known whether the phloem lectins possibly undergo *C*-terminal processing. If so, the propeptide must be short because the size of the mature lectin subunits is very similar to that of the primary translation products.

Sequence analysis of genomic clones from *Cucurbita maxima* revealed the occurrence of two introns in the coding sequence (Bostwick et al., 1994). In addition, the isolation of a genomic clone containing two lectin genes that are 99.8% identical over a region of 3055 nucleotides not only demonstrated the presence of two genes but also suggested a recent gene duplication event. Southern blot analysis of DNA from several *Cucurbita* species further indicated that the lectins are encoded by small gene families consisting of two to eight genes.

D. Carbohydrate-Binding Specificity

The Cucurbitaceae phloem lectins exhibit specificity toward oligomers of GlcNAc. Detailed studies of the *Luffa acutangula* lectin have demonstrated that the inhibitory potency of the GlcNAc-oligomers strongly increases with chain length up to five residues (Anatharam et al., 1986). The *Luffa acutangula* and the *Cucurbita maxima* lectins also recognize the internal di-*N*-acetylchitobiosyl sequences of *N*-linked glycan chains from fetuin, ovalbumin, and soybean lectin (Allen, 1979; Anatharam et al., 1986).

E. Molecular Evolution

The Cucurbitaceae phloem lectins are a small lectin family with no apparent sequence similarity to any other protein. No internal repeats have been found in the sequences of the lectin protomers, and therefore no conclusions can be drawn about their molecular evolution.

F. Physiological Role

The typical and exclusive location of the Cucurbitaceae phloem lectins in the phloem is probably indicative of their physiological

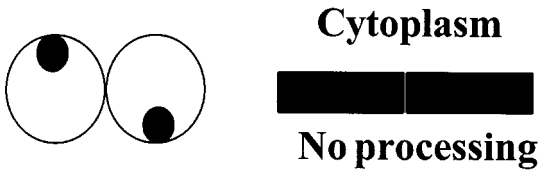

FIGURE 30. Schematic representation of the molecular structure of the Cucurbitaceae phloem lectins.

role. After cutting of the phloem vessels (e.g., after wounding of plant), the lectins react specifically with PP1 (which is another abundant phloem protein). This reaction involves the formation of intermolecular disulfide bridges between the lectin and PP1 and results in the formation of a rigid gel that blocks the phloem vessels. Most probably, this mechanism prevents infection of the damaged phloem by microorganisms (Read and Northcote, 1983). Hence, the Cucurbitaceae phloem lectins can be regarded as defense proteins.

XI. OTHER LECTINS

Several lectins cannot be classified yet because they have no apparent structural similarity with the above-described lectin families, and no sequence information is available to trace possible evolutionary relationships. Protein sequencing and/or molecular cloning are necessary to answer the question of whether these unclassified lectins possibly belong to novel, yet undiscovered families.

A. Occurrence and Molecular Structure

In the absence of clear criteria to group the unclassified lectins, they are discussed below in alphabetical order of the plant family in which they occur.

1. Apiaceae

The only known Apiaceae lectin was isolated from *Aegopodium podagraria* (ground elder; Peumans et al., 1985b). Rhizomes of ground elder contain a GalNAc-specific lectin consisting of eight identical glycosylated subunits of about 60 kDa. The lectin occurs exclusively in the rhizomes where it represents over 50% of the total soluble protein.

2. Araucariaceae

Seeds of *Araucaria brasiliensis* (pinhao) contain two mannose/glucose-specific lectins. Lectin I (200 kDa) is built up of 10 subunits of 20 kDa, whereas lectin II (200 kDa) is a hexamer composed of subunits of 34 kDa (Datta et al., 1991). Both lectins are glycosylated and have a similar amino acid composition. The pinhao lectins are the only gymnosperm lectins reported to date.

3. Celastraceae

Arils of the seeds of *Euonymus europaeus* (spindle tree) contain high levels of a lectin called *Euonymus europaeus* agglutinin or EEA (Petryniak and Goldstein, 1987). Native EEA has a molecular mass of 120 to 160 kDa and consists of 17 kDa and 35 kDa polypeptides (in a still unknown stochiometry). Similar lectins occur in the arils of other *Euonymus* species.

4. Cucurbitaceae

In addition to the typical Cucurbitaceae phloem lectins, several yet unclassified agglutinins have been isolated from different members of this family. Root stocks of *Bryonia dioica* (white bryony) and *Marah macrocarpus* (Californian wild cucumber) contain GalNAc-specific lectins consisting of two different disulfide bridge-linked glycosylated subunits of 35 and 30 kDa, respectively (Peumans et al., 1984b, 1987). The galactose-specific seed lectin of *Trichosanthes kirilowii* (serpent cucumber) also consists of

two different disulfide bridge-linked glycosylated subunits of 37 and 25 kDa (Falasca et al., 1989), whereas the galactose-specific seed lectin from *Telfaira occidentalis* (fluted gourd) behaves as a hexamer (180 kDa) composed of three pairs of disulfide-linked subunits of 30 kDa (Togun et al., 1994). All these lectins clearly resemble type 2 RIP with respect to their molecular structure. In addition, they also have a very similar amino acid composition. Based on these observations, it is tempting to speculate that these Cucurbitaceae lectins belong to the type 2 RIP lectin family. However, this presumed classification needs further confirmation by sequence data.

5. Euphorbiaceae

Several Euphorbiaceae lectins have been described (Nsimba-Lubaki et al., 1983, 1986). Most of these lectins are Gal- or GalNAc-specific agglutinins built up of two or four subunits of 30 to 35 kDa. For example, the GalNAc-specific *Euphorbia heterophylla* seed lectin consists of two identical glycosylated subunits of 32 kDa, whereas the GalNAc-specific *Hura crepitans* seed lectin is composed of four identical glycosylated subunits of 30 kDa (Barbieri et al., 1983). Similarly, the latex lectins from *Hura crepitans* and several *Euphorbia* species are GalNAc-specific agglutinins composed of either two or four identical or slightly different glycosylated polypeptides of 30 to 35 kDa. The seed lectin of *Croton tiglium* apparently differs from the other Euphorbiaceae lectins because it has a complex specificity and consists of two identical noncovalently linked glycosylated subunits of 56 kDa. Due to the lack of sequence information, the possible evolutionary relationships of all these Euphorbiaceae lectins remain unclear. In this respect the unclassified lectins from *Tetracarpidium conophorum* (Nigerian walnut) make an exception. Seeds of this particular Euphorbiaceae species contain two different lectins called TCA-I and TCA-II. TCA-I is a GalNAc-specific lectin consisting of two disulfide bridge-linked glycosylated subunits of 34 kDa, whereas TCA-II behaves as a monomeric lactose-specific lectin composed of a single glycosylated polypeptide of 34 kDa (Togun et al., 1988). As both TCA-I and TCA-II have some sequence similarity with the B chain of type 2 RIP they possibly belong to the type 2 RIP, lectin family.

6. Gramineae

In addition to the chitin-binding Gramineae lectins, two distinct lectins, called ZMA-I and ZMA-II, have been isolated from maize (*Zea mays*) kernels (Jankovic et al., 1990). ZMA-I is a GalNAc-specific lectin consisting predominantly of 23 kDa proteins. ZMA-II is a mannose-binding lectin consisting mainly of 12.5 kDa proteins. The maize lectins are definitely not related to the chitin-binding Gramineae lectins.

7. Labiatae

Within the Labiatae family lectins have been isolated from seeds of *Moluccella laevis* and *Salvia sclarea*. *Moluccella laevis* (Irish bell) seeds contain a major lectin of 130 kDa and a minor lectin of 250 kDa (Lis et al., 1988). Both lectins are GalNAc-specific and consist of three subunits: a 67-kDa subunit composed of two disulfide-bound polypeptides of 28 and 46 kDa, respectively, and two noncovalently linked subunits of 46 and 42 kDa. Seeds of *Salvia sclarea* (fetid clary sage) contain a GalNAc-specific lectin consisting of two identical disulfide-linked glycosylated subunits of 35 kDa. As no sequence information is available, it is not clear

whether the *Salvia* and *Moluccella* lectins share sequence similarity with each other or with other plant lectins.

XII. CONCLUSIONS

A comparative analysis of the molecular, biochemical, and structural data indicates that most of the currently known plant lectins can be classified in four large and three small families of structurally and evolutionary related proteins. In general, most members of a given lectin family consist of protomers with a similar amino acid sequence and overall three-dimensional fold. Differences in posttranslational processing and degree of oligomerization can give rise to different molecular forms of the native lectins. The chitin-binding lectins and the family of type 2 RIP and related lectins clearly differ in this respect because they comprise both chimero- and hololectins.

Most lectin families are homogeneous with respect to the overall specificity of their individual members. The only exceptions are the jacalin-related lectins, which comprise mannose and galactose-binding proteins, and the legume lectin family, which apparently covers a very broad range of specificities.

Lectins are divided in two groups according to their biosynthesis, processing, and topogenesis. Whereas the amaranthins, Cucurbitaceae phloem lectins, and the mannose-specific jacalin-related lectins are synthesized in the cytoplasm and undergo no or a very limited posttranslational modification, all other lectins follow the secretory pathway and are subject to co- and in many cases also posttranslational processing. Most plant lectins are encoded by either one or a few genes except the monocot mannose-binding lectins, which are under the control of extended gene families. Presently, introns have only been documented for the chitin-binding lectins, the Cucurbitaceae phloem lectins, and the jacalin-related lectins.

The seven lectin families are clearly different in horizontal and vertical evolutions. A given lectin family can be confined to a single plant family or a few taxonomically related families or may be widespread among the higher plants. Similarly, some lectin families developed only one type of protomer, whereas others comprise a wide range of structurally different subunits. A closer examination of the lectin structure also indicates that the chitin-binding lectins, the monocot mannose-binding lectins, the type 2 RIP, and the amaranthins consist of repetitive (sub)domains, suggesting that duplication and in tandem insertion of putative ancestral sugar-binding polypeptides of 30 to 45 amino acid residues played an important role in their molecular evolution. In addition to self-duplications, gene fusion of sugar-binding (sub)domains to structurally unrelated or enzymatic domains played an important role in the molecular evolution of especially the chimeric lectins. At several occasions lectins also evolved into proteins devoid of carbohydrate-binding activity. The occurrence of such lectin-related proteins, which has been documented for legume lectins, monocot mannose-binding lectins, and type 2 RIP, illustrates that loss of functionality is not a rare evolutionary event.

Structural, evolutionary, and functional data suggest that the different lectin families have a different physiological role. Most probably, all lectins arose from an ancestor with a highly specific recognition (and perhaps also signaling) function. During evolution of a lectin family, the original specific role has either been lost or conserved in only part of the modern lectins. All other members of the family are the result of an evolution to a less specific storage or defense or storage/defense function.

ACKNOWLEDGMENTS

Supported in part by grants from the Catholic University of Leuven (OT/90/19 and OT/94/17), CNRS, and the Conseil Régional de Midi-Pyrénées, and the Fund for Scientific Research-Flanders (grants 2.0046.93 and G.0223.97). W. P. is Research Director and E. V. D. Postdoctoral Fellow of this fund. We are grateful to Prof. L. Faye and Dr P. Lerouge (Université de Rouen, France) for providing information on the *N*-glycan structures of lectins.

REFERENCES

Adar, R. and Sharon, N. 1996. Mutational studies of the amino acid residues in the combining site of *Erythrina corallodendron* lectin. *FEBS Lett.* **239**: 668–674.

Agrawal, B. B. L. and Goldstein, I. J. 1968. Protein-carbohydrate interaction. VI. Isolation of concanavalin A by specific adsorption on cross-linked dextran gels. *Biochim. Biophys. Acta* **147**: 262–271.

Allen, A. K. 1979. A lectin from the exudate of the fruit of the vegetable marrow (*Cucurbita pepo*) that has a specificity for β-1,4–linked *N*-acetylglucosamine oligosaccharides. *Biochem. J.* **183**: 133–137.

Allen, A. K., Bolwell, G. P., Brown, D. S., Sidebottom, C., and Slabas, A. R. 1996. Potato lectin: a three-domain glycoprotein with novel hydroxyproline-containing sequences and sequence similarities to wheat-germ agglutinin. *Int. J. Biochem. Cell Biol*. **28**: 1285–1291.

Anantharamn V., Patanjali, S. R., Swamy, M. J., Sanadi, A. R., Goldstein, I. J., and Surolia, A. 1986. Isolation, macromolecular properties, and combining site of a chito-oligosaccharide-specific lectin from the exudate of ridge gourd (*Luffa acutangula*). *J. Biol. Chem.* **261**: 14621–14627.

Andersen, N. H., Cao, B., Rodriguez-Romero, A., and Arreguin, B. 1993. Hevein: NMR assignment and assessment of solution-state folding for the agglutinin-toxin motif. *Biochemistry* **32**: 1407–1422.

Asensio, J. L., Cañada, F. J., Bruix, M., Rodriguez-Romero, A., and Jimenez-Barbero, J. 1995. The interaction of hevein with *N*-acetyl-glucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose. *Eur. J. Biochem.* **230**: 621–633.

Aub, J. C., Tieslau, C., and Lankester, A. 1963. Reactions of normal and tumor cell surfaces to enzymes. I. Wheat-germ lipase and associated mucopolysaccharides. *Proc. Natl. Acad. Sci. USA* **50**: 613–619.

Baba, K., Ogawa, M., Nagano, A., Kuroda, H., and Sumiya, K. 1991. Developmental changes in the bark lectin of *Sophora japonica* L. *Planta* **183**: 462–470.

Balzarini, J., Schols, D., Neyts, J., Van Damme, E., Peumans, W., and De Clercq, E. 1991. α-(1–3)- and α-(1–6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. *Antimicrob. Agents Chemother.* **35**: 410–416.

Balzarini, J., Neyts, J., Schols, D., Hosoya, M., Van Damme, E., Peumans, W., and De Clercq, E. 1992. The mannose-specific plant lectins from *Cymbidium* hybrid and *Epipactis helleborine* and the $(N\text{-Acetylglucosamine})_n$-specific plant lectin from *Urtica dioica* are potent and selective inhibitors of human immunodeficiency virus and cytomegalovirus replication in vitro. *Antiviral Res.* **18**: 191–207.

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A., and Vijayan, M. 1994. Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 227–231.

Barbieri, L., Falasca, A., Franceschi, C., Licastro, F., Rossi, C. A., and Stirpe, F. 1983. Purification and properties of two lectins from the latex of the euphorbiaceous species *Hura crepitans* L. (sand-box tree) and *Euphorbia characias* L. (Mediterranean spurge). *Biochem. J.* **215**: 433–439.

Barbieri, L., Batelli, G. B., and Stirpe, F. 1993. Ribosome-inactivating proteins from plants. *Biochim. Biophys. Acta* **1154**: 237–282.

Barbieri, L., Valbonesi, P., Gorini, P., Pession, A., and Stirpe, F. 1996. Polynucleotide adenosine glycosidase activity of saporin-L1: effect on DNA, RNA and poly(A). *Biochem. J.* **319**: 507–513.

Barondes, S. H. 1988. Bifunctional properties of lectins: lectins redefined. *Trends Biochem. Sci.* **13**: 480–482.

Barre, A., Van Damme, E. J. M., Peumans, W. J., and Rougé, P. 1996. Structure-function relationship of monocot mannose-binding lectins. *Plant Physiol.* **112**: 1531–1540.

Barre, A., Van Damme, E. J. M., Peumans, W. J., and Rougé, P. 1997. Curculin, a sweet-tasting and taste-modifying protein, is a non functional mannose-binding lectin. *Plant. Mol. Biol.* **33**: 691–698.

Bausch, J. N. and Poretz, R. D. 1977. Purification and properties of the hemagglutinin from *Maclura pomifera* seeds. *Biochemistry* **16**: 5790–5794.

Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A., and Edelman, G. M. 1975. The covalent and three-dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides. *J. Biol. Chem.* **250**: 1513–1524.

Bednarek, S. Y., Wilkins, T. A., Dombrowski, J. E., and Raikhel, N. Y. 1990. A carboxy-terminal propeptide is necessary for proper sorting of barley lectin to vacuoles of tobacco. *Plant Cell* **2**: 1145–1155.

Bednarek, S. Y. and Raikhel, N. Y. 1991. The barley lectin carboxy-terminal propeptide is a vacuolar protein sorting determinant in plants. *Plant Cell* **3**: 1195–1206.

Beintema, J. J. and Peumans, W. J. 1992. The primary structure of stinging nettle (*Urtica dioica*) agglutinin: a two-domain member of the hevein family. *FEBS Lett.* **299**: 131–134.

Beintema, J. J. 1994. Structural features of plant chitinases and chitin-binding proteins. *FEBS Lett.* **350**: 159–163.

Bird, G. W. G. 1954. Observations on the interactions of the erythrocytes of various species with certain seed agglutinins. *Br. J. Exp. Pathol.* **35**: 252.

Blasco, E., Ngoc, L. D., Aucouturier, P., Preudhomme, J. L., and Barra, A. 1996. Mitogenic activity of new lectins from seeds of wild *Artocarpus* species from Vietnam. *Life Sci.* **319**: 405–409.

Bompard-Gilles, C., Rousseau, P., Rougé, P., and Payan, F. 1996. Structure of the pig pancreatic α-amylase in complex with the bean *Phaseolus vulgaris* inhibitor. Substrate mimicry in the active center of a mammalian α-amylase. *Structure* **4**: 1441–1452.

Bostwick, D. E., Dannehofer, J. M., Skaggs, M. I., Lister, R. M., Larkins, B. A., and Thompson, G. A. 1992. Pumpkin phloem lectin genes are specifically expressed in companion cells. *Plant Cell* **4**: 1539–1548.

Bostwick, D. E., Skaggs, M. I., and Thompson G. A. 1994. Organization and characterization of *Cucurbita* phloem lectin genes. *Plant Mol. Biol.* **26**: 887–897.

Boulter, D., Edwards, G. A., Gatehouse, A. M. R., Gatehouse, J. A., and Hilder, V. A. 1990. Additive protective effects of different plant-derived insect resistance genes in transgenic tobacco plants. *Crop Protect.* **9**: 351–354.

Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rougé, P., and Fontecilla-Camps, J. C. 1990a. X-ray crystal structure determination and refinement at 1.9 Å resolution of isolectin I from the seeds of *Lathyrus ochrus*. *J. Mol. Biol.* **214**: 571–584.

Bourne, Y., Roussel, A., Frey, M., Rougé, P., Fontecilla-Camps, J. C., and Cambillau, C. 1990b. Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: fine specificity of the monosaccharide-binding site. *Proteins Struct. Funct. Genet.* **8**: 365–376.

Bourne, Y., Mazurier, J., Legrand, D., Rougé, P., Montreuil, J., Spik, G., and Cambillau, C. 1994. Interaction of a legume lectin with the human lactotransferrin N2 fragment or with the isolated biantennary glycopeptide: role of the fucose moiety. *Structure* **2**: 209–219.

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.

Boyd, W. C. and Reguera, R. M. 1949. Studies on haemagglutinins present in seeds of some representatives of the family Leguminoseae. *J. Immunol.* **62**: 333–339.

Brewin, N. J. and Kardailsky, I. V. 1997. Legume lectins and nodulation by *Rhizobium. Trends Plant. Sci.* **2**: 92–98.

Broekaert, W. F. and Peumans, W. J. 1986. Lectin release from seeds of *Datura stramonium* and interference of the *Datura stramonium* lectin with bacterial motility. **In***: Lectins, Biology, Biochemistry, Clinical Biochemistry*. Vol. 5., pp. 57–66, Bøg-Hansen, T. C. and Van Driessche, E., Eds., W. De Gruyter, Berlin.

Broekaert, W. F., Allen, A. K., and Peumans, W. J. 1987. Separation and partial characterization of isolectins with different subunit compositions from *Datura stramonium* seeds. *FEBS Lett.* **220**: 116–120.

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.

Broekaert, W. F., Lee, H. -I., Kush, A., Chua, N.-H., and Raikhel, N. V. 1990. Wound-induced accumulation of mRNA containing a hevein sequence in laticifers of rubber tree (*Hevea brasiliensis*). *Proc. Natl. Acad. Sci. USA* **87**: 7633–7637.

Broekaert, W. F., Marien, W., Terras, F. R. G., De Bolle, M. F. C., Proost, P., Van Damme, J., Dillen, L., Claeys, M., Rees, S. B., Vanderleyden, J., and Cammue, B. P. A. 1992. Antimicrobial peptides from *Amaranthus caudatus* seeds with sequence homology to the cysteine/glycine-rich domain of chitin-binding proteins. *Biochemistry* **31**: 4308–4314.

Burger, M. M. and Goldberg, A. R. 1967. Identification of a tumor-specific determinant on neoplastic cell surfaces. *Proc. Natl. Acad. Sci. USA* **57**: 359–366.

Cammue, B., Stinissen, H. M., and Peumans, W. J. 1985. A new type of cereal lectin from leaves of couch grass (*Agropyrum repens* L.). *Eur. J. Biochem* **148**: 315–322.

Cammue, B. P. A., Peeters, B., and Peumans, W. J. 1986. A new lectin from tulip (*Tulipa*) bulbs. *Planta* **169**: 583–588.

Carrington, D. M., Auffret, A., and Hanke, D. A. 1985. Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* **313**: 64–67.

Chaddock, J. A. and Roberts, L. M. 1993. Mutagenesis and kinetic analysis of the active site Glu177 of ricin A-chain. *Protein Eng.* **6**: 425–431.

Chrispeels, M. J. and Raikhel, N. V. 1991. Lectins, lectin genes, and their role in plant defense. *Plant Cell* **3**: 1–9.

Clendennen, S. K. and May, G. D. 1997. Differential gene expression in ripening banana fruit. *Plant Physiol.* **115**: 463–469.

Collinge, D. B., Kragh, K. M., Mikkelsen, J. D., Nielsen, K. K., Rasmussen, U., and Vad, K. 1993. Plant chitinases. *Plant J.* **3**: 31–40.

Czapla, T. H. and Lang, B. A. 1990. Effect of plant lectins on the larval development of European corn borer (Lepidoptera: Pyralidae) and southern corn rootworm (Coleoptera:Chrysomelidae). *J. Econ. Entomol.* **83**: 2480–2485.

Datta, P. K., Figueroa, M. O. R., and Lajolo, F. M. 1991. Purification and characterization of two major lectins from *Araucaria brasiliensis* syn *Araucaria angustifolia* seeds (pinhao). *Plant Physiol.* **97**: 856–862.

De Bolle, M. F. C., David, K. M. M., Rees, S. B., Vanderleyden, J., Cammue, B. P. A., and Broekaert, W. F. 1993. Cloning and characterization of a cDNA encoding an antimicrobial chitin-binding protein from amaranth, *Amaranthus caudatus*. *Plant Mol. Biol.* **22**: 1187–1190.

Debray, H., Decout, D., Strecker, G., Spik, G., and Montreuil, J. 1981. Specificity of twelve lectins toward oligosaccharides and glycopeptides related to *N*-glycosylproteins. *Eur. J. Biochem.* **117**: 41–55.

Debray, H. and Rougé, P. 1984. The fine sugar specificity of the *Lathyrus ochrus* seed lectin and isolectins. *FEBS Lett.* **176**: 120–124.

Debray, H., Wieruszeski, J. M., Strecker, G., and Franz, H. 1992. Structural analysis of the car-

bohydrate chains isolated from mistletoe (*Viscum album*) lectin I. *Carbohydrate Res.* **236**: 135–143.

Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Pearlstone, J. R., Carpenter, M. R., Smillie, L. B., Nikrad, P. V., Spoohr, U., and Lemieux, R. U. 1990. Molecular recognition of a human blood group determinant by a plant lectin. *Can. J. Chem.* **68**: 1116–1121.

Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb (Gilboa), A. J., Dodson, E. J., Papiz, M. Z., Wan, T., and Campbell, J. 1989. The structure of the saccharide-binding site of concanavalin A. *EMBO J.* **8**: 2189–2193.

Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F., and Sacchettini, J. C. 1995. X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. *Biochemistry* **34**: 4933–4942.

Diaz, C., Melchers, L. S., Hooykaas, P. J. J., Lugtenberg, B. J. J., and Kijne, J. W. 1989. Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. *Nature* **338**: 579–581.

Dixon, H. B. F. 1981. Defining a lectin. *Nature* **338**: 192.

Dorland, L., van Halbeek, H., Vliegenthart, J. F. G., Lis, H., and Sharon, N. 1981. Primary structure of the carbohydrate chain of soybean agglutinin. *J. Biol. Chem.* **256**: 7708–7711.

Drickamer, K. 1995. Multiplicity of carbohydrate interactions. *Nature Struct. Biol.* **2**: 437–439.

Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J., and Wang, J. L. 1972. The covalent and three-dimensional structure of concanavalin A. *Proc. Natl. Acad. Sci. U.S.A.* **69**: 2580–2584.

Edelman, G. M. and Wang, J. L. 1978. Binding and functional properties of concanavalin A and its derivatives. III. Interactions with indoleactetic acid and other hydrophobic ligands. *J. Biol. Chem.* **253**: 3016–3022.

Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E., and Suddath, F. L. 1986. The crystal structure of pea lectin at 3.0-Å resolution. *J. Biol. Chem.* **261**: 16518–16527.

Elfstrand, M. 1898. Über blutkörperchenagglutinierende Eiweisse. **In**: *Görberdorfer Veröffentlichungen a. Band I.* pp. 1–159. Kobert, R., Ed., Enke, Stuttgart, Germany.

Endo, Y., Mitsui, K., Motizuki, M., and Tsurugi, K. 1987. The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28S ribosomal RNA caused by the toxins. *J. Biol. Chem.* **262**: 5908–5912.

Endo, Y., Tsurgi, K., and Lambert, J. M. 1988. The site of action of six different ribosome-inactivating proteins from plants on eukaryotic ribosomes: the RNA N-glycosidase activity of the proteins. *Biochem. Biophys. Res. Commun.* **150**: 1032–1036.

Etzler, M. E. 1986. Distribution and function of plant lectins. **In**: *The Lectins: Properties, Functions, and Applications in Biology and Medicine*. pp. 371–435. Liener, I. E., Sharon, N., and Goldstein, I. J., Eds., Academic Press, Orlando, FL.

Etzler, M. E., Murphy, J. B., Ewing, N. N., Kalsi, G., Stacey, G., Day, B. D., and Roberts, N. J. 1997. A NOD factor-binding root lectin/nucleotide phophatase: a possible receptor/signal transducer in the *Rhizobium*-legume symbiosis. *Glycoconjugate J.* **14**: S27.

Fabre, C., Causse, H., Mourey, L., Koninkx, J., Rivière, M., Hendriks, H., Puzo, G., Samama, J. P., and Rougé, P. 1998. Characterization and sugar binding properties of arcelin-1, an insecticidal lectin-like protein isolated from kidney bean (*Phaseolus vulgaris* L. cv RAZ-2) seeds. *Biochem. J.* **329**: 551–560.

Falasca, A. I., Abbondanza, A., Barbieri, L., Bolognesi, A., Rossi, C. A., and Stirpe, F. 1989. Purification and partial characterization of a lectin from the seeds of *Trichosanthes kirilowii* Maximowicz. *FEBS Lett.* **246**: 159–162.

Feldsted, R. L., Egorin, M. J., Leavitt, R. D., and Bachur, N. R. 1977. Recombinations of subunits of *Phaseolus vulgaris* isolectins. *J. Biol. Chem.* **252**: 2967–2971.

Foxwell, B. M. J., Donovan, T. A., Thorpe, P. E., and Wilson, G. 1985. The removal of carbohydrates from ricin with endoglycosidases H, F and D and α-mannosidase. *Biochim. Biophys. Acta* **840**: 193–203.

Frankel, A. E., Burbage, C., Fu, T., Tagge, E., Chandler, J., and Willingham, M. C. 1996. Ricin toxin contains at least three galactose-binding sites located in B chain subdomains 1 alpha, 1 beta and 2 gamma. *Biochemistry* **35**: 14749–14756.

Franz, H., Ziska, P., and Kindt, A. 1981. Isolation and properties of three lectins from mistletoe. *Biochem. J*. **195**: 481–484.

Fu, T., Burbage, C., Tagge, E. P., Brothers, T., Willingham, M. C., and Frankel, A. E. 1996. Ricin toxin contains three lectin sites which contribute to its in vivo toxicity. *Int. J. Immunopharmacol*. **18**: 685–692.

Gatehouse, A. M. R., Barbieri, L., Stirpe, F., and Croy, R. R. D. 1990. Effects of ribosome inactivating proteins on insect development — differences between Lepidoptera and Coleoptera. *Entomol. Exp. Appl*. **54**: 43–51.

Gatehouse, A. M. R., Powell, K. S., Peumans, W. J., Van Damme, E. J. M., and Gatehouse, J. A. 1995. Insecticidal properties of plant lectins: their potential in plant protection. **In**: *Lectins: Biomedical Perspectives.* pp. 35–57. Pusztai, A. and Bardocz, S., Eds., Taylor and Francis, London, UK.

Girdol, X., Chrestin, H., Tan, H. L., and Kush, A. 1994. Hevein, a lectin-like protein from *Hevea brasiliensis* (rubber tree) is involved in the coagulation of latex. *J. Biol. Chem*. **269**: 9278–9283.

Goldberg, R. B., Hoschek, G., and Vodkin, L. O. 1983. An insertion sequence blocks the expression of a soybean lectin gene. *Cell* **33**: 465–475.

Goldstein, I. J., Hughes, R. C., Monsigny, M., Osawa, T., and Sharon, N. 1980. What should be called a lectin? *Nature* **285**: 66.

Goldstein, I. J. and Poretz, R. D. 1986. Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. **In:** *The Lectins, Properties, Functions, and Applications in Biology and Medicine.* pp. 33–247. Liener, I. E., Sharon, N., and Goldstein, I. J., Eds., Academic Press, Orlando, FL.

Goossens, A., Geremia, R., Bauw, G., Van Montagu, M., and Angenon, G. 1994. Isolation and characterization of arcelin-5 proteins and cDNAs. *Eur. J. Biochem*. **225**: 787–795.

Habibi, J., Backus, E. A., and Czapla, T. H. 1993. Plant lectins affect survival of the potato leafhopper (Homoptera: Cicadellidae). *J. Econ. Entomol*. **86**: 945–951.

Hamelryck, T. W., Dao, M.-H., Poortmans, F., Chrispeels, M. J., Wyns, L., and Loris, R. 1996a. The crystallographic structure of phytohemagglutinin-L. *J. Biol. Chem*. **271**: 20479–20485.

Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Van Montagu, M., Wyns, L., and Loris, R. 1996b. Crystal structure of arcelin-5, a lectin-like defense protein from *Phaseolus vulgaris*. *J. Biol. Chem.* **271**: 32796–32802.

Hardman, K. D. and Ainsworth, C. F. 1972. Structure of concanavalin A at 2.4 Å resolution. *Biochemistry* **11**: 4910–4919.

Hatakeyama, T., Yamasaki, N., and Funatsu, G. 1986. Identification of the tryptophan residue located at the low-affinity saccharide binding site of ricin D. *J. Biochem.* **100**: 781–788.

Hervé, C., Dabos, P., Gallaud, J. P., Rougé, P., and Lescure, B. 1996. Characterization of an *Arabidopsis thaliana* gene that defines a new class of plant receptor kinases with an extracellular lectin-like domain. *J. Mol. Biol.* **258**: 778–788.

Hester, G., Kaku, H., Goldstein, I. J., and Wright, C. S. 1995. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nature Struct. Biol*. **2**: 472–479.

Hester, G. and Wright, C. S. 1996. The mannose-specific bulb lectin from *Galanthus nivalis* (snowdrop) binds mono- and dimannosides at distinct sites. Structure analysis of refined complexes at 2.3 Å and 3.0 Å resolution. *J. Mol. Biol.* **262**: 516–531.

Higgins, T. J. V., Chandler, P. M., Zurawski, G., Button, S. C., and Spencer, D. 1983. The biosynthesis and primary structure of pea seed lectin. *J. Biol. Chem*. **258**: 9544–9549.

Hilder, V. A., Powell, K. S., Gatehouse, A. M. R., Gatehouse, J. A., Gatehouse, L. N., Shi, Y., Hamilton, W. D. O., Merryweather, A., Newell, C., Timans, J. C., Peumans, W. J., Van Damme,

E. J. M., and Boulter, D. 1995. Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids. *Transgenic Res.* **4**: 18–25.

Hirai, M., Nakamura, K., Imai, T., and Sato, T. 1993. cDNAs encoding for storage proteins in the tubers of taro (*Colocasia esculenta* Schott.). *Jpn. J. Genet.* **68**: 229–236.

Hoedemaeker, F. J., Richardson, M., Diaz, C. L., de Pater, B. S., and Keijne, J. 1994. Pea (*Pisum sativum* L.) seed isolectins 1 and 2 and pea root lectin result from carboxypeptidase-like processing of a single gene product. *Plant Mol. Biol*. **24**: 75–81.

Hoffman, L. M., Ma, Y., and Barker, R.F. 1982. Molecular cloning of *Phaseolus vulgaris* lectin mRNA and use of cDNA as a probe to estimate lectin transcript levels in various tissues. *Nucl. Acids Res.* **10**: 7819–7828.

Hoffman, L. M. and Donaldson, D. D. 1985. Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. *EMBO J.* **4**: 883–889.

Hom, K., Gochin, M., Peumans, W. J., and Shine, N. 1995. Ligand-induced perturbations in *Urtica dioica* agglutinin. *FEBS Lett*. **361**: 157–161.

Hossaini, A. 1968. Hemolytic and hemagglutinating activities of 222 plants. *Vox Sang.* **15**: 410–417.

Huesing, J. E., Murdock, L. L., and Shade, R. E. 1991a. Effect of wheat germ isolectins on development of cowpea weevil. *Phytochemistry* **30**: 785–788.

Huesing, J. E., Murdock, L. L., and Shade, R. E. 1991b. Rice and stinging nettle lectin: insecticidal activity similar to wheat germ agglutinin. *Phytochemistry* **30**: 3565–3568.

Huesing, J. E., Shade, R. E., Chrispeels, M. J., and Murdock, L. L. 1991c. α-amylase inhibitor, not phytohemagglutinin, explains resistance of common bean seeds to cowpea weevil. *Plant Physiol.* **96**: 993–996.

Hung, C. H., Lee, M. C., Lee, T. C., and Lin, J. Y. 1993. Primary structure of three isoabrins determined by cDNA sequencing. *J. Mol. Biol*. **229**: 263–267.

Imberty, A., Casset, F., Gegg, C. V., Etzler, M. E., and Perez, S. 1994. Molecular modeling of the *Dolichos biflorus* seed lectin and its specific interactions with carbohydrates: α-D-*N*-acetylgalactosamine, Forssman disaccharide and blood group A trisaccharide. *Glycoconjugate J*. **11**: 400–413.

Ishiguro, M., Mitarai, M., Harade, H., Sekine, I., Nishimori, T., and Kikutani, M. 1983. Biochemical studies on the oral toxicity of ricin. I. Ricin administration orally can impair sugar absorption by the rat small intestine. *Chem. Pharm. Bull*. **32**: 3222–3227.

Ishimoto, M. and Kitamura, K. 1989. Growth inhibitory effects of an α-amylase inhibitor from kidney bean, *Phaseolus vulgaris* (L.) on three species of bruchids (Coleoptera: Bruchidae). *Appl. Ent. Zool*. **24**: 281–286.

Jankovic, M., Cuperlovic, M., and Hajdulovic, L. 1990. Salt-soluble lectins of corn grain. *Plant Physiol*. **93**: 1659–1662.

Jones, J. M., Cawley, L. P., and Teresa, G. W. 1967. Hemagglutinins (lectins) extracted from *Maclura pomifera*. *Vox Sang.* **12**: 211–214.

Jung, E., Fucini, P., Stewart, M., Noegel, A. A., and Schleicher, M. 1996. Linking microfilaments to intracellular membranes: the actin-binding and vesicle associated protein comitin exhibits a mannose-specific lectin activity. *EMBO. J.* **15**: 1238–1246.

Kabir, S., Aebersold, R., and Daar, A. S. 1993. Identification of a novel 4 kDa immunoglobulin -A-binding peptide obtained by the limited proteolysis of jacalin. *Biochim. Biophys. Acta* **1161**: 194–200.

Kabir, S. and Daar, A. S. 1994. The composition and properties of jacalin, a lectin of diverse applications obtained from the jackfruit (*Artocarpus heterophyllus*) seeds. *Immunol. Invest*. **23**: 167–188.

Kaku, H., Peumans, W. J., and Goldstein, I. J. 1990a. Isolation and characterization of a second lectin (SNA-II) present in elderberry (*Sambucus nigra* L.) bark. *Arch. Biochem. Biophys.* **277**: 255–262.

Kaku, H., Van Damme, E. J. M., Peumans, W. J., and Goldstein, I. J. 1990b. Carbohydrate-binding specificity of the daffodil (*Narcissus pseudonarcissus*) and amaryllis (*Hippeastrum* hybr.) bulb lectins. *Arch. Biochem. Biophys*. **279**: 298–304.

Kaku, H., Tanaka, Y., Tazaki, K., Minami, E., Mizuno, H., and Shibuya, N. 1996. Sialylated oligosaccharide-specific plant lectin from japanese elderberry (*Sambucus sieboldiana*) bark tissue has a homologous structure to type II ribosome-inactivating proteins, ricin and abrin. *J. Biol. Chem.* **271**: 1480–1485.

Kaminski, P. A., Buffard, D., and Strosberg, A. D. 1987. The pea lectin gene family contains only one functional gene. *Plant Mol. Biol*. **9**: 497–507.

Kataoka, J., Miyano, M., Habuka, N., Masuta, C., and Koiwai, A. 1993. A genomic gene for MAP, a ribosome-inactivating protein from *Mirabilis jalapa*, contains an intron. *Nucleic Acids Res*. **21**: 1035.

Katzin, B. J., Collins, E. J., and Robertus, J. D. 1991. Structure of Ricin A-chain at 2.5 Å. *Proteins Struct. Funct. Genet.* **10**: 251–259.

Kieliszewski, M. J., Showalter, A. M., and Leykam, J. F. 1994. Potato lectin: a modular protein sharing sequence similarities with the extensin family, the hevein lectin family, and snake venom disintegrins (platelet aggregation inhibitors). *Plant J.* **5**: 849–861.

Kim, Y. and Robertus, J. D. 1992. Analysis of several key active site residues of ricin A chain by mutagenesis and X-ray crystallography. *Protein Engin*. **5**: 775–779.

Kimura, Y., Hase, S., Kobayashi, Y., Kyogoku, Y., Ikenaka, T., and Funatsu, G. 1988. Structures of sugar chains of ricin D. *J. Biochem*. **103**: 944–949.

Kocourek, J. and Horejsi, V. 1983. A note on the recent discussion on definition of the term 'lectin'. **In**: *Lectins: Biology, Biochemistry and Clinical Biochemistry*. Vol. 3, pp. 3–6. Bøg-Hansen, T. C. and Spengler, G. A., Eds., Walter de Gruyter, Berlin, Germany.

Kocourek, J. 1986. Historical background. **In**: *The Lectins, Properties, Functions, and Applications in Biology and Medicine.* pp. 1–32. Liener, I. E., Sharon, N., and Goldstein, I. J., Eds., Academic Press, Orlando, FL.

Koeppe S. J. and Rupnov, J. H. 1988. Purification and characterization of a lectin from the seeds of amaranth (*Amaranthus cruentus*). *J. Food Sci*. **53**: 1412–1417.

Koike, T., Titani, K., Suzuki, M., Beppu, H., Kuzuya, H., Maruta, K., Shimpo, K., and Fujita, K. 1995. The complete amino acid sequence of a mannose-binding lectin from "Kidachi Aloe" (*Aloe arborescens* var. Natalensis Berger). *Biochem. Biophys. Res. Commun.* **214**: 163–170.

Kornfeld, K., Reitman, M. L., and Kornfeld, R. 1981. The carbohydrate-binding specificity of pea and lentil lectins. Fucose is an important determinant. *J. Biol. Chem*. **256**: 6633–6640.

Kraulis, P. J. 1991. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst*. **24**: 946–950.

Kumar, A. M., Timm, D. E., Neet, K. E., Owen, W. G., Peumans, W. J., and Rao, A. J. 1993. Characterization of the lectin from the bulbs of *Eranthis hyemalis* (winter aconite) as an inhibitor of protein synthesis. *J. Biol. Chem*. **268**: 25176–25183.

Ladin, B. F., Murray, E. E., Halling, A. C., Halling, K. C., Tilakaratne, N., Long, G. L., Houston, L. L., and Weaver, R. G. 1987. Characterisation of a cDNA encoding ricin E, a hybrid ricin-*Ricinus communis* agglutinin gene from the castor plant *Ricinus communis*. *Plant Mol. Biol*. **9**: 287–295.

Lamb, F. I., Roberts, L. M., and Lord, J. M. 1985. Nucleotide sequence of cloned cDNA coding for preproricin. *Eur. J. Biochem.* **148**: 265–270.

Landsteiner, K. and Raubitschek, H. 1907. Beobachtungen über Hämolyse und Hämagglutination. *Zentralbl. Bakteriol. Parasitenk. Infektionskr. Hyg. Abt. 1: Orig*. **45**: 660–667.

Lee, H.-I., Broekaert, W. F., and Raikhel, N. V. 1991. Co- and post-translational processing of the hevein preproprotein of latex of the rubber tree (*Hevea brasiliensis*). *J. Biol. Chem.* **266**: 15944–15948.

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

Lerner, D. R. and Raikhel, N. V. 1992. The gene for stinging nettle lectin (*Urtica dioica* agglutinin) encodes both a lectin and a chitinase. *J. Biol. Chem.* **267**: 11085–11091.

LeVine, D., Kaplan, M. J., and Greenaway, P. J. 1972. The purification and characterization of wheat-germ agglutinin. *Biochem. J*. **129**: 847–856.

Lin, J.-Y., Tserng, K.-Y., Chen, C.-C., Lin, L.-T., and Tung, T.-C. 1970. Abrin and ricin: new anti-tumour substances. *Nature* **227**: 292–293.

Lin, T. T. -S. and Li, S. S. -L. 1980. Purification and physicochemical properties of ricins and agglutinins from *Ricinus communis*. *Eur. J. Biochem*. **105**: 453–459.

Linthorst, H. J. M. 1991. Pathogenesis-related proteins of plants. *Crit. Rev. Plant Sci*. **10**: 123–150.

Lis, H., Latter, H., Adar, R., and Sharon, N. 1988. Isolation of two blood group type A and N specific isolectins from *Moluccella laevis* seeds. *FEBS Lett*. **233**: 191–195.

Lord, J. M. 1985. Precursors of ricin and *Ricinus communis* agglutinin. Glycosylation and processing during intracellular transport. *Eur. J. Biochem*. **146**: 411–416.

Lord, J. M., Roberts, L. M., and Robertus, J. D. 1994. Ricin: structure, mode of action, and some current applications. *FASEB J*. **8**: 201–208.

Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R., and Wyns, L. 1993. Crystal structure determination and refinement at 2.3–Å resolution of the lentil lectin. *Biochemistry* **32**: 9772–8781.

Maddison, W. P. and Maddison, D. R. 1992. *MacClade: Analysis of phylogeny and character evolution*, Version 3.0, Sinauer Associates, Sunderland, Massachusetts (USA).

Mandal, D. K., Nieves, E., Bhattacharyya, L., Orr, G. A., Roboz, J., Yu, Q., and Brewer, C. F. 1994. Purification and characterization of three isolectins of soybean agglutinin. Evidence for C-terminal truncation by electrospray ionization mass spectrometry. *Eur. J. Biochem.* **221**: 547–553.

Martins, J. C., Maes, D., Loris, R., Pepermans, H. A. M., Wyns, L., Willem, R., and Verheyden, P. 1996. ^{1}H NMR study of the solution structure of Ac-AMP2, a sugar binding antimicrobial protein isolated from *Amaranthus caudatus*. *J. Mol. Biol*. **258**: 322–333.

Mirkov, T. E. and Chrispeels, M. J. 1993. Mutation of Asn^{128} to Asp of *Phaseolus vulgaris* leucoagglutinin (PHA-L) eliminates carbohydrate-binding and biological activity. *Glycobiology* **3**: 581–587.

Mirkov, T. E., Wahlstrom, J. M., Hagiwara, K., Finardi-Filho, F., Kjemtrup, S., and Chrispeels, M. J. 1994. Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-α-amylase inhibitor family of the common bean and its relatives. *Plant Mol. Biol.* **26**: 1103–1113.

Montfort, W., Villafranca, J. E., Monzingo, A. F., Ernst, S., Katzin, B., Rutenber, E., Xuong, N. H., Hamlin, R., and Robertus, J. D. 1987. The three-dimensional structure of ricin at 2.8 Å. *J. Biol. Chem*. **262**: 5398–5403.

Moreira, R. A. and Ainouz, I. L. 1981. Lectins from seeds of jack fruit (*Artocarpus integrifolia* L.): Isolation and purification of two isolectins from the albumin fraction. *Bio. Plant (Praha)* **23**: 186–192.

Moreno, J. and Chrispeels, M. J. 1989. A lectin gene encodes the α-amylase inhibitor of the common bean. *Proc. Natl. Acad. Sci. USA* **86**: 7885–7889.

Mourey, L., Pédelacq, J. D., Birck, C., Fabre, C., Rougé, P., and Samama, J. P. 1998. Crystal structure of the arcelin-1 dimer from *Phaseolus vulgaris* at 1.9 Å resolution. *J. Biol. Chem.*, **273:** 12914–12922.

Murdock, L. L., Huesing, J. E., Nielsen, S. S., Pratt, R. C., and Shade, R. E. 1990. Biological effects of plant lectins on the cowpea weevil. *Phytochemistry* **29**: 85–89.

Murzin, A. G., Lesk, A. M., and Chothia, C. 1992. β-trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors, interleukins-1β, and 1α and fibroblast growth factors. *J. Mol. Biol.* **223**: 531–543.

Musette, P., Galelli, A., Truffa-Bachi, P., Peumans, W. J., Kourilsky, P., and Gachelin, G. 1996. The Jβ segment of the T cell receptor contributes to the Vβ-specific T cell expansion caused by staphylococcal enterotoxin B and *Urtica dioica* superantigens. *Eur. J. Immunol*. **26**: 618–622.

Nachbar, M. S. and Oppenheim, J. D. 1980. Lectins in the United States diet: a survey of lectins in commonly consumed foods and a review of the literature. *Am. J. Clin. Nutr*. **33**: 2338–2345.

Nagai, K and Yamaguchi, H. 1993. Direct demonstration of the essential role of the intramolecular high-mannose oligosaccharide chains in the folding and assembly of soybean (*Glycine max*) lectin polypeptides. *J. Biochem*. **113**: 123–125.

Nagata, Y. and Burger, M. M. 1972. Wheat germ agglutinin. Isolation and crystallization. *J. Biol. Chem*. **247**: 2248–2250.

Neuhaus, J.-M., Sticher, L., Meins, F., Jr., and Boller, T. 1991. A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. *Proc. Natl. Acad. Sci. USA* **88**: 10362–10366.

Nodari, R. O., Tsai, S. M., Gilbertson, R. L., and Gepts, P. 1993. Towards an integrated linkage map of common bean. 2. Development of an RFLP-based linkage map. *Theor. Appl. Genet.* **85**: 513–520.

Nowell, P. C. 1960. Phytohemagglutinin: an initiator of mitosis in cultures of normal human leukocytes. *Cancer Res*. **20**: 462–466.

Nsimba-Lubaki, M., Peumans, W. J., and Carlier, A. R. 1983. Isolation and partial characterization of a lectin from *Euphorbia heterophylla* seeds. *Biochem. J.* **215**: 141–145.

Nsimba-Lubaki, M. and Peumans, W. J. 1986. Seasonal fluctuations of lectins in the barks of elderberry (*Sambucus nigra*) and black locust (*Robinia pseudoacacia*). *Plant Physiol.* **80**: 747–751.

Nsimba-Lubaki, M., Allen, A. K., and Peumans, W. J. 1986. Isolation and partial characterization of latex lectins from three species of the genus *Euphorbia* (Euphorbiaceae). *Physiol. Plant*. **67**: 193–198.

Okamuro, J. K. and Goldberg, R. B. 1992. A mutant lectin gene is rescued from an insertion element that blocks its expression. *Plant Cell* **4**: 1141–1146.

Olsnes, S. 1978. Toxic and nontoxic lectins from *Abrus precatorius*. *Meth. Enzymol*. **50**: 323–330.

Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C., and Bliss, F. A. 1988a. Insecticidal activity and lectin homology of arcelin seed protein. *Science* **240**: 207–210.

Osborn, T. C., Burow, M., and Bliss, F. A. 1988b. Purification and characterization of arcelin seed protein from common bean. *Plant Physiol*. **86**: 399–405.

Osinaga, E., Tello, D., Batthyany, C., Bianchet, M., Tavares, G., Duran, R., Cervenansky, C., Camoin, L., Roseto, A., and Alzari, P. M. 1997. Amino acid sequence and three-dimensional structure of the Tn-specific isolectin B4 from *Vicia villosa*. *FEBS Lett*. **412**: 190–196.

Pak, J. H., Hendrickson, T., and Dobres, M. S. 1992. Predicted sequence and structure of a vegetative lectin in *Pisum sativum*. *Plant Mol. Biol*. **18**: 857–863.

Petryniak, J. and Goldstein, I. J. 1987. *Euonymus europaea* lectin. *Meth. Enzymol*. **138**: 552–561.

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.* pp. 99–116. Goldstein, I. J. and Etzler, M. E., Eds., Alan R. Liss Inc., New York.

Peumans, W. J., De Ley, M., and Broekaert, W. F. 1984a. An unusual lectin from stinging nettle (*Urtica dioica*) rhizomes. *FEBS Lett.* **177**: 99–103.

Peumans, W. J., Nsimba-Lubaki, M., Carlier, A. R., and Van Driessche, E. 1984b. A lectin from *Bryonia* root stocks. *Planta* **160**: 222–228.

Peumans W. J., De Ley, M., Stinissen, H. M., and Broekaert, W. F. 1985a. Isolation and partial characterization of a new lectin from seeds of the greater celandine (*Chelidonium majus*). *Plant Physiol.* **78**: 379–383.

Peumans, W. J., Nsimba-Lubaki, M., Peeters, B., and Broekaert, W. F. 1985b. Isolation and partial characterization of a lectin from *Aegopodium podagraria* rhizomes. *Planta* **164**: 75–82.

Peumans, W. J. and Cammue, B. 1986. Gramineae lectins: a special class of plant lectins. **In**:

Lectins; Biology, Biochemistry, Clinical Biochemistry, Vol. 5, pp. 31–38. Bøg-Hansen, T. C. and Van Driessche, E., Eds., W. De Gruyter, Berlin, Germany.

Peumans, W. J., Allen, A. K., Nsimba-Lubaki, M., and Chrispeels, M. J. 1987. Related glycoprotein lectins from root stocks of wild cucumbers. *Phytochemistry* **26**: 909–912.

Peumans, W. J. and Van Damme, E. J. M. 1995a. The role of lectins in plant defense. *Histochem. J.* **27**: 253–271.

Peumans, W. J. and Van Damme, E. J. M. 1995b. Lectins as plant defense proteins. *Plant Physiol.* **109**: 347–352.

Peumans, W. J., Verhaert, P., Pfüller, U., and Van Damme, E. J. M. 1996. Isolation and partial characterization of a small chitin-binding lectin from mistletoe (*Viscum album*). *FEBS Lett.* **396**: 261–265.

Peumans, W. J., Smeets, K., Van Nerum, K., Van Leuven, F., and Van Damme, E. J. M. 1997a. Lectin and alliinase are the predominant proteins in the nectar from leek (*Allium porrum*) flowers. *Planta* **201**: 298–302.

Peumans, W. J., Winter, H. C., Bemer, V., Van Leuven, F., Goldstein, I. J., Truffa-Bachi, P., and Van Damme, E. J. M. 1997b. Isolation of a novel plant lectin with an unusual specificity from *Calystegia sepium*. *Glycoconjugate J.* **14**: 259–265.

Peumans, W. J. and Van Damme, E. J. M. 1998. Plant lectins: versatile proteins with important perspectives in biotechnology. **In**: *Biotechnology and Genetic Engineering Reviews*, Vol. 15, 199–228.

Peumans, W. J., Roy, S., Barre, A., Rougé, P., Van Leuven, F., and Van Damme, E. J. M. 1998. Elderberry (*Sambucus nigra*) contains truncated Neu5Ac(α-2,6)Gal/GalNAc-binding type 2 ribosome-inactivating proteins. *FEBS Lett*., **425:** 35–39.

Potter, S., Uknes, S., Lawton, K., Winter, A. M., Chandler, D., DiMaio, J., Novitzky, R., Ward, E., and Ryals, J. 1993. Regulation of a hevein-like gene in *Arabidopsis*. *Mol. Plant Microbe Interact*. **6**: 680–685.

Powell, K. S., Gatehouse, A. M. R., Hilder, V. A., and Gatehouse, J. A. 1995a. Antifeedant effects of plant lectins and an enzyme on the adult stage of the rice brown planthopper, *Nilaparvata lugens*. *Entomol. Exp. Appl*. **75**: 51–59.

Powell, K. S., Gatehouse, A. M. R., Hilder, V. A., Van Damme, E. J. M., Peumans, W. J., Boonjawat, J., Horsham, K., and Gatehouse, J. A. 1995b. Different antimetabolic effects of related lectins toward nymphal stages of *Nilaparvata lugens*. *Entomol. Exp. Appl*. **75**: 61–65.

Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M., and Suguna, K. 1998. Carbohydrate specificity and quaternary structure in basic winged bean lectin: X-ray analysis of the lectin at 2.5 Å. *J. Mol. Biol*. **276**: 787–796.

Pueyo, J. J., Hunt, D. C., and Chrispeels, M. J. 1993. Activation of bean (*Phaseolus vulgaris*) α-amylase inhibitor requires proteolytic processing of the proprotein. *Plant Physiol.* **101**: 1341–1348.

Pusztai, A. 1991. *Plant Lectins*. Cambridge University Press, Cambridge, UK.

Pusztai, A., Ewen, S. W. B., Grant, G., Peumans, W. J., Van Damme, E. J. M., Rubio, L., and Bardocz, S. 1990. The relationship between survival and binding of plant lectins during small intestinal passage and their effectiveness as growth factors. *Digestion* **46**: 308–316.

Pusztai, A. and Bardocz, S. 1996. Biological effects of plant lectins on the gastrointestinal tract: metabolic consequences and applications. *Trends Glycosci. Glycotechnol*. **8**: 149–165.

Rahbé, Y., Sauvion, N., Febvay G., Peumans, W. J., and Gatehouse A. M. R. 1995. Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrthosiphon pisum. Entomol. Exp. Appl*. **76**: 143–155.

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., Lee, H.-I., and Broekaert, W. F. 1993. Structure and function of chitin-binding proteins. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**: 591–615.

Raina, A. and Datta, A. 1992. Molecular cloning of a gene encoding a seed-specific protein with nutritionally balanced amino acid composition from *Amaranthus*. *Proc. Natl. Acad. Sci. USA* **89**: 11774–11778.

Ramirez-Soto, D., and Poretz, R. D. 1991. The (1-3)-linked α-L-fucosyl group of the N-glycans of the *Wistaria floribunda* lectin is recognized by a rabbit anti-serum. *Carbohydrate Res*. **213**: 27–36.

Read, S. M. and Northcote, D. H. 1983. Subunit structure and interactions of the phloem proteins of *Cucurbita maxima* (pumpkin). *Eur. J. Biochem*. **134**: 561–569.

Reeke, G. N., Jr. and Becker, J. W. 1986. Three-dimensional structure of favin: saccharide binding-cyclic permutation in leguminous lectins. *Science* **234**: 1108–1111.

Renkonen, K. O. 1948. Studies on hemagglutinins present in seeds of some representatives of Leguminoseae. *Ann. Med. Exp. Biol. Fenn.* **26**: 66–72.

Richardson, J. S. 1981. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**: 167–339.

Richardson, M., Yarwood, A., and Rougé, P. 1987. The amino acid sequence of an atypical single-chain lectin from seeds of *Lathyrus sphaericus* (Retz.). *FEBS Lett.* **216**:145–150.

Rinderle, S. J., Goldstein, I. J., Matta, K. L., and Ratcliffe, R. M. 1989. Isolation and characterization of amaranthin, a lectin present in the seeds of *Amaranthus caudatus*, that recognizes the T- (or cryptic T)-antigen. *J. Biol. Chem.* **264**: 16123–16131.

Rini, J. M., Hofmann, T., and Carver, J. P. 1987. Amino acid sequence differences in the alpha chains of pea seed lectins: C-terminal processing. *Biochem. Cell Biol.* **65**: 338–344.

Rini, J. M., Hardman, K. D., Einspahr, H., Suddath, F. L., and Carver, J. P. 1993. X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 Å resolution. *J. Biol. Chem*. **268**: 10126–10132.

Rini, J. M. 1995. Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24**: 551–577.

Rojo, M. A., Yato, M., Ishii-Minami, N., Minami, E., Kaku, H., Citores, L., Girbes, T., and Shibuya, N. 1997. Isolation, cDNA cloning, biological properties, and carbohydrate-binding specificity of sieboldin-b, a type II ribosome-inactivating protein from the bark of elderberry (*Sambucus sieboldiana*). *Arch. Biochem. Biophys*. **340**: 185–194.

Roque-Bareira, M. C. and Campos-Neto, A. 1985. Jacalin: an IgA-binding lectin. *J. Immunol*. **134**: 1740–1743.

Rougé, P., Barre, A., Causse, H., Chatelain, C., and Porthé, G. 1993. Arcelin and α-amylase inhibitor from the seeds of common bean (*Phaseolus vulgaris*) are truncated lectins. *Biochem. Syst. Ecol*. **21**: 695–703.

Rutenber, E., Ready, M., and Robertus, J. D. 1987. Structure and evolution of ricin B chain. *Nature* **326**: 624–626.

Rutenber, E. and Robertus, J. D. 1991. Structure of Ricin B-chain at 2.5 Å resolution. *Proteins Struct. Funct. Genet.* **10**: 260–269.

Rutenber, E., Katzin, B. J., Collins, E. J., Mlsna, D., Ready, M. P. and Robertus, J. D. 1991. Crystallographic refinement of Ricin to 2.5 Å. *Proteins Struct. Funct. Genet.* **10**: 240–250.

Sabnis, D. D. and Hart, J. W. 1978. The isolation and some properties of a lectin (haemagglutinin) from *Cucurbita* phloem exudate. *Planta* **142**: 97–101.

Saito, K., Komae, A., Kakuta, M., Van Damme, E. J. M., Peumans, W. J., Goldstein, I. J., and Misaki, A. 1993. The alpha-mannosyl-binding lectin from leaves of the orchid twayblade (*Listera ovata*). Application to separation of α-D-mannans from α-D-glucans. *Eur. J. Biochem.* **217**: 677–681.

Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A., and Vijayan, M. 1996. A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a β-prism fold. *Nature Struct. Biol*. **3**: 596–603.

Sanz-Aparicio, J., Hermoso, J., Grangeiro, T. B., Calvete, J. J., and Cavada, B. S. 1997. The crystal structure of *Canavalia brasiliensis* lectin suggests a correlation between its quaternary conformation and its distinct biological

properties from Concanavalin A. *FEBS Lett*. **405**: 114–118.

Sarkar, M., Wu, A. M., and Kabat, E. A. 1981. Immunochemical studies on the carbohydrate specificity of *Maclura pomifera* lectin. *Arch. Biochem. Biophys*. **209**: 204–218.

Sastry, M. V. K., Banerjee, P., Patanjali, S. R., Swamy, M. J., Swarnalatha, G. V., and Surolia, A. 1986. Analysis of the saccharide binding to *Artocarpus integrifolia* lectin reveals specific recognition of T-antigen (β-D-Gal(1,3)D-GalNAc). *J. Biol. Chem*. **261**: 11726–11733.

Shaanan, B., Lis, H., and Sharon, N. 1991. Structure of a legume lectin with an ordered *N*-linked carbohydrate in complex with lactose. *Science* **254**: 862–866.

Shade, R. E., Schroeder, H. E., Pueyo, J. J., Tabe, L. M., Murdock, L. L., Higgins, T. J. V., and Chrispeels, M. J. 1994. Transgenic pea seeds expressing the α-amylase inhibitor of the common bean are resistant to bruchid beetles. *Bio/Technology* **12**: 793–796.

Sharma, V. and Surolia, A. 1997. Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. *J. Mol. Biol*. **267**: 433–445.

Sharon, N. and Lis, H. 1990. Legume lectins — a large family of homologous proteins. *FASEB J.* **4**: 3198–3208.

Shibuya, N., Goldstein, I. J., Shafter, J. A., Peumans, W. J., and Broekaert, W. F. 1986. Carbohydrate-binding properties of the stinging nettle (*Urtica dioica*) rhizome lectin. *Arch. Biochem. Biophys.* **249**: 215–224.

Shibuya, N., Goldstein, I. J., Broekaert, W. F., Nsimba-Lubaki, M., Peeters, B., and Peumans, W. J. 1987. The elderberry (*Sambucus nigra* L.) bark lectin recognizes the Neu5Ac(α2–6)Gal/GalNAc sequence. *J. Biol. Chem.* **262**: 1596–1601.

Shibuya, N., Goldstein, I. J., Van Damme, E. J. M., and Peumans, W. J. 1988. Binding properties of a mannose-specific lectin from the snowdrop (*Galanthus nivalis*) bulb. *J. Biol. Chem.* **263**: 728–734.

Shinshi, H., Neuhaus, J.-M., Ryals, J., and Meins, F., Jr. 1990. Structure of a tobacco endochitinase gene: evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain. *Plant Mol. Biol*. **14**: 357–368.

Smeets, K., Van Damme, E. J. M., and Peumans, W. J. 1994. Comparative study of the post-translational processing of the mannose-binding lectins in the bulbs of garlic (*Allium sativum*) and ramsons (*Allium ursinum*). *Glycoconjugate J.* **11**: 309–320.

Smeets, K., Van Damme, E. J. M., and Peumans, W. J. 1997a. Developmental regulation of lectin and alliinase synthesis in garlic bulbs and leaves. *Plant Physiol*. **113**: 765–771.

Smeets, K., Van Damme, E. J. M., Van Leuven, F., and Peumans, W. J. 1997b. Isolation, characterization and molecular cloning of a leaf-specific lectin from ramsons (*Allium ursinum* L.). *Plant Mol. Biol*. **35**: 531–535.

Smeets, K., Van Damme, E. J. M., Verhaert, P., Barre, A., Rougé, P., Van Leuven, F., and Peumans, W. J. 1997c. Isolation, characterization and molecular cloning of the mannose-binding lectins from leaves and roots of garlic (*Allium sativum* L.). *Plant Mol. Biol*. **33**: 223–234.

Smirnov, S., Shulaev, V., and Tumer, N. E. 1997. Expression of pokeweed antiviral protein in transgenic plants induces virus resistance in grafted wild-type plants independently of salicylic acid accumulation and pathogenesis-related protein synthesis. *Plant Physiol*. **114**: 1113–1121.

Smith, J. J. and Raikhel, N. V. 1989. Nucleotide sequences of cDNA clones encoding wheat germ agglutinin isolectins A and D. *Plant Mol. Biol.* **13**: 601–603.

Stillmark, H. 1888. Über Ricin ein giftiges Ferment aus den Samen von *Ricinus communis* L. und einige anderen Euphorbiaceen. *Inaugural Dissertation Dorpat*, Tartu.

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

Stratford, C. A. and Brown, S. S. 1985. Isolation of an actin binding protein from membranes of

Dictyostelium discoideum. *J. Cell. Biol.* **100**: 727–735.

Sturm, A. and Chrispeels, M. J. 1986. The high mannose oligosaccharide of phytohemagglutinin is attached to Asn^{12} and the modified oligosaccharide to Asn^{60}. *Plant Physiol.* **81**: 320–322.

Sturm, A., Bergwerff, A. A., and Vliegenthart, F. G. 1992. ^{1}H-NMR structural determination of the N-linked carbohydrate chains on glycopeptides obtained from the bean lectin phytohemagglutinin. *Eur. J. Biochem.* **204**: 313–316.

Summer, J. B. and Howell, S. F. 1936. The identification of the hemagglutinin of the Jack bean with concanavalin A. *J. Bacteriol.* **32**: 227–237.

Sweeney, E. C., Tonevitsky, A. G., Temiakov, D. E., Agapov, I .I., Saward, S., and Palmer, R. A. 1997. Preliminary crystallographic characterization of ricin agglutinin. *Protein Struct. Funct. Genet.* **28**: 586–589.

Tahirov, T. H., Lu. T. H., Liaw, Y. C., Chen, Y. L., and Lin, J. Y. 1995. Crystal structure of abrina at 2.14 Å. *J. Mol. Biol.* **250**: 354–367.

Togun, R. A., Animashaun, T., Benzie, C. R., and Kay, J. E. 1988. Purification and characterisation of a mitogenic lactose-binding lectin from *Tetracarpidium conophorum* seeds. **In**: *Lectins, Biology, Biochemistry, Clinical Biochemistry*. Vol. 6, pp. 369–372. Bøg-Hansen, T. C. and Freed, D. L. J., Eds., Sigma Chemical Co., St Louis, MO.

Togun, R. A., Animashaun, T., Kay, J. E., and Aboderin, A. A. 1994. A galactose binding T-cell mitogenic lectin from the seeds of *Telfairia occidentalis*. *Phytochemistry* **35**: 1125–1130.

Transue, T. R., Smith, A. K., Mo, H., Goldstein, I. J., and Saper, M. A. 1997. Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* lectin. *Nature Struct. Biol.* **10**: 779–783.

Tregear, J. W. and Roberts, L. M. 1992. The lectin gene family of *Ricinus communis*: cloning of a functional ricin gene and three lectin pseudogenes. *Plant Mol. Biol.* **18**: 515–525.

Van Damme, E. J. M. and Peumans, W. J. 1987. Isolectin composition of individual clones of *Urtica dioica*: evidence for phenotypic differences. *Physiol. Plant* **71**: 328–334.

Van Damme, E. J. M., Allen, A. K., and Peumans, W. J. 1987. Isolation and characterization of a lectin with exclusive specificity toward mannose from snowdrop (*Galanthus nivalis*) bulbs. *FEBS Lett.* **215**: 140–144.

Van Damme, E. J. M. and Peumans, W. J. 1988. Biosynthesis of the snowdrop (*Galanthus nivalis*) lectin in ripening ovaries. *Plant Physiol.* **86**: 922–926.

Van Damme, E. J. M., Allen, A. K., and Peumans, W. J. 1988. Related mannose-specific lectins from different species of the family Amaryllidaceae. *Physiol. Plant.* **73**: 52–57.

Van Damme, E. J. M. and Peumans, W. J. 1990a. Developmental changes and tissue distribution of lectin in *Galanthus nivalis* L. and *Narcissus* cv. Carlton. *Planta* **182**: 605–609.

Van Damme, E. J. M. and Peumans, W. J. 1990b. Isolectins in *Narcissus*: complexity, inter- and intraspecies differences and developmental control. *Physiol. Plant.* **79**: 1–6.

Van Damme, E. J. M., Goldstein, I. J., and Peumans, W. J. 1991a. A comparative study of related mannose-binding lectins from the Amaryllidaceae and Alliaceae. *Phytochemistry* **30**: 509–514.

Van Damme, E. J. M., Kaku, H., Perini, F., Goldstein, I. J., Peeters, B., Yagi, F., Decock, B., and Peumans, W. J. 1991b. Biosynthesis, primary structure and molecular cloning of snowdrop (*Galanthus nivalis* L.) lectin. *Eur. J. Biochem.* **202**: 23–30.

Van Damme, E. J. M., Goldstein, I .J., Vercammen, G., Vuylsteke, J., and Peumans, W. J. 1992a. Lectins of members of the Amaryllidaceae are encoded by multigene families which show extensive homologies. *Physiol. Plant.* **86**: 245–252.

Van Damme, E. J. M., Smeets, K., Torrekens, S., Van Leuven, F., Goldstein, I. J., and Peumans, W. J. 1992b. The closely related homomeric and heterodimeric mannose-binding lectins

from garlic are encoded by one-domain and two-domain lectin genes, respectively. *Eur. J. Biochem.* **206**: 413–420.

Van Damme, E. J. M., Smeets, K., Balzarini, J., Pusztai, A., Van Leuven, F., Goldstein, I. J., and Peumans, W. J. 1993a. Cloning and characterization of the lectin cDNA clones from onion, shallot and leek. *Plant Mol. Biol.* **23**: 365–376.

Van Damme, E. J. M., Smeets, K., Torrekens S., Van Leuven, F., and Peumans, W. J. 1993b. The mannose-specific lectins from ramsons (*Allium ursinum* L.) are encoded by three sets of genes. *Eur. J. Biochem.* **217**: 123–129.

Van Damme, E. J. M., Balzarini, J., Smeets, K., Van Leuven, F., and Peumans, W. J. 1994a. The monomeric and dimeric mannose-binding proteins from the Orchidaceae species *Listera ovata* and *Epipactis helleborine*: sequence homologies and differences in biological activities. *Glycoconjugate J.* **11**: 321–332.

Van Damme, E. J. M., Smeets, K., Torrekens, S., Van Leuven, F. and Peumans, W. J. 1994b. Characterization and molecular cloning of the mannose-binding lectins from three Orchidaceae species: *Listera ovata, Epipactis helleborine* and *Cymbidium* hybrid. *Eur. J. Biochem.* **221**: 769–777.

Van Damme, E. J. M., Barre, A., Bemer, V., Rougé, P., Van Leuven, F., and Peumans, W. J. 1995a. A lectin and a related non-carbohydrate-binding protein are the two most prominent proteins in the bark of yellow wood (*Cladrastis lutea*). *Plant. Mol. Biol*. **29**: 579–598.

Van Damme, E. J. M., Barre, A., Rougé, P., Van Leuven, F., and Peumans, W. J. 1995b. The seed lectins of black locust (*Robinia pseudoacacia*) are encoded by two genes which differ from the bark lectin genes. *Plant. Mol. Biol*. **29**: 1197–1210.

Van Damme, E. J. M., Barre, A., Smeets, K., Torrekens, S., Van Leuven, F., Rougé, P., and Peumans, W. J. 1995c. The bark of *Robinia pseudoacacia* contains a complex mixture of lectins: characterization of the proteins and the cDNA clones. *Plant Physiol.* **107**: 833–843.

Van Damme, E. J. M., Goossens, K., Smeets, K., Van Leuven, F., Verhaert, P., and Peumans, W. J. 1995d. The major tuber storage protein of Araceae species is a lectin: Characterization and molecular cloning of the lectin from *Arum maculatum* L. *Plant Physiol*. **107**: 1147–1158.

Van Damme, E. J. M., Smeets, K., and Peumans, W. J. 1995e. The mannose-binding monocot lectins and their genes. **In**: *Lectins: Biomedical Perspectives*. pp. 59–80. Pusztai, A. and Bardocz, S., Eds., Taylor and Francis, London, UK.

Van Damme, E. J. M. and Peumans, W. J. 1996. Progress in structural analysis of plant lectins through molecular cloning. **In**: *Lectins, Biology, Biochemistry, Clinical Biochemistry*. Vol. 11, pp.17–27. Van Driessche, E., Rougé, P., Beeckmans, S., Bøg-Hansen, T. C., Eds., Textop, Hellerup, Denmark.

Van Damme, E. J. M., Barre, A., Rougé, P., Van Leuven, F., Balzarini, J., and Peumans, W. J. 1996a. Molecular cloning of the lectin and a lectin related protein from common Solomon's seal (*Polygonatum multiflorum*). *Plant Mol. Biol.* **31**: 657–672.

Van Damme, E. J. M., Barre, A., Rougé, P., Van Leuven, F., and Peumans, W. J. 1996b. The NeuAc (α-2,6)-Gal/GalNAc binding lectin from elderberry (*Sambucus nigra*) bark is a type 2 ribosome inactivating protein with an unusual specificity and structure. *Eur. J. Biochem.* **235**: 128–137.

Van Damme, E. J. M., Barre, A., Rougé, P., Van Leuven, F., and Peumans, W. J. 1996c. Characterization and molecular cloning of SNAV (nigrin b), a GalNAc-specific type 2 ribosome-inactivating protein from the bark of elderberry (*Sambucus nigra*). *Eur. J. Biochem*. **237**: 505–513.

Van Damme, E. J. M., Barre, A., Rougé, P., Van Leuven, F., and Peumans, W. J. 1996d. Isolation and molecular cloning of a novel type 2 ribosome-inactivating protein with an inactive B chain from elderberry (*Sambucus nigra*) bark. *J. Biol. Chem*. **272**: 8353–8360.

Van Damme, E. J. M., Barre, A., Verhaert, P., Rougé, P., and Peumans, W. J. 1996e. Molecular cloning of the mitogenic mannose/maltose-specific rhizome lectin from *Calystegia sepium. FEBS Lett*. **397**: 352–356.

Van Damme, E. J. M., Briké, F., Winter, H. C., Van Leuven, F., Goldstein, I. J., and Peumans, W. J. 1996f. Molecular cloning of two different mannose-binding lectins from tulip bulbs. *Eur. J. Biochem.* **236**: 419–427.

Van Damme, E. J. M., Barre, A., Rougé, P., and Peumans, W. J. 1997a. Molecular cloning of the bark and seed lectins from the Japanese pagoda tree (*Sophora japonica*). *Plant Mol. Biol*. **33**: 523–536.

Van Damme, E. J. M., Roy, S., Barre, A., Citores, L., Mostafapous, K., Rougé, P., Van Leuven, F., Girbes, T., Goldstein, I. J., and Peumans, W. J. 1997b. Elderberry (*Sambucus nigra*) bark contains two structurally different Neu5Ac (α2,6)Gal/GalNAc-binding type 2 ribosome-inactivating proteins. *Eur. J. Biochem*. **245**: 648–655.

Van Damme, E. J. M., Roy, S., Barre, A., Rougé, P., Van Leuven, F., and Peumans, W. J. 1997c. The major elderberry (*Sambucus nigra*) fruit protein is a lectin derived from a truncated type 2 ribosome-inactivating protein. *Plant J*. **12**: 1251–1260.

Van Damme, E. J. M., Van Leuven, F., and Peumans, W. J. 1997d. Isolation, characterization and molecular cloning of the bark lectins from *Maackia amurensis*. *Glycoconjugate J*. **14**: 449–456.

Van Damme, E. J. M., Peumans, W. J., Pusztai, A., and Bardocz, S. 1998. *Handbook of Plant Lectins: Properties and Biomedical Applications.* John Wiley & Sons, Chichester, UK.

Van Driessche, E. 1988. Structure and function of Leguminosae lectins. **In**: *Advances in Lectin Research*. Vol. 1, pp. 73–134, Franz, H., Ed., VEB Verlag Volk und Gesundheit, Berlin, Germany.

Van Eijsden, R. R., Hoedemaeker, F. J., Diaz, C. L., Lugtenberg, B. J. J., De Pater, B. S., and Kijne, J. W. 1992. Mutational analysis of pea lectin. Replacement of Asn^{125} by Asp in the monosaccharide-binding site eliminates mannose/glucose binding activity. *Plant Mol. Biol*. **20**: 1049–1058.

Van Parijs, J., Broekaert, W. F., Goldstein, I. J., and Peumans, W. J. 1991. Hevein: an antifungal protein from rubber-tree (*Hevea brasiliensis*) latex. *Planta* **183**: 258–262.

Van Parijs, J., Joosen, H. M., Peumans, W. J., Geuns, J. M., and Van Laere, A. J. 1992. Effect of the lectin UDA (*Urtica dioica* agglutinin) on germination and cell wall formation of *Phycomyces blakesleeanus* Burgeff. *Arch. Microbiol*. **158**: 19–25.

Venkatesh, Y. P. and Lambert, J. M. 1997. Galactose-induced dimerization of blocked ricin at acidic pH: evidence for a third galactose-binding site in ricin B-chain. *Glycobiology* **7**: 329–335.

Vierheilig, H., Iseli, B., Alt, M., Raikhel, N., Wiemken, A., and Boller, T. 1996. Resistance of *Urtica dioica* to mycorrhizal colonization: a possible involvement of *Urtica dioica* agglutinin. *Plant and Soil* **183**: 131–136.

Vodkin, L. O., Rhodes, P. R., and Goldberg, R. B. 1983. A lectin gene insertion has the structural features of a transposable element. *Cell* **34**: 1023–1031.

Waljuno, K., Scholma, R. A., Beintema, J., Mariono, A., and Hahn, A. M. 1975. Amino acid sequence of hevein. *Proc. Int. Rubber Conf., Kuala Lumpur* **2**: 518–531.

Wang, M.-B., Boulter, D., and Gatehouse, J. A. 1994. Characterization and sequencing of cDNA clone encoding the phloem protein PP2 of *Cucurbita pepo*. *Plant Mol. Biol*. **24**: 159–170.

Wantyghem, J., Platzer, N., Giner, M., Derappe, C., and Goussault, Y. 1992. Structural analysis of the carbohydrate chain of glycopeptides isolated from *Robinia pseudoacacia* seed lectins. *Carbohydrate Res*. **236**: 181–193.

Watkins, W. M. and Morgan, W. T. J. 1952. Neutralization of the anti-H agglutinin in eel by simple sugars. *Nature* **169**: 825–826.

Weiner, O. H., Murphy, J., Griffiths, G., Schleicher, M., and Noegel, A. A. 1993. The actin-binding protein comitin (p24) is a component of the Golgi apparatus. *J. Cell. Biol.* **123**: 23–34.

Weis, W. I. and Drickamer, K. 1996. Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem*. **65**: 441–473.

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

Wongkham, S., Wongkham, C., Boonsiri, P., Simasathiansophon, S., Trisonthi, C., and Atisook, K. 1995. Isolectins from seeds of *Artocarpus lakoocha*. *Phytochem*. **40**: 1131–1134.

Wood, K. A., Lord, J. M., Wawrzynczak, E. J., and Piatak, M. 1991. Preproabrin: genomic cloning, characterisation and the expression of the A-chain in *Escherichia coli*. *Eur. J. Biochem.* **198**: 723–732.

Wright, C. S. 1977. The crystal structure of wheat germ agglutinin at 2.2 Å resolution. *J. Mol. Biol*. **111**: 439–457.

Wright, C. S. 1987. Refinement of the crystal structure of wheat germ agglutinin isolectin 2 at 1.8 Å resolution. *J. Mol. Biol.* **194**: 501–529.

Wright, C. S., Kaku, H. and Goldstein, I. J. 1990. Crystallization and preliminary X-ray diffraction results of snowdrop (*Galanthus nivalis*) lectin. *J. Biol. Chem*. **265**: 1676–1677.

Wright, H. T., Sandrasegaram, G., and Wright, C. S. 1991. Evolution of a family of *N*-acetylglucosamine binding proteins containing the disulfide-rich domain of wheat germ agglutinin. *J. Mol. Evol*. **33**: 283–294.

Wright, C. S. 1992. Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. *J. Biol. Chem*. **267**: 14345–14352.

Wright, C. S. and Jaeger, J. 1993. Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J. Mol. Biol*. **232**: 620–638.

Wright, C. S. and Hester, G. 1996. The 2.0 Å structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: evidence for two unique binding modes. *Structure* **11**: 1339–1352.

Wright, C. S. and Kellogg, G. E. 1996. Differences in hydropathic properties of ligand binding at four independent sites in wheat germ agglutinin-oligosaccharide crystal complexes. *Protein Sci*. **5**: 1466–1476.

Wright, C. S. 1997. New folds of plant lectins. *Curr. Opin. Struct. Biol*. **7**: 631–636.

Yagi, F., Hidaka, M., Minami, Y., and Tadera, K. 1996. A lectin from leaves of *Neoregelia flandria* recognizes D-glucose, D-mannose and *N*-acetylglucosamine, differing from the mannose-specific lectins from other monocotyledonous species. *Plant Cell Physiol*. **37**: 1007–1012.

Yamaguchi, H., Funaoka, H., and Iwamoto, H. 1992. Structures of sugar chains of the subunits of an α-amylase inhibitor from *Phaseolus vulgaris* white kidney beans. *J. Biochem*. **111**: 388–395.

Yamaguchi, K. I., Mori, A., and Funatsu, G. 1995. The complete amino acid sequence of lectin-C from the roots of pokeweed (*Phytolacca americana*). *Biosci. Biotechnol. Biochem.* **59**: 1384–1385.

Yamaguchi, K. I., Mori, A., and Funatsu, G. 1996. Amino acid sequence and some properties of lectin-D from the roots of pokeweed (*Phytolacca americana*). *Biosci. Biotechnol. Biochem.* **60**: 1380–1382.

Yamaguchi, K. I., Yurino, N., Kino, M., Ishiguro, M., and Funatsu, G. 1997. The amino acid sequence of mitogenic lectin-B from the roots of pokeweed (*Phytolacca americana*). *Biosci. Biotechnol. Biochem.* **61**: 690–698.

Yamamoto, K., Ishida, C., Saito, M., Konami, Y., Osawa, T., and Irimura, T. 1994. Cloning and sequence analysis of the *Maackia amurensis* haemagglutinin cDNA. *Glycoconjugate J.* **11**: 572–575.

Yamamoto, K., Konami, Y., and Irimura, T. 1997. Sialic acid-binding motif of *Maackia amurensis* lectins. *J. Biochem.* **121**: 756–761.

Yamasaki, N., Hatakeyama, T., and Funatsu, G. 1985. Ricin D-saccharide interaction as studied by ultraviolet difference spectroscopy. *J. Biochem.* **98**: 1555–1560.

Yamashita, H., Theerasilp, S., Aiuchi, T., Nakaya, K., Nakamura, Y. and Kurihara, Y. 1990. Purification and complete amino acid sequence of a new type of sweet protein with taste-modify-

ing activity, curculin. *J. Biol. Chem.* **265**: 15770–15775.

Yang, H. and Czapla, T. H. 1993. Isolation and characterization of cDNA clones encoding jacalin isolectins. *J. Biol. Chem.* **268**: 5905–5910.

Young, N. M., Johnston, R. A. Z., and Watson, D. C. 1991. The amino acid sequences of jacalin and the *Maclura pomifera* agglutinin. *FEBS Lett.* **282**: 382–384.

Young, N. M. and Oomen, R. P. 1992. Analysis of sequence variation among legume lectins. A ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. *J. Mol. Biol*. **228**: 924–934.

Young, N. M., Watson, D. C., and Thibault, P. 1995. Mass spectrometric analysis of genetic and post-translational heterogeneity in the lectins jacalin and *Maclura pomifera* agglutinin. *Glycoconjugate J.* **12**: 135–141.

Zenteno, E. and Ochoa, J.-L. 1988. Purification of a lectin from *Amaranthus leucocarpus* by affinity chromatography. *Phytochemistry* **27**: 313–317.