Assigned Reading:

Review - History of lectins: from hemagglutinins to biological recognition molecules

Review - Protein-Carbohydrate Interactions as Part of Plant Defense and Animal Immunity

Chapter 24, Essentials to Glycobiology (3rd edition)
History of lectins: from hemagglutinins to biological recognition molecules
Introduction

Toward the end of the 19th century, evidence started to accumulate for the presence in nature of proteins possessing the ability to agglutinate erythrocytes. Such proteins were referred to as hemagglutinins, or phytoagglutinins, because they were originally found in extracts of plants. It is generally believed that the earliest description of such a hemagglutinin was by Peter Hermann Stillmark in his doctoral thesis presented in 1888 to the University of Dorpat (now Tartu, Estonia), one of the oldest universities in czarist Russia (reviewed by Franz, 1988). This hemagglutinin, which was also highly toxic, was isolated by Stillmark from seeds of the castor tree (*Ricinus communis*) and was named ricin. Subsequently, H. Hellin, also at Tartu, demonstrated the presence of a toxic hemagglutinin, abrin, in extracts of the jequirity bean (*Abrus precatorius*). Ricin and abrin soon became commercially available, which prompted Paul Ehrlich, at the Royal Institute of Experimental Therapy (Frankfurt), to employ them as model antigens for immunological studies. Although the preparations available to him were very crude by present criteria (we know now that the ricin and the abrin each contained a weakly agglutinating, powerful toxin and a poorly toxic but strong agglutinin, all galactose-specific), he was able to establish with them in the 1890s several of the fundamental principles of immunology. Thus Ehrlich found that mice were rendered immune to a lethal dose of ricin or abrin by repeated small (sublethal), subcutaneous injections of the lectin and that anti-ricin did not protect the animals against the toxic effects of abrin, nor did anti-abrin protect against ricin. This provided clear evidence for the specificity of the immune response. Ehrlich also showed that immunity to the toxins is transferred from a mother to her offspring by blood during pregnancy and by milk after birth. By studying the inhibitory effect of the anti-ricin immune serum on the agglutinating activity of ricin, he demonstrated that there was a quantitative relationship between the amount of antiserum and that of antigen it could neutralize and on this basis performed the first quantitative determination of an antibody *in vitro*. These studies thus demonstrated the specificity of the antibody response, the phenomenon of immunological memory, and the transfer of humoral immunity from a mother to her offspring.

The general public became aware of ricin in 1978, following its use as a weapon in the notorious politically motivated “umbrella murder” of Georgi Markov, Bulgarian opposition writer and broadcaster in exile. Attempts to employ ricin as a potential weapon of war have been carried out by the United States during World War I; during World
War II a ricin bomb was developed and tested by the British military, but it has never been deployed as a weapon for mass destruction. More recently, ricin has found its way into the arsenals of extremist individuals, groups, and governments.

Sugar binding and blood type specificity

In 1919, James B. Sumner at Cornell University (Ithaca, New York), well known for being the first to crystallize in 1926 an enzyme, urease (for which he was awarded the Nobel Prize 21 years later), isolated from jack bean (Canavalia ensiformis) a crystalline protein that he named concanavalin A and in this way obtained a pure hemagglutinin for the first time. However, nearly two decades passed before Sumner and Howell (1936) reported that concanavalin A agglutinates cells such as erythrocytes and yeasts and also precipitates glycogen from solution. They further showed that hemagglutination by concanavalin A was inhibited by sucrose, demonstrating for the first time the sugar specificity of lectins. With much foresight, they suggested that the hemagglutination induced by concanavalin A might be a consequence of a reaction of the plant protein with carbohydrates on the surface of the red cells.

Already the early results obtained by Stillmark indicated some selectivity in the ricin-induced agglutination of red cells from different animals. This observation was corroborated and further extended by Karl Landsteiner from the University of Vienna, the discoverer of the human A, B, and O blood groups in 1900. Nearly a decade later he reported that the relative hemagglutinating activities of various seed extracts were quite different when tested with red blood cells from different animals (Landsteiner and Rabbitsche, 1907). Because of this specificity, Landsteiner concluded that the actions of plant hemagglutinins “resemble antibody reactions in all essentials.” He therefore used these proteins to illustrate the specificity concept in the introductory chapter of his classic book The Specificity of Serological Reactions (1936).

The 1940s saw the discovery, made independently by William C. Boyd at Boston University and by Karl O. Renkonen at the University of Helsinki, Finland, of the human blood group (or blood type) specificity of the hemagglutinins. They found that crude extracts of the lima bean, Phaseolus limensis, and the tufted vetch, Vicia cracca, agglutinated blood type A erythrocytes but not blood type B or O cells, whereas an extract of the asparagus pea, Lotus tetragonolobus, agglutinated specifically blood type O erythrocytes. Olavi Mäkelä (1957), a doctoral student of Renkonen, examined in 1954–56 extracts from seeds representing 743 plant species and 165 genera, all of the family Leguminosae, and detected hemagglutinating activity in more than one-third of them; close to one-tenth of the hemagglutinins exhibited blood type specificity. Although several of the latter were specific either for blood type O or type A, or both type A and B erythrocytes, and one, from Dolichos biflorus, reacted much better with A1 erythrocytes than with A2, the only extract from Griffonia simplicifolia (previously known as Bandeiraea simplicifolia) exhibited almost exclusively type B specificity. Since then, additional hemagglutinins specific for blood types A and O (but not B) have been discovered, as well as several for other blood types, such as N (Vicia graminea lectin), T (peanut agglutinin, PNA) and Tn (the lectins of Vicia villosa and Moluccella laevis).

The blood type–specific hemagglutinins played a crucial role in early investigations on the structural basis of the specificity of the antigens associated with the ABO blood group system. In the 1950s, Walter J. T. Morgan and Winifred M. Watkins at the Lister Institute, London, found that the agglutination of type A red cells by lima bean lectin was best inhibited by α-linked N-acetyl-D-galactosamine and that of type O cells by the lectin of L. tetragonolobus was best inhibited by α-linked L-fucose. They concluded that α-N-acetyl-D-galactosamine and α-L-fucose are the sugar determinants conferring A and H(O) blood group specificity, respectively. Both conclusions have been substantiated by subsequent investigations (for a recent review, see Morgan and Watkins, 2000). The pioneering work of Watkins and Morgan was among the earliest evidence for the presence of sugars on cell surfaces and their potential roles as identity markers, an accepted theme in modern glycochemistry. It took a while, however, before the counterreceptors for surface sugars, that is, the endogenous lectins that recognize these sugars, were identified, the first being the mammalian hepatic asialoglycoprotein receptor to be described later.

The ability of plant agglutinins to distinguish between erythrocytes of different blood types led Boyd and Shapleigh (1954) to propose for them the name lectins, from the Latin legere, to pick out or choose. This term was generalized by us to embrace all sugar-specific agglutinins of nonimmune origin, irrespective of source and blood type specificity (Sharon and Lis, 1972).

Mitogenic stimulation of lymphocytes and agglutination of cancer cells

Two major discoveries made in the early 1960s were instrumental in bringing lectins into the limelight. The first of these was by Peter C. Nowell (1960) at the University of Pennsylvania, Philadelphia, who found that the lectin of the red kidney bean (Phaseolus vulgaris), known as phytohemagglutinin (PHA), is mitogenic, that is, it possesses the ability to stimulate lymphocytes to undergo mitosis. This discovery had a revolutionary impact on immunology in that it shattered the view, held until then, that lymphocytes are dead-end cells incapable of dividing or differentiating further. Within a short time, several other lectins were proven to be mitogenic. Of special significance was the finding that concanavalin A acts as a mitogen because, in contrast to PHA, its activity could be inhibited by low concentrations of monosaccharides, for example, mannose. This finding provided proof that mitogenic stimulation is the result of binding of lectins to sugars on the surface of the lymphocytes and was among the earliest demonstrations for a biological role of cell surface sugars. Mitogenic lectins soon became tools for the study of signal transmission into cells and for the analysis of the biochemical events that occur during lymphocyte stimulation in vitro. A most
valuable outcome of such studies was the discovery in the 1970s by Robert C. Gallo and his associates at the National Institutes of Health (Bethesda) of T cell growth factor, now known as interleukin-2, in conditioned medium of normal human lymphocytes stimulated by PHA (Morgan et al., 1976).

The second discovery was made by Joseph C. Aub at the Massachusetts General Hospital in Boston (Aub et al., 1963, 1965). He found that wheat germ agglutinin (WGA) has the ability to preferentially agglutinate malignant cells. This was followed by the reports of Max M. Burger at Princeton University and Leo Sachs and Michael Inbar at the Weizmann Institute (Rehovot) that concanavalin A exhibits the same ability. Together with Sachs and Ben-Ami Sela, we subsequently found that soybean agglutinin (SBA) also possesses the same property. Such investigations provided early evidence that changes in cell surface sugars are associated with the development of cancer and led to the assumption that high susceptibility to agglutination by lectins was a property shared by all malignant cells. Unfortunately, this is now known not to be generally true.

Lectins galore

Until the early 1970s, the presence of hemagglutinins had been reported in numerous organisms, primarily plants, but only very few had been purified, almost all by conventional techniques. In addition to concanavalin A, they included the plant lectins from soya beans, green peas, Dolichos biflorus seeds, wheat germ, and mushroom (Agaricus campestris) (reviewed in Sharon and Lis, 1972) and the animal lectins of eel (Springer and Desai, 1971), snail (Hammarstrøm and Kabat, 1969), and horseshoe crab (Marchalonis and Edelman, 1968). The pace of lectin isolation increased dramatically with the introduction of affinity chromatography for lectin purification by Irwin J. Goldstein and Bipin B. L. Agrawal of the University of Michigan, originally for the isolation of concanavalin A on immobilized dextran (Sephadex) (Agrawal and Goldstein, 1967). Numerous lectins have thus become available, for a time still mainly from plants, the number of the latter being now about 500. The interest in these lectins was greatly stimulated by the demonstration that they are invaluable tools for the detection, isolation, and characterization of glycoconjugates, primarily of glycoproteins, for histochemistry of cells and tissues and for the examination of changes that occur on cell surfaces during physiological and pathological processes, from cell differentiation to cancer (Table I).

The occurrence of hemagglutinins in animals was noted quite early, almost all in invertebrates or lower vertebrates, but until the middle of the 1970s, only the three of these mentioned (of eel, snail, and horseshoe crab) were isolated and characterized. The first of the animal lectins shown to be specific for a sugar (L-fucose) was from the eel (Watkins and Morgan, 1952). The isolation in 1974 of the first mammalian lectin, the galactose-specific hepatic asialoglycoprotein receptor, was an outcome of the investigation by Gilbert Ashwell at the NIH together with Anatol G. Morell at the Albert Einstein Medical School (New York) of the mechanisms that control the lifetime of glycoproteins in

<table>
<thead>
<tr>
<th>Table I. Major applications of lectins&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell identification and separation</td>
</tr>
<tr>
<td>Detection, isolation, and structural studies of glycoproteins</td>
</tr>
<tr>
<td>Investigation of carbohydrates on cells and subcellular organelles; histochemistry and cytochemistry</td>
</tr>
<tr>
<td>Mapping of neuronal pathways</td>
</tr>
<tr>
<td>Mitogenic stimulation of lymphocytes&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Purging of bone marrow for transplantation&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Selection of lectin-resistant mutants</td>
</tr>
<tr>
<td>Studies of glycoprotein biosynthesis</td>
</tr>
</tbody>
</table>

<sup>a</sup>Lectins from sources other than plants are rarely in use.
<sup>b</sup>In clinical use.

blood circulation (Hudgin et al., 1974; Stockert et al., 1974). At the same time, Vivian Teichberg from our department reported (Teichberg et al., 1975) the isolation from the electric eel of the first member of the family of the β-galactose-specific lectins, designated galectins (Barondes et al., 1994), of which over a dozen members have now been characterized. Since the beginning of the 1980s, the number of purified animal lectins also started to grow quickly, largely thanks to the advent of recombinant techniques.

From primary to 3D structures

The 1970s also witnessed the intensification of studies of the molecular properties of individual lectins, a prerequisite for a deep understanding of their activities at the molecular level. These studies ranged from the determination of the main physicochemical parameters of lectins to complete amino acid sequencing and elucidation of their 3D structure. Until the advent of recombinant techniques, determination of the primary structure of lectins proceeded rather slowly, and by the end of that decade the complete sequences of only half a dozen lectins, all from plants, were known. In this case, too, concanavalin A led the field, being the first lectin whose primary sequence has been established (Edelman et al., 1972). Concurrently, Edelman’s group and independently Karl Hardman with Clinton F. Ainsworth at Argonne National Laboratories (Argonne, Illinois), solved the 3D structure of concanavalin A by high resolution X-ray crystallography, another first for this lectin (Edelman et al., 1972; Hardman and Ainsworth, 1972). This was soon followed by the determination of the structure of WGA as well as of its complexes with its ligands (N-acetylneuraminic acid and β4-linked N-acetylglucosamine oligomers) by Christine Schubert Wright at the Virginia Commonwealth University (Richmond) even before the complete amino sequence of this lectin had become available (Wright, 1977). The striking difference between the structure of concanavalin A and WGA fully corroborated the suggestion presented by us at the 1973 International Glycoconjugate Symposium in Lille, France, that was based largely on compositional data of these two
proteins, that although “lectins have many biological properties in common, they represent a diversified group of proteins with respect to size, composition and structure” (Sharon et al., 1974).

The availability of the primary structure of numerous lectins allowed the identification of homologies between the sequences of lectins from taxonomically related sources, as originally demonstrated for the legume lectins by one of us (N.S.) in collaboration with Donny Strosberg at the Free University of Brussels (Foriers et al., 1977). By the end of the following decade, homologies were found also for lectins from other families, such as the galectins and the C-type (Ca\(^{2+}\) requiring) lectins (Drickamer, 1988).

During the past few years, the number of lectin primary and 3D structures has increase dramatically, with some 200 of the latter having been elucidated (www.cermav.cnrs.fr/lectines). In addition, many structures of lectin–carbohydrate complexes have been solved. Quite surprisingly, remarkable similarities have been noticed between the tertiary structures of lectins from diverse sources, in spite of the lack of primary sequence similarities (Figure 1). One such common tertiary structure, first observed in the legume lectins, and referred to

---

**Fig. 1.** Structures of different lectins represented as ribbon diagrams. (A) Upper row shows monomers of lectins from different sources that share the jelly roll or lectin fold. First three lectins (from left) in the lower row all exhibit the β-trefoil fold. (B) Variations in quaternary lectins structures. The gray spheres represent metal ions; bound carbohydrate is shown in ball-and-stick representation. All diagrams, except for that of SAP, reprinted with permission from R. Loris (2002) *Biochim. Biophys. Acta*, 1572, 198–208. The diagram of SAP is from PDB entry sac.
as the lectin fold, consists characteristically of an elaborate jelly roll, derived from antiparallel β-strands, arranged as two β-sheets (Srinivasan et al., 1996). This fold has been found in the legume lectins, the galectins, and in several other animal lectins, such as the pentraxins (Crennell et al., 1994) and ERGIC-53 (Irin et al., 1996; Velloso et al., 2002), as well as in nonlectin molecules, for example, several glycosidases, among them Vibrio cholerae sialidase.

Starting in the late 1980s, considerable information has become available, by X-ray crystallography and site-directed mutagenesis, of the chemical groups on the lectin and on the carbohydrates that interact with each other and of the types of bond formed, primarily hydrogen bonds and hydrophobic interactions. It has been concluded that lectins recognize sugars in diverse ways, just like other proteins recognize their ligands (Sharon, 1993).

Carbohydrate recognition domains

Based on an analysis of the then known amino acid sequences of animal lectins, Kurt Drickamer from Columbia University (New York) proposed in 1988 that the carbohydrate-binding activity of most of them resides in a limited polypeptide segment, designated by him as the carbohydrate-recognition domain (CRD) (Drickamer, 1988). He named the CRD found in the galectins S-CRD and that found in C-type lectins C-type CRD. By now several types of CRD have been discerned, in addition to those just mentioned, each of which shares a pattern of invariant and highly conserved amino acid residues at a characteristic spacing. On this basis it was possible to divide the majority of the animal lectins into structurally related families and superfamilies, the most widely occurring of which is that of the C-type lectins (CTLs). Other families of special interest are the P-type lectins and the siglec.

The majority of the CTLs are large, asymmetric transmembrane glycoproteins, in which the CRD is attached to a variable number of structurally and functionally different polypeptide domains. In contrast, the galectins are generally small, soluble, nonglycosylated proteins and, unlike the CTLs, do not require Ca²⁺ for their activity. Members of the CTL superfamily are grouped into three families—selectins (the most celebrated one), collectins, and endocytic lectins. The story of the selectins started with attempts to elucidate the molecular basis of lymphocyte homing. These attempts greatly benefited from the availability of an in vitro assay for measuring the interaction of lymphocytes with postcapillary high-endothelial venules (HEVs), a known site of lymphocyte exit from the blood stream (Stampfer and Woodruft, 1977). Using this assay, which reflects the in vivo homing of lymphocytes, Eugene C. Butcher and colleagues at Stanford University obtained a monoclonal antibody (MEL-14) against a murine lymphocyte antigen (Gallatin et al., 1983). The antibody inhibited the binding of the lymphocytes to HEV in vitro and their homing in vivo, suggesting that the MEL-14 antigen has a direct role in these phenomena. From inhibition experiments of the lymphocyte-HEV binding, Steven D. Rosen and Lloyd M Stoolman at the University of California, San Francisco, have concluded that sugars of the endothelial cell might also be involved in this binding and that the lymphocytes should have a membrane-bound lectin with specificity for fucose and Man-6-P (Stoolman et al., 1984). This lectin was subsequently shown to be identical with the MEL-14 antigen.

In 1987 Bevilacqua and co-workers of Harvard Medical School have developed two monoclonal antibodies that identified a second cell-surface antigen, designated ELAM (endothelial-leukocyte adhesion molecule)-1, expressed on stimulated human endothelial cells but not on unstimulated ones (Bevilacqua et al., 1987). Another vascular cell adhesion molecule was originally isolated from activated platelets independently by Rodger McEver at the Oklahoma Medical Research Foundation, Oklahoma City (McEver and Martin, 1984) and by Bruce and Barbara C. Furie at Tufts University, Boston (Berman et al., 1986; Hsu-Lin et al., 1984), and designated GPM-140 and PADGEM, respectively.

These three cell adhesion molecules, collectively known for a while as LEC-CAMS, were identified as a discrete family of CTLs after the virtually simultaneous publication in 1989 of their primary sequences (Bevilacqua et al., 1989; Johnston et al., 1989; Lasky et al., 1989; Siegelman et al., 1989); these go now under the names L-selectin, E-selectin, and P-selectin, respectively (reviewed in Lasky, 1995). They were all shown to have a similar domain structure, with an extracellular part that consists of an amino terminal CRD, an epidermal growth factor-like domain, and several short repeating units related to complement-binding protein. They bind specifically to the trisaccharide NeuAcα(2-3)Galβ(1-4)(Fucα(1-3))GlcNAc, known as siaI-Lewisx (siaLeα in brief) and its positional isomer, NeuAcα(2-3)Galβ(1-3)(Fucα(1-4))GlcNAc (siaLeβ), with both fucose and sialid acid required for binding (Brandley et al., 1990; Stoolman, 1989). The selectins recognize the carbohydrate ligands only when the latter are present on particular glycoproteins, such as cell surface mucins, pointing to the role of the carrier molecule in lectin-carbohydrate interactions; one of the best characterized of such carriers is the P-selectin glycoprotein ligand (Moore et al., 1992).

The paradigm of the endocytic lectins is the mammalian hepatic asialoglycoprotein receptor already mentioned. The collectins, represented by the soluble mannose-binding proteins of mammalian serum and liver, first detected by chance as a contaminant of a preparation of α-mannosidase from human liver (Robinson et al., 1975), subsequently purified and characterized by Toshiaki Kawasaki and Ikuo Yamashina at Kyoto University, Japan (Kawasaki et al., 1978; Kozutsumi et al., 1980), are characterized by an NH₂-terminal collagen-like stretch of repeating Gly-X-Y- triplets (where X and Y are any amino acid). The structural unit of the mannose-binding proteins is a trimer of identical subunits with a triple-stranded collagen helix and three CRDs (Weis and Drickamer, 1994). This arrangement of CRDs at a fixed spacing has important biological implications, to be discussed later.

A different kind of CRD has been identified in the siglec. This family of sialic acid-binding Ig-like lectins, a member of the Ig superfamily, was discovered when the cloning of a macrophage lectin-like adhesion molecule named sialoadhesin (siglec-1) revealed striking structural similarities to a B cell restricted member of the Ig superfamily, CD22 (siglec-2) and to two other members of the Ig superfamily,
CD33 (siglec-3) and the myelin-associated glycoprotein (siglec-4) (Crocker et al., 1994). Members of this family, 11 of which have been identified in humans, are type I transmembrane proteins with an extracellular part consisting of a CRD-containing N-terminal V-set Ig-like domain, followed by variable numbers of C2-set Ig-like domains. Except for myelin-associated glycoprotein (siglec-4), exclusively expressed in the nervous system, they are all found on cells of the hematopoietic system. Each siglec has a distinct expression pattern in different cell types, indicating that they perform highly specific functions.

A recent addition to the growing list of mammalian lectins is dectin-1, a β-glucan receptor, identified by Gordon Brown and Siamon Gordon (2001) at Oxford by screening a cDNA library of a macrophage cell line with zymosan. It is a small type II transmembrane receptor containing one CRD, which recognizes β1,3 and/or β1,6-glucans and intact yeasts.

In protection and symbiosis

The question of the possible physiological role of lectins has intrigued investigators from the start and focused on plant lectins, which for long time were virtually the only ones known (reviewed by Etzler, 1986). It was speculated, for example, that lectins may function as antibodies to protect plants against harmful soil bacteria, control seed germination, or be involved in the transport and storage of sugars, but no evidence for these speculations could be found. However, two proposals put forward in the 1970s still hold. According to first one, lectins protect plants against phytopathogenic microorganisms and insects as well as against predatory animals. The second theory assumes that they are involved in the association between leguminous plants and their symbiotic nitrogen-fixing bacteria. Probably the earliest publication on the insecticidal action of lectins came in 1976 from the laboratory of Irvin E. Liener at the University of Minnesota, at St. Paul (Minnesota) in which it was reported that feeding bruchid beetles with a diet containing the black bean lectin resulted in the death of the bruchid larvae (Janzen et al., 1976). On this basis the authors concluded that the major role of lectins in legumes is to protect them from attack by insect seed predators. In subsequent studies, several other lectins were shown to be insecticidal, among them WGA, Galanthus nivalis lectin and jacalin.

The proposal that lectins may be involved in the protection of plants against pathogenic microorganisms was originally based on the observation made at Rehovot that WGA, PNA, and SBA inhibited the sporulation and growth of fungi such as Trichoderma viride, Penicillium notatum, and Aspergillus niger (see Barkai-Golan et al., 1976). Potato lectin was subsequently shown to act in a neurosporan spores of the major carbohydrate specificity groups were all found to cause growth disruption during germination of spores of Neurospora crassa, Aspergillus amstelodami, and Botryodiplodia theobromae (Brambl and Gade, 1985). It was also shown that recombinant Urtica dioica agglutinin that has a similar specificity to that of WGA (Broekaert et al., 1989) inhibited the growth of fungal phytopathogens.

The idea that lectins are responsible for the specific association between nitrogen-fixing rhizobia and leguminous plants, which provides the plant with the needed nitrogen, was advanced nearly three decades ago (Bohlool and Schmidt, 1974; Hamblin and Kent, 1973). It was based on the finding that a lectin from a particular legume bound in a carbohydrate-specific manner to the surface polysaccharides or lipopolysaccharides of the corresponding rhizobial species but not to bacteria that are symbionts of other legumes. For instance, SBA agglutinated most strains of Bradyrhizobium japonicum that nodulate soybeans but not nonnodulating bradyrhizobial strains. The suggestion has therefore been made that rhizobial attachment to plant roots occurs by interaction between the bacterial surface carbohydrates and lectins present in the roots of the leguminous plants. This became known as the lectin recognition hypothesis, which is still the subject of controversy, because of the lack of unequivocal evidence and of some inconsistencies. Thus for most host–symbiont systems examined, there is no proof for the presence of lectins and their ligands on plant roots and bacteria, respectively, at precisely the right time and location. Moreover, the correlation between the specificity of the host lectin and its ability to recognize the nodulating bacteria of that host is not very strict. Also, several lines of soybeans with no detectable lectin in their seeds or vegetative tissues were nodulated normally by the corresponding rhizobial symbiont.

Application of the techniques of molecular genetics gave results that bolstered the lectin recognition hypothesis but did not fully settle the controversy (reviewed by Kijne, 1996; Hirsch, 1999).

Recently, a variant of the lectin recognition hypothesis has been proposed, that postulates that the host-specific attachment of the rhizobium is achieved through the interaction between species-specific lipo-chitooligosaccharide signal molecules produced by the bacteria, named nodulation factors (Nods), and a new type of a plant root lectin found in different leguminous plants but not in plants of different families (Kalsi and Etzler, 2000).

Recognition molecules

In a broader sense, the foregoing discussion implies that lectins possess the ability to act as recognition molecules inside cells, on cell surfaces, and in physiological fluids (Figure 2 and Table II). This is in fact the current view of the biological function of lectins, which also evolved during the 1970s (Ashwell and Morell, 1972; Ofek et al., 1978).

Although indications that lectins may function in recognition had appeared in the literature prior to that time, their significance was not appreciated then. A case in point is the demonstration in the 1950s, mainly by of Alfred Gottschalk at the Walter and Eliza Hall Institute (Melbourne, Australia) that the influenza virus hemagglutinin is responsible for the attachment of the virus to the host cells as a prerequisite for infection. However, it was only following the isolation in 1974 of the asialoglycoprotein receptor, a hepatic lectin, and the discovery of its unexpected ability to
recognize and to bind terminal galactose residues on serum glycoproteins (reviewed by Ashwell and Morell, 1974) that the role of lectins in biological recognition started to gain popularity. Support came soon with the identification, by William S. Sly at St. Louis University (Kaplan et al., 1977) of the mannose-6-phosphate receptor crucially involved in intracellular trafficking of lysosomal enzymes. The demonstration that hepatic lectins may also mediate the clearance of bacteria from blood in the absence of opsonins (antibodies and complement), was an early indication of the participation of lectins in non-immune defense, or innate immunity (see later discussion).

Another key finding was made in 1979, when our group, together with others, demonstrated that urinary tract infection in mice by mannose-specific Escherichia coli could be prevented by methyl α-D-mannoside (Aronson et al., 1979). It was the first direct evidence for the involvement of bacterial lectins in the initiation of infection, the basis for the present attempts in academia and industry, to apply carbohydrates for antiadhesion therapy of such diseases (reviewed by Mulvey et al., 2001).

Together with Itzhak Ofek, we demonstrated at the same time that the mannose-specific bacterial surface lectins may also mediate attachment of the bacteria to phagocytic cells in the absence of opsonins, leading to engulfment and killing of the bacteria. This process, another example of innate immunity, which we named lectinophagocytosis, may be of importance in the clearance of bacteria from nonimmune patients or from opsonin-poor sites, such as renal medulla or the peritoneal cavity (Ofek and Sharon, 1988). Additional lectins have been implicated in innate immunity. A prominent example is the mannose-specific receptor present on the surface of macrophages; it binds infectious organisms that expose mannose-containing glycans on their surface, leading to their ingestion and killing. Another, recently discovered one, is dectin-1, specific for β1,3 and/or β1,6-glucans, present on fungi.

A similar function, albeit by a different mechanism, is performed by the soluble mannose-binding lectins (MBLs) of mammalian serum and liver (Epstein et al., 1996; Turner, 1996). These proteins bind to oligomannosides of infectious microorganisms, causing activation of complement without participation of antibody, and subsequent lysis of the pathogens, thus acting in innate immunity. The spatial arrangement of the CRDs in the MBLs provides a structural basis for their ability to bind ligands with repetitive, mannose-rich structures, such as

---

### Table II. Functions of lectins

<table>
<thead>
<tr>
<th>Lectin</th>
<th>Role in</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microorganisms</strong></td>
<td></td>
</tr>
<tr>
<td>Amoeba</td>
<td>Infection</td>
</tr>
<tr>
<td>Bacteria</td>
<td>Infection</td>
</tr>
<tr>
<td>Influenza virus</td>
<td>Infection</td>
</tr>
<tr>
<td><strong>Plants</strong></td>
<td></td>
</tr>
<tr>
<td>Various</td>
<td>Defense</td>
</tr>
<tr>
<td>Legumes</td>
<td>Symbiosis with nitrogen-fixing bacteria</td>
</tr>
<tr>
<td><strong>Animals</strong></td>
<td></td>
</tr>
<tr>
<td>Calnexin, calreticulin,</td>
<td>Control of glycoprotein biosynthesis</td>
</tr>
<tr>
<td>ERGIC-53</td>
<td></td>
</tr>
<tr>
<td>Collectins</td>
<td>Innate immunity</td>
</tr>
<tr>
<td>Dectin-1</td>
<td>Innate immunity</td>
</tr>
<tr>
<td>Galectins</td>
<td>Regulation of cell growth and apoptosis; regulation of the cell cycle; modulation of cell-cell and cell-substratum interactions</td>
</tr>
<tr>
<td>Macrophage mannose</td>
<td>Innate immunity; clearance of sulfated glycoprotein hormones</td>
</tr>
<tr>
<td>receptor</td>
<td></td>
</tr>
<tr>
<td>Man-6-P receptors</td>
<td>Targeting of lysosomal enzymes</td>
</tr>
<tr>
<td>L-selectin</td>
<td>Lymphocyte homing</td>
</tr>
<tr>
<td>E- and P-selectins</td>
<td>Leukocyte trafficking to sites of inflammation</td>
</tr>
<tr>
<td>Siglecs</td>
<td>Cell-cell interactions in the immune and neural system</td>
</tr>
<tr>
<td>Spermadhesin</td>
<td>Sperm-egg interaction</td>
</tr>
</tbody>
</table>

---

Fig. 2. Cell surface lectin–carbohydrate interactions. Lectins serve as means of attachment of different kinds of cell as well as viruses to other cells via the surface carbohydrates of the latter. In some cases, cell-surface lectins bind particular glycoproteins (e.g., asialoglycoproteins), whereas in other cases the carbohydrates of cell surface glycoproteins or glycolipids serve as sites of attachment for biologically active molecules that themselves are lectins (e.g., carbohydrate-specific bacterial and plant toxins, or galectins). Based on an original diagram from BioCarbAB (Lund, Sweden).
found on fungal and microbial surfaces, but not to the oligomannose units of mammalian glycoproteins (Weis and Drickamer, 1994).

The discovery of the selectins and the demonstration that they play a crucial role in the control of lymphocyte homing and of leukocyte trafficking to sites of inflammation was a landmark in lectin research. Indeed, the selectins provide the best paradigm for the role of sugar–lectin interactions in biological recognition. They mediate the binding of leukocytes to endothelial cells and thereby initiate a rolling phase, in which the lectins interact transiently with glycan ligands, leading eventually to their extravasation. Prevention of adverse inflammatory reactions by inhibition of leukocyte–endothelium interactions, another application of anti-adhesion therapy, has become a major aim of the biomedical and pharmacological industry. There are also indications that the selectins may function in the spread of cancer cells from the main tumor to other sites in the body and that by blocking their sugar-recognition sites it may be possible to prevent the formation of metastases.

From the late 1980s, evidence started to accumulate that several lectins of different types direct intracellular glycoprotein traffic, by acting as chaperones and sorting receptors in the secretory pathway. Calnexin, a membrane-bound lectin of the endoplasmic reticulum (ER), functions in parallel with calreticulin, its soluble homolog, as part of a quality control system that ensures proper folding of glycoproteins destined to the cell surface. The mannose-specific intracellular lectin, ERGIC-53, first identified as a resident of the ER–Golgi intermediate compartment (Schweizer et al., 1988) carries a specific subset of nascent glycoproteins between the two compartments. Two distinct mannose-6-phosphate receptors, the only members of the P-type lectin family, mediate the targeting of newly synthesized hydrolases from the rough ER to the lysosomes (Hoflack and Kornfeld, 1985). Both receptors bind their ligands, oligosaccharides bearing terminal Man-6-P residues, most efficiently at pH 6–7, with such oligosaccharides in the trans-Golgi network, receptors bind their ligands, oligosaccharides bearing terminal Man-6-P residues, most efficiently at pH 6–7, and to release them in the more acidic environment of the lysosomes.

The galectins are believed to act as modulators of cell–substratum interactions and to be essential for the normal differentiation and growth of all multicellular animals. They are capable of inducing cell proliferation, cell arrest, or apoptosis (physiological cell death) and have been implicated in organ morphogenesis, tumor cell metastasis, leukocyte trafficking, immune response, and inflammation, as well as recognition of extracellular matrix.

Epilogue

As we have shown in this article, during 120+ years, lectins have come a long way since their first detection in plants as hemagglutinins to their present status as ubiquitous recognition molecules with myriad exciting functions and applications.

Abbreviations

CRD, carbohydrate recognition domain; CTL, C-type lectin; ER, endoplasmic reticulum; HEV, high-endothelial venule; MBL, mannose-binding lectin; PHA, phytohemagglutinin; PNA, peanut agglutinin; SBA, soybean agglutinin; WGA, wheat germ agglutinin.

References

In addition to the references listed here, readers are referred to those in Kocourk (1986) and Sharon and Lis (2003).


Protein-Carbohydrate Interactions as Part of Plant Defense and Animal Immunity
Protein-Carbohydrate Interactions as Part of Plant Defense and Animal Immunity

Kristof De Schutter and Els J. M. Van Damme *

Lab Biochemistry and Glycobiology, Department of Molecular Biotechnology, Ghent University, Coupure links 653, B-9000 Ghent, Belgium; E-Mail: Kristof.DeSchutter@UGent.be

* Author to whom correspondence should be addressed; E-Mail: ElsJM.VanDamme@UGent.be; Tel.: +32-9-264-6086; Fax: +32-9-264-6219.

Academic Editor: Derek J. McPhee

Received: 7 April 2015 / Accepted: 14 May 2015 / Published: 19 May 2015

Abstract: The immune system consists of a complex network of cells and molecules that interact with each other to initiate the host defense system. Many of these interactions involve specific carbohydrate structures and proteins that specifically recognize and bind them, in particular lectins. It is well established that lectin-carbohydrate interactions play a major role in the immune system, in that they mediate and regulate several interactions that are part of the immune response. Despite obvious differences between the immune system in animals and plants, there are also striking similarities. In both cases, lectins can play a role as pattern recognition receptors, recognizing the pathogens and initiating the stress response. Although plants do not possess an adaptive immune system, they are able to imprint a stress memory, a mechanism in which lectins can be involved. This review will focus on the role of lectins in the immune system of animals and plants.

Keywords: lectin; carbohydrate; plant defense; animal immunity

1. Introduction

Protein glycosylation is generally recognized as one of the major co- and post-translational modifications. Furthermore, many proteins that specifically recognize and bind these glycosylated structures have been identified in all kingdoms of life. In recent years, it has been clearly established that interactions between carbohydrate chains and their partner proteins can mediate many important
biological events. These carbohydrate-binding proteins will not only interact with the glycan structures on proteins, but will also recognize and bind to carbohydrate chains on glycolipids and proteoglycans, or can interact with polysaccharides and free sugars.

Proteins, of non-immune origin, that can recognize and bind specific carbohydrate structures are classified as lectins. The term lectin (from the Latin ‘lectus’—chosen, selected) was first used by Boyd and Shapleigh [1] to emphasize the selectivity in the carbohydrate binding. More recently, Peumans and Van Damme [2] defined lectins as “all proteins possessing at least one non-catalytic domain which binds reversibly to a specific mono- or oligosaccharide”. The importance of these carbohydrate-binding proteins is shown by their occurrence in all kingdoms of life.

The whole group of lectins can be divided into multiple families of structurally and evolutionary related proteins based on a conserved carbohydrate-recognition domain (CRD). In plants, lectins can be classified into 12 families [3] (Table 1). This classification system proved to be superior to the many attempts to group plant lectins based on sugar specificity and allows to classify the majority of all putative lectin sequences based on the sequence and evolutionary relationships between the CRDs. The assignment of animal lectins to several groups is less straightforward. Originally animal lectins were divided into the C-type (Ca\(^{2+}\) dependent) lectins and the S-type (sulfydryl-dependent or B-galactoside binding) lectins [4], and all remaining lectins were classified in a heterogeneous group, the N-type glycans (not C nor S-type). However, as more knowledge became available with respect to the three-dimensional structure, lectin activity was found to be associated with a diversity of structural motifs. At present, there are at least 14 families grouping structurally related carbohydrate-binding domains (Table 1) and many other “orphan” lectins with a unique structure. Some of these orphan lectins belong to well-established families that are generally not linked to sugar-binding activity [5]. Furthermore, many animal lectins will also interact with macro-molecules through protein-protein, protein-lipid or protein-nucleic acid interactions [6]. Of all lectin domains distinguished in animals only the families with calnexin and calreticulin and the R-type lectins are also present in plants.

In addition to the classification based on structural domains, lectins can be grouped based on their expression patterns. In plants, lectins with a constitutive expression are often present at high concentrations in some specific cells or organs. Examples of these abundant lectins are the storage proteins found in seeds or specialized vegetative tissues. These lectins accumulate during a certain developmental stage and offer a source of nitrogen or amino acids that can readily be degraded when needed [3]. In addition to a role as storage lectins, the localization of these abundant proteins in the vacuoles or extracellular space, suggests a role in plant defense. Most probably, these lectins serve a role in protecting the seeds from pathogens and diseases, or herbivory.

In contrast to the highly abundant constitutively expressed lectins, the inducible plant lectins are present at a low basal level, often too low for detection by western blot analysis. However, in response to biotic or abiotic stresses, the expression of these lectins is upregulated. Although lectin levels rise significantly, they remain low abundant proteins. Interestingly, most of these stress inducible lectins are located in the nucleocytoplasmic compartment of the plant cell and are often expressed throughout the plant. Based on these observations, it is proposed that lectin-mediated protein–carbohydrate interactions in the cytoplasm and the nucleus play an important or possibly even crucial role in the stress physiology of the plant cell [7,8]. Similar to plants, several animal lectins are known to have an inducible expression pattern upon detection of stress [9]. Especially for the lectins with low expression levels, the study of
protein-carbohydrate interactions can be challenging. Furthermore, the diversity of carbohydrates and glycosidic linkages complicates a quantitative study of carbohydrates using traditional methods [10]. However, novel computational methods can circumvent some of these difficulties and provide new tools to study carbohydrate interactions [10].

Table 1. Lectin families in plants and animals.

<table>
<thead>
<tr>
<th>Plant Lectin Family</th>
<th>Typical Saccharide Ligands</th>
<th>Predicted Localization</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Agaricus bisporus</em> lectin family</td>
<td>GlcNAc/GalNAc, Galactose</td>
<td>Nucleus, cytosol</td>
</tr>
<tr>
<td>Amaranthin family</td>
<td>GalNAc</td>
<td>Nucleus, cytosol</td>
</tr>
<tr>
<td>Chitinase related agglutinin family</td>
<td>High mannose N-glycans</td>
<td>Vacuole, membrane bound</td>
</tr>
<tr>
<td>Cyanovirin family</td>
<td>Mannose</td>
<td>Nucleus</td>
</tr>
<tr>
<td><em>Euonymus europaeus</em> lectin family</td>
<td>Galactosides, high-mannose N-glycans</td>
<td>Nucleus, cytosol</td>
</tr>
<tr>
<td><em>Galanthus nivalis</em> lectin family</td>
<td>Mannose</td>
<td>Vacuole, nucleus, cytosol or membrane bound</td>
</tr>
<tr>
<td>Hevein family</td>
<td>Chitin</td>
<td>Vacuole</td>
</tr>
<tr>
<td>Jacalin family</td>
<td>Mannose- and galactose-specific subgroup</td>
<td>Vacuole, nucleus, cytosol, vacuole</td>
</tr>
<tr>
<td>Legume family</td>
<td>Mannose</td>
<td>Vacuole, nucleus, cytosol or membrane bound</td>
</tr>
<tr>
<td>LysM family</td>
<td>Chitin, peptidoglycan</td>
<td>Vacuole, nucleus, cytosol or membrane bound</td>
</tr>
<tr>
<td><em>Nicotiana tabacum</em> lectin family</td>
<td>(GlcNAc)$_n$, high-mannose and complex N-glycans</td>
<td>Nucleus, cytosol</td>
</tr>
<tr>
<td><em>Ricin-B</em> family</td>
<td>Gal/GalNAc, Sialylated Gal/GalNAc</td>
<td>Vacuole, nucleus, cytosol</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Animal Lectin Family</th>
<th>Typical Saccharide Ligands</th>
<th>Predicted Localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calnexin and Calreticulin</td>
<td>Glc$_1$Man$_8$</td>
<td>ER</td>
</tr>
<tr>
<td>M-type lectins</td>
<td>Man$_8$</td>
<td>ER</td>
</tr>
<tr>
<td>L-type lectins</td>
<td>Various</td>
<td>ER, Golgi</td>
</tr>
<tr>
<td>P-type lectins</td>
<td>Man$_6$-phosphate</td>
<td>Secretory pathway</td>
</tr>
<tr>
<td>C-type lectins</td>
<td>Mannosides, galactosides, sialic acids and others</td>
<td>Membrane bound, extracellular</td>
</tr>
<tr>
<td>S-type lectins (galectins)</td>
<td>$\beta$-galactosides</td>
<td>Cytosol, extracellular</td>
</tr>
<tr>
<td>I-type lectins (siglecs)</td>
<td>Sialic acid</td>
<td>Membrane bound</td>
</tr>
<tr>
<td>R-type lectins</td>
<td>Various</td>
<td>Golgi, membrane bound</td>
</tr>
<tr>
<td>F-box lectins</td>
<td>GlcNAc$_2$ of N-glycans</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td>Fibrinogen-type lectin</td>
<td>GlcNAc, GalNAc</td>
<td>Membrane bound, extracellular</td>
</tr>
<tr>
<td>Chi-lectins</td>
<td>Chito-oligosaccharides</td>
<td>Extracellular</td>
</tr>
<tr>
<td>F-type lectins</td>
<td>Fucose terminating oligosaccharides</td>
<td>Extracellular</td>
</tr>
<tr>
<td>Intelectins</td>
<td>Galactose, galactofuranose, pentoses</td>
<td>Membrane bound, extracellular</td>
</tr>
<tr>
<td>Annexins</td>
<td>Glycosaminoglycans, heparin and heparin sulfate</td>
<td>Membrane bound</td>
</tr>
</tbody>
</table>

* Table redrafted from [6,11].

The binding of a lectin with a specific carbohydrate structure is mediated through hydrophobic (Van der Waals) interactions and hydrogen bonds. The specificity of this interaction results from the formation of specific hydrogen bonds and metal coordination bonds to key hydroxyl groups [12]. In
addition, unwanted recognition is sometimes excluded by steric exclusion. However, the sequence or the threedimensional conformation of the CRD is no indication for its specificity, since structurally unrelated lectins can recognize similar carbohydrate structures [12]. In addition some lectins with similar CRDs can recognize different carbohydrates. These phenomena can be attributed to the shallowness of the sugar-binding site and the limited number of contacts with the sugar that allows the CRD to recognize multiple carbohydrate structures, further referred to as ‘the promiscuity of the CRD’. Another interesting feature in carbohydrate-binding sites is that within a lectin family, there is a common mechanism for binding of a core monosaccharide in the primary binding site, but diversity in binding of oligosaccharides or glycoconjugates is achieved through extended and secondary binding sites unique to individual lectins [13]. Most likely, a number of lectin domains from animals and plants descended from a common ancestor through divergent evolution [13]. One lectin family that is represented both in mammalia and plants is the R-type lectin family (Ricin-B family in plants). The β-trefoil structure of the R-type lectin was first identified in ricin. In animals a structurally related domain was found in fibroblast growth factors and the cysteine-rich domain of the mannose receptor [14]. Although both ricin and the mannose receptor can bind glycan structures containing galactose ((sialylated) galactose/N-acetylgalactosamine and Lewis^a^ structures respectively), their carbohydrate binding site differs significantly [14]. While the β-trefoil structure is conserved, the sequence of their CRDs differs, and amino acids involved in the binding of the carbohydrate structure are not conserved, accounting for the promiscuity of the CRDs.

Animal and plant lectins tend to play a role in a wide variety of biological processes, with some lectins having more than one function. Unfortunately for many lectins their physiological importance remains enigmatic. Several CRDs are involved in the recognition of invaders, and thus are part of the immune system, which consists of various types of cells and molecules that specifically interact with each other to initiate the host defense mechanism. To operate properly, the immune system must be able to detect a wide variety of pathogenic agents, and distinguish them from the organism’s own healthy cells. Apart from their role in the immune system CRDs are important for a multitude of cellular processes like cell-cell interactions, self/non-self-recognition and intracellular routing. In addition, lectins also play an important role as molecular chaperones for glycoprotein quality control.

This review wants to offer an anthology of the multitude of functions carried out by lectins in the immune system of plants and animals. We will discuss the role of lectins in the recognition of pathogens in the innate immune system as pattern recognition receptors, their involvement in autophagy and their importance in stress signaling and other processes in the immune system. Furthermore we will also touch on their contribution in the vertebrate adaptive immune response and their potential involvement in epigenetic stress imprinting in plants. However, in view of the vast amount of lectins known today this review will only focus on a few examples to illustrate the functions of lectins in the immune system, and does not claim to present a complete overview of all the lectins involved.

2. Lectins as Pattern Recognition Receptors in the Innate Immune System

The first line of defense against infection by other organisms is the innate immune system. This non-specific immune system consists of cells and molecules that can recognize the pathogens and initiate a generic defense response. This system does not offer long-lasting protective immunity, as does the adaptive immune system, but activation of the innate system is required to initiate an adaptive response.
In the innate immune system, pathogenic microorganisms are recognized through highly conserved structures, pathogen associated molecular patterns (PAMPs). These structures are recognized by the pattern recognition receptors (PRRs) of the host [15]. Since many of the PAMPs recognized by the PRRs are carbohydrate structures, lectins play an important role as PRRs. These lectin PRRs are highly variable in structure and can occur in a soluble as well as in a membrane-associated form.

2.1. Animal Lectin PRRs

One of the best studied examples of an inducible lectin with a role as a PRR in the immune response is the mannose-binding protein (MBP), also known as mannose binding lectin (MBL) 2. This lectin is a family member of the collectins which represent a group of soluble lectins belonging to the Ca$^{2+}$-dependent C-type lectin superfamily [16]. Collectins are organized as oligomers of a trimeric subunit, composed of a C-terminal lectin domain and an N-terminal collagen-like domain which enables the formation of the triplets [16]. Under normal conditions, MBL2 is synthesized at a basal level by the liver and secreted into the serum. However, after exposure to pathogenic microorganisms, the level of MBL2 mRNA transcripts and protein increases [17]. In response to infection, local inflammatory cells secrete cytokines into the bloodstream that stimulate the liver to produce large numbers of acute phase proteins, including MBL2. In the innate immune system, MBL2 functions as a PRR binding a range of carbohydrates including N-acetylglucosamine, mannose, N-acetylmannosamine, fucose and glucose, enabling the lectin to interact with surface glycans on a wide selection of viruses, bacteria, yeasts, fungi and protozoa (Figure 1).

When the C-terminal recognition portion of MBL2 binds to carbohydrates on the pathogen surface, the N-terminal domain can interact with collectin receptors on macrophages, which in turn leads to phagocytosis (Figure 1). In addition, MBL2 can activate the complement pathway through a unique pathway, i.e., the lectin pathway, independent from the classical pathway [18] (Figure 1). In the classical pathway of complement fixation, the binding of antibodies to pathogens leads to the recruitment of the first component of complement, C1q. In turn, C1q associates with two serine proteases, C1r and C1s, which initiates a proteolytic cascade. In the lectin pathway, MBL2 activates the same cascade through direct activation of serine proteases, MBP-associated serine proteases (MASPs), without the involvement of C1q. In essence, MBL2 is substituting for C1q [19]. The structural organization of collectins resembles that of C1q, being composed of an N-terminal collagenous domain and a globular C-terminal domain. However, C1q differs from MBL2 in that it contains a C-terminal immunoglobulin binding domain. However, at present MBL2 is the only collectin known to activate the complement pathway [16]. Activation of C2-C4 complexes through activation of the proteolytic cascade from either pathway leads to the cleavage of the C3 complement component. The resulting C3b fragment will insert into the surface of the target, initiating a lytic pathway in which additional complement components insert into the membrane to form a pore. In addition, C3b mediates phagocytosis of the pathogen through interaction with receptors on macrophages.

The pentraxins (belonging to the L-type lectin superfamily) represent another class of soluble lectins involved in the innate immune response, among these are C-reactive proteins (CRP) and serum amyloid P components (SAP) [20,21]. Similar to MBL2, they are acute phase proteins whose expression levels are increased upon stimulation by cytokines (IL-1 and IL-6). Their biological functions include
activation of the complement pathway and stimulation of phagocytic leukocytes [16]. Furthermore, ficolins are complement activating soluble PRR able to sense molecular patterns on both pathogens and apoptotic cell surfaces [6].

The mannose receptor is a membrane bound PRR on the surface of macrophages [22]. This glycoprotein belongs to the C-type superfamily of lectins and possesses eight CRDs [23, 24], which bind terminal mannose, fucose and N-acetylglucosamine (GlcNAc) residues [25]. Binding of the mannose receptor to components in the cell wall of pathogens leads to the internalization of the pathogen by macrophages (Figure 1). After phagocytosis by the macrophage, the phagosomes will fuse with lysosomes, killing the pathogen. However, Mycobacteria, the causal agent of tuberculosis and leprosy, make use of the mannose receptor to gain access into the cell. They prevent the fusion of the phagosome with lysosomes, creating a habitat to survive and proliferate within macrophages [26].

In addition to its role as a PRR in the innate immune response, several other functions have been attributed to this mannose receptor, including the attachment of sperm to oocytes [27], endocytosis for antigen presentation [22] and clearance of glycoproteins [28, 29]. Interestingly, the expression of the mannose receptor is repressed during the early stages of inflammation and is induced during the resolution phase [30]. This expression profile is in accordance with a role in clearing inflammatory agents [30].

Recognition of yeasts through binding of the β-glucans on their surface by Dectin-1, a natural killer (NK)-cell-receptor-like C-type lectin, leads to the uptake and killing of these yeasts. In addition, Dectin-1 will induce a signaling cascade leading to the production of cytokines and chemokines (Figure 1). This signaling occurs cooperatively with Toll-like receptors (TLR) and the use of spleen tyrosine kinases (SYK) [31, 32].

The nucleotide-binding oligomerization domain (NOD) proteins NOD1 and NOD2 provide the cell with another level of microbial surveillance. Being cytoplasmic PRRs they recognize conserved fragments in particular muramyl peptides in the cell wall of many types of bacteria [33]. These muramyl peptides are derived from the bacterial peptidoglycan. The mechanisms by which the muramyl fragments enter the cell and activate the NOD-like receptors NOD1 and NOD2 remain poorly understood but multiple routes of entry have been reported [33]. Upon sensing distinct peptidoglycan fragments, NOD1 and NOD2 activate intracellular signaling pathways that drive pro-inflammatory and antimicrobial responses [33]. In addition, polymorphisms in NOD2 have been identified as the strongest genetic risk factor in the development of Crohn’s disease [33].
Figure 1. Model of animal innate immunity by lectin PRRs. Detection of pathogen/microbe-associated molecular patterns (P/MAMPs) by membrane bound pattern recognition receptors (PRRs) with lectin domains (lectin receptor kinases and lectin-like proteins) will initiate an intracellular signaling cascade on the one hand and lead to phagocytosis of the pathogen on the other hand. Soluble lectin PRRs can recognize P/MAMPs and subsequently bind to receptors that will trigger phagocytosis or can activate the complement system. Next to the recognition of P/MAMPs at the surface, cytoplasmic PRRs can sense the presence of bacterial peptidoglycan fragments which leads to activation of the intracellular signaling cascade. The intracellular signaling cascade can include downstream protein phosphorylation (e.g., MAPK cascade, TRAF/IRAK signaling), transcription factor activation (e.g., NF-kB, AP1, NFAT), or reactive oxygen signaling with cross-communication between the different components. This signaling cascade will lead to activation of stress-responsive or developmental signal-responsive genes or the production of chemokines and cytokines. In addition, the innate immune response can activate the adaptive immune response.
2.2. Plant Lectin PRRs

Despite obvious differences in the immune system of plants and animals, there are also some striking similarities [34]. While the adaptive immunity is unique to vertebrates, the innate immune response most probably has ancient origins [35]. In contrast to animals where specialized cell types (macrophages, neutrophils and dendritic cells) in the blood circulatory system play a key role in the detection of pathogens and the activation of the immune system, most plant cells are autonomously capable of sensing the presence of pathogens and activating a defense response [36]. Similar to animals, plants have evolved systems for non-self-recognition and anti-microbial defense [36]. Like animals, plants have acquired specialized PRRs for their defense against pathogens [37]. These PRRs are able to recognize the damage-associated molecular patterns (DAMPs) and the pathogen- or microbe-associated molecular patterns (PAMPs/MAMPs). As in animals, many of these PRRs carry lectin domains able to recognize and interact with carbohydrate structures from microbial organisms or saccharides derived from plant cell wall damage. Upon PAMP/MAMP and DAMP perception by the PRRs, an intracellular response is activated, referred to as the PAMP/MAMP-triggered immunity (PTI/MTI) (Figure 2). Besides the components to sense the pathogens, also the building blocks of PAMP-induced signaling cascades leading to transcriptional activation of response genes are shared between the two kingdoms. In particular, nitric oxide as well as mitogen-activated protein kinase (MAPK) cascades have been implicated in triggering innate immune responses, which ultimately lead to the production of anti-microbial compounds [36]. In addition, this response includes ion fluxes across the plasma membrane, increase of cytosolic Ca\(^{2+}\) levels, production of reactive oxygen species and protein phosphorylation. This complex response of the plant finally leads to profound transcriptional changes, stomatal closure and cell wall reinforcement [38], and will ultimately limit the growth of the pathogen [39].

The best studied example of the perception of PAMPs by plant PRRs is the recognition of bacterial flagellin, through the conserved flg22 epitope, by the Arabidopsis FLS2 receptor-like kinase [38]. For more detailed information on FLS2 signaling and some other examples involving recognition through protein-protein interactions we refer to some recent review papers focused on this topic [40,41]. A significant part of the patterns recognized by the plant PRRs are carbohydrate structures, which are either present on the surface of the pathogen (e.g., lipopolysaccharides, peptidoglycans and chitin molecules) or are derived from plant own molecules (i.e., DAMPs), released during invasion of the pathogen (e.g., cellulose fragments, arabinogalactan proteins and oligogalacturonides).

For the recognition of the D/M/PAMPs, a large diversity of membrane-bound or soluble PRRs with a lectin domain have been identified in plants. However, only a limited number of them have been functionally characterized. The membrane-bound PRRs carrying one or more extracellular lectin domains are often coupled to an intracellular Ser/Thr kinase domain (Figure 2). These lectin receptor kinases (LecRK) can be classified into four types: G-, C-, L- and LysM-type [42,43].

Due to their sessile lifestyle, plants are not only subjected to biotic stresses but also to multitude of abiotic stress factors (e.g., cold, excess water, increased salt concentrations). To be able to respond to these detrimental environmental conditions, plants have evolved defense pathways able to sense and respond to these abiotic stresses. Several of the LecRKs have been reported to act both upon biotic and abiotic stresses [44–49].
Figure 2. Model of plant innate immunity by lectin PRRs. After perceiving extracellular signals (pathogen/microbe-associated molecular patterns (P/MAMPs), damage associated molecular patterns (DAMPs) or pathogen-derived effector proteins) by membrane bound pattern recognition receptors (PRRs) with lectin domains (lectin receptor kinases LecRKs and lectin-like proteins), an intracellular signaling cascade is initiated in plants. This signaling cascade can include downstream protein phosphorylation, transcription factor activation, or modulation of hormonal pathways, ultimately leading to activation of stress-responsive or developmental signal-responsive genes and pathogen/microbe triggered immunity (P/MTI). In addition, perception of intracellular effectors by nucleocytoplasmic receptors (potentially including nucleocytoplasmic lectins), leads to activation of the effector triggered immunity (ETI). However, the precise signaling pathways leading to P/MTI and ETI are not fully known and need to be elucidated. Figure redrafted from [40,42].

In contrast to animal systems, where the C-type lectins represent a major player in the recognition of pathogens and the induction of the immune response, C-type lectins in plants are rather rare. Only one C-type LecRK has been identified in Arabidopsis and rice, and its function is not fully elucidated [45,50,51].
The G-type LecRKs carry a lectin domain belonging to the *Galanthus nivalis* agglutinin (GNA) family of plant lectins. Although 32 G-type LecRKs have been identified in *Arabidopsis* and 100 in rice [51], it still needs to be confirmed whether the lectin domain plays a role in the interaction with the pathogen. G-type LecRKs have been shown to function in self-incompatibility reactions and defense to both biotic and abiotic stresses [44,46,49].

The 42 L-type LecRKs identified in the genome of *Arabidopsis thaliana* (72 in rice) are grouped into nine clades on a dendrogram. These kinase receptors with a Legume-like (L-type) lectin domain show variable expression patterns in different tissues and developmental stages in response to stimuli [45]. The amino acid sequence of the legume-like lectin domain is only poorly conserved in the L-type LecRKs, therefore it is doubtful whether it possesses true lectin activity. Despite the fact that it remains to be proven that L-type LecRKs possess lectin activity, several of these proteins were reported to be involved in plant resistance to pathogens [47,52–54]. In addition, some L-type LecRKs have been described to act in hormone signaling and stomatal immunity [47]. Although the overall legume-like lectin domain is poorly conserved, a hydrophobic site within this lectin domain is highly preserved [53]. This site can play a role in the detection of extracellular ATP by L-type LecRKs. Recently, extracellular ATP was shown to play a role as a signaling molecule in plant stress responses [55,56]. The best studied LecRKs in plants belong to the LysM-type LecRKs. These receptors recognize the GlcNAc moieties in various types of bacterial peptidoglycans and fungal chitins [57,58]. Next to their role in recognizing pathogenic microorganisms, the LysM LecRKs are also involved in the recognition of beneficial microorganisms such as mycorrhiza and rhizobacteria [58–60].

### 2.3. Plant Effector Triggered Immunity

In addition to the PTI/MTI responses, plants have evolved another defense mechanism, referred to as the effector triggered immunity (ETI) (Figure 2). In contrast with the former responses, ETI will mainly act inside the plant cell. Successful pathogens have evolved mechanisms to counter the PTI response of the plant by the delivery of specific elicitors or effectors, also called avirulence (Avr) proteins [61]. These effectors are usually directly secreted into the plant cell by a type III secretion system and suppress or block the PRR-dependent signaling [62]. These effectors are recognized by a class of intracellular plant receptors containing a nucleotide-binding and leucine-rich repeat domain (NB-LRR), thereby activating the ETI response. In the ETI response, plants express specific resistance genes upon recognition of these pathogenic effectors to overcome the action of these effectors. Both the PTI/MTI and the ETI response can lead to programmed cell death of the host cell through activation of the hypersensitive response (HR), but can also result in systemic acquired resistance (SAR) which activates defense mechanisms throughout the plant [38,63].

### 3. Autophagy

Recognition of pathogens by PRRs leads to the induction of the immune responses and will also result in phagocytosis of the pathogen. These phagocytic vesicles will subsequently fuse with lysosomes leading to the degradation of the pathogen. However, some bacteria are capable of escaping this process by releasing effectors into the cytosol. *Mycobacterium tuberculosis* secretes an effector (EsxH) that prevents the maturation of the phagosome and fusion with lysosomes, thereby escaping degradation and
creating a niche to proliferate [64]. In the case of the bacterium *Salmonella enterica* serovar *Thyphimurium*, the major cause of food poisoning, the pathogen can invade and grow in the gut epithelial cells after digestion by its host. *S. thyphimurium* is recognized by the Toll-like receptor 4 on the gut epithelial cells, which binds to the lipopolysaccharides on the outer membrane of the Gram-negative bacterium [65]. *Salmonella* uses this interaction and its subsequent phagocytosis to invade the host cell. Early after cell invasion, the bacterium resides in a vesicle known as the *Salmonella* containing vacuoles (SCV). The bacteria then use a type III secretion system to generate pores in the SCV membrane, through which they can deliver effectors into the cell’s cytoplasm [66]. These effectors modulate the activity of the host cell machinery to prevent degradation and promote growth of the intracellular pathogen. Moreover, some bacteria can escape the damaged SCVs and replicate in the cytosol. However, cells have developed a defense mechanism against this kind of intracellular pathogens. Microbes in damaged vesicles as well as microbes residing in the cytosol are targeted by the autophagy system that will engulf them in autophagosomes and restrict bacterial growth [67,68].

Autophagy is a strictly regulated and mainly non-selective process used by the cell to capture cytoplasmic cargoes in a double membrane vesicle, the autophagosome, to maintain a balance between synthesis and degradation of their own components. The uptake of the intracellular cargo in these vesicles is mediated through evagination of the rough endoplasmatic reticulum associated membrane [69]. When this vesicle fuses with lysosomes its contents become degraded and nutrients and building blocks get recycled. Malfunctioning of the autophagy system has been linked to human diseases such as cancer, neurodegenerative disorders, diabetes and inflammatory bowel disease [66]. In addition to its function in the housekeeping of the cell, autophagy can also help to defend the mammalian cytosol against bacterial infection [70]. Efficient pathogen engulfment is mediated by cargo-selecting autophagy adaptors that rely on pattern recognition or danger receptors to label the invading pathogen as autophagy cargo [71]. This labeling is typically mediated through polyubiquitin coating [72]. Recently the carbohydrate-dependent galectin-8 pathway was discovered, representing a novel mechanism that allows to get rid of bacteria that have been taken up by the cell [71]. Galectins are β-galactoside binding lectins that accumulate in the cytosol before being secreted via a leader peptide-independent pathway [71]. When these galectins occur extracellularly, they bind to cell surface glycans in order to modulate cellular behavior. Inside the cytosol, where complex carbohydrates are absent under normal physiological conditions, galectins play a role as danger and/or pattern recognition receptors. Thurston et al. [71] reported that galectin-8 is recruited to damaged SCVs and the absence of galectin-8 leads to an increase in the growth rate of the intracellular bacteria. However, in contrast to the direct tagging of *Shigella flexneri* by autophagy related gene 5 (ATG5) [73], it was observed that galectin-8 does not bind to the bacterial glycans but to the host cell’s own complex glycans that become accessible in the damaged vesicles. Supporting this observation, galectin-8 was also found to decorate vesicles with other vesicle damaging bacteria such as *Listeria monocytogenes* and *Shigella flexneri*. Moreover, osmotic damage of cytoplasmic vesicles also leads to the recruitment of galectin-8 [71]. Galectin-8 interacts with the autophagy receptor NDP52, which subsequently recruits LC3 and other components of the autophagy machinery to the damaged vesicles. Although, other galectins (galectin-3 and -9) are also recruited to the damaged vesicles, their role in cellular defense to infection remains enigmatic [71].

Similar to animals and yeast, essential autophagic machineries are conserved in plants. The role of autophagy related genes (ATGs) in the lifecycle of plants is similar to that in animals. In *Arabidopsis*,
autophagy is activated in certain stages of development, senescence or in response to starvation, or after environmental stress [74]. In addition, autophagy in plants is suggested to be required for proper regulation of the hypersensitive response programmed cell death [75,76]. However the question remains whether autophagy is involved in the elimination of bacterial pathogens in plants and if any of the cytoplasmic lectins contribute to this process. Compared to animals, not many examples of intracellular microorganisms are known in plant cells since due to their rigid cell wall, plant cells are usually not phagocytic [77]. One well studied example in plants is the endosymbiosis of *Rhizobium*, where the engulfment of the microorganism by a plant-derived membrane occurs by a phagocytic mechanism [78]. These endosymbionts counteract their degradation [79] but it is unknown if they trigger an autophagosome response. One indication that cytoplasmic plant lectins can be involved in autophagy comes from the observation of an interaction between the *Arabidopsis* jacalin-related lectins and proteins of the ER-bodies [80]. These novel ER-derived organelles have been shown to play a role in plant defense against pathogens [81] and are linked to autophagocytosis [82].

Next to a role in the defense against intracellular bacteria, autophagy has also been reported to play a role in the elimination of intracellular and extracellular viruses. Overexpression of *beclin 1*, a protein required for autophagosome formation and recruitment of autophagosome components, in neurons reduced Sindbis virus replication [83], while in cells deprived of *beclin 1*, Sindbis virus and *Herpes simplex* virus replicate to higher titers than in wild-type cells [84,85]. Similarly, the plant tobacco mosaic virus accumulated to a higher level in *Beclin 1*-silenced plants than in wild-type plants [75]. These results indicate that autophagy functions as an antiviral defense mechanism in which the ATG proteins might target cellular factors or pathways required for virus replication and spread [86].

4. Non-PRR Function of Lectins in the Immune Response

Many of the animal PRRs that detect microbial infection induce the innate immune response by triggering an intracellular signaling cascade which stimulates expression of chemokines, cytokines and other immune mediators. The role of protein-carbohydrate interactions is not limited to the interaction between the pathogen and the host cell, but they can also play an important part in the signaling pathway induced upon pathogen recognition or in other processes of the immune response.

Inflammation is one of the responses of the innate immune system upon detection of stress. Induction of stress signaling in macrophages upon recognition of microorganisms leads to the secretion of chemokines and cytokines which initiate inflammation. A key function of inflammation is the recruitment of leukocytes through vascular tissue, a process in which C-type lectins play a key role. Selectins are present on the surface of leukocytes (L-selectins), endothelial cells (E- and P-selectins) and platelets (P-selectins) and bind to sialyl Lewis structures (can also bind other carbohydrates such as heparin) [87]. The interaction between these selectins and their ligand will result in a process called leukocyte rolling, which will slow down the movement of the leukocytes in the bloodstream at the sites of inflammation and increases the contact between the leukocyte and the epithelial cells. This allows the tight interaction between Integrin (MAC-1) and its ligand (IMAC-1). In addition, MAC-1 is also able to bind to carbohydrates such as zymosan and lipopolysaccharides [16].

Sialic acid binding Ig-like lectins or Siglecs are a good example of lectins involved in regulating cellular activation within the immune system. Sialoadhesin (siglec-1), a macrophage lectin-like adhesion
molecule, has been implicated in key interactions with T-lymphocytes, leading to generation of cytotoxic T-cells [88]. Since siglec-1 lacks the cytoplasmic tyrosine-based motif present in other siglecs, it is suggested to mediate extracellular rather than intracellular functions [89]. The B-cell restricted Siglec-2 (CD22) plays an important role in regulating B-cell activity [89]. The balance between positive and negative signals is an important factor in the immune response [90]. Upon cellular activation by receptors, counteracting inhibitory signals can be delivered through distinct mechanisms, resulting in a higher activation threshold [89]. CD22 is described as an inhibitory receptor. Although the precise role of the sialic acid binding activity of CD22 remains to be elucidated, mice lacking CD22 show a higher prevalence of autoantibodies with increasing age [91].

Similar to animals, several families of plant lectins have been identified with a putative role in plant defense against biotic or abiotic stress other than the recognition of stress agents. Under non-stress conditions, the expression levels for these lectins are low, often under the detection limit. But after perception of biotic and/or abiotic stress signals, their transcript levels are upregulated. Although the precise role of these lectins in plant defense is not fully understood, several of them have been reported to increase plant resistance to biotic or abiotic agents.

Lectins with an *Euonymus* lectin-like (EUL) domain are ubiquitous in plants, indicating a universal role for these lectins [92]. Despite their widespread distribution, the EUL proteins were only recently described as a new lectin family. Although all characterized lectins from this family show a similar localization in the nucleocytoplasmic compartment of the cell, their ligand specificity varies depending on the protein and species studied [93]. Most EULs are chimeric proteins in which the lectin domain is linked to an unknown domain. In contrast to the *Euonymus europaeus* agglutinin, the prototype of this family, which is expressed at high concentrations in the arils surrounding the seeds, the EUL proteins from *Arabidopsis* and rice are low abundant proteins, the expression of which is induced after the plant was subjected to various biotic (bacterial and fungal infection) and abiotic (dehydration, salinity, osmotic stress and abscisic acid (ABA) treatment) stresses. *Arabidopsis* EULS3 was shown to interact with the nucleocytoplasmic ABA receptor RCAR1 [94] and a Ca\(^{2+}\) dependent kinase involved in ABA response in stomata guard cells [95], suggesting a role for ArathEULS3 in ABA signaling and stomatal closure.

Jacalin-related lectins can be divided into two subgroups based on their substrate specificity: the galactose-binding and the mannose-binding jacalins [3]. Structural analysis demonstrated that the relatively large size of the carbohydrate binding site, which is extended in the galactose-binding jacalins compared to the mannose-binding jacalins, allows this promiscuity of the CRD [96]. There are some striking differences between the two subgroups: First, polypeptides of the mannose-binding jacalins remain intact whereas those of the galactose-binding subgroup are cleaved into two dissimilar polypeptides [3]. Second, galactose-binding jacalins are synthesized on ER bound ribosomes as a proprotein whereas the mannose-binding lectins are synthesized without signal peptide [3], resulting in a different subcellular localization for both subgroups of jacalin-related lectins. Whereas the galactose-binding jacalins mainly reside in the vacuole, the mannose-binding lectins are located in the nucleocytoplasmic compartment of the cell [97]. Several jacalin-related lectins have been associated with disease resistance, abiotic stress signaling, wounding and insect damage [98]. The jacalin domain is often present as part of chimeric proteins, where the lectin domain is fused to an unrelated domain. Interestingly, several of these chimeric proteins contain domains related to stress response and defense. In Poaceae species (wheat, rice and maize), mannose-binding jacalin-related proteins in which the
C-terminal lectin domain is fused to an N-terminal dirigent or disease-response domain have been identified. In wheat (*Triticum aestivum*), nearly half of the jacalin-related lectins are chimeric proteins containing this dirigent motif, and those already studied in more detail were all reported to have increased expression levels after application of some stress factor [98]. Rice (*Oryza sativa*) expresses a protein composed of a C-terminal jacalin domain fused to a leucine rich repeat (LRR) and an N-terminal NB-ARC domain. The NB-ARC domain is a novel signaling motif found in bacteria and eukaryotes, and is shared between plant resistance gene products and regulators of cell death in animals. Furthermore, also proteins consisting only of one or more jacalin domain(s) can be upregulated after stress perception. For example, Orysata is a salt inducible mannose-binding lectin from rice shoots. In addition, Orysata is also induced upon hormone treatment (jasmonic acid (JA) and ABA), infection with *Magnaporthe grisea* and during senescence [99–101]. The expression of the wheat TaJRL1 lectin, a component of the defense signaling, is induced after fungal infection in a hormone (salicylic acid (SA) and JA) dependent manner [102]. In contrast to the plant hormone dependent regulation, jacalin-related lectins can also be regulated independently from hormones. The *Arabidopsis JAX1* offers resistance to potex virus by inhibiting accumulation of viral RNA in a hormone independent manner [103]. In addition to a signaling role, jacalin-related lectins can have an effector function. The Helja lectin from sunflower (*Helianthus tuberosus*) was shown to possess antifungal properties. Furthermore, it induces morphological changes and reactive oxygen species production in yeast as well as alteration of membrane permeability [104].

Recently, proteins containing the jacalin-related lectin domain were also identified outside the plant kingdom. The human lectin ZG16p is a soluble mammalian lectin with a jacalin-related β-prism fold. ZG16p has been reported to bind both to glycosaminoglycans and α-mannose terminating short glycans [105]. Recognition of a broad spectrum of ligands by ZG16p may account for the multiple functions of this lectin in the formation of zymogen granules via glycosaminoglycan binding, and in the recognition of pathogens in the digestive system through α-mannose-related recognition [105]. Furthermore, the jacalin-related domain has also been reported in a few prokaryotes [3].

The expression of the nucleocytoplasmic tobacco lectin, referred to as the *Nicotiana tabacum* agglutinin or Nictaba, is enhanced after treatment with JA, a plant hormone important for plant development and involved in signaling of abiotic and biotic stress, such as insect herbivory [40]. Nictaba specifically recognizes GlcNAc oligomers but also reacts with the core structure of N-glycans. The Nictaba domain is widespread in the plant kingdom and several Nictaba-related lectins were shown to have a role in plant defense [40]. For instance, Nictaba was shown to possess entomotoxic activity on Lepidopteran insects and can protect plants against insect herbivory [40]. The phloem proteins (PP) represent a subgroup of the Nictaba-related lectins isolated from *Cucurbitaceae*. In contrast to Nictaba, the PP2 lectins are constitutively expressed in the companion cells of the phloem and are subsequently translocated to the phloem sap. Expression of the *Arabidopsis* homolog PP2-A1 is enhanced by ethylene treatment and *Pseudomonas* infection [40]. In a computational study [106] *Arabidopsis* PP2-A1 was shown to exhibit specificity for a diverse range of glycans including chitin oligomers and the MansGlcNAc2 core of high mannose N-glycans [106,107]. Consistent with a role in plant defense, PP2-A1 represses feeding of the aphid *Myzus persicae* and aphid borne yellow virus transmission [40].

The ricin-B family of lectins contains the first lectin discovered in plants, in particular ricin. Ricin, isolated from the seeds of castor bean (*Ricinus communis*), is a chimeric protein composed of a lectin
domain and a domain with RNA N-glycosidase activity, enabling the protein to cleave an adenine residue from the large rRNA. This enzymatic activity results in the inactivation of ribosomes and the arrest of protein translation. As a result ricin is classified in the group of ribosome inactivating proteins. The uptake of these proteins by the host cell is mediated through the ricin-B lectin domain which binds to specific glycan structures on the cell surface. The ricin-B related proteins accumulate in the vacuole or are secreted to the extracellular space [3]. Similar to the jacalin-related proteins, proteins containing a ricin-B domain are widespread and have also been identified outside the plant kingdom in animals, fungi and bacteria. These lectins are classified as R-type lectins [108]. Although the carbohydrate specificity of the ricin-B lectins can vary [109], evidence supports that many of the ricin-B related lectins play a role in defense against pathogens and insects [40].

5. Adaptive Immunity and Epigenetic Imprinting

5.1. Adaptive Immunity

In vertebrates an adaptive immune response is induced following the innate immune response. When naive T-cells interact with their specific antigen presented on the surface of antigen-presenting cells, they become activated and differentiate into effector T-cells. A subset of these effector T-cells, the T-helper cells, will subsequently stimulate the proliferation and differentiation of B-cells.

The adaptive immune response depends on the innate immune response. It was observed that several PRRs, involved in the innate immunity, also play a role in triggering the adaptive response. Toll-like receptors, nucleotide-oligomerization domain (Nod)-like receptors, retinoic acid-inducible gene-1 (RIG-1)-like receptors and some C-type lectin receptors can trigger both the innate and adaptive immune responses [110]. Whether these receptors contribute to the innate or the adaptive response is dependent of the cell type where they are expressed [111]. Moreover, lectins can have similar functions in the adaptive response and in the innate response. They play a role in cell-cell interactions and can contribute to the activation and differentiation of cells (e.g., Siglecs).

5.2. Epigenetic Imprinting

Plants, unlike animals, lack the adaptive immune mechanisms and rely solely on their multilayered innate immune system to prevent pathogen infection. However, plants are not completely defenseless against reoccurring stresses.

When higher animals are exposed to a specific pathogen, the adaptive immune system creates an immunological memory after the initial response. This leads to an enhanced response to subsequent encounters with that pathogen. Although there is no adaptive immune system in plants, plants must respond and adapt to recurring biotic and abiotic stress factors especially since their sessile lifestyle does not allow to escape from these stresses. It is a well-known phenomenon that stress tolerance in plants can be improved through pre-exposure to an abiotic or biotic stress stimulus. Primed (against biotic stress) or hardened (against abiotic stress) plants display a faster and/or stronger activation to the various defense responses that are induced following infection by pathogens, attack by insects or in response to abiotic stress [112]. This priming can be elicited by either exposure to the stress clues themselves or by the exogenous application of chemical treatments [113]. The latter process can be compared to
vaccination in vertebrates. The advantage of priming is that it offers the plant an enhanced protection without the costs of constitutively expressing their defense genes. However, the molecular mechanism of this stress-imprinting on plant immunity is not fully elucidated. Though several mechanisms have been proposed, Conrath et al. [112] suggested that two potential mechanisms can be involved. The first one involves the accumulation of signaling proteins in an inactive configuration that are activated upon stress, potentially through MAPK signaling [112,114]. The second mechanism involves the accumulation of transcription factors that enhance defense gene transcription after stress recognition [112]. Jung et al. [115] proposed a system in which the accumulation of secondary metabolites such as SA and azelaic acid was involved. Another mechanism was described by Bruce et al. [113] in which priming occurs through epigenetic changes. Epigenetics refers to the heritable changes in gene expression that do not involve changes to the DNA sequence. These epigenetic changes consist of DNA methylations, histone modifications and RNA interference that lead to changes in gene expression by activation or silencing. The links and interactions between the different mechanisms for priming are largely unknown but due to the long-lasting stress resistance, it is most likely that epigenetic mechanisms are largely in control [116].

Epigenetic changes in response to cold stress have been reported during vernalisation of Arabidopsis. The long-term exposure to low temperatures in winter is memorized and needed for the plant to be able to flower in the following spring [117]. This ‘memory’ involves a change in the chromatin structure of the flowering locus through modification of histones by several vernalisation (VRN) genes. Recent data suggest that also some of the stress inducible plant lectins play a role in epigenetic priming during stress response. One of the proteins involved in vernalisation in wheat is the VER2 cold inducible jacalin-related protein. Interaction with the O-GlcNAc modified TaGRP2 RNA-binding protein allows accumulation of the VRN1 transcription factor through stabilization of its mRNA [118]. Treatment of tobacco with JA, a plant hormone involved in signaling of biotic and abiotic stress and in plant development, leads to the enhanced expression of the Nictaba protein in the nucleus and the cytosol [119,120]. In addition, this lectin is also induced upon insect herbivory [120]. The accumulation of Nictaba is not limited to the treated leaf but a systemic response is observed in all leaves of the plant [121]. Screening for interaction partners revealed that the core histones H2A, H2B and H4 are the major binding partners for Nictaba in the nucleus [122]. This interaction is established through the specific binding of the lectin with O-GlcNAc modified histone proteins [123]. It was hypothesized that the interaction between Nictaba and histones bound to DNA might result in chromatin remodeling, and subsequent alteration of gene expression as a response to stress or changing environmental conditions [124].

6. Conclusions

Although plants and animals shared their last common ancestor until at least one billion years ago, throughout the eons they have been facing common threats such as bacteria, fungi and viruses trying to invade them. To protect themselves against these threats, both plants and animals evolved a defense system, the immune system. While for a long time it was presumed that these defense systems in plants and animals were very different, it has become clear that both plants and animals respond to infection by pathogens using similar regulatory modules. Common features include receptors that detect molecular
signatures of infectious organisms (PAMPs/MAMPs) and of tissue damage (DAMP), conserved signal transduction pathways (e.g., MAPK pathway) and the production of antimicrobial molecules.

Lectins or carbohydrate-binding proteins play an important role in the immune system of animals and plants as PRRs. Carbohydrate structures on the surface of pathogens or released from host cells due to damage provoked by the pathogen are recognized by soluble or membrane-bound lectins that trigger a signaling cascade resulting in the induction of the defense mechanisms, the pathogen/microbial triggered immunity (P/MTI). Although in animals several lectins were found to exert a function as PRR, in plants only the LysM-domain lectins have been shown unambiguously to be dependent on carbohydrate binding for their PRR function. Other identified LecRKs or lectin receptor proteins may depend on protein-protein interactions for the recognition of their target.

In addition to the lectins with a role as PRR, lectins can play a role in stress signaling or as an effector or mediator in the stress response. In plants, several lectin families have been described that are transcriptionally upregulated upon biotic or abiotic stresses. These stress-inducible lectins mainly reside in the cytoplasm and the nucleus, and are suggested to play a role in signal transduction in several stress response pathways.

Successful pathogens manage to overcome the innate immune response through secretion of toxins or effectors into the host cell. Plants have therefore developed another defense layer to detect these effectors via different immune receptors (resistance or R proteins) and initiate effector triggered immunity (ETI). Another method for successful pathogens to overcome the immune response is to escape degradation by the host cell. Recognition of pathogens by PRRs often leads to the phagocytosis of these pathogens by the host cell. These phagocytic vesicles are subsequently fused with lysosomes, resulting in the degradation of the pathogen. This degradation is required in vertebrate systems to activate the adaptive immunity by antigen presenting cells. However, some pathogens are capable of escaping this degradation by the delivery of effector proteins that prevent maturation of the phagocytic vacuoles. These pathogens create a niche inside these (damaged) vesicles or manage to escape these vesicles and proliferate intracellularly. Plant and animal cells have developed a mechanism to defend themselves against these intracellular parasites through autophagy. Lectins are shown to be involved in the recognition of damaged phagocytic vacuoles in animals. Since in plants the essential machineries of autophagy are conserved, lectins are suggested to play a role in this process, but at present their precise function is not fully elucidated.

Although plants lack an adaptive immune system, they are able to create long lasting stress resilience against reoccurring stresses. Epigenetic changes can facilitate quicker and more potent responses to subsequent attacks. One of the stress inducible plant lectins from tobacco, referred to as Nictaba, was shown to interact with O-GlcNAc modified histones and can potentially act as a chromatin remodeler allowing an alteration of gene expression as a result of biotic or abiotic stresses.

Although plants and animals differ significantly, their defense system relies for an important part on comparable mechanisms. This is supported by recent data which demonstrated that lectin domains that descended from common ancestors through divergent evolution appear widely across the kingdoms of life (including prokaryotes, fungi, plants and animals) [13]. However these conserved lectin domains show promiscuity in their sugar binding specificity. These variations might be linked to a directed evolution to recognize carbohydrate structures on the pathogenic agents threatening the plant/animal cell.
Despite our growing knowledge and insight into the world of lectins, the number of fully characterized lectins for which the physiological relevance has been resolved is still rather low. The presence of lectin domains as part of chimeric proteins increases the structural diversity among lectins but will also expand the functional variation within a lectin family. Future research will have to elucidate the importance of protein carbohydrate recognition for different biological processes and the key role of lectin domains in these events.

Acknowledgments

This work was funded primarily by the Fund for Scientific Research-Flanders (FWO grant G006114N) and the Research Council of Ghent University (project BOF15/GOA/005).

Author Contributions

K.D.S. wrote the first draft of the manuscript. E.V.D. supervised the manuscript and was involved manuscript corrections and discussions.

Conflicts of Interest

The authors declare no conflict of interest.

References


© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).
Chapter 24, Essentials to Glycobiology (3rd edition)

Viridiplantae and Algae
Chapter 24: VIRIDIPLANTAE AND ALGAE

Malcolm A. O’Neill, Alan G. Darvill, Marilyn E. Etzler, Debra Mohnen, and Serge Perez

<table>
<thead>
<tr>
<th>Old #</th>
<th>New #</th>
<th>Text File Name</th>
<th>Original (Words)*</th>
<th>Current (Words)*</th>
<th>% Cut</th>
<th>Original Figures</th>
<th>Current Figures</th>
<th>Original Tables</th>
<th>Current Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>24</td>
<td>3edch24Final</td>
<td>4418</td>
<td>5474</td>
<td>+24 %</td>
<td>7</td>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

3edch24f1MOD.tif (MOD= modified, from 2nd ed Fig 1Ch22.)
3edch24f2MOD.tif (MOD= modified, from 2nd ed Fig 2Ch22.)
3edch24f3MOD.tif (MOD= modified, from 2nd ed Fig 3Ch22.)
3edch24f4MOD.tif (MOD= modified, from 2nd ed Fig 4Ch22.)
3edch24f5MOD.tif (MOD= modified, from 2nd ed Fig 5Ch22.)
3edch24f6MOD.tif (MOD= modified, from 2nd ed Fig 6Ch22.)
Recent research on plant glycan structure and function has typically emphasized model plants such as Arabidopsis and plants of commercial importance. However, there is increasing interest in studying the glycans produced by plants from all the major orders of the viridiplantae. Such studies, together with the availability of transcriptomic data for numerous green algae and land plants, have begun to reveal a rich diversity in glycan structures and insight into how some of these structures have changed during the evolution of the viridiplantae. In this chapter we provide an overview of the current knowledge of green plant glycan structures with an emphasis on the features that are unique to land plants.

INTRODUCTION

Viridiplantae (green plants) are a clade of photosynthetic organisms that contain chlorophylls a and b, produce and store their photosynthetic products inside a double-membrane-bounded chloroplast, and have cell walls that typically contain cellulose. The Viridiplantae are comprised of two clades - the Chlorophyta and the Streptophyta. The Chlorophyta contain most of the organisms typically referred to as “green algae”. The term "algae" is also used for several other groups of photosynthetic eukaryotes including Diatoms and the red, brown, golden and yellow-green algae. The Streptophyta comprise several other lineages that are also referred to as “green algae" and the land plants. Land plants include the liverworts, mosses, hornworts, lycopods, ferns, gymnosperms and flowering plants.

PLANT GLYCAN DIVERSITY

Green plants synthesize diverse glycans that vary in their structural complexity and molecular size. Raffinose oligosaccharides (raffinose, stachyose, and verbascose) are almost ubiquitous in the plant kingdom and rank second only to sucrose in abundance as soluble carbohydrates. These oligosaccharides are derivatives of sucrose that contain one or more α-Gal residues. Plants also synthesize numerous low molecular weight glycoconjugates that contain either aromatic (e.g. phenolic glycosides) or aliphatic (glycolipids) aglycones.

Plant polysaccharides are linear or branched polymers composed of the same or different monosaccharides. For example, cellulose is composed of 1-4-linked β-D-Glc residues (Figure 24.1A), whereas the structurally complex plant cell wall pectic polysaccharide referred to as rhamnogalacturonan II (RG-II) contains 12 different monosaccharides linked together by up to
21 different glycosidic linkages (Figure 24.2). Plant proteoglycans are structurally diverse glycans in which carbohydrate accounts for up to 90% of the molecule and is O-linked to the protein via hydroxyamino acids (Figure 24.3). Plant glycoproteins typically contain 15% or less of carbohydrate in the form of N-linked oligomannose, complex, hybrid, and paucimannose oligosaccharides (Figure 24.4). Land plants also form O-GlcNAc-modified nuclear and cytosolic proteins (Chapter 19).

**NUCLEOTIDE SUGARS - THE BUILDING BLOCKS**

Nucleotide sugars are the donors used for synthesis of glycoconjugates and glycosylated secondary metabolites (Chapter 5). In plants, the majority of these activated monosaccharides exist as their nucleotide-diphosphates (e.g. UDP-Glc), although at least one monosaccharide, Kdo, exists as its cytidine monophosphate derivative (CMP-Kdo). Nucleotide sugars are formed from the carbohydrate generated by photosynthesis, from the monosaccharides released by hydrolysis of sucrose and storage carbohydrates, and by recycling monosaccharides from glycans and the cell wall. Nucleotide sugars are also formed by interconverting pre-existing activated monosaccharides. To date 30 different nucleotide-sugars and at least 100 genes encoding proteins involved in their formation and interconversion have been identified in plants.

**PLANT GLYCOSYLTRANSFERASES AND GLYCAN-MODIFYING ENZYMES**

Plant genomes contain numerous genes encoding proteins involved in the synthesis and modification of glycans. These proteins are spread across many enzyme classes in the Carbohydrate Active enZymes (CAZy) database (Table 24.1). Many of these proteins may be involved in the formation and modification of the polysaccharide-rich cell wall. Indeed, the unicellular alga Ostreococcus tauri, which is one of the few plants that does not form a cell wall, has a much smaller number of genes predicted to be involved in glycan metabolism.

**PLANT CELL WALLS**

A substantial portion of the carbohydrate formed by photosynthesis is used to produce the polysaccharide-rich walls that surround plant cells. Primary and secondary cell walls are distinguished by their composition, architecture and functions. A primary wall surrounds growing and dividing plant cells and non-growing cells in the soft tissues of fruits and leaves.
These walls are capable of controlled extension to allow the cell to grow and expand yet are sufficiently strong to resist the cells internal turgor pressure. A much thicker and stronger secondary wall is often formed once a cell has ceased to grow. This wall is deposited between the plasma membrane and the primary wall and is composed of layers that differ in the orientation of their cellulose microfibrils. The secondary walls of vascular tissues involved in the movement of water and nutrients through the plant are further strengthened by the incorporation of lignin. The ability to form conducting tissues with lignified and rigid secondary walls was an indispensable event in the evolution of vascular land plants, as it facilitated the transport of water and nutrients and allowed extensive upright growth. Secondary cell walls account for most of the carbohydrate in plant biomass being considered as feedstock for the production of biofuels and value-added chemicals (Chapter 59).

**PRIMARY CELL WALL GLYCANS**

Primary cell walls are composites that resemble fiber-reinforced porous gels. The complex structures and functions of these walls result from the assembly and interactions of a limited number of structurally-defined polysaccharides and proteoglycans. Wall structure and organization may change during cell division and development and in response to biotic and abiotic challenges by the differential synthesis and modification of the non-cellulosic components or by the addition of new components.

Primary walls of land plants contain cellulose, hemicellulose and pectin, in different proportions. They also contain structural proteins/proteoglycans, enzymes, phenolics and minerals. Pectin and hemicellulose are present in approximately equal amounts in the so-called type I primary walls of gymnosperms, dicots and non-graminaceous monocots, whereas hemicellulose is far more abundant than pectin in the type II walls of the grasses. Much less is known about the composition of the walls of avascular plants and green algae, although it is likely that the walls of these plants contain cellulose, pectin and hemicellulose.

**Cellulose**

Cellulose, the most abundant biopolymer in nature, is a linear polymer composed of 1-4 linked \(\beta-D\)-Glc residues (Figure 24.1A). Several of these chains are hydrogen bonded to one another to form paracrystalline microfibrils. Each microfibril is predicted to contain between 18 and 24
glucan chains. The glucan chain is synthesized by a cellulose synthase (CESA) complex that exists as a hexameric rosette structure on the cell's plasma membrane. Three CESAs, encoded by three different genes, are believed to interact to form a trimeric complex, which in turn assembles into a hexameric rosette. The catalytic site of each cellulose synthase is located in the cytosol and transfers glucose from UDP-Glc onto the elongating glucan chain. The mechanisms involved in the formation of a microfibril from individual glucan chains are not well understood, although it may involve a self-assembly process that is facilitated by specific proteins. The newly formed microfibrils are then deposited in the wall of a growing cell with an orientation that is transverse to the axis of elongation. This orientation may be guided in part by protein-mediated interactions between CESA proteins and cortical microtubules.

Naturally occurring cellulose is a mixture of two crystalline forms, Iα and Iβ, together with surface chains and less crystalline material. The Iα and Iβ polymorphs differ mainly in the packing arrangement of their hydrogen-bonded sheets. Many properties of native cellulose depend on interactions that occur at the surface of the microfibrils. The surface chains are accessible and reactive whereas the hydroxyl groups of the internal chains in the crystal participate in extensive intra- and inter-molecular hydrogen bonding. Cellulose is insoluble in water and somewhat resistant to hydrolysis by endo- and exo-glucanases because of this highly packed arrangement of the glucan chains.

Several types of enzymes including endoglucanases, cellobiohydrolases, and β-glucosidases are required for cellulose depolymerization. Many of these enzymes consist of a catalytic domain connected to a cellulose-binding module. This module facilitates binding of the enzyme to the insoluble substrate. Some microorganisms also produce copper-dependent oxidases that render crystalline cellulose more susceptible to hydrolysis. Cellulases and other enzymes involved in cellulose hydrolysis often exist as macromolecular complexes referred to as cellulosomes. Improving the effectiveness of cellulosomes is an area of active research, to increase the conversion of plant biomass to fermentable sugar (Chapter 59).

**Hemicelluloses**

Hemicelluloses are branched polysaccharides with a backbone composed of 1-4-linked β-D-pyranosyl residues with an equatorial O-4 (Glc, Man, and Xyl). Xyloglucan, glucuronoxylan, arabinoxylan and glucomannan (Figure 24.1B-E) are included under this chemical definition of
Hemicelluloses. Hemicelluloses and cellulose have structural and conformational similarities that allow them to form strong, noncovalent associations with one another in the cell wall, although the biological significance of these interactions is a subject of debate.

Xyloglucans are distinguished from one another by the number of 1-4-linked β-D-GlcP backbone residues that are branched. XXXG-type xyloglucans are composed of subunits in which three consecutive backbone residues bear an α-Xyl substituent at O-6 and a fourth, unbranched backbone residue (Figure 24.1B). Each Xyl residue (X side chain) may itself be extended by the addition of one or more monosaccharides including β-D-Gal, α-L-Fuc, α-L-Ara and β-D-GalA. Eighteen structurally unique sidechains have been identified to date, although only a subset of these are synthesized by a single plant species. XXXG-type xyloglucan is present in the primary walls of hornworts, lycopods, ferns, gymnosperms, a diverse range of dicots and all monocots with the exception of the grasses. XXGG-type xyloglucans, which have only two consecutive backbone residues bearing an α-D-Xyl substituent at O-6, are present in the primary walls of the grasses, some Solanaceae, mosses and liverworts. The Xyl may be extended by the addition of Gal, Ara or GalA, but rarely if ever with Fuc.

Early models of dicot primary walls predicted that xyloglucan acted as tethers between cellulose microfibrils and that controlled enzymatic cleavage of the xyloglucan facilitated wall expansion and thus plant cell growth. However, the identification of an Arabidopsis mutant that is unable to synthesize xyloglucan yet exhibits near normal growth and development challenged this notion and led to the suggestion that pectin may have a more important role in controlling wall expansion than previously believed.

Glucuronoxylan (GAX, Figure 24.1C) is the predominant non-cellulosic polysaccharide in the type II walls of the grasses. Its backbone is composed of 1-4-linked α-D-Xyl residues many of which are substituted at O-3 with α-L-Araf residues. These Araf residues may be further substituted at O-2 with an α-L-Araf or a β-D-Xylp residue. A small number of the backbone residues are substituted at O-2 with α-D-GlcP and its 4-O-methylated counterpart (MeGlcP).

The presence of 1-3, 1-4-linked β-glucans (also referred to as mixed-linkage glucans) in the walls of grasses was once considered to be a unique feature of these plants. However, structurally related mixed-linkage β-glucans have also been identified in the walls of Selaginella
(lycopod) and Equisetum (horrsetails), although the evolutionary relationship between these β-glucans is not known.

**Pectins**

Pectins are structurally complex polysaccharides that contain 1-4-linked α-GalA. Three structurally distinct pectins - homogalacturonan, substituted galacturonan, and rhamnogalacturonan - have been identified in plant cell walls (Figure 24.2). Homogalacturonan, which may account for up to 65% of the pectin in a primary wall, is composed of 1-4-linked α-GalA. The carboxyl group may be methyl-esterified and the glucose itself may be acetylated at O-2 or O-3. The extent of methyl-esterification is controlled by pectin methyl-esterases present in the wall and affects the ability of homogalacturonan-containing glycans to form ionic calcium cross-links with themselves and with other pectic polymers. Such interactions alter the mechanical properties of the wall and may influence plant growth and development.

Rhamnogalacturonans are polysaccharides with a backbone composed of GalA and rhamnose (Rha) residues in the repeating disaccharide 4-α-D-GalpA-1-2-α-L-Rhap-1. Many of the GalAs are acetylated at O-2 and/or O-3. Depending on the plant, between 20 and 80% of the Rha residues may be substituted at O-4 with linear or branched side chains composed predominantly of Ara and Gal, together with smaller amounts of Fuc and GlcA (Figure 24.2). Little is known about the functions of these side chains and their contribution to the properties of the primary wall.

Substituted galacturonans have a backbone composed of 1-4-linked α-D-GalA acid residues that are substituted to varying degrees with mono-, di- or oligosaccharides. Xylogalacturonans contain single β-Xyl residues linked to O-3 of some of the backbone residues (Figure 24.2) whereas apiogalacturonans have β-D-apiose (Api) and apiobiose linked to O-2 of some of the backbone residues. Apiogalacturonans have only been identified in the walls of duckweeds and seagrasses.

Rhamnogalacturonan-II, which accounts for between 2% and 5% of the primary cell wall, is the most structurally complex polysaccharide yet identified in nature. It is composed of 12 different linked together by up to 21 different glycosidic linkages (Figure 24.2). Four structurally different sidechains are attached to the galacturonan backbone. Two structurally conserved disaccharides
(sidechains C and D) are linked to O-2 of the backbone. The A and B sidechains, which contain between 7 and 9 monosaccharides, are linked to O-3 of the backbone. Several of the monosaccharides in RG-II are O-methylated and/or O-acetylated.

Virtually all of the RG-II exists in the primary wall as a dimer cross-linked by a borate ester. The ester is formed between the Api residue in side chain A of each RG-II monomer (Figure 24.2). The dimer forms rapidly in vitro when the RG-II monomer is reacted with boric acid and a divalent cation. However, the mechanism and site of dimer formation in planta has not been determined. Borate cross-linking of RG-II is likely to have substantial effects on the properties of pectin and the primary wall as RG-II is itself linked to homogalacturonan (Figure 24.2). Indeed, mutations that affect RG-II structure and cross-linking result in plants with abnormal walls and severe growth defects. Swollen primary walls and abnormal growth together with reduced RG-II cross-linking are also a characteristic of boron deficient plants.

Pectin is believed to exist in the cell wall as a macromolecular complex comprised of structural domains - homogalacturonan, rhamnogalacturonan and substituted galacturonan - that are covalently linked to one another. However, there is only a limited understanding of how these structural domains are organized (Figure 24.2). Molecular modeling of a pectin (~50 kDa) containing homogalacturonan (degree of polymerization ~100) and rhamnogalacturonan with arabinogalactan sidechains together with modeling of RG-II conformation have begun to provide insights into the conformations and relative dimensions of each pectin structural domain.

The conformation of the homogalacturonan chain is largely unaffected by the conformation at the glycosidic linkage, by changes in its degree of methyl-esterification or the presence of counter ions. The homogalacturonan region has a persistence length of ~20 GalA residues, which is likely to be sufficient to stabilize junction zones formed with Ca$^{++}$. In vitro studies suggest that the maximum stability of such junction zones is obtained with oligomers containing ~15 non-esterified GalA residues. Thus controlling the distribution of methyl-ester groups along the homogalacturonan backbone provides a mechanism to regulate the physical properties of pectin, including its ability to form gels. For a gel to form and not to be brittle, other features including sequences that interrupt inter-chain associations in the pectin macromolecule may be important. For example the structural diversity and the conformational flexibility of the oligosaccharide sidechains of the rhamnogalacturonan domain will limit or prevent inter-chain pairing. The
The presence of 1-2-linked Rha residues does not introduce kinks into the backbone geometry of rhamnogalacturonan and thereby limit inter-chain associations. Rather, it is the side chains linked to these residues that are responsible for preventing or limiting inter-chain associations.

The conformations of the four side chains attached to the homogalacturonan backbone may lead RG-II to adopt a “disk-like” shape. Well-defined tertiary structures are predicted for the monomer and the dimer. In the dimer, borate-ester cross-linking and Ca$^{++}$ inter-chain pairing further stabilizes the two disks. The apparent resistance of RG-II to wall modifying enzymes together with the formation of a cation-stabilized RG-II dimer likely results in a structure that resists temporal changes. By contrast, homogalacturonan is continually modified by the action of wall enzymes and its contribution to wall architecture is therefore time dependent.

Increased knowledge of the physical properties of primary wall polysaccharides and proteoglycans is required to understand how modulating the amounts and structural features of a few common polysaccharides and glycan domains lead to primary walls with diverse properties and functionalities. Further research is also needed to determine if wall structure and function results from the non-covalent interactions of polysaccharides and proteoglycans or from the formation of glycan-containing architectural units with specific structural and functional roles. The later scenario is analogous to the organization of proteoglycans and O-linked mucins in the extracellular matrix of animal cells (Chapters 10, 16 and 17).

**PLANT SECONDARY CELL WALL GLYCANS**

The secondary walls of woody tissue and grasses are composed predominantly of cellulose, hemicellulose, and lignin. The inclusion of lignin results in a hydrophobic composite that is a major contributor to the structural characteristics of secondary walls.

Heteroxylans are the major hemicellulosic polysaccharide present in the secondary (lignified) cell walls of flowering plants. These heteroxylans are classified according to the type and abundance of the substituents on the 1-4-linked β-D-Xylp residues of the polysaccharide backbone. Glucuronoxylans (GX), which are major components in the secondary walls of woody and herbaceous eudicots, have a α-D-GlcA or MeGlcA substituent at O-2 (Figure 24.1C). Gymnosperm secondary walls contain arabinoglucuronoxylans (AGXs), which in addition to MeGlcA substituents, have Araf residues attached to O-3 of some of the backbone residues. The
GAX in the secondary walls of grasses typically contain less Ara/f residues than their primary wall counterpart (Figure 24.1D). Ferulic or coumaric acids are often esterified to the Ara/f residues of GAX in grass primary and secondary cells walls.

Eudicot and gymnosperm secondary wall GX and AGX have a well-defined glycosyl sequence -4-β-D-Xylp-1-3-α-L-Rhap-1-2-α-D-GalpA-1-4-4-Xylp at their reducing end (Figure 24.1C). This sequence is required for normal xylan synthesis during secondary cell wall formation and may have a role in regulating the polymers chain length. This sequence is present at the reducing end of heteroxylans of all monocots except the grasses.

Hemicellulose and Pectin Biosynthesis

Genes that encode polysaccharide biosynthetic enzymes, including many of those required for xyloglucan, glucuronoxylan, arabinoxylan and cellulose synthesis and some of those required for pectin synthesis have been identified. This information, together with improved methods to generate recombinant plant GTs with high enzymatic activity is providing a framework for an increased understanding of how plant cell wall polysaccharides are synthesized. There is general consensus that cellulose is synthesized by GT complexes localized at the plasma membrane and that most pectins and hemicellulose are synthesized in the Golgi apparatus. Members of the cellulose synthase-like gene families, CSLF and CSLH are likely involved in 1-3, 1-4-linked β-glucan biosynthesis. However, there is considerable debate about whether this polysaccharide is formed in the Golgi apparatus or at the plasma membrane.

Despite advances in understanding how polysaccharides are synthesized we still do not know how many of the wall polymers are synthesized by Golgi-localized multi-enzyme complexes or if they are assembled by GTs localized in different regions of the Golgi apparatus. We also do not understand how the newly synthesized polymers are assembled into a functional cell wall.

Plants Produce Proteoglycans Containing O-Linked Oligosaccharides and O-Linked Polysaccharides

Plants produce glycoproteins and proteoglycans that contain oligo- or polysaccharides that are linked to hydroxyproline (Hyp) and serine (Ser). Hyp is formed post-translationally by endoplasmic reticulum-localized prolyl hydroxylases and is O-glycosylated in the ER and in the Golgi apparatus. The degree and type of Hyp glycosylation is determined to a large extent by the
protein's primary sequence and the arrangement of Hyp residues. Hyp glycosylation is initiated by the addition of an Ara or a Gal residue. Contiguous Hyp residues are arabinosylated, whereas clustered but noncontiguous Hyp residues are galactosylated. Ser residues and occasionally Thr residues may also be O-glycosylated in these proteins.

Three structurally distinct plant proteoglycans containing glycosylated Hyp and Ser – the extensins, proline/hydroxyproline-rich proteoglycans and arabinogalactan proteins – have been identified. Extensins are hydroxyproline-rich proteoglycans with Ser(Hyp)₄ repeat sequences and contain between 50 and 60% (w/w) glycan. Most of the carbohydrate exists as oligosaccharides containing one to four Ara residues linked to Hyp together with a small number of single Gal residues α-linked to Ser. The proline/hydroxyproline-rich proteoglycans, which contain from 3% to 70% (w/w) carbohydrate, are distinguished from the extensins by amino acid sequence. Both of these families of HRGPs likely have a structural role in the cell wall. The expression of genes involved in their synthesis is developmentally regulated and is often induced by wounding and fungal attack of plant tissues.

The glycopeptide signaling molecules PSY1, CLE2 and CLV3 contain arabinosylated hydroxyproline and have numerous roles in plant growth and development. A highly glycosylated Hyp-rich domain in which three or four Ara residues are attached to Hyp and a single Gal residue is linked to Ser is present in potato, tomato, and thorn-apple lectins.

Arabinogalactan proteins (AGPs) have a glycan content of up to 90% (w/w). Chains of between 30 and 150 monosaccharides are linked to the protein by Gal-O-Ser and Gal-O-Hyp linkages. These chains have a 1-3-linked β-Gal backbone that is extensively substituted at O-6 with sidechains of 1-6-linked β-Gal. These sidechains are terminated with Ara, GlcA and Fuc residues. Some AGPs may contain homogalacturonan, RG-I and xylan covalently linked to the arabinogalactan (Figure 24.3) thereby forming a protein-hemicellulose-pectin complex referred to as APAP1. The location of this complex in the plant and its biological function remains to be determined.

Several AGPs are secreted into the cell wall whereas others are linked to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor. Plant GPI anchors contain a phosphoceramide core. The glycan portion of the GPI anchor of pear cell AGP has the sequence α-d-Man1-2α-d-Man1-6-α-d-Man1-4-GlcN-inositol. At least 50% of the Man attached to the GlcN (see
Chapter 12) is itself substituted at O-4 with a β-Gal, a feature that may be unique to plants. Many functions have been proposed for the AGPs including their participation in signaling, development, cell expansion, cell proliferation, and somatic embryogenesis.

**THE N-LINKED GLYCANS OF PLANT GLYCOPROTEINS HAVE UNIQUE STRUCTURES**

Many of the proteins that have passed through the plant secretory system contain N-linked oligomannose, complex, hybrid, or paucimannose-type glycans (Figure 24.4). The initial stages of the synthesis of these N-glycans, including the transfer of the oligosaccharide precursor from its dolichol derivative and the control of protein folding in the endoplasmic reticulum are comparable in plants and animals (Chapter 9). However, two modifications of N-glycans during passage through the Golgi are unique to plants.

Oligomannose-type N-glycans are often trimmed in the cis-Golgi and then modified in the medial-Golgi by N-GlcNAc transferase I (GnT-I) catalyzed addition of GlcNAc to the distal Man of the core. In reactions that are unique to plants, a β-Xyl is often added to O-2 of the core Man. In the trans-Golgi, α-Fuc may be added to O-3 of the GlcNAc residue that is itself linked to asparagine (Figure 24.5). The XylT and FucT that catalyze these reactions act independently of one another but do require at least one terminal GlcNAc residue for activity. The FucT is related to the Lewis FucT family, whereas the XylT is unrelated to other known GTs.

The xylosylated and fucosylated N-glycans are often trimmed by α-mannosidase II. A second GlcNAc may then be added by GnT-II. Some plant N-glycans do not undergo further mannose trimming and proceed through the Golgi as hybrid-type N-glycans. Complex and hybrid-type N-glycans may be further modified by the addition of Gal and Fuc in the trans-Golgi. Plant glycoproteins are either secreted from the cell or transported to the vacuoles. Many of the glycoproteins present in the vacuoles contain paucimannose type glycans, suggesting that they are trimmed by vacuolar glycosidases (Figure 24.5).

The presence of sialic acid in the N-glycans of plant glycoproteins was claimed but likely represented environmental contamination. Plants do have genes that encode proteins containing sequences similar to sialyltransferase motifs but their functions have not been established.
ALGAL GLYCANS

Only a few glycans of green algae have been studied in detail. For example, the cell wall of *Chlamydomonas reinhardii* is a crystalline lattice formed from hydroxyproline-rich glycoproteins. A sulfated polysaccharide composed of Kdo and GalA is the major glycan in the cell walls of *Tetraselmis striata*. Some but not all of the glycans present in the cell walls of land plants may also be present in the cell walls of some green algae. Several of the polysaccharides produced by red and brown algae are used in the food industry as gelling agents, stabilizers, thickeners and emulsifiers. They are also used in paints, cosmetics and in paper and as reagents for scientific research.

Agarose (agar) and carrageenan are sulfated galactans obtained from red seaweeds. The polysaccharide is comprised of the repeating disaccharide 3-β-D-Galp-1-4-3,6-anhydro-α-L-Galp-1, unit. Some of the D-Gal and L-Gal units are O-methylated. Pyruvate and sulfate groups may also present in small quantities. There are three main types of carrageenan - κ-carrageenan has one sulfate group per disaccharide, τ-carrageenan has two sulfates per disaccharide, and λ-carrageenan has three sulfates per disaccharide.

Alginic acid, a linear polysaccharide composed of 1-4-linked β-D-ManA and its C-5 epimer 1-4-linked α-L-GulA, is obtained from various species of brown seaweed. These monosaccharides are typically arranged in blocks of either ManA or GulA separated by regions comprised of 4-ManpA-1-4-GulpA-1 sequences.

Brown seaweeds produce polysaccharides that have potential in the treatment of diseases. Laminaran, a linear storage polysaccharide composed of 1-3 and 1-6-linked β-D-Glc residues. There are reports that laminaran has anti-apoptotic and anti-tumor activities. Fucoidans are a group of sulfated polysaccharides isolated from several brown algae that have been reported to have anticoagulant, anti-tumor, anti-thrombosis, anti-inflammatory and antiviral properties. Fucoidans have a backbone of 1-3-linked α-Fuc that is substituted at O-2 with fucose and at O-4 with sulfate or fucose. Other fucoidans have backbones of alternating 1-3- and 1-4-linked α-Fuc residues.

PLANT GLYCOLIPIDS
Glycoglycerol lipids are the most abundant glycolipids in plants. Mono- and digalactosyldiacylglycerol have been identified in all plants, whereas tri- and tetragalactosyldiacylglycerol have a more restricted taxonomic distribution (Figure 24.6). The synthesis of these galactolipids is initiated by the formation of diacylglycerol in the endoplasmic reticulum (ER) membrane and the chloroplast membrane. Galactolipids formed in the ER membrane contain predominantly C16 fatty acids at the sn2 position and C18 fatty acids at the sn3 position. The chloroplast pathway produces C18 fatty acids at both positions. Each of these fatty acids is then desaturated to 16:3 or 18:3 acyl groups. Monogalactosyldiacylglycerol (MGDG) is synthesized by the transfer of Gal from UDP-Gal to diacylglycerol by an MGDG synthase. Digalactosyldiacylglycerol (DGDG) is formed from MGDG by the transfer of Gal from UDP-Gal by a DGDG synthase. These reactions occur primarily in the outer chloroplast membrane. The products are then transported to the inner membrane and the thylakoid membranes of the chloroplast. The presence and abundance of MGDG in the chloroplast thylakoid membrane is important for normal photosynthesis to occur. Sulfoquinovosyldiacylglycerol, which is formed from diacylglycerol is also abundant in the thylakoid membrane, and may also have a role in photosynthesis.

Small amounts of MGDG and DGDG are present in the plasma membranes of cells, although the mechanism of galactolipid exchange among the membranes is not understood. Inositolphosphoceramides that contain GlcN and GlcA have been identified in plants, although their function remains to be determined. Glycosphingolipids containing Gal and Fuc (Chapter 11) are present in plant cell plasma membranes, but none have been fully characterized. No gangliosides have been identified in plants. There is however evidence that Kdo-containing lipids with homology to bacterial Lipid A are present in plant organelle membranes.

**OTHER PLANT GLYCOCONJUGATES**

Plants produce numerous phenolics, terpenes, steroids, and alkaloids that are collectively referred to as secondary metabolites. Many of these compounds are O-glycosylated or contain sugars linked via N, S or C atoms. Glycosylated secondary metabolites often have important roles in a plants response to biotic and abiotic challenges and may also have value as pharmaceuticals.

In general, the addition of a single sugar or an oligosaccharide may increase water solubility, enhance chemical stability, or alter both chemical and biological activity. For example, the
activity of several plant hormones may be regulated by converting them to their glucose esters or their glucosides. Digoxin and oleandrin are potent cardiac glycosides isolated from Foxglove and Oleander, respectively. Myrosinase-catalyzed cleavage of S-linked Glc from glucosinolates leads to the formation of pungent mustard oils when mustard and horseradish are damaged. The steviol glycosides, which are far sweeter than sucrose, are used as natural sugar substitutes. The bitter taste of citrus fruits is due to naringin, a glycosylated flavanoid.

**PLANT MUTANTS PROVIDE CLUES TO GLYCAN FUNCTION**

The availability of plant lines carrying mutations in specific genes has yielded considerable insight into glycan biosynthesis and function. Arabidopsis is widely used as a model dicot as it is easy to grow, has a short life cycle and its relatively small genome has been sequenced and extensively annotated. Numerous databases and resources are available for Arabidopsis, most notably TAIR, which provides access to a large number of plants carrying chemically induced or T-DNA insertion mutations in genes of both known and unknown function. Collections of other plant mutant lines including soybean and Brachypodium exist, but are not yet as well developed as TAIR.

Several Arabidopsis cell wall mutants were identified by screening 5200 chemically mutagenized Arabidopsis plants for changes in the glycosyl residue compositions of their cell walls. One of these mutants (mur1) lacks an isoform of GDP-Man-4,6-dehydrase (Chapter 6) and is deficient in Fuc. The mutant partially compensates for this deficiency by adding l-Gal, a homolog of Fuc (6-deoxy-l-Gal), to xyloglucan, RG-I, and RG-II. The addition of l-Gal to RG-II results in incomplete synthesis of the A sidechain and a decrease in borate cross-linking of RG-II, which is likely responsible for the dwarf stature of mur1 plants. Two additional mutants (mur2 and mur3) identified in the same screen were found to be defective in xyloglucan sidechain synthesis. Subsequently, all the other Arabidopsis genes encoding the GTs required for xyloglucan sidechain formation have been identified. Some of the genes involved in xyloglucan synthesis in rice and tomato have also been identified and functionally characterized.

Genes involved in the synthesis of the GX backbone (IRX10 and IRX10-L), the addition of GlcA (GUX), the O-methylation of GlcA (GXMT1), and O-acetylation of the backbone (XOAT) have been identified and functionally characterized. There has also been progress in identifying grass
genes and GTs involved in GAX synthesis. No genes involved in the synthesis of the Rha-containing reducing end sequence of secondary wall GX have yet been identified.

A family of genes that encode GTs involved in the synthesis of 1-4-linked α-galacturonans have been identified. Two of the encoded proteins, GAUT1 and GAUT7, exist as a complex that synthesizes homogalacturonan. Only one GT involved in RG-II synthesis has been identified. Similarly, few genes have been identified and shown to have a role in the synthesis of the backbone and side chains of RG-I.

Fucosyltransferases, glucuronosyltransferases galactosyltransferases, and arabinosyltransferases involved in arabinogalactan and extensin biosynthesis have been identified. Plants carrying loss-of-function mutations in some of the genes encoding these GTs exhibit growth defects and reduced fertility.

Plants carrying mutations at different steps along the N-glycosylation pathway have begun to provide insight into the role of protein glycosylation in plants. Mutants (dgII) defective in the oligosaccharyltransferase complex have phenotypes that range from reduced cell elongation to embryonic lethality. However, a mutant (cgl), which lacks GnT-I activity has no discernible developmental or growth defects even though it produces glycoproteins enriched in Man₅GlcNAc₂ but lacking complex N-glycans.

Plant mutants defective in O-GlcNAc modification of proteins exhibit numerous changes in growth and development processes (Chapter 19). It is not known how this modification affects cellular processes as only a small number of O-GlcNAc modified proteins have been identified.

The plant-specific modifications of N-glycans result in glycoproteins that are often highly immunogenic and cause allergic responses in humans. The demonstration that complex N-glycans are not essential for plant growth initiated studies to engineer plant N-glycosylation pathways to produce glycoproteins that that do not activate the mammalian immune system. Plants lacking the GTs that add Xyl and Fuc to N-linked glycans produce glycoproteins lacking immunogenic glyco-epitopes. Other glycosylation pathways involved in the addition of sialic acid and Gal must be introduced to fully "humanize" the glycoproteins if plants are to be used to produce recombinant therapeutic glycoproteins.
Glycolipids are important for chloroplast development and for photosynthesis. The \textit{mgd1} mutant, which contains 50% of the normal amounts of MGDG, is deficient in chlorophyll production and has altered chloroplast ultrastructure. Additional evidence for the role of MGDG has been obtained using \textit{galvestine-1} (2-oxobenzo[d]imidazol-3-yl) piperidine-1-carboxylate), a chemical inhibitor of monogalactosyldiacylglycerol synthases, which has been shown to impair chloroplast development in Arabidopsis.

\textbf{ACKNOWLEDGEMENTS}

The authors appreciate helpful comments and suggestions from Corinna Landig, Robert Townley, and James Prestegard.
FURTHER READING


Figure Legends

FIGURE 24.1 Glycosyl sequences of cellulose and selected hemicelluloses present in plant cell walls.

FIGURE 24.2 Schematic structure of pectin showing the three main pectic polysaccharides: rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II), and homogalacturonan (HG). The borate ester is formed between the apiosyl residue in side chain A of each RG-II monomer. A region of substituted galacturonan, referred to as xylogalacturonan (XGA), is also shown. XGA is not present in most cell wall pectins. The relative abundance of each pectin domain (HG, RG-I and RG-II) is dependent of the plant species.

FIGURE 24.3 Schematic structure of the proteoglycan referred to as Arabinoxylan Pectin Arabinogalactan Protein1 (APAP1).

FIGURE 24.4 Types of N-glycans identified in plants.

FIGURE 24.5 Processing of N-glycans in the plant secretory system. Only those events that are unique to plants are shown in detail.

FIGURE 24.6 The most abundant plant galactolipids.
Table 24.1. Estimated number of genes encoding proteins involved in the synthesis and modification of glycans in plants and humans

<table>
<thead>
<tr>
<th>Organism</th>
<th>GTs(^1)</th>
<th>Hydrolases</th>
<th>Lyases</th>
<th>Esterases</th>
<th>CBMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis</td>
<td>466 (43)</td>
<td>399 (35)</td>
<td>34 (2)</td>
<td>85 (4)</td>
<td>126 (10)</td>
</tr>
<tr>
<td>Rice</td>
<td>568 (43)</td>
<td>429 (34)</td>
<td>16 (2)</td>
<td>53 (4)</td>
<td>122 (9)</td>
</tr>
<tr>
<td>Human</td>
<td>236 (43)</td>
<td>98 (30)</td>
<td>ni(^2)</td>
<td>1 (1)</td>
<td>40 (5)</td>
</tr>
<tr>
<td>Ostreococcus</td>
<td>61 (25)</td>
<td>29 (12)</td>
<td>ni(^2)</td>
<td>2 (2)</td>
<td>2 (2)</td>
</tr>
</tbody>
</table>

\(^1\) Values in parenthesis are the number of CAZy classes. \(^2\)ni, none identified to date.
Fig 24.1 Glycosyl sequences of cellulose and selected hemicelluloses present in plant cell walls.

Fig 24.2 Schematic structure of pectin showing the three main pectic polysaccharides: rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II), and homogalacturonan (HG). The borate ester is formed between the apiosyl residue in side chain A of each RG-II monomer. A region of substituted galacturonic, referred to as xylogalacturan (XGA), is also shown. XGA is not present in most cell wall pectins. The relative abundance of each pectin domain (HG, RG-I and RG-II) is dependent on the plant species.

Fig. 24.3 Schematic structure of the proteoglycan referred to as Arabinoxylan Pectin Arabinogalactan Protein I (APAP1).

Fig 24.4 Types of N-glycans identified in plants.

Fig 24.5 Processing of N-glycans in the plant secretory system. Only those events that are unique to plants are shown in detail.

Fig 24.6 The most abundant plant galactolipids.