32

L-Type Lectins

Richard D. Cummings, Marilynn E. Etzler, T. N.C. Ramya, Koichi Kato, Gabriel A. Rabinovich, and Avadhesha Surolia

HISTORICAL BACKGROUND, 1
COMMON FEATURES OF L-TYPE LECTINS, 2
PLANT L-TYPE LECTINS, 3 Distribution and Localization, 3 Structure, 4 Glycan-Binding Activities, Functions, and Applications, 6
L-TYPE LECTINS IN PROTEIN QUALITY CONTROL AND SORTING, 6 Calnexin/Calreticulin, 6 ERGIC-53 and Its Related Proteins ERGL, VIP36, and VIPL, 7
OTHER L-TYPE LECTINS, 8 Pentraxins and Related Proteins, 9 VP4 and VP8 in Rotaviruses, 9
OTHER PROTEINS WITH JELLY-ROLL MOTIFS AND L-TYPE LECTIN DOMAINS, 9
ACKNOWLEDGMENTS, 10
FURTHER READING, 10

The L-type lectins occur in the seeds of leguminous plants. They have structural motifs present in a variety of glycan-binding proteins (GBPs) from other eukaryotic organisms. The structures of many of these lectins have been characterized, and many L-type lectins are used in a wide range of biomedical and analytical procedures and implicated in biomedical processes. This chapter discusses the structure-function relationships of these lectins and the various biological roles they have in different organisms.

HISTORICAL BACKGROUND

The L-type lectins have a rich history going back to the end of the 19th century when it was found that extracts from the seeds of leguminous plants (Fabaceae family but also called the Leguminosae family) could agglutinate red blood cells. These agglutinins (later named lectins) were found to be soluble proteins abundant in the seeds of leguminous plants. Different species of legumes were found to differ in hemagglutination specificity. Much work on these proteins was performed in the early part of the 20th century, including the crystallization of concanavalin A (ConA; the hemagglutinin [HA] from jack beans), which was first isolated by James Sumner in 1919 and crystallized in 1936 and was the first commercially available lectin. The legume

Copyright \bigcirc 2022 The Consortium of Glycobiology Editors, La Jolla, California. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial License, which permits reuse and distribution, except for commercial purposes, provided that the original author and source are credited.

Published by Cold Spring Harbor Laboratory Press; doi: 10.1101/glycobiology.4e.32

2 | CHAPTER 32

lectins were also found to have hemagglutinating properties because of their ability to bind glycans on the cell surface.

The abundance of these proteins in the soluble extracts of legume seeds (up to $\sim 5\%-10\%$ total protein) enabled many lectins to be isolated and characterized. Such seed lectins have considerable amino acid sequence homology, and the variety of carbohydrate-binding specificities found among these lectins made them useful tools in a wide variety of analytical and biomedical procedures.

The crystal structures of a number of legume seed lectins allowed the identification of the carbohydrate-binding sites. Structural similarities in the tertiary structures were identified among these lectins and several other lectins, including the galectins (see Chapter 36), as will be discussed in this chapter. For this reason, the term "L-type lectins" has recently been designated as a classification for all proteins with this legume seed lectin-like protein structure.

COMMON FEATURES OF L-TYPE LECTINS

The L-type lectins are distinguished from other lectins primarily based on tertiary structure and not the primary sequence. In general, either the entire lectin monomer or the carbohydrate-recognition domains (CRDs) of the more complex lectins is composed of antiparallel β -sheets connected by short loops and β -bends, and they usually lack any α -helices. These sheets form a dome-like structure related to the "jelly-roll fold," which is often called a "lectin fold." The carbohydrate-binding site is generally localized toward the apex of this dome. The tertiary structure of the monomer of ConA, the lectin from the seeds of the legume *Canavalia ensiformis*, is shown in Figure 32.1. The crystal structures of at least 20 other legume L-type lectin monomers have been determined by high-resolution X-ray crystallography and are almost superimposable on this structure. Thus, it is not surprising that the primary amino acid



FIGURE 32.1. Structure of concanavalin A (ConA), a legume seed lectin in complex with a branched pentasaccharide GlcNAcβ1-2Manα1-3(GlcNAcαβ1-2Manα1-6)Man to 2.7 Å. The pentasaccharide is depicted in the *center*. The tetrameric Con A (monomer 26.5 kDa and 235 amino acids) is shown (*left*) each with the pentasaccharide and the orthogonal view is shown (*right*). The bound metal is indicated by the *purple* and *green balls* (Mn⁺⁺ and Ca⁺⁺). The monomer is best described as a "jelly-roll fold," consisting of a flat six-stranded antiparallel "back" β-sheet, a curved seven-stranded "front" β-sheet, and a five-stranded "top" sheet linked by loops of various lengths. ConA dimers involve antiparallel side-by-side alignment of the flat six-stranded "back" sheets, giving rise to a contiguous 12-stranded sheet. The tetramerization of ConA occurs by a back-to-back association of two dimers. (Taken from the PDB ftp://ftp.wwpdb.org; Moothoo DN, Naismith JH. 1998. *Glycobiology* **8:** 173–181.)

L-TYPE LECTINS | 3



FIGURE 32.2. Comparison of the subunit structures of soybean agglutinin (A) complexed with a pentasaccharide containing Gal β 1-4GlcNAc-R and human galectin-3 at 1.4 Å (*B*) complexed with Gal β 1-4GlcNAc. Both lectins display a related β -barrel configuration. (Soybean agglutinin structure [1sbd] available at PDB and based on Olsen LR, Dessen A, Gupta D, Sebesan S, Sacchettini JC, Brewer CF. 1997. *Biochemistry* **36**: 15073–15080. Human galectin-3 structure [1kjl] available at PDB and based on Sorme P, Arnoux P, Kahl-Knutsson B, Leffler H, Rini JM, Nilsson UJ. 2005. J Am Chem Soc **127**: 1737–1743.) For PDB structures go to ftp://ftp.wwpdb.org/.

sequences of legume lectins show remarkable homology with one another and of the sequences of many other legume seed lectins sequenced but not yet crystallized. Fewer, but significant, homologies in primary structure have been found between legume L-type lectins compared with some L-type lectins from far distant sources, such as ERGIC-53 and VIP36, found in vertebrates. Yet, in other L-type lectins, no homology is found with the seed lectins, although they contain similar lectin folds. For example, a comparison of the tertiary structure of the legume soybean lectin with the structure of human galectin-3 shows that both proteins contain the typical L-type lectin fold, but no amino acid sequence homology exists between the two lectins (Figure 32.2).

Relationships between sequences of L-type lectins from legumes compared with the phylogeny of the various species within the Fabaceae family of plants suggest that these lectins most probably arose from divergent evolution. It remains an open possibility that the tertiary structures of some of the other members of the "L-type lectin family" arose by convergent evolution. It must also be noted that for a protein to be firmly placed in the L-type lectin category it must have the lectin fold and glycan-binding activity.

All soluble L-type lectins found to date are multimeric proteins, although all do not have the same quaternary structure. Thus, these lectins are multivalent with more than one glycanbinding site per lectin molecule. The same multivalent principle applies to the membrane bound L-type lectins because the presence of two or more molecules on a membrane surface essentially presents a multivalent situation. In addition to increasing the avidity of the lectins for branched and/or cell-surface glycans, this multivalence can have great biological significance. Binding of the lectins to the cell surface can lead to aggregation of specific glycan receptors, promoting a variety of biological responses such as mitogenesis and various signal transduction processes.

PLANT L-TYPE LECTINS

Distribution and Localization

Plant L-type lectins are primarily found in the seeds of leguminous plants and are synthesized during seed development several weeks after flowering; they are transported to the vacuole where they become condensed into specialized vesicles called protein bodies. They are stable during desiccation of the seeds and can remain in that state indefinitely until the seeds

germinate. They represent one of several classes of proteins stored in high concentrations in the seeds and are often called storage proteins. During seed germination, the storage bodies become the vacuoles of the cotyledons, which appear as the first leafy appendages of the plant. During the first week of development, these cotyledons provide food for the plant and eventually shrivel up and disappear. L-type lectins have also been found in the bark of some leguminous trees, and very low amounts of these lectins are also found in other vegetative tissues of legumes. In some cases, these latter lectins are encoded by separate but very similar genes. More than 100 of the seed legume L-type lectins have been characterized and are the most extensively studied proteins of this class. It also should be noted that the L-type fold has been observed in noncarbohydrate binding proteins, such as lectin-like receptor kinases in plants that are important in development and stress responses.

Structure

A common feature of the legume L-type lectins is their monomeric structure. The structures of the monomers consist of three antiparallel β -sandwich: a flat six-stranded "back" sheet, a concave seven-stranded "front" sheet, and a short "top" sheet that keeps the two major sheets together (Figure 32.1A,B). All of these lectins require Ca⁺⁺ and a transition metal ion (usually Mn⁺⁺) for their carbohydrate-binding activity. The glycan-binding and metal-binding sites are localized in close proximity to each other at the top of the "front" sheet.

The glycan-binding site is composed of four loops: A, B, C, and D (Figure 32.3, top). These loops contain four invariant amino acids that are essential for carbohydrate binding (Figure 32.3, bottom). Loop A contains an invariant aspartate, which forms hydrogen bonds between its side chain and the glycan ligand. This amino acid is linked to its preceding amino acid (usually alanine) by a rare *cis*-peptide bond, which is stabilized by the metal ions and is necessary for the proper orientation of the aspartate in the combining site. Loop B contains an invariant glycine, which also forms hydrogen bonds with the ligand. An exception to this case is found in two lectins (ConA and the closely related *Dioclea grandiflora* lectin) in which the glycine is replaced with an arginine. Both the glycine and arginine form hydrogen bonds with the ligand via their main-chain amides. Loop C contains an invariant asparagine, which forms a hydrogen bond with the ligand via the side chain, and an invariant hydrophobic amino acid. Besides these invariant amino acids, backbone atoms of the residues of loop D also contribute to monosaccharide recognition.

The legume L-type lectins are generally classified into groups based on their carbohydrate specificities as often identified by the ability of monosaccharides to inhibit their agglutinating activity. These differences in specificities are brought about by variability in the conformation and size of the D loop and to some extent by the C loop. Although the main specificity regions of the legume lectins are determined by the loops, there are sites other than these that contribute to lectin specificity. There are several additional modes of refining these specificities, such as interaction with water, posttranslational modifications, and state of oligomerization.

Legume L-type lectins are oligomeric—mostly dimeric or tetrameric in nature—and adopt a variety of quaternary structures. The back β -sheet in the monomeric unit is involved in oligomerization. Small differences in the monomeric structure of the different legume lectins result in different modes of association of the β -sheet during oligomerization. For instance, although both ConA and PNA are tetrameric—being dimers of dimers—the quaternary association in ConA involves the association of the two six-stranded back β -sheets to form an extended 12-stranded β -sheet in each dimer, whereas the quaternary association in PNA involves a back-to-back arrangement of the back β -sheets. The tetrameric structure of ConA is shown in Figure 32.1C,D. Although some of the other lectins occur as dimeric and tetrameric structures, several other different orientations of the β -sheets account for the variability in dimeric and tetrameric structures of other lectins in this class. Interestingly, some legume lectins have a

L-TYPE LECTINS 5



becilicity/ hibitor
alNAc
alNAc
omplex
omplex
alNAc
al
JC
JC
IcNAc
IcNAc
an/Glc
an/(an/(an/(an/(

FIGURE 32.3. (*Top*) Three-dimensional structure of a peanut agglutinin (PNA) monomer showing the four loops involved in sugar binding: loops A, B, C, and D. The bound sugar (lactose) is shown as a "*ball-and-stick*" model. Calcium and manganese ions are required for ligand binding. (*Bottom*) Sequence alignment of loops A – D in other legume lectins. The size of binding-site loop D and monosaccharide specificity show an explicit correlation. Monosaccharide specificity and number of gaps are indicated at the *right*. Key residues are highlighted in *blue* and highly conserved residues have been indicated with an *asterisk*. (Adapted, with permission of Elsevier, from Vijayan M, Chandra N. 1999. *Curr Opin Struct Biol* **9**: 707–714; and the PBD deposited structure 1V6l.) EcorL, *Erythrina corallodendron* lectin; DBL, *Dolichos biflorus* lectin; E-PHA, erythroagglutining phytohemagglutinin from *Phaseolus vulgaris*; L-PHA, leukoagglutinating phytohemagglutinin from *P. vulgaris*; SBA, soybean agglutinin *Glycine max*; PNA, peanut agglutinin *Arachis hypogaea*; LTA, *Lotus tetragonolobus* agglutinin; UEA-I, *Ulex europaeus* agglutinin-II; LAA-I, *Laburnum alpinum* agglutinin-I; LSL, *Lonchocarpus sericeus* lectin; ConA, concanavalin A *Canavalia ensiformis*; LenL, lentil lectin *Lens culinaris*; PSL, pea lectin *Pisum sativum*; Favin, *Vicia faba* lectin.

hydrophobic binding site that binds adenine and adenine-derived plant hormones with micromolar affinity; this is two to three orders of magnitude higher than their affinity for monosaccharides. Three of these lectins (the soybean agglutinin, phytohemagglutinin L [PHAL], and *Dolichos biflorus* lectin) have been crystallized and found to have a unique tetrameric structure, in that the dimer–dimer interface creates a channel running through the center of the tetramer. Two identical adenine-binding sites are found at opposite ends of this channel.

Another common feature of the legume lectins is that they are secretory proteins and undergo cotranslational signal peptide removal, which accompanies their entry into the secretory system. All but peanut agglutinin (PNA) are N-glycosylated as precursors; the N-glycans undergo normal posttranslational modification as they transit the Golgi apparatus. The lectins vary from one another as to whether the mature proteins contain oligomannose-type, complextype, or a mixture of both types of N-glycans. The lectins may also undergo a variety of proteolytic modifications as they transit through the secretory system. Some of the lectins are cleaved to generate a β -chain, corresponding to the amino terminus and an α -chain corresponding to the carboxyl terminus. For example, the pea lectin and favin (the lectin from *Vicia faba*) are tetrameric glycoproteins that contain two types of subunits, α and β , which are ~ 5 kDa and ~ 21 kDa, respectively. These two lectins are each synthesized as single polypeptide precursors that contain the sequences of both chains in the following orientation: β -chain– α -chain. The chains associate to form dimers; they are then proteolytically processed in the protein bodies to form tetramers containing two separate α - and β -chains. Other lectins may undergo carboxyterminal trimming of only some of their subunits. For example, the soybean agglutinin, phytohemagglutinin E (E-PHA), and *D. biflorus* lectins are tetramers of equimolar mixtures of intact and trimmed subunits.

The most intriguing proteolytic modification occurs in the case of ConA. A small segment is removed from the interior of the protein and the original amino terminus is ligated with the original carboxyl terminus. This forms what is termed a circularly permuted protein. The N-glycosylated segment of the protein is removed during this transpeptidation process; thus, the mature ConA is not a glycoprotein, in contrast to most seed lectins in the Fabaceae family, which have N-glycans. Thus, the protein sequence of isolated ConA aligns with other seed lectins, whereas the alignment of the DNA encoding the protein with other lectin genes indicates that the gene is circularly permuted.

Glycan-Binding Activities, Functions, and Applications

Despite years of research, the intrinsic biological roles of L-type lectins are poorly understood. The glycan-binding specificity can vary significantly; ConA binds mannose and glucosecontaining glycans, whereas the lectins from the trees *Maackia amurensis* and *Sambucus nigra* bind to sialylated glycans (see Chapters 15, 29, and 31). There is evidence from many approaches that the legume lectins are insecticidal, antifungal, and antimicrobial and can be toxic to animals that eat raw seeds. Thus, the seed lectins may function in plant immunity as a type of pattern-recognition receptor (PRR) to protect the offspring of the plant. As highly abundant seed constituents, the lectins could also serve as a storage protein for the plant development. L-type lectins are involved in the symbiosis of plants with nitrogen-fixing bacteria, such as in Rhizobium-legume symbiosis, but the precise function of lectins in this regard is unclear. A recent study has shown that the *D. biflorus* seed lectin is also a lipoxygenase. It will be of interest to see how many other L-type lectins have this activity, which is necessary to initiate the wound-induced defense pathway in plants.

Regardless of their physiological function, legume L-type lectins might be useful in applications requiring agents with antitumor, antiviral, antibacterial, antifungal or antinociceptive properties. In addition, given their well-established glycan recognition preferences, these plant lectins are useful for glycophenotyping different cell types in physiological and pathological settings by means of immunofluorescence or immunohistochemical analysis.

L-TYPE LECTINS IN PROTEIN QUALITY CONTROL AND SORTING

Calnexin/Calreticulin

Calnexin (CNX) and calreticulin (CRT) are homologous molecular chaperones that mediate quality control of proteins in the endoplasmic reticulum (ER) (see Chapter 39). Although CRT is a soluble ER luminal component, CNX is membrane-bound and is perhaps closely

L-TYPE LECTINS | 7

associated with the protein-translocating channel that imports nascent proteins into the ER. Both CNX and CRT bind to monoglucosylated, high-mannose-type glycans and prevent their exit from the ER until they are properly folded and assembled into correct quaternary structures (Chapter 39). For example, generation of a functional major histocompatibility complex class I (MHC I) depends on its association with CRT or CNX as component of a peptide-loading complex and the antigen presentation machinery. During the binding and dissociation from CRT or CNX, if the glycoprotein folds correctly, then glucose removal by glucosidase-II allows its passage out of the ER. If a glycoprotein misfolds or aggregates, it is reglucosylated by UDP-Glc:glycoprotein glucosyltransferase (UGGT) containing four tandem thioredoxin-like domains; this enzyme only recognizes misfolded or aggregated glycoproteins. Following reglucosylation, the monoglucosylated protein binds again to CRT or CNX. Thus, there is a cycle of glucose removal and addition by the alternating actions of glucosidase-II and UGGT and interactions with CNX/CRT.

Both CRT and CNX are Ca^{++} -binding proteins, and their carbohydrate-binding activity is sensitive to changes in Ca^{++} concentration. CNX is a type I membrane protein with its carboxy-terminal end in the cytoplasm. The lumenal portion of the protein is divided into three domains: a Ca^{++} -binding domain (which is adjacent to the transmembrane domain), a proline-rich long hairpin loop called the P domain, and the amino-terminal L-type lectin domain. CRT has a similar structure, but it is missing the cytoplasmic and transmembrane regions; it is retained in the ER through its KDEL-retrieval signal at the carboxyl terminus (Figure 32.4).

ERGIC-53 and Its Related Proteins ERGL, VIP36, and VIPL

ERGIC-53 (human gene LMAN1) and its sequence-related proteins ERGL and VIP36 are type I membrane proteins that participate in vesicular protein transport in the secretory system (see Chapter 39). All share an L-type lectin folding motif. Orthologs of ERGIC-53 are found in plants and all animals and have been independently identified as being important in production of infectious viruses (e.g., coronavirus and filovirus). ERGIC-53 occurs in the ER-Golgi intermediate compartment (ERGIC), and its cytoplasmic carboxyl terminus contains the dilysine/diphenylalanine KKFF retention/retrieval motif. The dilysine is recognized by the COPI coatomer complex; this binding enables the coated vesicles to be recycled from the ERGIC back to the ER. The diphenylalanine helps to direct the COPII-coated vesicles to ER export sites by binding to the COPII coatomer. The location of VIP36 (vesicular integral membrane protein 36) is uncertain; overexpressed protein has been found in both the ER and ERGIC, as well as *cis*-Golgi. ERGIC 53 can associate with the soluble ER partner MCFD2 (multiple coagulation factor deficiency protein 2), which forms a cargo receptor complex important in factor V and VIII biosynthesis. ERGIC-53 binds to N-linked glycans and allosterically activates MCFD2, which thereby becomes able to capture specific polypeptide segments of cargo glycoproteins. Human mutations in ERGIC-53 are associated with deficiencies of circulating blood clotting factors V and VIII, which are glycoproteins with multiple N-glycans. ERGIC-53-deficient mice show FV and FVIII deficiencies and liver ER distension, with accumulation of α 1-antitrypsin and GRP78. Partially penetrant, perinatal lethality occurs, dependent on inbred genetic background, suggesting potential roles for as yet unidentified ERGIC-53-dependent cargo proteins.

Both ERGIC-53 and VIP36 bind to oligomannose-type glycans and require Ca⁺⁺ for their carbohydrate-binding activity. These two proteins were the first animal lectins found to share some sequence and structural homology with the legume seed lectins. Although the overall sequence identity of these proteins to the seed lectins is only \sim 19%–24%, those amino acids important for metal and carbohydrate binding in the seed lectins are conserved, including the invariant aspartate, glycine, and asparagine. The invariant aspartate also participates in a *cis*-peptide bond with its preceding amino acid; this is similar to the case of the legume seed lectins

8 | CHAPTER 32



FIGURE 32.4. Schematic representations of calreticulin (CRT) and calnexin (CNX) showing the lectin domain, the P domain (containing the proline repeats), and the calcium-binding domain (*A*,*B*). Structure of CNX based on crystallographic data (*C*). (Adapted, with permission of Elsevier, from Schrag JD, et al. 2001. *Mol. Cell* **8**: 633–644.) Structures of the lectin domains of ERGIC-53 and VIP36 based on crystallographic data shown in cartoon and surface representation with the ligands indicated in *stick representation* (*D*,*E*).

discussed in the above section. The crystal structure of the lectin domain of ERGIC-53 has been determined and confirms the structural similarity of these lectins (Figure 32.4). In contrast to CRT/CNX, ERGIC-53 and VIP36 can bind the deglucosylated branch of high-mannose-type glycans and thereby contribute to vesicular transport of correctly folded glycoproteins in the early secretory pathway. During transport from the ER to the Golgi, release of cargo glycoproteins from ERGIC-53 is caused by the environmental acidification and a decrease in Ca⁺⁺. Like ERGIC-53, ERGL can associate with MCFD2 and is probably involved in FV and FVIII transport. VIPL (VIP36-like) is an ER-resident protein and functionally able to bind deglucosylated high-mannose-type glycans, and associate with ERGIC-53, which may be important in regulating ERGIC-53 localization.

OTHER L-TYPE LECTINS

A variety of other proteins have been described that have carbohydrate-binding domains with tertiary structures similar to the L-type lectin fold and may be considered as members of this family. Members of the galectin family of lectins also fit into this category and are the subject of a separate chapter in this volume (see Chapter 36). Other GBPs that may fit into this category are briefly discussed below.

L-TYPE LECTINS | 9

Pentraxins and Related Proteins

The pentraxins are a superfamily of plasma proteins that are involved in innate immunity in invertebrates and vertebrates and are designated as PRRs. They contain L-type lectin folds and require Ca⁺⁺ ions for ligand binding. Their name is based on the pentameric arrangement of their subunits. The short PTX C-reactive protein (CRP), which binds phosphocholine residues on polysaccharides and on phospholipids, and the serum amyloid P (SAP) component, which also binds carbohydrate derivatives on bacterial polysaccharides and to amyloid fibrils, are acute-phase proteins in humans and mice, respectively. This family also contains long PTX that have an unrelated long amino-terminal domain coupled to the PTX domain and includes neuronal PTX 1, neuronal PTX 2, neuronal PTX receptor, PTX 3, and PTX 4. A prototypical member, PTX3 is a soluble pattern recognition molecule involved in innate immunity and has been implicated in the pathogenesis of autoimmune diseases, cardiovascular inflammation, and cancer. In addition, PTX3 may also help in the assembly of a hyaluronan-rich extracellular matrix (ECM), which may involve binding to the hyaluronan-heavy chain-1 complex (inter- α -trypsin inhibitor or I α I or ITI) and regulation by TNF-stimulated gene-6 (TSG-6).

Laminin G domain-like (LG) modules are made of 180–200 amino acid residues and were first identified in laminins. The carboxyl terminus of the laminin α -chain has five tandem laminin G domains, which are important in heparin and sulfatide binding and cell/basement membrane adhesion, such as to the novel glycans of α -dystroglycan (α -DG, see Chapter 13). Some LG modules share binding properties for cellular receptors and carbohydrate ligands, indicating that the LG fold may have evolved from the L-type lectin fold for participation in related functions.

VP4 and VP8 in Rotaviruses

VP4 is the surface spike protein in rotaviruses and is proteolytically cleaved to generate VP8, which is a GBP, and VP5, a hydrophobic region required for membrane entry. VP8 has an L-type lectin fold and structural features in common with galectins. This domain is required for infectivity of most animal rotaviruses. In some rotavirus strains, VP8 binds sialylated glycans, whereas most rotavirus VP8 recognizes type 1 and/or type 2 N- or O-glycans lacking sialic acid (Sia) and can especially bind to human milk oligosaccharides with appropriate sequences.

OTHER PROTEINS WITH JELLY-ROLL MOTIFS AND L-TYPE LECTIN DOMAINS

A number of proteins are known to contain the jelly-roll motif found in L-type lectins. These include the binding domain of *Clostridium* neurotoxins, which can bind to gangliosides (e.g., GT1b and GD1b, see Chapter 11); *Pseudomonas aeruginosa* exotoxin A in which one of its three domains, the amino-terminal domain Ia, displays an L-type lectin structure; *Vibrio cholerae* sialidase, a three-domain protein with a six-bladed β -propeller neuraminidase domain flanked by two L-type lectin domains; and the Leech intramolecular *trans*-sialidase (TS), which has a multidomain architecture with a lectin-like domain II and an irregular β -stranded domain III that is built around a canonical catalytic domain C. Domain II may be involved in carbohydrate recognition through sugar ring and aromatic side-chain interactions, as observed in many lectins.

Although the L-type lectin domain occurs in many proteins, in many cases evidence for carbohydrate binding and/or specificity is lacking. Examples include the L-type lectin receptor kinases (LecRKs) in plants, which is a large family of several thousand members comprising membrane receptors that are key to plant development, immunity, and adaptive responses to stimuli; and mammalian thrombospondins, a family of matricellular proteins implicated in

10 | CHAPTER 32

angiogenesis and inflammation, including TSP-1, which has a carboxy-terminal domain that shows the typical β -sandwich of two curved antiparallel β -sheets, a feature of the jelly-roll topology, in which the L-type lectin domain is compactly assembled with three calcium-binding type 3 (T3) repeats

ACKNOWLEDGMENTS

The authors acknowledge helpful comments and suggestions from Yasuhiro Ozeki.

FURTHER READING

- Lis H, Sharon N. 1998. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev* 98: 637–674. doi:10.1021/cr940413g
- Moothoo DN, Naismith JH. 1998. Concanavalin A distorts the beta-GlcNAc-(1-2)-Man linkage of β-GlcNAc-(1-2)-α-Man-(1-3)-[β-GlcNAc-)-(1-2)-α-Man-(1-6)]-Man upon binding. Glycobiology 8: 173-181. doi:10.1093/glycob/8.2.173
- Hamelryck TW, Loris R, Bouckaert J, Dao-Thi M.-H, Strecker G, Imberty A, Fernandez E, Wyns L, Etzler ME. 1999. Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. J Mol Biol 286: 1161–1177. doi:10.1006/jmbi .1998.2534
- Loris R, Bouckaert J, Hamelryck T, Wynn L. 1999. Legume lectin structure. *Biochim Biophys Acta* 1383: 9–36. doi:10.1016/s0167-4838(97)00182-9
- Vijayan M, Chandra N. 1999. Lectins. Curr Opin Struct Biol 9: 707-714. doi:10.1016/s0959-440x(99) 00034-2
- Schrag JD, Vergeron JJ, Li Y, Borisova S, Hahn M, Thomas DY, Cygler M. 2001. The structure of calnexin, and ER chaperone involved in quality control of protein folding. *Mol Cell* 8: 633–644. doi:10.1016/ s1097-2765(01)00318-5
- Srinivas VR, Reddy GB, Ahmad N, Swaminathan CP, Mitra N, Surolia A. 2001. Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. *Biochim Biophys Acta* **1527**: 102–111. doi:10.1016/s0304-4165 (01)00153-2
- Sorme P, Arnoux P, Kahl-Knutsson B, Leffler H, Rini JM, Nilsson UJ 2005. Structural and thermodynamic studies on cation-Pi interactions in lectin–ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine–arene interaction. J Am Chem Soc 127: 1737–1743. doi:10.1021/ ja043475p
- Gouget A, Senchou V, Govers F, Sanson A, Barre A, Rouge P, Pont-Lezica R, Canut H. 2006. Lectin receptor kinases participate in protein–protein interactions to mediate plasma membrane–cell wall adhesions in *Arabidopsis. Plant Physiol* 140: 81–90. doi:10.1104/pp.105.066464
- Bouwmeester K, Govers F. 2009. Arabidopsis L-type lectin receptor kinases: phylogeny, classification, and expression profiles. J Exp Bot 60: 4383-4396. doi:10.1093/jxb/erp277
- Kouno T, Watanabe N, Sakai N, Nakamura T, Nabeshima Y, Morita M, Mizuguchi M, Aizawa T, Demura M, Imanaka T, et al. 2011. The structure of *Physarum polycephalum* hemagglutinin I suggests a minimal carbohydrate recognition domain of legume lectin fold. *J Mol Biol* 405: 560–569. doi:10.1016/j.jmb.2010.11.024
- Zhang B, Zheng C, Zhu M, Tao J, Vasievich MP, Baines A, Kim J, Schekman R, Kaufman RJ, Ginsburg D. 2011. Mice deficient in LMAN1 exhibit FV and FVIII deficiencies and liver accumulation of α1-antitryhpson. Blood 118: 3384–3391. doi:10.1182/blood-2011-05-352815
- Klaus JP, Eisenhauer P, Russo J, Mason AB, Do D, King B, Taatjes D, Cornillez-Ty C, Boyson JE, Thali M, et al. 2013. The intracellular cargo receptor ERGIC-53 is require for the production of infectious arenavirus, coronavirus, and filovirus particles. *Cell Host Microbe* 14: 522–534. doi:10.1016/ j.chom.2013.10.010

- Croci DO, Cerliani JP, Dalotto-Moreno T, Méndez-Huergo SP, Mascanfroni ID, Dergan-Dylon S, Toscano MA, Carmelo JJ, García-Vallejo JJ, Ouyang J, et al. 2014. Glycosylation-dependent lectin-receptor interactions preserve angiogenesis in anti-VEGF refractory tumors. *Cell* **156**: 744– 758. doi:10.1016/j.cell.2014.01.043
- Satoh T, Suzuki K, Yamaguchi T, Kato K. 2014. Structural basis for disparate sugar-binding specificities in the homologous cargo receptors ERGIC-53 and VIP36. *PLoS ONE* **9**: e87963. doi:10.1371/journal.pone.0087963
- Grandhi NJ, Mamidi AS, Surolia A. 2015. Pattern recognition in legume lectins to extrapolate amino acid variability to sugar specificity. *Adv Exp Med Biol* 842: 199–215. doi:10.1007/978-3-319-11280-0_13
- Kim DJ, Christofidou ED, Keene DR, Hassan Milde M, Adams JC. 2015. Intermolecular interactions of thrombospondins drive their accumulation in extracellular matrix. *Mol Biol Cell* 26: 2640–2654. doi:10.1091/mbc.e14-05-0996
- Wang Y, Weide R, Govers F, Bouwmeester K. 2015. L-type lectin receptor kinases in *Nicotiana benthamiana* and tomato and their role in *Phytophthora* resistance. *J Exp Bot* **66**: 6731–6743. doi:10.1093/jxb/erv379

Lamriben L, Graham JB, Adams BM, Hebert DN. 2016. N-glycan-based ER molecular chaperone and protein quality control system: the calnexin binding cycle. *Traffic* 17: 308–326. doi:10.1111/tra.12358

- Magrini E, Mantovani A, Garlanda C. 2016. The dual complexity of PTX3 in health and disease: a balancing act? *Trends Mol Med* 22: 497–510. doi:10.1016/j.molmed.2016.04.007
- Doni A, D'Amico G, Morone D, Mantovani A, Garlanda C. 2017. Humoral innate immunity at the crossroad between microbe and matrix recognition: the role of PTX3 in tissue damage. *Semin Cell Dev Biol* **61:** 31–40. doi:10.1016/j.semcdb.2016.07.026
- Satoh T, Kato K. 2020. Recombinant expression and purification of animal intracellular L-type lectins. *Methods Mol Biol* **2132**: 21–28. doi:10.1007/978-1-0716-0430-4_3

