NCBI Bookshelf. A service of the National Library of Medicine, National Institutes of Health.

Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.

Chapter 26 Discovery and Classification of Glycan-Binding Proteins

Ajit Varki, Marilynn E Etzler, Richard D Cummings, and Jeffrey D Esko.

This chapter provides an overview regarding naturally occurring glycan-binding proteins (GBPs), with an emphasis on their discovery and current classification schemes. Some general principles regarding the structure and function of GBPs are also considered, as well as aspects of their glycan-binding properties. Further information regarding most of the major classes of GBPs can be found in Chapters 28–35. For details regarding the analysis of glycan–protein interactions and the physical principles involved, see Chapter 27.

GLYCAN RECOGNITION BY PROTEINS: A KEY TO SPECIFICITY IN GLYCOBIOLOGY

Glycans can mediate a wide variety of biological roles by virtue of their mass, shape, charge, or other physical properties. However, many of their more specific biological roles are mediated via recognition by GBPs. Nature appears to have taken full advantage of the vast diversity of glycans expressed in organisms by evolving protein modules to recognize discrete glycans that mediate specific physiological or pathological processes. Indeed, there are no living organisms in which GBPs have not been found.

TWO MAJOR CLASSES OF GLYCAN-BINDING PROTEINS—LECTINS AND GLYCOSAMINOGLYCAN-BINDING PROTEINS

Excluding glycan-specific antibodies, it is possible to classify GBPs broadly into two major groups—lectins and glycosaminoglycan-binding proteins (see Table 26.1). Most lectins are members of families with defined "carbohydrate-recognition domains" (CRDs) that apparently evolved from shared ancestral genes, often retaining specific features of primary amino acid sequence or three-dimensional structure. Thus, new family members can be identified by searching protein sequence or structural databases. Despite this ability to predict new GBPs, the structures of glycans recognized by members of a single lectin family can be quite diverse. Single-site-binding affinities in many lectins appear to be low (with K_d values in the micromolar range), although some lectins recognize glycans with much higher affinity (with K_d values in the nanomolar range). For those lectins with low affinity, multivalent interactions between multiple CRDs and multiple glycans are often required to produce the high-avidity binding interactions that are relevant in vivo. Lectins tend to recognize specific terminal aspects of glycan chains by fitting them into shallow, but relatively well-defined, binding pockets. In contrast, protein interactions with sulfated glycosaminoglycans seem to involve surface clusters of positively charged amino acids that line up against internal regions of extended anionic glycosaminoglycan chains (see Chapter 35). Thus, despite the fact that the glycosaminoglycan structural motifs recognized by a given molecule can be quite specific, most glycosaminoglycan-binding proteins do not seem to be evolutionarily related to each other. Rather, their defining feature (the ability to recognize sulfated glycosaminoglycans) seems to have emerged by convergent evolution. Hyaluronan-binding proteins (hyaladherins) are exceptions, having a characteristic evolutionarily conserved fold, which docks with short segments of the hyaluronan chain (see Chapter 15). Thus, although hyaluronan is structurally a nonsulfated glycosaminoglycan, hyaladherins may be better classified with the lectins, rather than with other glycosaminoglycan-binding proteins.

DISCOVERY AND HISTORY OF LECTINS

Lectins were first discovered more than 100 years ago in plants (see below and Chapters 28 and 29), but they are now known to be present throughout nature. Lectins are also prevalent in the microbial world, wherein they tend to be called by other names, such as hemagglutinins and adhesins (see below and Chapter 34).

Plant Lectins

In 1888, Stillmark found that extracts of castor bean seeds contained a protein that could agglutinate animal red blood cells. Soon thereafter, a number of other plant seeds were found to contain such "agglutinins," but interest in them began to wane with the advent and development of the field of immunochemistry. It was not until the Second World War and the resulting interest in blood typing for blood transfusion that some lectins were found to be specific for various ABO blood types (see Chapter 13) and others were found to have specificities for different glycans. These agglutinins were thus renamed "lectins," a term derived from the Latin word "legere," meaning "to select." Lectins have since been found in almost every plant species studied and are particularly abundant in the seeds of leguminous plants. These "L-type" lectins have been intensively studied and are discussed in further detail in Chapter 29. Plant lectins that fit into the "R-type" category are described in Chapter 28. Other categories of plant lectins have been found and are included in Table 26.2.

Viral Lectins (Hemagglutinins)

The influenza virus hemagglutinin was the first GBP isolated from a microorganism (~1950), and it is now one of the most thoroughly studied of all lectins. Wiley and associates crystallized the viral hemagglutinin, determined its structure in 1981, and later solved the structure of cocrystals prepared with sialyllactose. Since then, the crystal structures for several other viral hemagglutinins have been determined as well. Like animal cell lectins, most viral lectins bind to terminal sugar residues, but some can bind to internal sequences found in linear or branched glycans. The specificity of these interactions can be highly selective. For example, the human influenza viruses bind primarily to cells containing Siaα2-6Gal linkages, whereas other animal and bird influenza viruses preferentially bind to Siaα2-3Gal termini (see book cover figure). Influenza C, in contrast, binds preferentially to glycoproteins containing terminal 9-O-acetylated sialic acids. Many other viruses (e.g., reovirus, rotavirus, Sendai, and polyomavirus) also appear to use sialic acids in specific linkages for infection. Other viruses display glycosaminoglycan-binding proteins that can bind to heparan sulfate proteoglycans, often with high specificity for certain sulfated sequences (e.g., the gD protein of herpes simplex virus).

Bacterial Lectins (Adhesins and Toxins)

Many bacterial lectins have been described, and they fall into two classes: (1) lectins (adhesins) that reside on the bacterial surface and facilitate bacterial adhesion and colonization and (2) secreted bacterial toxins. Many bacterial lectins bind to membrane glycolipids, whereas only a few bind to glycoproteins. In some cases, binding specificity can explain the tissue tropism of the organism; for example, urinary tract infection by specific serotypes of *Escherichia coli* depends on binding mannose or P blood group structures. Not all of these interactions are pathogenic, and some glycan–protein interactions between bacteria and host tissues play important roles in symbiosis. For instance, the columnar epithelium that lines the large intestine expresses Galα1-4Galβ1-4Glcβ-ceramide, whereas the cells lining the small intestine may not. Thus, *Bacterioides, Clostridium*, *E. coli*, and *Lactobacillus* species that recognize Galα4Gal only colonize the large intestine under normal conditions. The granulocytotropic bacterium

Anaplasma phagocytophilum infects leukocytes, particularly neutrophils, by binding to the cell surface mucin P-selectin glycoprotein ligand-1 (PSGL-1) and recognizing both the peptide of PSGL-1 and the associated sialyl Lewis^x antigen.

Many bacterial surface lectins are present in the form of long, hairy appendages known as fimbriae or pili that extend away from the cell. The presence of multiple glycan-binding subunits in the fimbriae allows multivalent interactions, thus increasing the avidity of bacterial host-cell binding. Several examples of heparan sulfate–binding adhesins are also now known.

Many heat-sensitive secreted bacterial toxins bind to glycans. The best-studied example is cholera toxin from *Vibrio cholerae*, which consists of A and B subunits, in the ratio AB_5. The B subunits bind to multiple GM_1 ganglioside receptors through glycan-recognition domains located on the base of the subunits, facilitating delivery of the toxic A subunit into the cytoplasm (see Chapter 34). The structures of related toxins from *Shigella dysenteria*, *Bordetella pertussis*, and *E. coli* have also been solved.

Animal Lectins

The first suggestion that lectins existed in animals was inferred from the reaggregation of dissociated marine sponge cells, a form of species-specific recognition. However, at least some of these phenomena appear to be mediated by glycan–glycan interactions. Agglutinating proteins, similar to the hemagglutinins of viruses, were described in the body fluids of various crustaceans and arachnids. The first suggestion that there might be lectins endogenous to vertebrates came from the studies of Victor Ginsburg and colleagues in the 1960s, in which blood leukocytes from rats were treated with bacterial glycosidases and then injected back into the circulation. The treatments resulted in changes in the homing of cells to different internal organs. Given the nature of the tools available at that time, it was difficult to ascertain whether these phenomena were mediated by an endogenous lectin. In retrospect, these studies were probably demonstrating the generation of ligands for a hepatic galactose receptor (due to sialidase action) or, possibly, loss of ligands for the selectins.

The first direct evidence for a mammalian lectin arose serendipitously during work by Gilbert Ashwell and colleagues, who were studying mechanisms that controlled the turnover of glycoproteins in blood circulation. Looking for an improved way to introduce a radioactive tracer into purified glycoproteins before reinjection, they attempted to transfer the tritium label from tritiated borohydride into the sugar chains of these proteins, which they knew contained the terminal sequence Sia-Gal-GlcNAc. This was successfully done either following enzymatic oxidation of the 6-position of galactose residues (which required first removing the outer sialic acid residues) or following mild periodate oxidation of sialic acid side chains (which left the rest of the sialic acid molecule intact). Surprisingly, there was a dramatic difference between the circulating half-lives of these two preparations, which terminated in either sialic acid or galactose residues. The molecules that retained labeled sialic acids remained in the circulation for many days, whereas those that had lost them (and now terminated in galactose residues) disappeared within minutes. The importance of these terminal β-linked galactose residues was confirmed by in vitro resialylation or by β-galactosidase treatment, both of which partially restored stability in circulation. The site of accumulation of the desialylated glycoproteins was found to be primarily the liver. This led to the discovery of the "asialoglycoprotein receptor," a hepatocyte membrane protein complex that specifically recognizes terminal β-linked galactose or *N*-acetylgalactosamine on circulating glycoproteins or cells. This lectin activity was markedly inhibited by plasma from birds and reptiles. This, in turn, led to the discovery of a chicken hepatic lectin, which preferentially recognizes terminal *N*-acetylglucosamine. Thus, chickens were found to have a *N*-acetylglucosamine-specific hepatic receptor

instead of a galactose-specific one, permitting the circulation of asialoglycoproteins, but not glycoproteins that also have galactose removed. Affinity columns of immobilized asialoglycoproteins were then used to purify the mammalian hepatic asialoglycoprotein receptor.

The same approach was then applied by investigators working in other systems, uncovering a variety of galactose-binding lectins in cell types ranging from the slime mold *Dictyostelium discoideum* to mammalian tissues such as the heart. Most of these lectins eventually proved to be quite different from the original hepatic asialoglycoprotein receptor, being water-soluble and of relatively low molecular weight; they are now known as the galectins (see Chapter 33). Robert Hill and colleagues reported yet another type of hepatic uptake system that seemed to involve recognition of fucose moieties. In the early 1970s, Elizabeth Neufeld and colleagues reported a carbohydrate-dependent system that mediates the uptake of lysosomal enzymes by fibroblasts. In 1977, William Sly and colleagues demonstrated specific blockade of this uptake by the monosaccharide Man-6-P. Further work by the groups of Stuart Kornfeld, Bill Jourdian, Kurt von Figura, and others led to the discovery of Man-6-P receptors, which recognize phosphorylated high-mannose-type glycans that are selectively expressed on lysosomal enzymes (see Chapter 30). Encouraged by the prospect of using this uptake system to correct lysosomal enzyme deficiency diseases in humans, other investigators infused labeled lysosomal enzymes into intact animals and followed their fate. As it turned out, mature lysosomal enzymes were not rich in Man-6-P (the phosphate esters are mostly removed after initial targeting to lysosomes; see Chapter 30). Instead, rapid clearance mostly involved recognition of terminal mannose and *N*-acetylglucosamine residues. This led to the discovery of the macrophage mannose receptor.

Thus, by the beginning of the 1980s, the concept of vertebrate lectins that recognize specific endogenous ligands had become firmly established. Several circulating soluble lectins were also discovered in the blood plasma of various species, with varied glycan-binding specificities. Initially, it had been thought that sialic acids, while serving as ligands for exogenous microbial pathogens, generally acted as “masks” within vertebrates, preventing binding by endogenous lectins that recognized underlying saccharides. The discovery of some arachnid and crustacean lectins that could recognize sialic acids in vitro did not change this impression because these organisms did not themselves express endogenous sialic acids. The first indication that sialic acids might serve as endogenous ligands within vertebrates came from the discovery that binding of the complement regulatory H protein to “self” cell surfaces was dependent on sialic acid residues. Steve Rosen and colleagues then showed that sialidase treatment of rat lymph node sections abolished binding of lymphocytes to high endothelial venules. This was the first demonstration of the involvement of glycans in recognition by what turned out to be the selectin family of vascular receptors (see Chapter 31).

In 1972, Tim Hardingham and Helen Muir showed that hyaluronan could aggregate cartilage proteoglycans. Studies by Vincent Hascall and Richard Heinegard in 1974 documented that there is specific binding between hyaluronan, the amino-terminal globular part of the proteoglycan, and a link protein (see Chapter 15). By the late 1980s, it was realized that the primary amino acid sequence of a protein could be used to predict glycan-recognition properties of proteins (see below). This led to the recognition of the hyaluronan-binding properties of CD44 and the correct prediction that selectins would recognize carbohydrates. More recently, a specific clearance system was also discovered that recognized the sulfated *N*-acetylgalactosamine residues on pituitary glycoprotein hormones (see Chapter 28). In addition, the discovery of sialic acid–dependent binding by the B-cell molecule CD22 and the cloning of the macrophage sialic acid–dependent receptor sialoadhesin led to the identification of another new family of lectins (the Siglecs), which belong to the immunoglobulin superfamily (I-type lectins; see Chapter 32). Most recently, several lectins have been discovered within the endoplasmic reticulum (ER)-Golgi pathway itself, where glycan

biosynthesis occurs (see Chapter 36).

Sulfated Glycosaminoglycan-binding Proteins

As mentioned above, a large group of GBPs that defy classification based on sequence data or general structure are those that recognize sulfated glycosaminoglycans such as heparin, chondroitin sulfate, and dermatan sulfate (see Chapter 35). The best-studied example is the interaction of heparin with antithrombin. Heparin was actually discovered in 1916 by a medical student named Jay McLean, but it was not until 1939 that heparin was shown to be an anticoagulant in the presence of heparin cofactor, which was then identified as antithrombin in the 1950s. The interaction of antithrombin and heparin depends on specific positively charged amino acid residues provided by two opposing α-helices in antithrombin, which bind to a specific pentasaccharide sequence of unusual structure within a heparin chain. Most other sulfated glycosaminoglycan-binding proteins have clusters of positively charged residues arrayed along the surface of the protein. Many of these proteins can also interact with sulfated glycosaminoglycans in a looser fashion—that is, they do not always show a high level of specificity like that demonstrated by antithrombin. However, in most cases, the actual sequence of sulfated sugars required for biological function is unknown, leaving open the possibility that specific sequences mediate the formation of higher-order complexes, for example, between a ligand and its receptor. In many cases, the sulfated glycosaminoglycan acts as a template for homo-oligomerization or for approximation of two proteins.

CLASSIFICATION OF LECTINS BASED ON SEQUENCE AND STRUCTURAL HOMOLOGY

The first classifications came from the plant lectin field and were based on the glycan sequences to which they bound best (e.g., β-galactoside-binding lectins). Only with the advent of molecular cloning did a more consistent classification emerge based on amino acid sequence homology and evolutionary relatedness. The first such classification was proposed by Kurt Drickamer using some highly conserved amino acid sequence motifs in the CRDs of two groups of lectins. One group required calcium for recognition and the members were therefore called C-type lectins; the other group required free thiols for stability and the members were termed S-type lectins (later renamed the galectins, as not all of them were thiol-dependent). Meanwhile, the two lectins that recognized Man-6-P were sequenced and found to be homologous but distinct from all the others, justifying their designation as P-type lectins. Although some classes of lectins, such as the P-type and galectins, appear to recognize a single class of sugars (Man-6-P and β-galactosides, respectively), others like the C-type lectins recognize a variety of molecules that share only a lectin protein module in common. Figure 26.1 shows generic structures of several of these classes of lectins. With the availability of sequence data and crystal structures, it also became possible to classify the plant lectins into several distinct groups (see Table 26.2). Interestingly, it turns out that several of these groups have structural or sequence similarity with animal lectins, revealing how evolutionarily ancient these CRDs are.

A major breakthrough occurred when the independent cloning of three homologous vascular adhesin receptors revealed a common amino-terminal C-type lectin motif; these three molecules eventually turned out to be the selectins (see Chapter 31). This was the first time that glycan recognition had been predicted on the basis of the primary amino acid sequence of a cloned protein, validating the concept of classification based on sequence homology. The cloning of a variety of circulating soluble lectins also led to the recognition of a subset of C-type lectins designated as the "collectin family." In addition, two Ca^{++}-binding lectins (calnexin and calcireticulin) are unrelated to the C-type lectins (not all Ca^{++}-requiring lectins are C-type lectins), and they specifically recognize glucose residues on newly synthesized glycoproteins (see Chapter 36).

Studies in the 1990s also revealed that immunoglobulin superfamily members can recognize carbohydrates, leading to a new group of I-type lectins (see Chapter 32). A subgroup of these molecules, which specifically recognizes sialic acids, has been designated the "Siglecs" (for sialic acid–recognizing immunoglobulin superfamily lectins) (see Chapter 32). These and other general groupings are based primarily on sequence homologies and probable evolutionary relatedness and include the majority of known animal lectins. However, many others do not show any obvious sequence homologies or evolutionary relationships. For example, a class of evolutionarily very ancient circulating soluble lectins called the pentraxins are recognizable not so much by primary sequence homologies, but by a consistent pentameric structural organization and a role in the primary host immune response.

At the present time, there is no single universally accepted classification of lectins. Table 26.2 summarizes currently known lectins into two categories and makes some suggestions for naming of some of the groups in category I, which are characterized by sequence homologies or evolutionary relationships. As with all nomenclature issues, it will be up to investigators within the relevant fields to communicate with each other and to decide whether these suggestions are acceptable.

As mentioned earlier, glycosaminoglycan-binding proteins cannot be classified based on shared evolutionary origins, as they seem to have mostly evolved by convergent evolution. Table 16.8 in Chapter 16 presents a limited listing of examples of these types of GBPs. For further details, see Chapter 35.

NATURE OF GLYCAN–PROTEIN INTERACTIONS

Several crystal structures of GBPs with their cognate ligands have been determined, allowing an understanding of these interactions at the level of atomic resolution (see Chapter 27 for details). These can be divided into two general groups: those involving sulfated glycosaminoglycan chains (mostly mediated by ordered arrays of surface charge contacts; see Chapter 35) and those involving other classes of glycans (e.g., N- and O-glycans). Several principles have emerged about the latter group of glycan–protein interactions. First, the binding sites for low-molecular-weight glycans are of relatively low affinity (K_d values in the high micromolar to low millimolar range) and comprise shallow indentations on the surface of the proteins. Second, selectivity is mostly achieved via a combination of hydrogen bonds (involving the hydroxyl groups of the sugars) and by van der Waals packing of the hydrophobic face of monosaccharide rings against aromatic amino acid side chains. Third, further selectivity and enhanced affinity can be achieved by additional contacts between the glycan and the protein, sometimes involving bridging water molecules or divalent cations. Finally, the actual region of contact between the saccharide and the polypeptide typically involves only one to three monosaccharide residues, although there are many exceptions to this observation. As a consequence of all of the above, these lectin-binding sites tend to be of relatively low affinity, although they can exhibit high specificity. The ability of such low-affinity sites to mediate biologically relevant interactions in the intact system thus usually requires multivalency. Glycosaminoglycan-binding proteins differ from other GBPs in that binding usually involves five or more sugar residues. In some cases, the affinity can be quite high (K_d values in nanomolar to micromolar range), but the specificity in terms of linear sequence of sulfated residues is often relaxed.

MANY GLYCAN-BINDING PROTEINS ARE FUNCTIONALLY MULTIVALENT

Until the 1990s, all of the plant and animal lectins discovered were found to be naturally multivalent, either because of their defined multisubunit structure or by virtue of having multiple carbohydrate-binding sites within a single polypeptide. Indeed, increased avidity generated by multivalent binding of lower-affinity single sites appears to be a common mechanism for optimizing lectin function in nature, and a traditional definition for a lectin was"a multivalent

carbohydrate-binding protein that is not an antibody." Notable exceptions to this general rule are the selectins, which have only a single CRD site within their extracellular polypeptide domains (see Chapter 31). The same situation applies to the Siglecs (see Chapter 32) and many glycosaminoglycan-binding proteins (see Chapter 35). However, in these instances, these molecules can become functionally multimeric by noncovalent association and clustering on cell surfaces. Since many lectins may function as signaling molecules, their multivalency may promote cross-linking of relevant cell-surface receptors and may be required for signaling. Whether biologically significant binding by any plant or animal lectin can arise from a strictly monovalent interaction remains a difficult problem to address. It is also of note that a single lectin can carry multiple binding sites for multiple ligands; for example, the macrophage mannose receptor is now known to bind not only to mannans, but also via a distinct CRD to the 4-O-sulfated GalNAc residues of pituitary glycoprotein hormones.

DEFINING NATURAL LIGANDS FOR GLYCAN-BINDING PROTEINS

Despite the fact that they were discovered first, remarkably little is known about the natural intrinsic ligands for plant lectins. Indeed, to date the only definitive example is that of the recognition of *Rhizobium* Nod factors by LNP (lectin-nucleotide phosphohydrolase) lectins (see Chapter 37). On the other hand, plant lectins recognize many animal or prokaryotic glycans with a high degree of specificity (see Chapters 28, 29, 31, and 45), and some (but not all) have toxic properties. This suggests that some plant lectins were probably evolved for protection against other species and/or for mutually beneficial (symbiotic) reasons. Details about the natural ligands for animal lectins can be found in other chapters in this volume.

Despite their specificity, monosaccharides or small oligosaccharide units tend to be weak inhibitors of lectin interactions. The natural ligands for most lectins are typically complex glycoconjugates that carry clustered arrays of the cognate carbohydrate or unique glycan structures, thus cooperating with clustered lectin-binding sites to generate high-avidity binding, which is further enhanced by mass transport effects (high local concentrations of ligands). In some instances (e.g., the selectins), the nature of this clustering is not easily defined, and cooperation with other aspects of the underlying polypeptide may be necessary to generate optimal binding. For example, optimal binding of P- and L-selectin to the ligand PSGL-1 requires a precise ordered combination of an amino-terminal peptide sequence bearing an O-glycan with a terminal sialyl Lewis^x motif adjacent to a peptide rich in acidic residues and sulfated tyrosines (see Chapter 31). Partly for this reason, it is common to see the names of the underlying polypeptide backbone used to define the nature of a ligand, for example, "PSGL-1 is the ligand for P-selectin." However, it should be recognized that unless it is correctly glycosylated or otherwise modified (e.g., sulfated), the PSGL-1 polypeptide is not itself the ligand. Typically, these polypeptides are simply carriers of the true ligands for lectins, which are made up of combinations of glycan units.

Recombinant lectins that are often used to identify potential biological ligands are usually multimeric in structure or are presented in multivalent clustered arrays in soluble complexes or on solid supports (see Chapters 45 and 48). Thus, although a variety of molecules may be found to bind to a given recombinant lectin in a glycosylation-dependent manner, only a few of these "ligands" may actually be involved in mediating biologically significant interactions. The challenge then is to tell the difference between what *can* bind to a recombinant lectin in an in vitro experiment and what actually *does* bind to the native lectin in a biologically relevant manner in vivo. Indeed, the term "ligand" should probably be reserved for the latter type of biologically relevant structures. It should also be kept in mind that the natural ligands of some animal lectins may be present primarily on foreign invaders. For example, the circulating soluble mannan-binding protein may serve to bind and opsonize (mark for phagocytosis) microorganisms

that bear high densities of mannose, such as yeasts and other fungi. All of these issues are dealt with in more detail in Chapters 27–35.

A more complex variation is the concept of a "clustered saccharide patch" (CSP), which is generated by a peptide carrying multiple closely spaced glycans. Free glycans in solution or at single peptide attachment sites have significant freedom of motion in an aqueous environment. Thus, lectins that recognize linear glycan chains do so by capturing and "freezing" one of the many possible solution conformations of the glycan. However, glycans packed closely together (e.g., O-glycans in mucins or clustered glycosphingolipids in cell-surface rafts) have less mobility. Such clustering of common glycans could present uncommon CSPs, generated by multiple glycans closely spaced enough to restrict their motion and together they could form a unique epitope, perhaps aided or stabilized by peptide interactions with adjacent glycans. Releasing the glycans from the polypeptide backbone would result in complete loss of recognition by the lectin. There are examples of antibodies and enzymes whose specificities cannot be explained by single glycan chain recognition and which likely recognize such CSPs. Selectin recognition of heterogeneous carcinoma mucins and some features of plant lectin-recognition phenomena are also probably explained by CSPs. However, this concept has been more difficult to visualize and define at the scale of atomic resolution.

BIOSYNTHESIS, TRAFFICKING, AND REGULATION OF GLYCAN-BINDING PROTEINS

From a functional point of view, it is worth considering GBPs in two physical classes: soluble and membrane-bound. Cell membrane–bound GBPs are more likely to be involved in endocytosis, cell adhesion, or cell signaling and to stay confined to the cell type of their original synthesis. On the other hand, soluble GBPs are capable of diffusing locally in tissues and/or entering the blood circulation. Although useful in functional terms, this type of physical classification is confounded by two issues: First, some GBPs that start out as membrane-bound proteins can be proteolytically shed into the extracellular fluid. Second, soluble GBPs can become attached to cell surfaces or matrices via their glycan-binding sites. Figure 26.2 indicates some examples of these relationships and the nature of the potential interactions with natural ligands, which, in turn, can also be soluble or membrane-bound.

All membrane-bound and many soluble lectins are synthesized on ER-bound ribosomes and delivered to their eventual destinations via the ER-Golgi pathway. Thus, the GBPs themselves are often glycoproteins. However, a significant subset of soluble GBPs (galectins, heparin-binding growth factors, and some cytokines) are synthesized on free ribosomes in the cytoplasm and then delivered directly to the exterior of the cell by an as yet poorly understood mechanism. This makes functional sense, since several of these lectins can recognize biosynthetic intermediates that occur in the Golgi-ER pathway (e.g., galactosides and high-mannose oligosaccharides). The circumvention of the conventional pathway of secretion may allow these molecules to avoid unwanted premature interactions with potential ligands that are synthesized within the same cell.

Some GBP genes are expressed constitutively, whereas others are induced by gene activation under specific biological circumstances. E-selectin is a good example of a lectin that is transcriptionally regulated and expressed by endothelial cells only after activation by inflammatory mediators (see Chapter 31). Many lectins are regulated posttranslationally or after secretion. For example, some members of the galectin family of GBPs are sensitive to the redox state of the environment and can remain active only in the reducing environment of the cytoplasm. Upon entering the oxidizing environment of the extracellular space, they must therefore bind to ligands or become progressively inactivated. Another form of regulation occurs when the lectin binds to cognate sugar chains present on the same molecule or the same cell surface and hence becomes functionally inactive (e.g., the Siglecs, where sialic acid–bearing ligands from

the same cell surface can restrict the availability of the lectin-binding sites to external ligands) (see Chapter 32). Some membrane-bound lectins are also internalized upon binding to ligands, with delivery to internal acidic compartments (endosomes). The cargo can be released, allowing some of the receptors to recycle back to their original location, for example, the mannose 6-phosphate receptor. In some cases, the lectin is then transferred to the lysosome, where it undergoes degradation.

FURTHER READING

1. Stillmark H. Inaugural dissertation. University of Dorpat, Dorpat (now Tartu); Estonia: 1888. Uber Ricin, Ein giftiges Ferment aus den Samen von *Ricinus communis* L. und einigen anderen Euphoribiaceen.
2. Goldstein IJ, Hughes RC, Monsigny M, Osawa T, Sharon N. What should be called a lectin? Nature. 1980;285:66.
3. Ashwell G, Harford J. Carbohydrate-specific receptors of the liver. Annu Rev Biochem. 1982;51:531–554. [PubMed: 6287920]
4. Drickamer K. Two distinct classes of carbohydrate-recognition domains in animal lectins. J Biol Chem. 1988;263:9557–9560. [PubMed: 3290208]
5. Kornfeld S. Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. Annu Rev Biochem. 1992;61:307–330. [PubMed: 1323236]
6. Drickamer K, Taylor ME. Biology of animal lectins. Annu Rev Cell Biol. 1993;9:237–264. [PubMed: 8280461]
7. Varki A. Selectin ligands. Proc Natl Acad Sci. 1994;91:7390–7397. [PMC free article: PMC44407] [PubMed: 7519775]
8. Powell LD, Varki A. I-type lectins. J Biol Chem. 1995;270:14243–14246. [PubMed: 7782275]
9. McEver RP, Cummings RD. Role of PSGL-1 binding to selectins in leukocyte recruitment. J. Clin. Invest. . 1997;100:485–492. [PMC free article: PMC508213] [PubMed: 9239393]
10. Etzler ME. Oligosaccharide signaling of plant cells. J. Cell Biochem. Suppl. 1998:30–31. 123–128. [PubMed: 9893263]
11. Cebo C, Vergoten G, Zanetta JP. Lectin activities of cytokines: Functions and putative carbohydrate-recognition domains. Biochim. Biophys. Acta. 2002;1572:422–434. [PubMed: 12223284]
12. Esko JD, Selleck SB. Order out of chaos: Assembly of ligand binding sites in heparan sulfate. Annu Rev Biochem . 2002;71:435–471. [PubMed: 12045103]
13. Helenius A, Aebi M. Roles of N-linked glycans in the endoplasmic reticulum. Annu Rev Biochem. 2004;73:1019–1049. [PubMed: 15189166]
14. Varki A, Angata T. Siglecs—The major subfamily of I-type lectins. **Glycobiology**. 2006;16:1R–27R. [PubMed: 16014749]
15. Rabinovich GA, Toscano MA, Jackson SS, Vasta GR. Functions of cell surface galectin-glycoprotein lattices. Curr Opin Struct Biol. 2007;17:513–520. [PMC free article: PMC2100406] [PubMed: 17950594]
16. Sharon N. Lectins: Carbohydrate-specific reagents and biological recognition molecules. J Biol Chem. 2007;282:2753–2764. [PubMed: 17145746]
17. Varki A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature. 2007;446:1023–1029. [PubMed: 17460663]

Figures

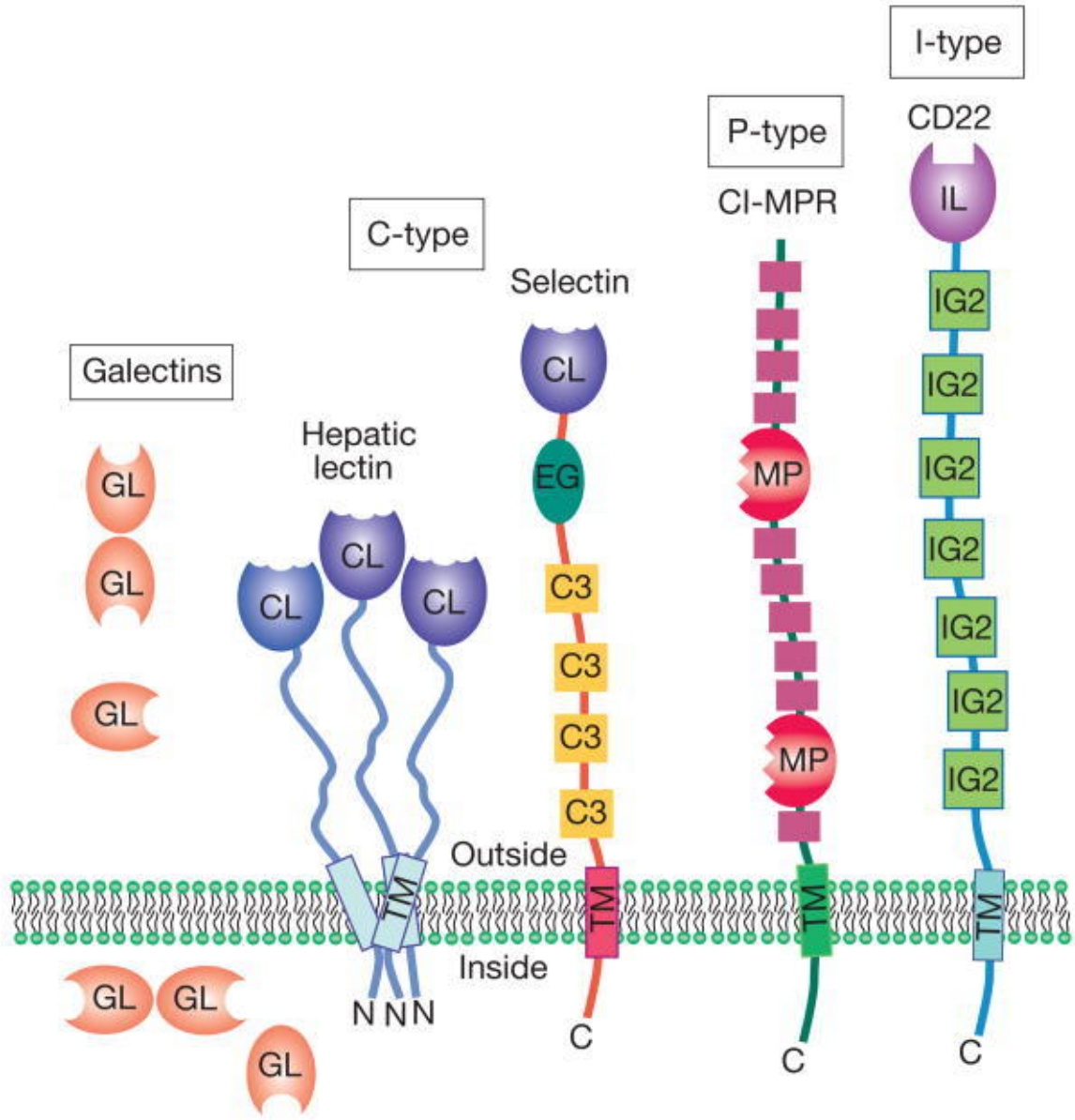

FIGURE 26.1

Schematic examples of major types of animal lectins, based on protein structure. Examples of some of the major families are shown. The emphasis is on the extracellular domain structure and topology. The following are the defined carbohydrate-binding domains (CRDs) shown: (CL) C-type lectin CRD; (GL) S-type lectin CRD; (MP) P-type lectin CRD; (IL) I-type lectin CRD. Other domains are (EG) EGF-like domain; (IG2) immunoglobulin C2-set domain; (TM) transmembrane region; and (C3) complement regulatory repeat. The number of domains underlying the CRD can vary among family members.

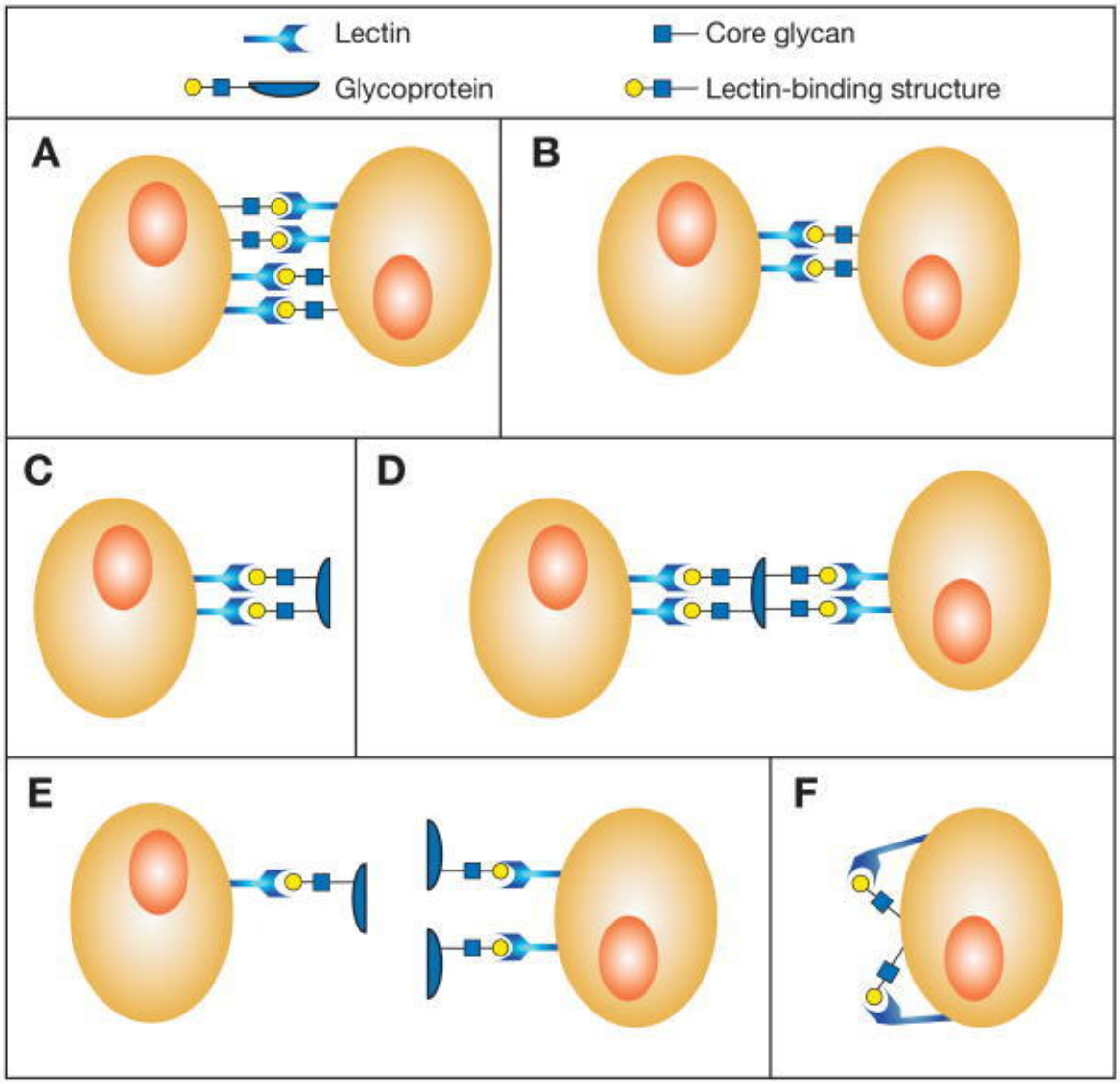

FIGURE 26.2

Possible mechanisms of regulation of an animal lectin by cognate ligands. Potential ligands can be on the cell surface and/or on soluble glycoproteins (including sugar chains attached to the lectin itself). As discussed in the text, the "ligand" depicted in this cartoon can either be a simple linear or terminal glycan sequence or a very complex motif that includes more than one monosaccharide and/or additional components. Direct cell–cell interactions could occur among lectin-positive cells (*A*) or between lectin-positive cells and other cell types bearing cognate ligands (*B*). Soluble glycoprotein ligands could interact directly with lectin-positive cells (*C*), bridge between two such cells (*D*), or inhibit cell–cell interactions involving the lectin (*E*). Expression of ligands on lectin-positive cells could inactivate the lectin function by either inter- or intramolecular interactions (*F*).

Tables

TABLE 26.1

Comparison of the two major classes of glycan-binding proteins

	Lectins[a]	**Glycosaminoglycan-binding proteins**[b]
Shared evolutionary origins	yes (within each group)	no
Shared structural features	yes (within each group)	no
Defining AA residues involved in binding	often typical for each group	patch of basic amino acid residues
Type of glycans recognized	N-glycans, O-glycans, glycosphingo-lipids (a few also recognize sulfated glycosaminoglycans)	different types of sulfated glycosaminoglycans
Location of cognate residues within glycans	typically in sequences at outer ends of glycan chains	typically in sequences internal to an extended sulfated glycosaminoglycan chain
Specificity for glycans recognized	stereospecificity high for specific glycan structures	often recognize a range of related sulfated glycosaminoglycan structures
Single-site binding affinity	often low; high avidity generated by multivalency	often moderate to high
Valency of binding sites	multivalency common (either within native structure or by clustering)	often monovalent
Subgroups	C-type lectins, galectins, P-type lectins, I-type lectins, L-type lectins, R-type lectins etc.	heparan sulfate–binding proteins, chondroitin sulfate–binding proteins, dermatan sulfate–binding proteins
Types of glycans recognized within each group	can be similar (e.g., galectins) or variable (e.g., C-type lectins)	classification itself is based on type of glycosaminoglycan chain recognized

Modified from Varki A. and Angata T. 2006. ***Glycobiology*** **16:** 1R–27R.

a There are other animal proteins that recognize glycans in a lectin-like manner and do not appear to fall into one of the well-recognized classes (e.g., various cytokines).

b Hyaluronan (HA)-binding proteins (hyaloadherins) fall in between these two classes. On the one hand, some (but not all) of the hyaloadherins have shared evolutionary origins. On the other hand, recognition involves internal regions of HA, which is a nonsulfated glycosaminoglycan.

TABLE 26.2

A general classification of lectins and lectin-like proteins

Category I—Defined lectin families with structural and/or evolutionary sequence similarities[a]
β-prism lectins—Jacalin-related (B-type?)
C-type lectins (e.g., calcium-dependent lectins such as selectins, collectins, etc.)
eel fucolectins (E-type?)
ficolins—fibrinogen/collagen-domain-containing lectins (F-type?)
garlic and snowdrop lectins and related proteins (G-type?)
galectins (formerly S-type lectins)
hyaluronan-binding proteins or hyaladherins (H-type?)
I-type lectins—immunoglobulin superfamily members, including the Siglec family
amoeba lectins—Jacob and related chitin-binding proteins (J-type?)
L-type lectins, plant legume seed lectins, ERGIC-53 in ER-Golgi pathway, calnexin family
M-type lectins—α-mannosidase-related lectins (e.g., EDEM)
N-type lectin nucleotide phosphohydrolases (LNPs) with glycan-binding and apyrase domains
P-type (i.e., mannose-6-phosphate receptors)
R-type (e.g., ricin, other plant lectins, GalNAc-SO_4 receptors)
tachylectins from horseshoe crab *Tachypleus tridentatus* (T-type?)
haevin-domain lectins (e.g., wheat germ agglutinin, haevin, etc.) (W-type?)
Xenopus egg lectins/eglectins (X-type?)
Category II—Lectin-like proteins without established evolutionary classification
some annexins
pentraxins with pentavalent domain structure
some laminin G domains, recognizing glycans on α-dystroglycan
CD11b/CD18 (β3-integrin, CR3) recognizes fungal glucans and exposed GlcNAc residues on glycoproteins
complement factor H, recognizes cell-surface sialic acids as “self”
tumor necrosis factor-α, binds to oligomannose N-glycans
interleukins IL-1α, IL-1β, IL-4, IL-6, and IL-7 bind various glycans
amphoterin binds carboxylated N-glycans

a Category I includes all lectin families with generally agreed-upon names (e.g., C-type and R-type). In addition, the question marks (e.g., F-type? and X-type?) indicate suggested names for other families. Final acceptance of the latter terms will require a consensus among those scientists who study each respective family.

Copyright © 2009, The Consortium of Glycobiology Editors, La Jolla, California.

Bookshelf ID: NBK1923 PMID: 20301249