Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition†

Halina Lis and Nathan Sharon*

Department of Membrane Research and Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel

Received July 25, 1997 (Revised Manuscript Received January 20, 1998)

Contents

I. Introduction

Proteins that interact with carbohydrates noncovalently occur widely in nature. Prominent examples are carbohydrate-specific enzymes and anticarbohydrate antibodies. In recent years, another class of such proteins, the lectins,[1−5] has come into the forefront of biological research. Lectins bind mono- and oligosaccharides reversibly and with high specificity, but are devoid of catalytic activity, and in contrast to antibodies, are not products of an immune response. Each lectin molecule contains typically two or more carbohydrate-combining sites, i.e., they are di- or polyvalent. Therefore, when they react with cells, for example erythrocytes, they will not only combine with the sugars on their surfaces, but will also cause cross-linking of the cells and their subsequent precipitation, a phenomenon referred to as cell agglutination. The erythrocyte agglutinating, or hemagglutinating, activity of lectins is a major attribute of these proteins and is used routinely for their detection and characterization. Lectins also form cross-links between polysaccharide or glycoprotein molecules in solution and induce their precipitation. Both the agglutination and precipitation reactions of lectins are inhibited by the sugar ligands for which the lectins are specific.

Lectins are found in most organisms, ranging from viruses and bacteria to plants and animals. They are readily obtainable in purified form, mostly by affinity chromatography on the immobilized ligand, and more recently also by recombinant DNA techniques. They represent a heterogeneous group of oligomeric proteins that vary widely in size, structure, molecular organization, as well as in the constitution of their combining sites. Nonetheless, many of them belong to distinct protein families with similar sequences and structural features. In fact, sequence similarity with known lectins provide a novel guideline for the detection and identification of new ones.

† Abbreviations used: CRD, carbohydrate recognition domain; ECorL, *Erythrina corallodendron* lectin; GNA, snowdrop agglutinin; Ig, immunoglobulin; LOL, *Lathyrus ochrus* lectin; MBP, mannose binding protein; PDP, protein data bank; PHA, kidney bean lectin; PNA, peanut agglutinin; RCA, *Ricinus communis* agglutinin; SAP, serum amyloid P component; SBA, soybean agglutinin; WGA, wheat germ agglutinin. For abbreviation of oligosaccharide names, see Table 5.

* Address for correspondence: Dr. Nathan Sharon, Department of Membrane Research and Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel. Telephone: 972-8-9343605. Fax: 972-8-9468256. E-mail: bfsharon@weizmann.weizmann.ac.il.

S0009-2665(94)00413-9 CCC: $30.00
Published on Web 03/19/1998

Halina Lis

Nathan Sharon

Halina Lis and Nathan Sharon have been collaborating in research on lectins since the early 1960s, mainly on those from soybean, *Erythrina corallodendron*, and *Moluccella laevis*. They published extensively on the subject, including some two dozen major reviews, several of which are widely cited, as well as a book. Lis did graduate work with Arne Tiselius at the University of Uppsala, where she received her Ph.D. in 1957. After two years of research at the University of Rome and the Istituto di Sanitá, also in Rome, she joined the Department of Biophysics of the Weizmann Institute, headed by Ephraim Katchalski-Katzir, and was promoted to Associate Professor in 1986.

Sharon did his graduate studies with Aharon Katchalsky, and gained a Ph.D. degree from the Hebrew University, Jerusalem, in 1953. He subsequently joined the Department of Biophysics of the Weizmann Institute as Research Associate, did postdoctoral research in the laboratories of Fritz Lipmann, Roger Jeanloz, and Dan Koshland, was promoted to Associate Professor in 1965 and to Professor in 1968. Between 1973 and 1990, he held intermittently the position of Department Head, and for several years he served as Dean of the Faculty of Biophysics and Biochemistry. Sharon was visiting Professor at many institutions in the United States and Europe. He was President of the Israel Biochemical Society (1969–1970), of the Federation of European Biochemical Societies (FEBS) (1980–81) and of the International Glycoconjugate Organization (1989–1991). He is the recipient of a number of awards and honors, among them Membership of the European Molecular Biology Organization (EMBO) and of the Israel Academy of Sciences and Humanities; Honorary Membership of the American Society of Biological Chemists; Bijvoet Medal, Utrecht University; Docteur Honoris Causa, Université René Descartes, Paris; and Israel Prize in Biochemistry and Medicine. In addition to research and teaching, Sharon has been active in communicating science (in Hebrew) to the Israeli public, on the radio and in the daily press.

Although lectins were first described at the turn of the century, their study started to gain momentum only in the 1960s.[2,6,7] They were then shown to be invaluable tools for the structural and functional investigation of complex carbohydrates, especially glycoproteins, and for the examination of changes that occur on cell surfaces during physiological and pathological processes, from cell differentiation to cancer.[8,9] At present, they are the focus of intense attention because of the realization that they act as recognition determinants in diverse biological processes.[10,11] These include clearance of glycoproteins from the circulatory system, control of intracellular traffic of glycoproteins, adhesion of infectious agents to host cells, recruitment of leukocytes to inflammatory sites, as well as cell interactions in the immune system, in malignancy and metastasis. Investigation of lectins and their role in cell recognition, as well as the application of these proteins for the study of carbohydrates in solution and on cell surfaces, are making marked contributions to the advancement of glycobiology.[12] Developments in the latter field are having a significant impact on lectin research, so that the two are now moving ahead hand in hand.

During the past decade, there has been remarkable progress in elucidating the features of lectins that are important for carbohydrate binding. This was made possible by the refinement of old techniques and development of new ones. In particular, high-resolution X-ray crystallography of lectins in complex with their ligands allowed the identification of the chemical groups on the protein and on the carbohydrate that interact with each other and of the types of bond formed. Further information on the contribution of individual amino acids to the activity of a lectin has been obtained by site-directed mutagenesis experiments and also by molecular modeling. Of special interest are the studies of lectin−oligosaccharide complexes, since they provide a basis for the understanding of how lectins recognize their natural ligands.

In this article we deal mainly with the specificity and structure of lectins, with emphasis on their carbohydrate binding sites and the mechanism of lectin−carbohydrate interactions, and we also discuss briefly their roles and applications. For recent reviews on the subject, see refs 4 and 13−18. Bacterial toxins that are carbohydrate binding proteins, although sometimes considered as lectins,[19] will not be covered.

II. Carbohydrate Specificity

A. Monosaccharides

On the basis of their specificity, lectins are classified into five groups, according to the monosaccharide for which they exhibit the highest affinity: mannose, galactose/*N*-acetylgalactosamine, *N*-acetylglucosamine, fucose, and *N*-acetylneuraminic acid (sugars are of the D configuration except for fucose which is L) (Table 1). Relevant for the biological activities of lectins is the fact that of the numerous monosaccharides found in nature, only those listed above are typical constituents of surfaces of eukaryotic cells. Only in exceptional cases does one find lectins that exhibit affinity for other sugars. One example is the human serum amyloid P component (SAP) (see section III.A.5), a lectin specific for the 4,6-cyclic pyruvate acetal of galactose;[20] to date, this rare

Table 1. Specificity Groups[a]

lectin			
source	name/abbrev	preferred oligosaccharide[b]	RA[c]
mannose[d]			
jackbean	concanavalinA/ConA	Manα6(Manα3)Man	130
Escherichia coli[e]	type 1 fimbriae		
fava bean	favin		
Galanthus nivalis (snowdrop)[e]	GNL	Manα6(Manα3)Man	
Lathyrus ochrus	LOL	octassaccharide	
lentil	LCL		
rat serum	MBP-A[f]		
pea	PSL	fucose-containing hexasaccharide	
N-acetylglucosamine			
Griffonia simplicifolia	GSII		
wheat germ	WGA	$(GlcNAc\beta4)_3$	3000
galactose/*N*-acetylgalactosamine			
Artocarpus integrifolia (jackfruit)	jacalin	Galβ3GalNAc	
Dolichos biflorus[g]	DBL	GalNAcα3GalNAc	36[h]
Erythrina corallodendron (coral tree)	ECorL	Galβ4GlcNAc	30−50[i]
Helix pomatia (snail)[g]			
lima bean[g]	LBA	GalNAcα3(Fucα2)Gal	43[h]
Moluccella laevis[g] (bells of Ireland)	MLL		
peanut[j]	PNA	Galβ3GalNAc	50[i]
ricin	RCA II		
soybean[d]	SBA		
fucose			
Anguilla anguilla (eel)			
Lotus tetragonolobus	LTA		
Ulex europeus	UEA I	Fucα2Galβ4GlcNAcβ6R	900
sialic acid			
Sambucus nigra (elderberry)		NeuAcα2,3Gal	30−80
		NeuAcα2,6Gal	1600
Limulus polyphemus (horseshoe crab)		NeuAcα2,6GalNAc	30

[a] For references, see ref 8. [b] For structures of oligosaccharides not shown, see Table 3. [c] Relative affinity compared to that of the monosaccharide; usually based on hemagglutination inhibition assays. [d] Most lectins in this group bind also glucose, often with similar affinity. [e] Lectin does not bind glucose. [f] Although termed mannose binding, this lectin binds mannose, *N*-acetylglucosamine and fucose with roughly equal affinities. A similar protein, designated MPB−C is found in mammalian liver. [g] Lectin exhibits pronounced preference for *N*-acetylgalactosamine. [h] With *N*-acetylgalactosamine as reference monosaccharide. [i] With galactose as reference monosaccharide. [j] Does not bind *N*-acetylgalactosamine.

carbohydrate was found in certain algal polysaccharides, in a marine sponge, and in a yeast, but not in bacteria nor higher organisms.

The affinity of the lectins for monosaccharides is usually weak, with association constants in the millimolar range, yet it is often highly selective.[2,21] In particular, lectins specific for galactose do not react with glucose (its 4 epimer) or mannose (the 2 epimer of glucose), nor do those specific for mannose bind galactose. Similarly, with the exception of wheat germ agglutinin (see below), members of the *N*-acetylglucosamine specificity group do not combine with *N*-acetylgalactosamine (and vice versa). However, the selectivity of lectins for monosaccharides is not always so high. Thus, many lectins tolerate variations at C-2 of the pyranose ring and those of the mannose specificity group may bind the epimeric glucose as well. Most lectins that bind galactose interact also with *N*-acetylgalactosamine, in some cases preferentially, e.g., soybean agglutinin (SBA), the affinity of which for the latter monosaccharide is 25−50 times higher than that for galactose. Others bind both monosaccharides with nearly the same affinity, as is the case with the *Erythrina corallodendron* (coral tree) lectin (ECorL). For this reason they are classified as one specificity group, Gal/GalNAc, even though certain of them (e.g., peanut agglutinin, PNA) do not bind *N*-acetylgalactosamine at all. Occasionally, lectins combine with monosaccharides that appear structurally unrelated, but that present similar topographical features when appropriately viewed. For instance, wheat germ agglutinin binds both *N*-acetylglucosamine and *N*-acetylneuraminic acid, and in contrast to other *N*-acetylglucosamine-specific lectins *N*-acetylgalactosamine as well, although more weakly. Consideration of the three-dimensional structures of these monosaccharides reveals similarity at positions C-2 (acetamide group) and C-3 (hydroxyl group) of the pyranose ring of the two hexosamines with those of C-5 and C-4 on *N*-acetylneuraminic acid, respectively (Figure 1); these are the positions critical for productive contact with the combining site of the lectin (cf. section IV.A.2). Also, mannose-specific animal lectins (e.g., the rat mannose binding proteins, MBP's) bind fucose too (Figure 1).

Certain lectins belonging to the same specificity group combine preferentially, or almost exclusively, either with the α- or β-glycosides of the respective monosaccharide, whereas others lack anomeric specificity. The properties of the aglycon may markedly influence the interaction of a glycoside with a lectin. In particular aromatic glycosides bind to many lectins much more strongly than aliphatic ones, attesting to the presence of a hydrophobic region close to the carbohydrate-combining site. The hydrophobic effect is at times so strong that lectins that show a marked preference for methyl α-glycosides over the corre-

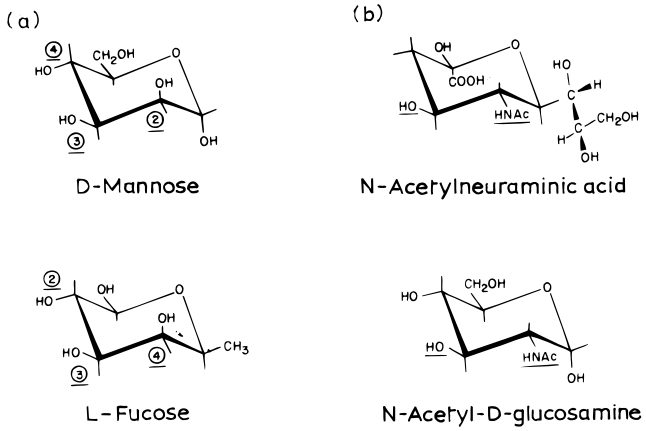


Figure 1. Common structural features of mannose and fucose (A) and of *N*-acetylneuraminic acid and *N*-acetylglucosamine (B). Groups that occupy the same position in space are underlined. (A) Rotation of the fucose molecule by 180° allows superimposition of its ring oxygen, 4-OH, 3-OH, and 2-OH with the ring oxygen, 2-OH, 3-OH, and 4-OH of mannose, respectively. (B) Conformational similarity of *N*-acetylglucosamine and *N*-acetylneuraminic acid at the underlined positions (acetamide and hydroxyl) of the pyranose rings is observed when the sialic acid molecule is suitably rotated. The conformation of *N*-acetylgalactosamine (not shown) at the relevant positions is identical with that of *N*-acetylglucosamine. (Reprinted by permission from ref 2. Copyright 1989 Chapman and Hall Ltd.)

sponding β anomers exhibit an inverse anomeric specificity when tested with the corresponding *p*-nitrophenyl glycosides.[22] The lectins within each group may also differ markedly in their affinity for other derivatives. For example, concanavalin A and favin, the lectin from fava bean, both bind glucose equally well. However, while the affinity of the 3-O methyl or phenyl ethers of glucose to concanavalin A is 10−20 times weaker than that of glucose, for favin it is 3−4 times higher. Quite unusually, concanavalin A interacts also with peptides that contain the Tyr-Pro-Tyr motif, with an affinity close to that of methyl α-mannoside.[23,24] Such peptides may bind to the lectin by hydrogen bonding with the hydroxyl groups of the tyrosines (which mimic sugar oxygens) and hydrophobic interactions with carbons on the aromatic side chain (which mimic sugar cabons). The peptide and carbohydrate ligands were shown to bind to the lectin at the same site, thus representing a case of true glycomimetics.[25]

B. Oligosaccharides

The classification of lectins according to their monosaccharide specificity masks the fact that they often exhibit an exquisite specificity for di-, tri-, and tetrasaccharides (with association constants up to 1000-fold higher as compared with the monosaccharide) (Table 1) and that certain lectins interact only with oligosaccharides (Table 2). Moreover, lectins of the same specificity group may differ markedly in their affinities for different oligosaccharides. From the functional point of view, binding of oligosaccharides is of special significance since, as mentioned earlier, they are most likely the natural ligands of lectins. The affinities of lectins to oligosaccharides may be influenced by the shape of the latter compounds which are flexible molecules with considerable freedom of rotation around the glycosidic bonds connecting the individual monosaccharide constituents. This has been demonstrated by molecular modeling, as well as by high-resolution nuclear magnetic resonance (NMR) studies of oligosaccharides in solution.[26−31] For instance, in the oligosaccharide Man(α1−3)[Man(α1−6)]Man(β1−4)GlcNAc-(β1−4)GlcNAc (the pentasaccharide core, present in all asparagine-linked carbohydrate chains of glycoproteins)[12,32,33] and many of its derivatives, the α1−6-linked mannose can form two rotational isomers relative to the C5−C6 bond of the β1−4-linked mannose. The prevalence of either of the two isomers depends on the type of substitution on the mannose residues of the core. In particular, attachment of a *N*-acetylglucosamine linked β1−4 to Man(β1−4) ("bisecting" *N*-acetylglucosamine) fixes the orientation of the Man(α1−6)Man arm into one of the two possible conformations and markedly decreases the binding of the oligosaccharide to concanavalin A (Figure 2). Because of their flexibility, oligo-

Table 2. Lectins Specific for Oligosaccharides Only

lectin	abbrev	oligosaccharide
Escherichia coli	type P fimbriae	Galα4Gal
	K99 fimbriae	NeuGcα2,3Galβ4GlcNAc
galectins		Galβ4Glc; Galβ4GlcNAc
Griffonia simplicifolia IV	GSIV	Fucα2Galβ3(Fucα4)GlcNAc
Phaseolus vulgaris	E-PHA	Galβ4GlcNAcβ2Manα6) GlcNAcβ4Manβ4-R[a] GlcNAcβ2Manα3)
	L-PHA	Galβ4GlcNAcβ6 Man Galβ4GlcNAcβ2)
potato		(GlcNAcβ4)$_{2-4}$
selectins		Neu5Acα2,3Galβ4GlcNAc Fucα3 (sLe[x])
tomato		(GlcNAcβ4)$_{3-4}$

[a] R = GlcNAc(β1−4)GlcNAc.

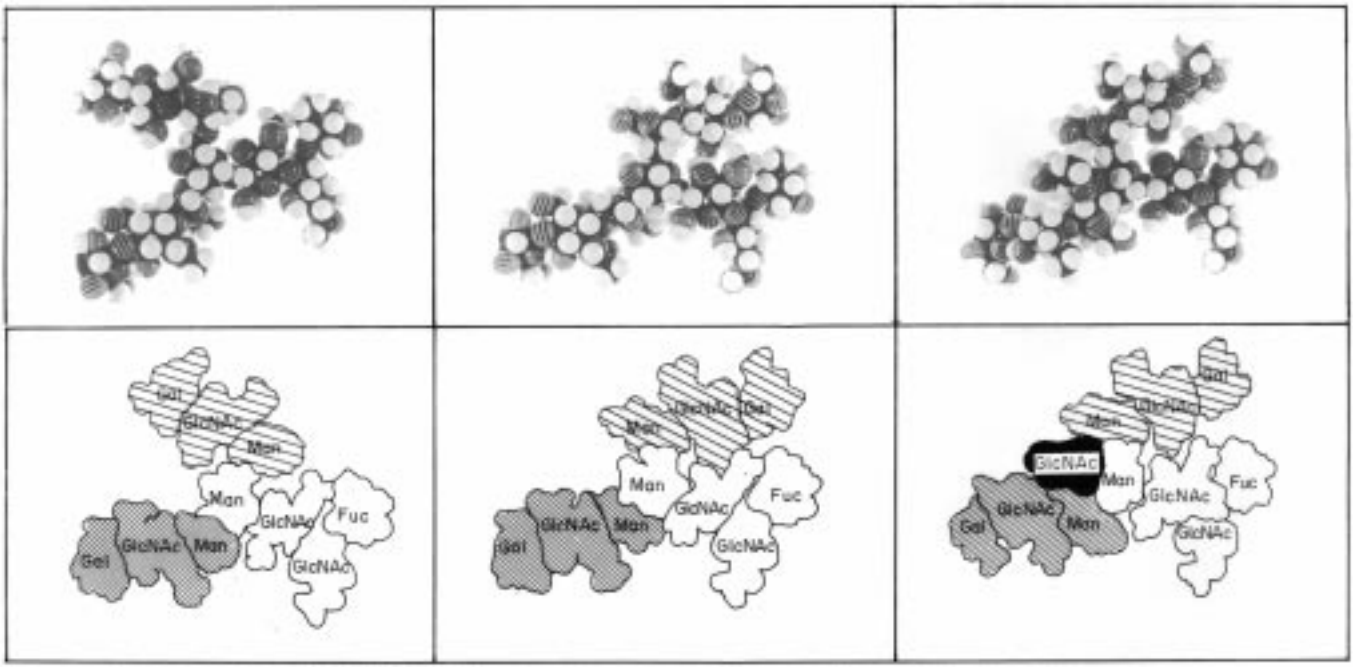


Figure 2. Space-filling models (top) and silhouettes (bottom) of the dibranched complex decasaccharide (entry 14 in Table 5) (left and center) and of the same disaccharide with a bisecting *N*-acetylglucosamine (entry 15 in Table 5) (right). The unbisected structure can adopt two orientations about the α1−6 linkage, whereas the bisected analogue can adopt only a single orientation about this linkage. Fine shaded area, α1−3 arm; striped area, α1−6 arm; black area, bisecting *N*-acetylglucosamine. (Reprinted by permission from ref 2. Copyright 1989 Chapman and Hall Ltd.)

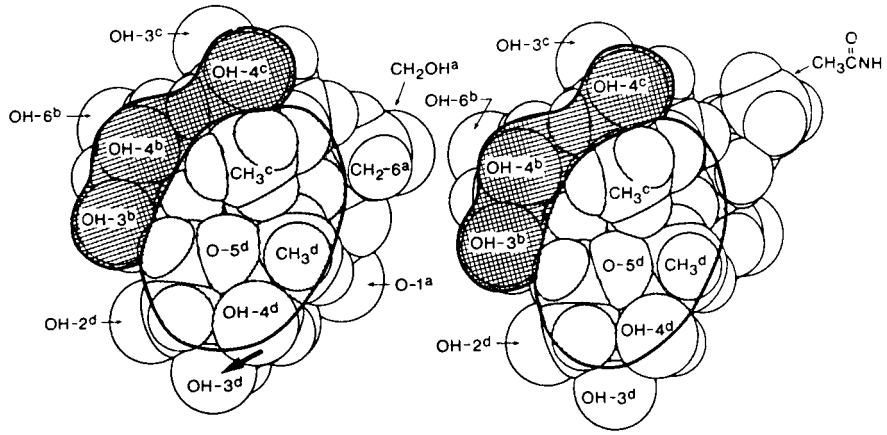
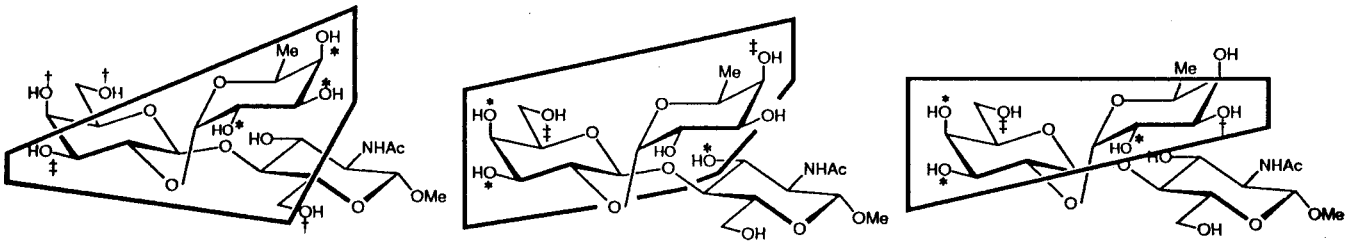


Figure 3. Computer drawn models (top) and chemical structures (bottom) of the Le^b and Le^y blood group determinants. The a, b, c and d superscripts refer to atoms on the *β*-*N*-acetylglucosamine, *β*-galactose, α1−4-linked fucose, and α1−2-linked fucose residues, respectively. The shaded areas represent regions that are bound by *Griffonia simplicifolia* lectin IV. These surfaces are present in both tetrasaccharides, which react almost equally well with the lectin. (Reprinted by permission from Spohr, U.; Hindsgaul, O.; Lemieux, R. U. *Can. J. Chem.* **1985**, *63*, 2644−2652. Copyright 1985 National Research Council of Canada.)

saccharides that differ in their chemical structure may have substantial topographic features in common and, as a result of this similarity, will bind to the same lectin (Figure 3). On the other hand, different lectins specific for the same oligosaccharide, may recognize different regions of its surface (Figure 4).

Flexibility around glycosidic linkages leads to conformational heterogeneity. However, lectins bind branched oligosaccharides in a single conformation, which is not necessarily the most populated one in solution.[26−31] Upon binding, the rotational freedom of the oligosaccharide becomes restricted, resulting



Ulex europaeus I *Galactia tenuiflora* *Psophocarpus tetragonolobus II*

Figure 4. The involvement of the eight hydroxyl groups of the H-type trisaccharide MeβFuc(α1−2)Gal(β1−4)GlcNAc in its complex with three different lectins. (Reprinted by permission from Du, M.-H.; Spohr, U.; Lemieux, R. *Glycoconjugate J.* **1994**, *11*, 443−461. Copyright 1994 Chapman and Hall, Ltd.)

Table 3. Simple Lectins and Lectin−Carbohydrates Complexes with Known Tridimensional Structure

family	lectin	abbrev	ligand[a]	Å	ref
		Plant Lectins			
legume	concanavalin A	ConA[b]		2.0	*c*
				1.2	*d*
			MeαMan	2.0	123
			MeαGlc	2.0	*e*
			Manα3(Manα6)Man	2.3	129
	Erythina corallodendron	ECorL	Galβ4Glc	1.7	45
	fava bean	favin	MeαMan		*f*
	Griffonia simplicifolia	GSIV	Fucα2Galβ3(Fucα4)GlcNAc	2.0	*g*
	red kidney bean	PHA	complex pentasaccharide	2.8	46
	Lathyrus ochrus	LOL I		1.9	*h*
			Manα3Manβ4GlcNAc, complex	2.1	132
			octasaccharide	2.3	133
		LOL II	decasaccharide	3.3	134
	lentil	LCL		1.8	*i*
			MeαMan, MeαGlc	2.0, 2.2	125
	pea	PSL		3.0	*j*
			Manα3(Manα6)Man	2.6	*k*
	peanut	PNA	Galβ4Glc	2.25	47
	soybean	SBA	biantennary pentasaccharide	2.6	48
cereal	wheat germ	WGA	NeuAc(α2−3)Galβ4Glc GlcNAcβ4GlcNAc	2.2	50
			sialoglycopeptide	2.0	51
Amaryllidaceae	snowdrop	GNA	MeαMan	2.3	55
			mannopentaose	2.0	56
Moraceae	*Artocarpus integrifolia*	Jacalin	MeαGal	2.43	58
		Animal Lectins			
Galectins	Human heart	Galectin 1	Galβ4GlcNAc	1.9	69
			octasaccharide	2.15−2.45	165
	Rat liver	Galectin 2	Galβ4Glc	2.9	70

[a] The sugars are bound to the protein with conformations in the $^4C_1$ or $^1C_4$ pyranose form. [b] Structures of demetalized lectin,[41] as well as lectin in which the Ca^{2+} and Mn^{2+} were substituted with other metals are also available (Naismith, J. H.; Habash, J.; Harrop, S.; Helliwell, J. R.; Hunter, W. N.; Wan, T. C. M.; Weisgerber, S.; Kalb (Gilboa), A. J.; Yariv, J. *Acta Crystallogr.* **1993**, *D49*, 561. Emmerich, C.; Helliwell, J. R.; Redshaw, M.; Naismith, J. H.; Harrop, S. J.; Raftery, J.; Kalb (Gilboa), A. J.; Yariv, J.; Dauter, Z.; Wilson, K. S. *Acta Crystallogr.* **1994**, *D50*, 749). [c] Weisgerber, S.; Helliwell, J. R. *J. Chem. Soc. Faraday Trans.* **1993**, *89*, 2667. [d] Parkin, S.; Rupp, B.; Hope, H. *Acta Crystallogr.* **1996**, *D52*, 1161. [e] Harrop, S. J.; Naismith, J. H.; Emmerich, C.; Habash, J.; Weisgerber, S.; Kalb (Gilboa), A. J.; Yariv, J.; Helliwell, J. R. *Acta Crystallogr.* **1993**, *A49*, C-94. [f] Reeke, G. N., Jr.; Becker, J. W. *Science* **1986**, *234*, 1108. [g] Delbaere, L. T. J.; Vandonselaar, M.; Prasad, L.; Wilson, K. S.; Dauter, Z. *J. Mol. Biol.* **1993**, *230*, 950. [h] Bourne, Y.; Abergel, C.; Cambillau, C; Frey, M.; Rougé, P.; Fontecilla-Camps, J.-C. *J. Mol. Biol.* **1990**, *214*, 571−584. [i] Loris, R.; Van Overberge, D.; Dao-Thi, M.-H.; Poortmans, F.; Maene, N.; Wyns, L. *Proteins Struct. Funct. Genet.* **1994**, *20*, 330−346. [j] Einspahr, H.; Parks, E. H.; Suguna, K.; Subramanian, E. Suddath, F. L. *J. Biol. Chem.* **1986**, *261*, 16518. [k] Rini, J. M.; Hardman, K. D.; Einspahr, H.; Suddath, F. L.; Carver, J. P. *J. Biol. Chem.* **1993**, *268*, 10126.

in a decrease in the entropy of the system (see section IV.C).

III. Molecular Structure

The amino acid sequences of several hundreds of lectins, and in addition the three-dimensional structures of some two dozen of them, almost all in complex with a ligand (Tables 3 and 4), have been elucidated and new sequences and structures are being added at an increasing rate. This makes it possible to replace the traditional division of lectins according to their origin, i.e., plants, animals, and microorganisms, by a classification based on common structural features. Most lectins fall clearly into one of the following three classes: (a) simple, (b) mosaic (or multidomain), and (c) macromolecular assemblies, although borderline cases exist. Within each class, lectins can be grouped into distinct families with similar sequences and structural properties. Oc-

Table 4. Multidomain (Mosaic) Lectins and Lectin−Carbohydrate Complexes with Known Structure

family	lectin	abbrev	source	ligand	Å	ref
viral lectins						
influenza virus	hemagglutinin	HA				
wild type			human isolates	NeuAc(α2−6)Galβ4Glc	3.0	139
wild type				Meα4-O-Ac-NeuAc,	2.9	74
wild type				NeuAc derivatives	2.15−3.0	139
mutant				NeuAc(α2-3)Galβ4Glc	2.9	139
polyoma virus				NeuAc(α2-3)Galβ4Glc	3.65	143
				NeuAc(α2-3)Galβ3GlcNAcβ3Galβ4Glc / NeuAc(α2-6)	1.9	144
polyoma virus VP1				NeuAc(α2-3)Galβ3GlcNAcβ3Galβ4Glc / NeuAc(α2-6)	1.9	75
animal lectins						
C-type	mannose-binding protein	MBP-A	rat serum			
wild type				Oligomannose glycopeptide	1.7	145
mutant				Galβ4Glc	1.9−2.0	149
	mannose-binding protein	MBP-C	rat liver	MeαMan, MeαGlcNAc, MeαFuc, MeβFuc, Gal	1.7−1.9	145
pentraxins	serum amyloid protein	SAP	human serum	Me β-4,6-*O*-(1-carboxyethylidene)-Gal	2.0	20

casionally, similar families are found in phylogenetically unrelated organisms, e.g., plants and animals, and are considered as belonging to one superfamily.

A. Simple Lectins

Simple lectins consist of a small number of subunits, not necessarily identical, of molecular weight usually below 40 kDa, which may contain an additional domain besides their carbohydrate binding site(s). This class comprises practically all known plant lectins as well as the galectins (formerly S-lectins), a family of galactose-specific animal lectins.

1. Legume

The largest and most thoroughly studied family of the simple lectins is that of legumes, of which close to 100 members have been characterized, almost all isolated from seeds of the plants.[8,34−36] Concanavalin A from Jack bean, the prototype member of this family, was first isolated in 1919 by James Sumner (of urease fame) and shown by him, in 1936, to be specific for mannose and glucose. Other well-studied legume lectins are phytohemagglutinin (PHA) from the red kidney bean, SBA, PNA, and ECorL. In some cases, different lectins have been isolated from seeds of the same plant, e.g., those of *Griffonia simplicifolia* afforded lectins specific for galactose/*N*-acetylgalactosamine, *N*-acetylglucosamine, or a complex oligosaccharide (Tables 1 and 2). Also, single lectins from legumes (or other plants) often occur as a mixture of closely related proteins known as isolectins. Typically, legume lectins consist of two or four identical, or almost identical, subunits (or protomers) of 25−30 kDa, each with a single, small carbohydrate combining site with the same specificity. They also contain a tightly bound Ca^{2+} and a transition metal ion, predominantly Mn^{2+}, per subunit which are required for carbohydrate binding.[37] In addition to their carbohydrate combining site, several of the legume lectins possess a hydrophobic site that binds nonpolar compounds such as adenine and indoleacetic acid.

The subunits of the legume lectins are commonly made up of single polypeptide chains of about 250 amino acids that may carry one or two *N*-linked oligosaccharides. In some lectins (e.g., those from pea and lentil) the polypeptides are fragmented into a light (α) and heavy (β) chain. Legume lectins exhibit remarkable sequence homologies, with about 20% of invariant amino acids, and close to 20% of similar ones. The conserved amino acids include several of those that participate in hydrogen-bonding or hydrophobic interactions with the monosaccharide held in the combining site, and almost all the residues that coordinate the metal ions. Concanavalin A occupies a special position, since it exhibits an unusual homology, referred to as "circular homology", with the other legume lectins. This homology is obtained by aligning residue 119 with the amino terminal residue of the other lectins, proceeding to the carboxyl end of concanavalin A and continuing along its amino terminal region. It is the result of an unusual rearrangement of the peptide chain that occurs in the last step of the synthesis of the lectin[38] (Figure 5). Quite surprisingly, two mannose-specific animal lectins (MR60/ERGIC-53 and VIP36) are homologous with those of the legumes and also contain the two key monosaccharide-binding residues (asparagine and aspartic acid, cf. section IV.A.1).[39,40]

The three-dimensional structures of 10 legume lectins, mostly in complex with carbohydrate ligands, have been elucidated by high-resolution X-ray crystallography (Table 3), and in one case constructed by molecular modeling.[41] The subunits are in the shape of a dome, made up largely of two antiparallel β-sheets, one of six strands and the other of seven (Figure 6). The structures are nearly superimposable, irrespective of the specificity of the lectins. The strands of the sheets form jellyrolls, also referred to as the lectin fold.[42] The majority of residues not included in the β structures are in loops and β bends that connect the strands of the β sheets. The six-stranded sheet is almost flat, while the other is

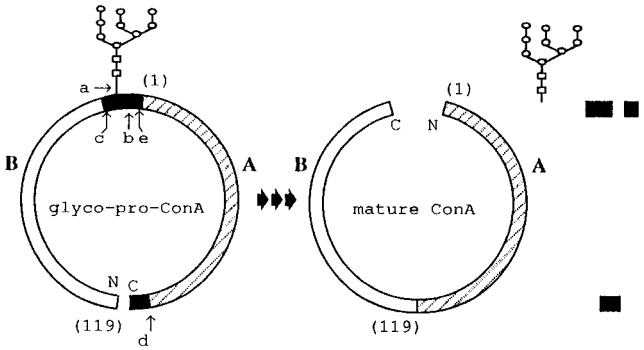


Figure 5. Posttranslational modifications during concanavalin A synthesis. Summary of processing events converting glycosylated pro-concanavalin A to the mature lectin. Amino and carboxy termini are indicated by N and C, and the numbers in brackets are residue positions in mature concanvalin A. During processing in the plant inactive glycosylated pro-lectin is deglycosylated (arrow a), resulting in appearance of lectin activity. An endopeptidase then cleaves (arrows b, c, d, and e) a carboxy terminal nonapetide and the glycosylated spacer (shown as solid infills) and residues 118 (arrow d) and 119 are ligated enzymatically. Splicing thus results in a transposition of the linear arrangement of the protein sections designated B and A. (Reprinted with permission from Jones, D. H. In *Perspectives on Protein Engineering & Complementary Techniques*; Geisow, M. J., Epton, R., Eds.; Mayflower Worldwide Limited: United Kingdom, p 70. Copyright 1995 Mayflower Worldwide Limited.)

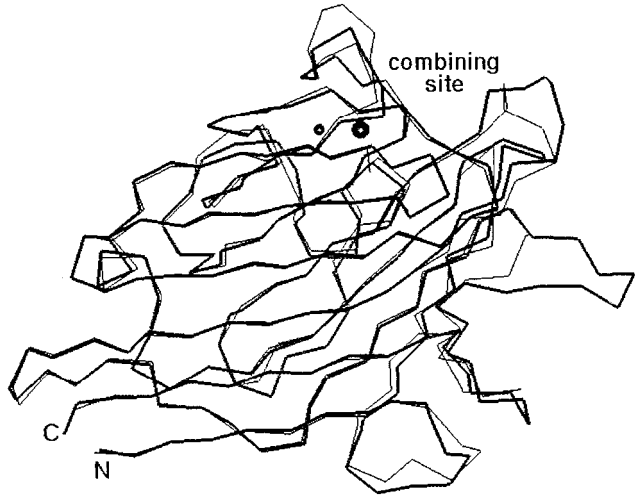


Figure 6. Computer model of the α-carbon chain of the subunit of *Lathyrus ochrus* lectin 1(LOLI, thin line), superimposed on that of *Erythrina corallodendron* lectin (thick line).

concave. The combining sites of the carbohydrate and of the metal ions are located mostly in the β folds of the seven-chain, curved sheet. The concavity of the face provides a shallow carbohydrate binding site, located at the top of each protomer, that is easily accessible not just to monosacharides, but to oligo- and polysaccharides as well. The Ca^{2+} and Mn^{2+} are situated 4.25 Å apart and are in close proximity (9−13 Å) to the carbohydrate binding site; they help to position the amino acids that form contacts with the carbohydrate, but do not bind it directly. Each of the two metals ions is linked to four amino acid side chains, two of which belong to aspartic acid residues that are shared by both metals (Figure 7). Four water molecules that are conserved in all legume

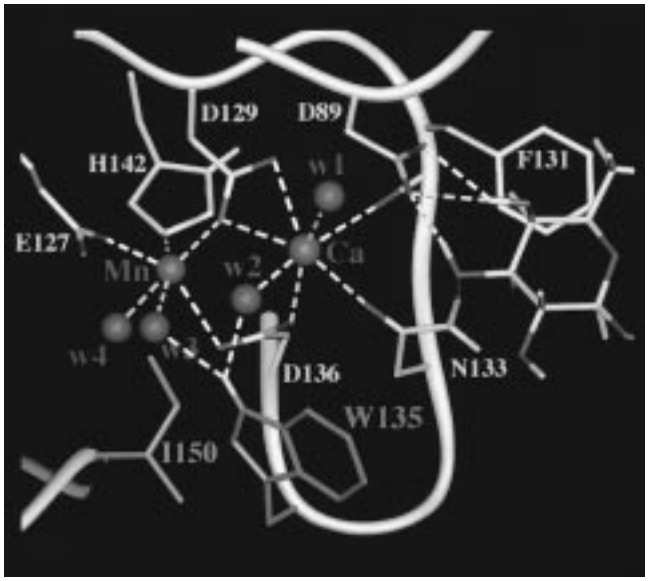


Figure 7. The metal binding site in ECorL. w1−4 denote conserved water molecules. (Courtesy of Drs. J. Ångström and E. Moreno, Department of Medical Biochemistry, University of Göteborg, Sweden.)

lectins also participate, directly or indirectly, in metal binding.[43] Large changes have been observed in the crystallographic structure of concanavalin A upon demetalization of the lectin, which results in loss of its carbohydrate binding ability.[44] They are apparently initiated mainly by the removal of the calcium ion, which causes the destruction of both the Ca^{2+} and the carbohydrate combining sites.

The most common mode of subunit dimerization, as exemplified by concanavalin A, involves the antiparallel side-by-side alignment of the two flat β sheets, leading to the formation of a contigous 12-stranded β sheet that extends across the dimer interface[13,34] (Figure 8A). A clearly different mode of dimerization was observed in ECorL, in which the covalently linked carbohydrate units interfere with the canonical dimeric interface[45] (Figure 8B). In the crystal, the *N*-linked carbohydrate is tethered by an intricate network of intra- and intermolecular hydrogen bonds and as a result its structure could be resolved—the first time a glycoprotein oligosaccharide has been clearly seen in the electron density map (Figure 9). The covalently bound carbohydrate imposes a noncanonical mode of dimerization also on the lectin from *Griffonia simplicifolia* (GS IV),[46] although it differs from that seen in ECorL. Surprisingly, PNA which is not glycosylated, forms dimers very similar to those in GSIV, with a back to back association of the two subunits through the flat β sheet.[47]

The tetrameric legume lectins can be considered as "dimers of dimers". Three different modes of dimer−dimer association to form tetramers have been discerned, two of which are illustrated in Figure 10. In concanavalin A, the formation of the tetramers involves the central parts of both dimers and contacts are mainly through loop interactions.[13] In SBA[48] and PHA,[49] the two curved 12-stranded β sheets interact through contacts between their two outermost strands, creating a chanel between them. PNA has an un-

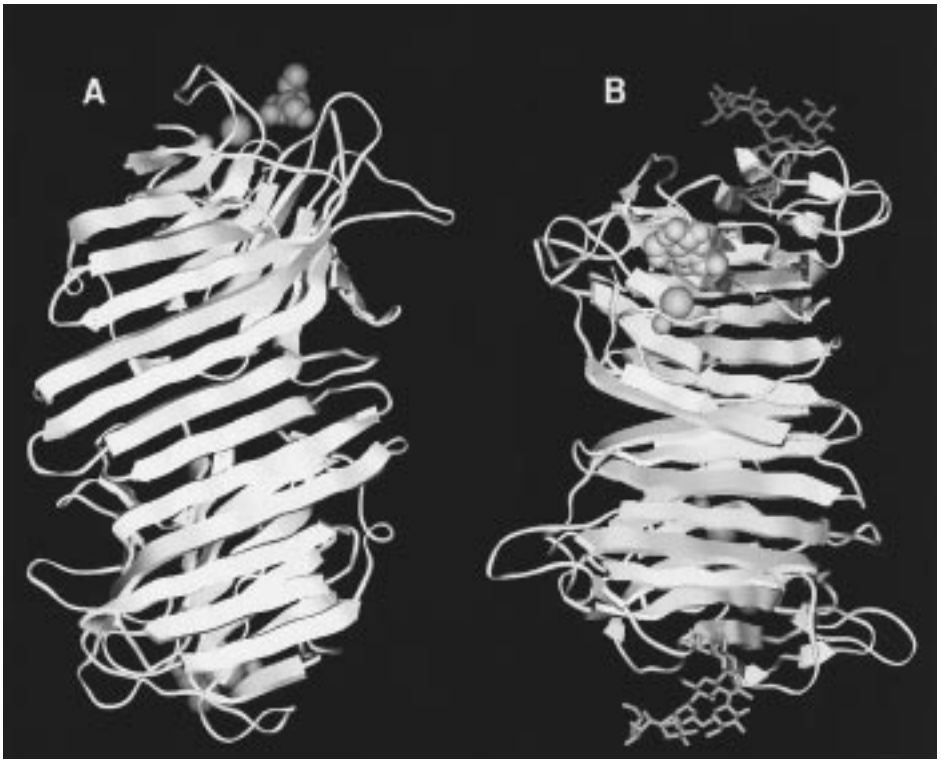


Figure 8. Two types of dimer found in legume lectins: (A) the canonical dimer, represented by concanavalin A (PDB entry 5CNA) and (B) *Erythrina corallodendron* lectin (PDB entry LTE). The CPK models represent the bound sugar—mannose in A and lactose in B. The two spheres close to the bound sugar depict the bound metal ions (Ca^{2+} and Mn^{2+}) present in all legume lectins. A stick model of the N-linked heptasaccharide of *Erythrina corallodendron* lectin is also shown. (This figure, as well as all computer drawn figures, were prepared by program MSI/BIOSYM INC., San Diego, CA.)

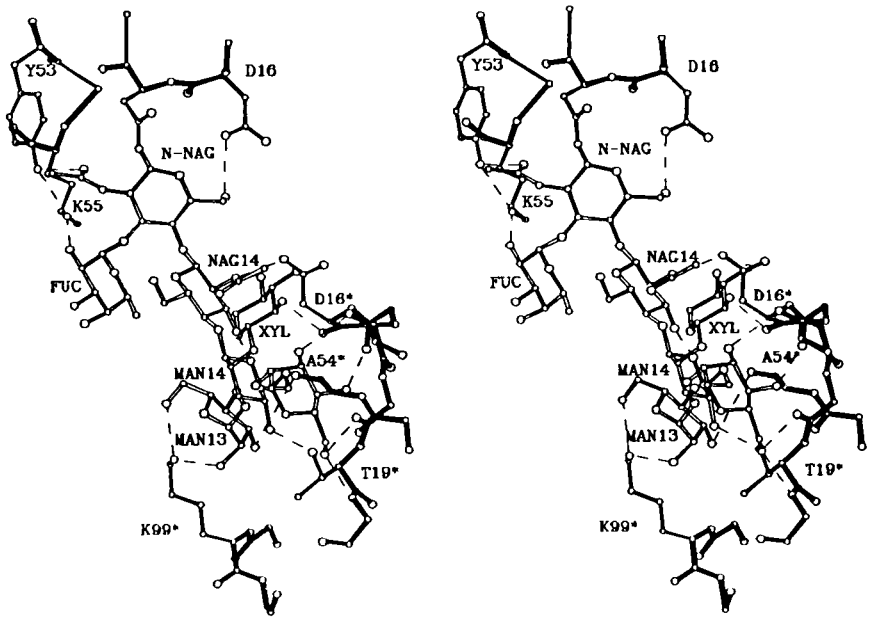


Figure 9. Network of hydrogen bonds (dashed lines), stabilizing the *N*-linked carbohydrate of ECorL (stereo). Water molecules omitted for clarity. (Reprinted with permission from ref 45. Copyright 1991 American Association for the Advancement of Science.)

usual, open quaternary structure, the most interesting aspect of which is the absence of symmetry in the tetramer.

2. Cereal

Another family of simple lectins is that of the cereals, which includes wheat germ agglutinin (WGA) and barley and rice lectins. Members of this family, too, consist mostly of two identical subunits, although they differ markedly from the legume lectins. For instance, they are exceptionally rich in cysteine, which in legume lectins is almost always absent. WGA, the only member of the cereal lectin family characterized in molecular detail, is a mixture of

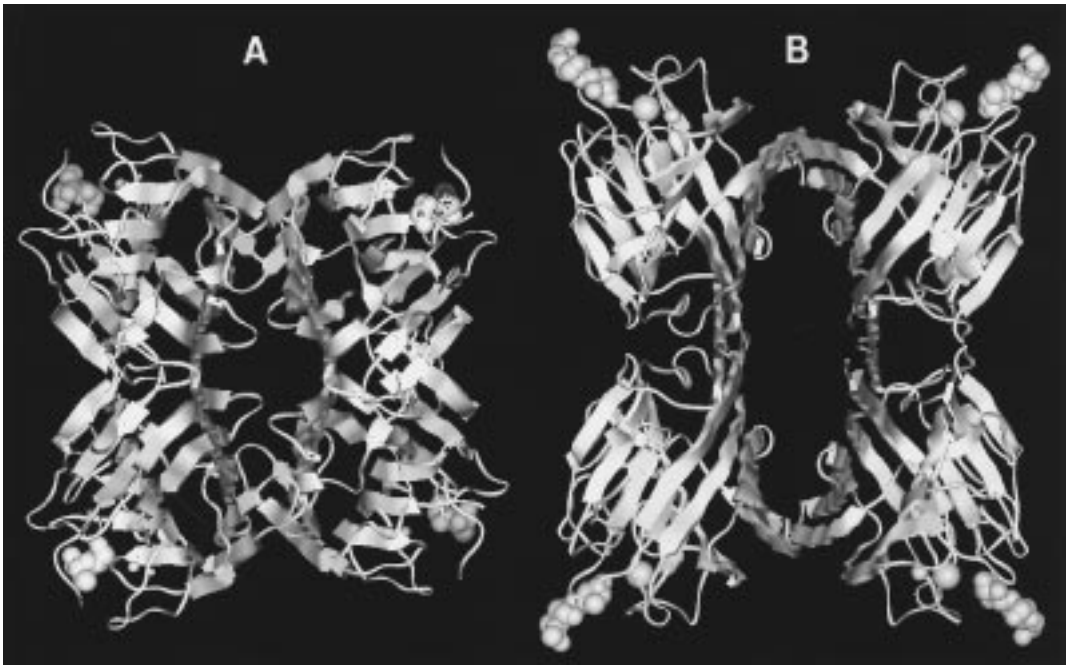


Figure 10. Ribbon representation of the teramers of (A) concanavalin A (PDB entry SCNA) and (B) soybean agglutinin (PDB entry 1SBA).

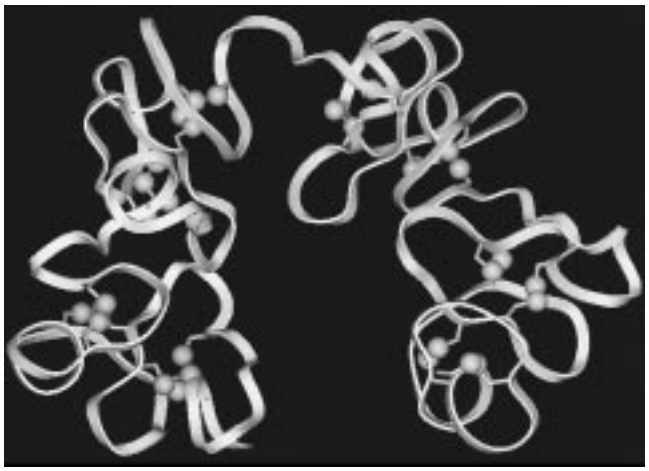


Figure 11. Ribbon representation of wheat germ agglutinin (PDB entry WGC). The sulfur atoms of the disulfide bridges are shown as small balls.

three isolectins that differ slightly in their amino acid composition, being the product of distinct, yet closely related genes. The isolectins are dimers of two identical 17 kDa subunits and are devoid of metals.[8,20] Each subunit is made up of four homologous subdomains (A to D) of 43 amino acids; the domains are similarly folded, with four identically positioned disulfide bridges[50] (Figure 11). There are thus 16 such bridges per WGA subunit, resulting in a highly stable molecule. In the dimer, the subunits associate in a head to tail fashion, resulting in subdomain pairs (A−D and B−C), each partner originating from a different subunit (Figure 12). The protein is devoid of the commonly occurring secondary structural elements, the β sheet and α helix. Other unusual features are the presence of multiple binding sites due to the internal 4-fold structure repeat and their location at the interface between the subunits that form the molecular dimer of the lectin. There are eight such sites per dimer, four of which are unique.[51]

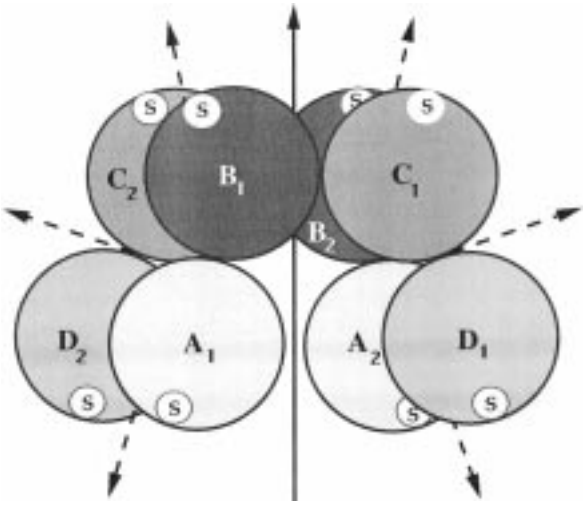


Figure 12. Schematic representation of the wheat germ agglutinin dimers. Domains are shown as large shadowed circles and labeled A1, B1, C1, D1, etc. The position of the molecular 2-fold axis is indicated by an arrow. Broken arrows represent the two types of pseudo-2-fold axes generated in the dimer interface between domains of different dimers. "S" refers to the aromatic carbohydrate-binding pocket. (Reprinted by permission from ref 51. Copyright 1996 Cambridge University Press.)

Because 2-fold related sites (e.g., B1/C2 and C2/B1) are equivalent, there are two of each type of site present in the dimer (see section IV.A.2).

3. Amaryllidaceae and Related Families

The bulbs of plants of the amaryllis, orchid, and garlic families contain lectins that bind mannose (but not glucose), whose sequences are highly conserved, exhibiting 80−90% homology.[52,53] They are distinguished by their small monomer size (12 kDa), presence of 3-fold internal repeats of 36 amino acids, lack of metal requirement, and weak affinity for the monosaccharide ligand ($K_a < 10^2$ M^{-1}).[54]

Table 5. Structures of Oligosaccharides Mentioned in This Review[a]

1	Galβ4Glc	lactose
2	Galβ4GlcNAc	*N*-acetyllactosamine
3	Neu5Acα2,3Galβ4GlcNAc Fucα3	sLe^x (sialyl-Le^x)
4	Neu5Acα2,3Galβ3GlcNAc Fucα4	sLe^a (sialyl-Le^a)
5	Manα6Manβ4GlcNAc	linear trisaccharide
6	Manα6 Man Manα3	branched trisaccharide
7	Manα6 Manβ4GlcNAcβ4GlcNAc Manα3	pentasaccharide core
8	Galβ4GlcNAcβ2 GalOR^b Galβ4GlcNAcβ3	2,3 biantennary pentasaccharide (glycoside)
9	Galβ4GlcNAcβ2 GalOR^b Galβ4GlcNAcβ4	2,4 biantennary pentasaccharide (glycoside)
10	Galβ4GlcNAcβ2 GalOR^b Galβ4GlcNAcβ6	2,6 biantennary pentasaccharide (glycoside)
11	Galβ4GlcNAcβ3 GalOR^b Galβ4GlcNAcβ6	3,6 biantennary pentasaccharide (glycoside)
12	Galβ4GlcNAcβ2Manα6 Manβ4GlcNAc Galβ4GlcNAcβ2Manα3	octasaccharide
13	Manα6 Manβ4GlcNAcβ4GlcNAc-Asn Manα3 Fucα6	fucose-containing glycopeptide
14	Galβ4GlcNAcβ2Manα6 Manβ4GlcNAcβ4GlcNAc Galβ4GlcNAcβ2Manα3 Fucα6	branched decasaccharide
15	Galβ4GlcNAcβ2Manα6 GlcNAcβ4Manβ4GlcNAcβ4GlcNAc Galβ4GlcNAcβ2Manα3 Fucα6	bisected decasaccharide
16	Manα6 Manα6 Manα3 Man Manα3	branched pentamannose
17	Manα6 Manα6 Manα3 Manβ4GlcNAcβ4GlcNAcβ1-Asn Manα2Manα3	Asn-oligomannose
18	NeuAcα2,3Galβ3 GalNAc-Thr NeuAcα2,6	sialoglycopeptide
19	NeuAcα2,3Galβ3 GlcNAcβ3Galβ4Glc NeuAcα2,6	sialohexasaccharide

[a] Almost all the above oligosaccharides are present in N-linked (and/or O-linked) glycoproteins. [b] $R = -(CH_2)_5COOCH_3$.

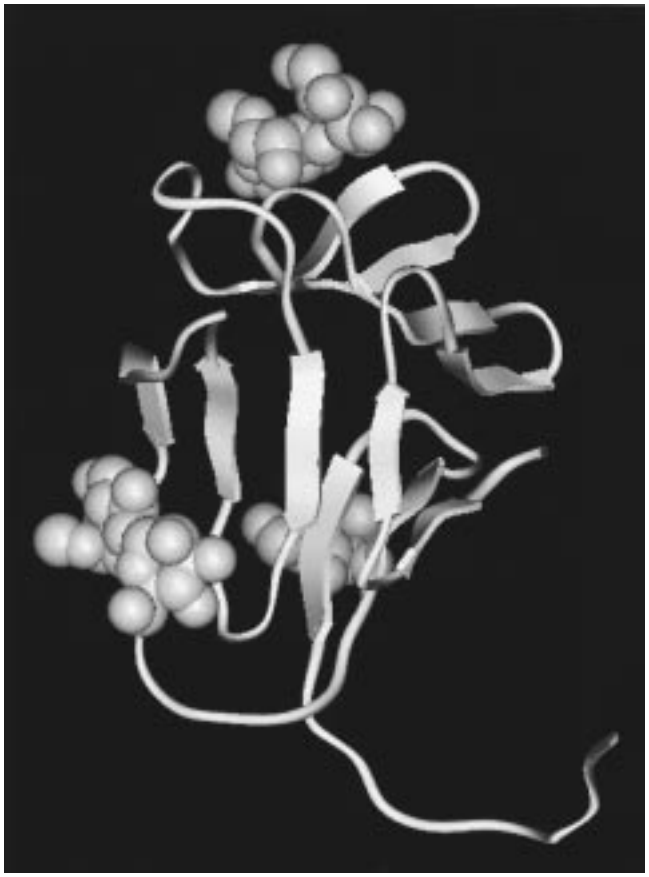


Figure 13. Ribbon representation of the monomer of *Galanthus nivalis* lectin, showing also CPK models of three molecules of the bound mannose. The monomer contains three subdomains with antiparallel four-stranded β-sheet structures. There is one inter-subdomain disulfide bond, between the second and third subdomains, and the interior of the monomer is stabilized by conserved hydrophobic residues. The carboxy terminal subdomain 1 includes one strand from the adjacent subunit. PDB entry 1NNIV.

The three-dimensional structure of only one of these lectins—that of the snowdrop (*Galanthus nivalis*) (GNA)—in complex with methyl α-mannoside[55] and with an octasaccharide[56] (Table 5) has been elucidated. It is a flat tetrameric molecule with a central opening 16 Å wide. Each monomer contains three subdomains (1, 2, and 3) with antiparallel, four-stranded, β-sheet structures (Figure 13). There is one inter-subdomain disulfide bond, between the second and third subdomains, and the interior of the monomer is stabilized by conserved hydrophobic residues. The carboxy terminal subdomain 1 includes one strand from the adjacent subunit, and the four subunits hence form two pairs of dimers (A−D and B−C) within the tetramer. The pairs are also stabilized by a rare symmetric interaction between the Arg101 residues in each monomer. The lectin is further unusual in that it apparently has one carbohydrate combining site per subdomain, i.e., the tetramer is dodecavalent.

4. Moraceae

Jacalin, the galactose-specific lectin from the seeds of jackfruit (*Artocarpus integrifolia*, a plant of the *Moraceae* family), is a tetrameric glycoprotein with

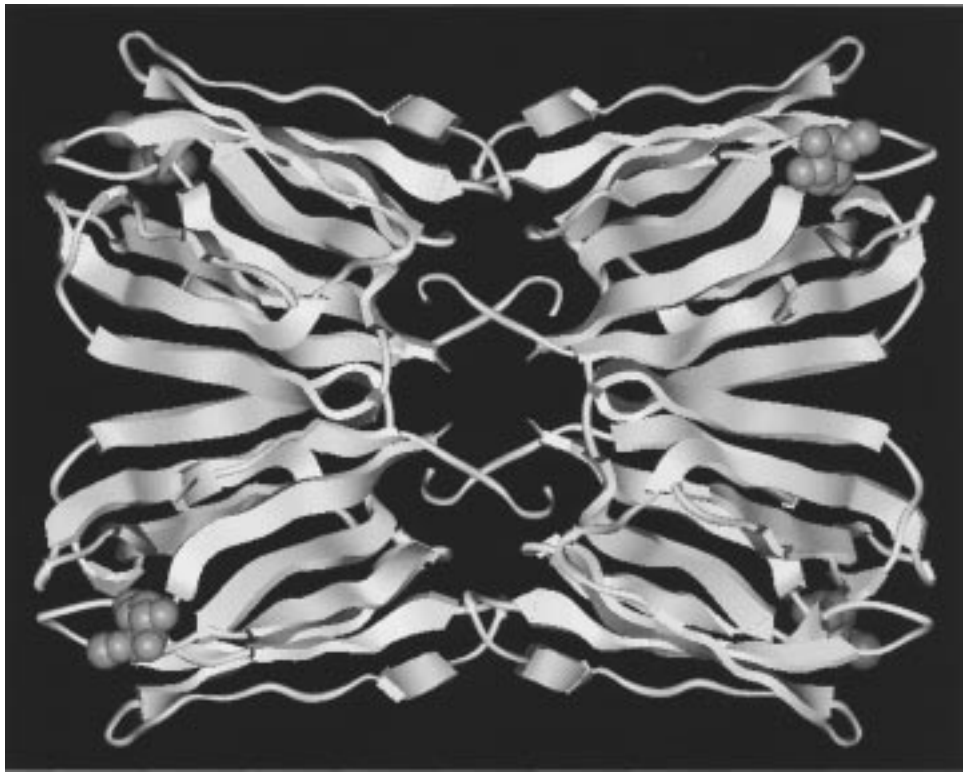


Figure 14. Ribbon representation of the tetramer of jacalin, showing also the CPK model of the bound galactose. PDB entry 1JAC.

a molecular weight of about 66 kDa. Each of its subunits consists of a heavy chain (α) of 133 amino acids and a light chain (β) of 20 residues.[57] The primary structure of jacalin shows no significant similarity with any other lectin, except that from *Maclura pomifera*, also a member of the *Moraceae* family. Crystallographic studies have shown that the subunits of jacalin are made up of three four-stranded antiparallel β sheets, arranged like the faces of a triangular prism, with loops connecting strands in the sheets[58] (Figure 14). It is stabilized by hydrophobic interactions in the core of the subunit and a small number of hydrogen bonds involving main-chain, as well as side-chain, atoms. This recently discovered arrangement, classified as the β-prism fold, has been found in a few proteins, but not in any other lectin.

5. Euphorbiaceae

Beans of the castor tree (*Ricinus communis*) contain two closely related lectins, ricin and *Ricinus communis* agglutinin, RCA;[21] the former is one of the deadliest poisons known: it is by weight about 10 times as toxic as cobra venom and, according to some estimates, a single molecule is sufficient to kill a cell. Although classified by us as simple, ricin and RCA represent borderline cases: their structure is more complex than that of the lectins discussed hitherto, but they do not fulfill the criteria that define the other two classes of lectin. Ricin is a heterodimeric protein with a MW of 60 kDa, made up of two S−S linked chains, A and B. The latter contains two carbohydrate binding sites specific for galactose, whereas the cytotoxic activity resides in the A chain which acts by enzymatically inactivating the RNA involved in protein synthesis. The B chain is made up of two globular domains, each of which comprises a link domain and three homologous 40-residue subdomains. Like the WGA subunit, the B chain is stabilized by several disulfide-linked cysteines. RCA is a dimer of two subunits, each of which is similar to ricin but it is not toxic. The three-dimensional structure of ricin has been determined by X-ray crystallography at 2.5−2.6 Å resolution[59,60] (Figure 15). The A chain is a globular protein with extensive secondary structures, both β pleated sheet and α helix, and a reasonably prominent cleft, assumed to be the active site responsible for the toxic action of ricin. The B chain folds into two topologically similar domains, each binding lactose in a shallow cleft. Preliminary crystallographic characterization of RCA has shown that it forms an elongated molecule of 120 Å × 60 Å × 40 Å with two A chains at the center and a B chain at each end. The A chains are covalently associated, with a disulfide bridge between Cys157 of each of the chains. Additional contacts at residues 114−115 stabilize the dimer interface.[61]

6. Galectins

The galectins constitute a family of soluble, β-galactoside specific lectins that combine preferentially with lactose (entry 1 in Table 5) and *N*-acetyllactosamine[62−64] (entry 2 in Table 5). They occur predominantly in mammals, but have been found also in other vertebrates (e.g., frog,[65]) and in some invertebrates (for instance sponges[66]), but not in plants. Indeed, the first member of this family has been isolated some 20 years ago from the electric organ of electric eel.[67] Their structure is relatively simple and they share a highly homologous domain (known as

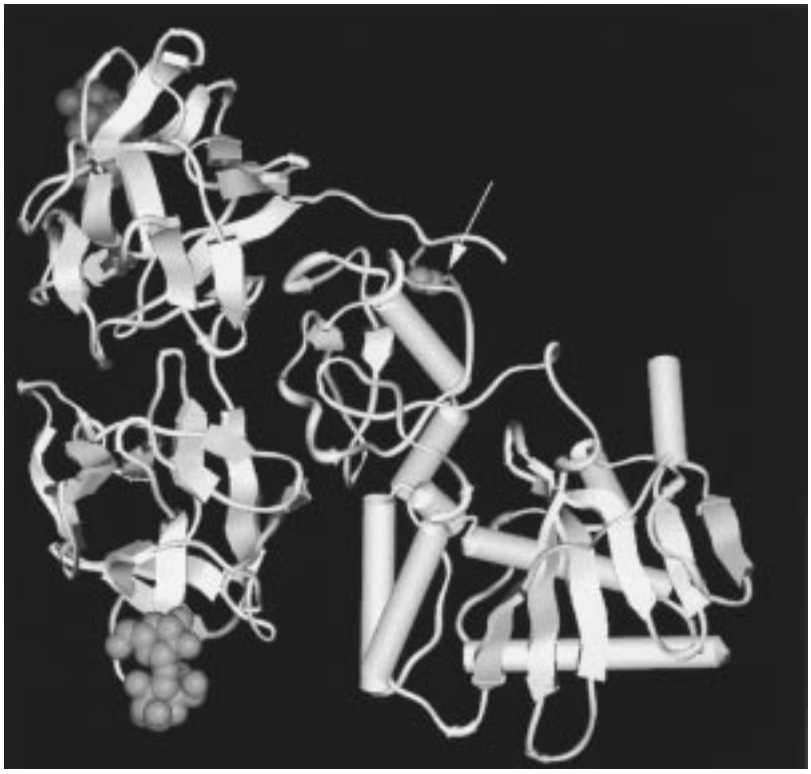


Figure 15. Ribbon representation of ricin. The cylindrical strutures represent α-helices. The arrow points to the S−S bond between the A and B subunits. CPK models of two galactose molecules in the combining sites of the B subunit are also shown. PDB entry 1AAI.

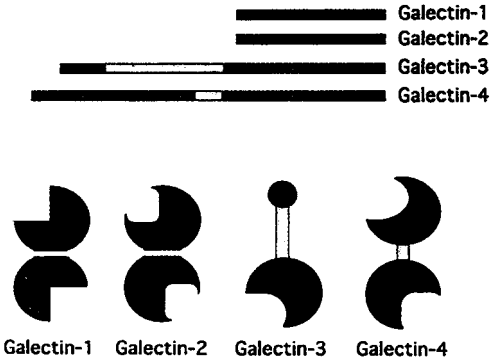


Figure 16. Schematic representation of the different types of galectin. The proteins are shown as linear diagrams corresponding to single peptide chains (top) and as assembled proteins. (Reprinted with permission from ref 62. Copyright 1994 American Society for Biochemistry and Molecular Biology.)

the S-carbohydrate recognition domain, or S-CRD). At least four arrangements of the S-CRD have been identified, including monomers and homodimers of subunits (MW about 14 kDa), as well as larger polypeptides (MW 30−35 kDa) containing one or two copies of the S-CRD in association with an accessory region, or linker, characterized by proline- and glycine-rich repeat sequences (Figure 16). Preparations of 14 kDa lectins that have been truncated from either end by molecular genetic techniques were inactive; the functional CRD corresponds therefore to nearly the entire lectin.[68] Remarkably, the three-dimensional structure of the galectins also exhibits the jellyroll topology (or lectin fold) found in the legume lectins, despite the absence of significant sequence similarity and a different location of the

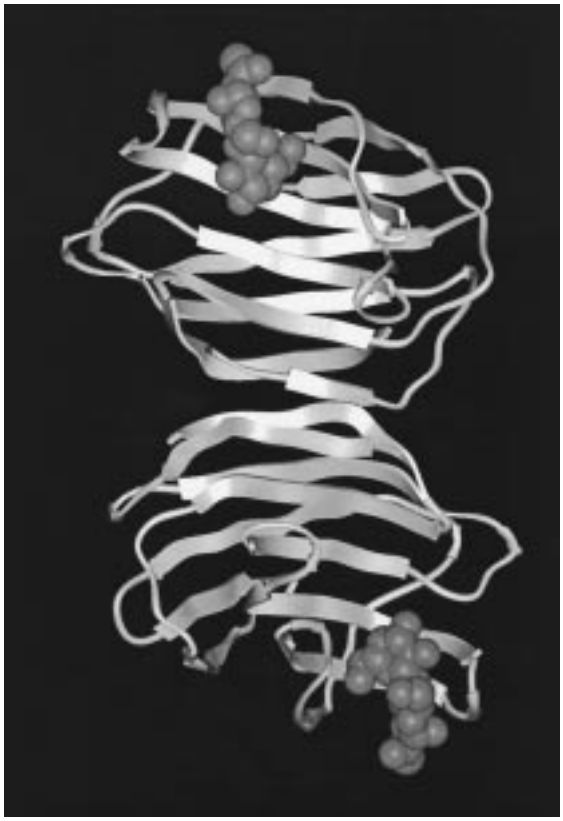


Figure 17. Ribbon represenation of the dimeric bovine spleen galectin-1, showing also the CPK models of the bound *N*-acetyllactosamine. PDB entry 1SLT.

combining site[69,70] (Figure 17). It has therefore been suggested that both lectin families may be considered as members of the same superfamily.[42] However the

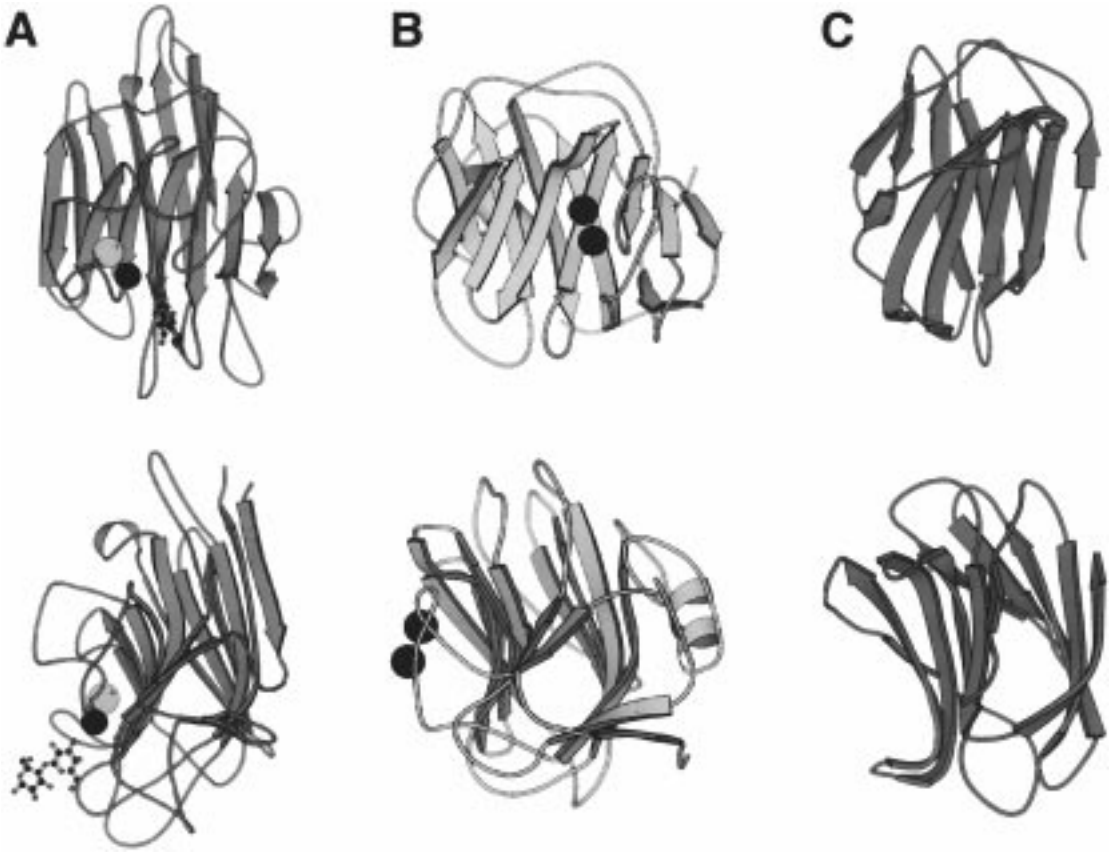


Figure 18. Similarity in the overall folding of lectins from diverse sources: (A) *Erythrina corallodendron* lectin; (B) human serum amyloid P component; (C) bovine spleen galectin 1, viewed from different directions. The black spheres represent Ca^{2+}; the gray one, Mn^{2+}. In A, the stick model of bound lactose is also shown. Modified from Crennell et al. *Structure* **1994**, *2*, 535.

question of the evolutionary origin, whether by divergence from a common ancestor or by convergence to a similar, compact structural framework with similar functional characteristics, can at present not be answered unambiguosly.[16]

7. Pentraxins

A topology very similar to the lectin fold has also been observed in SAP (see section II.A.1)[20,71] (Figure 18). SAP is a member of the pentraxins—a family of oligomeric plasma proteins with the capacity of calcium-dependent ligand binding, named for the pentameric arrangement of their subunits (Figure 19). In this case too there is very little sequence similarity with the legume lectins and the nature of carbohydrate binding is different as well.

B. Mosaic (Multidomain)

Included in this group are diverse proteins from different sources—viral hemagglutinins on one hand and animal lectins of the C-, P-, and I-type on the other. They are all composite molecules with a wide range of molecular weights, consisting of several kinds of protein modules or domains, only one of which possesses a carbohydrate binding site. Table 4 lists the mosaic lectins for which the three-dimensional structure, in most cases in complex with carbohydrate, has been elucidated. Many of these lectins are monovalent, but since they are embedded in membranes, they act in a multivalent fashion.

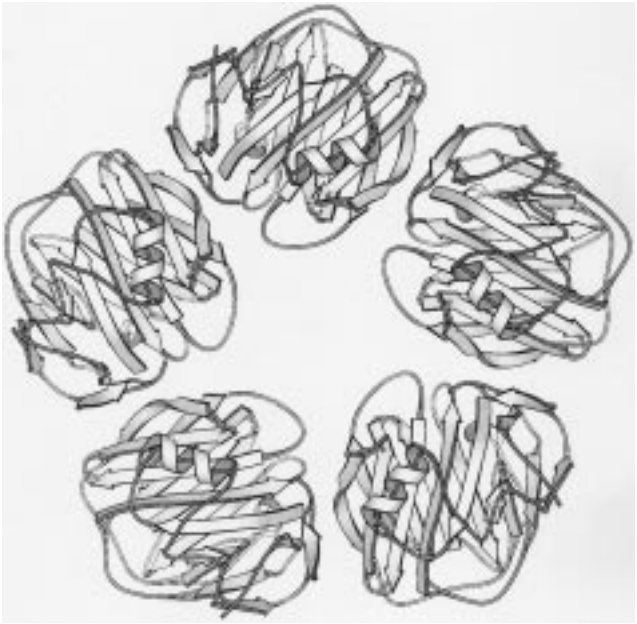


Figure 19. Structure of the pentamer of serum amyloid P component (SAP) viewed along the noncrystallographic 5-fold axis of symmetry. (Figure prepared using MOLSCRIPT according to Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946−950. Reprinted by permission from ref 20. Copyright 1994 Macmillan Magazines Limited.)

1. Viral Hemagglutinins

a. Influenza Virus Hemagglutinin. The influenza virus hemaggutinin, first described in the 1950s, is the most thoroughly investigated of the multidomain lectins.[72,73] Its subunit is composed of two polypeptides, HA1 and HA2, with molecular masses of 36 and 26 kDa, respectively, covalently linked by a single disulfide bond. Each subunit consists of a hydrophilic, C-terminal domain located on the inner side of the membrane, a hydrophobic membrane spanning region of 24−28 residues, an elongated α-helical stem and a globular domain projecting 135 Å from the membrane. The globular domain is made up of HA1 only, and contains the carbohydrate binding site of the lectin (Figure 20). The subunits associate noncovalently to form trimers.

b. Murine Polyoma Virus. This virus is a nonenveloped, icosahedrically symmetrical particle, with circular, double-stranded DNA genomes. The outer shell (capsid) of the virion contains 360 copies of the viral protein VP1 (MW ∼42 kDa) arranged in pentamers.[74] Each subunit of VP1 has two antiparallel β sheets with a topology which resembles the lectin fold; some loops that connect the β strands are extensive and contain additional secondary structure elements. The most striking feature of the capsid structure is the way the individual pentamers are tied together by the carboxy terminal arms of the monomers. The last 63 residues emerge from each monomer and "invade" a subunit of another pentamer, where they form a β strand that augments a sheet in the target subunit.[75]

2. C-Type Lectins

This class of lectins has been so named because they require Ca^{2+} for activity.[3,76,77] It includes over 50 members, all characterized by an extracellular carbohydrate recognition domain (C-CRD) consisting of 115−130 amino acids, of which 14 are invariant and 18 highly conserved. To the CRD is attached a variable number of domains of different kinds, which form the bulk of the molecule. Lectins included in this class have been grouped into three families—endocytic lectins, collectins, and selectins, each sharing a common overall architecture defined by the nature of their domains (Figure 21). An exception is the mannose-specific macrophage surface lectin which has an unique structure but is included among the endocytic lectins because it shares with them a common function.

a. Endocytic Lectins. The prototype of this family is the galactose/*N*-acetylgalactosamine specific lectin from rabbit hepatocytes (RHL), also known as hepatic asialoglycoprotein receptor (or hepatic binding protein, HBP), the first mammalian lectin to be described in the early 1970s.[78−80] Similar lectins have subsequently been found on hepatocytes of other mammals. Examples of other endocytic lectins are the avian hepatic lectin specific for *N*-acetylglucosamine, also present on hepatocytes,[78] a galactose-specific lectin on peritoneal macrophages, and a fucose-specific receptor (lectin) found on the Kupffer cells of the liver.[81] The endocytic lectins are type II transmembrane proteins, consisting of a short amino terminal cytoplasmic domain, a hydrophobic, membrane-spanning domain, and a neck region to which a carboxy terminal CRD is linked. The mammalian hepatic asialoglycoprotein receptors are usually composed of two types of subunit: a smaller but more abundant (MW 40−46 kDa) and a larger, less abundant (up to 60 kDa), that occur in varying proportions in the different lectins. In the rat, the larger subunit exists in two forms with an identical amino acid sequence, one glycosylated and the other devoid of carbohydrates (Figure 22). The subunits possess

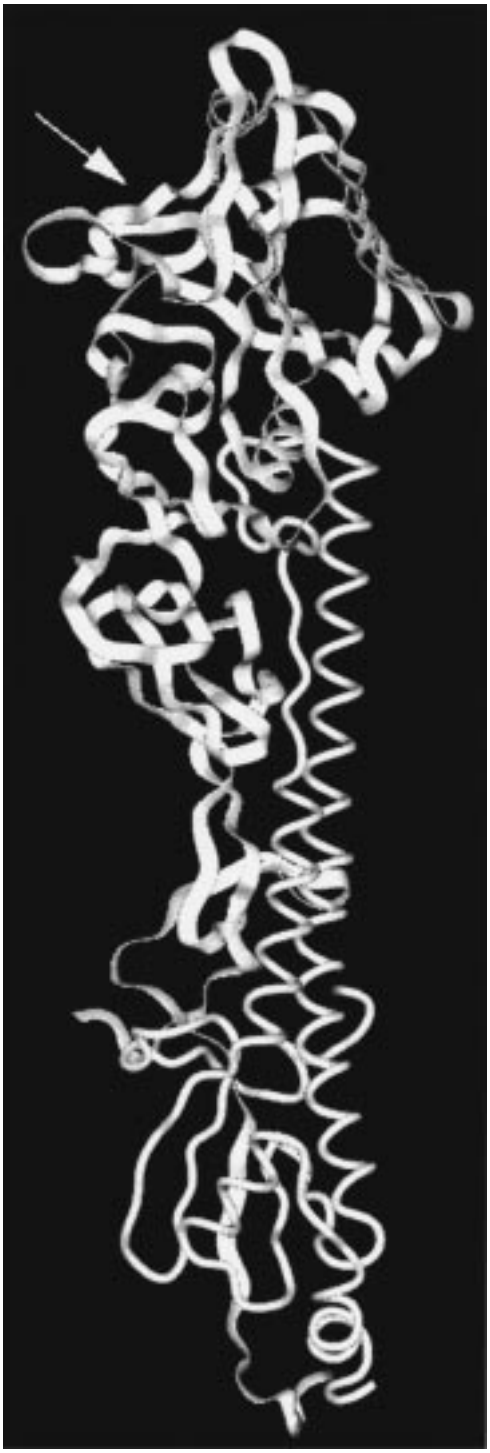


Figure 20. Ribbon presentation of the influenza virus monomer. The broad ribbon represents subunit HA1; the narrow one, subunit HA2. The arrow indicates the location of the carbohydrate binding site. PDB entry HGF.

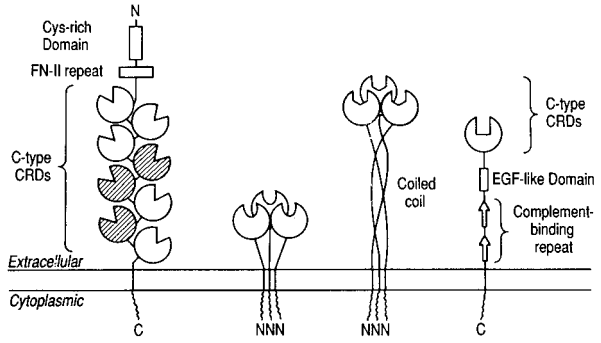


Figure 21. Organization of membrane-bound C-type animal lectins: (from left to right) the mannose macrophage receptor, a type I membrane protein; two examples of type II endocytic receptors (the chicken hepatic lectin and the Kupffer cell receptor); and L-selectin. (Reprinted with permission from ref 3. Copyright 1993 Annual Reviews Inc. URL: http://www.AnnualReviews.org.)

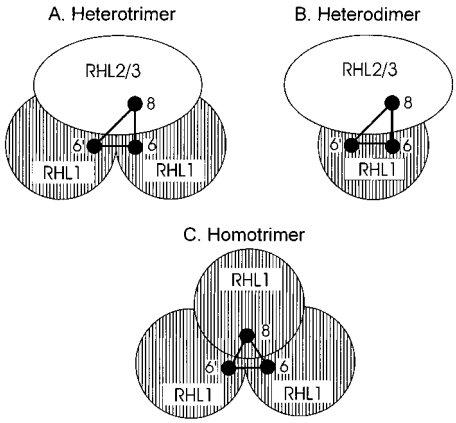


Figure 22. Models of the subunit organization of the rat asialoglycoprotein receptor: (A) heterodimer incorporating two RHL1 subunits and one RHL2 or RHL3 subunit and (B) heterodimer of one RHL1 and one RHL2/3. Homotrimer of RHL1 obtained by solubilization. (Reprinted from Rice, K. G.; Lee, Y. C. *Adv. Enzymol.* **1993**, *66*, 41. Copyright 1993 John Wiley & Sons Inc.)

similar primary structures, except for their neck regions, where considerable differences occur. The two types of subunit are organized in a sterically specific and rigid orientation and both have to be present on the cell surface in order to form a functional receptor. These mammalian lectins have a strong tendency to associate and in purified form appear to exist as hexamers, although the exact stoichiometry of the subunits in the heteropolymer is unclear. The galactose binding sites in such a complex are tightly packed and arranged to best accommodate a triantennary oligosaccharide with nonreducing galactose residues that are 1.5−3.1 nm apart, which binds to the receptor with an affinity ~6 orders of magnitude higher than monovalent ligands.

The mannose-specific macrophage surface lectin (MW 175 kDa) differs from the other endocytic lectins in that it is a type I transmembrane protein (i.e., its carboxy terminal is in the cytoplasm and the amino terminal is outside the cell) (Figure 21). Moreover, the extracellular part of the molecule consists of three domains: a unique cysteine-rich segment, a region similar to the type II repeats of fibronectin and a domain, closest to the membrane, containing eight CRDs.[3,81−83] For a while this was the only known case of a C-type protein with more than one CRD within a single polypeptide chain. More recently, other proteins (e.g., the phospholipase A_2 receptor from muscle) have been shown to possess the same architecture.[84]

b. Collectins. The collectins are soluble proteins, composed of an amino terminal cysteine-rich domain, followed by a number of collagen-like repeats, an α-helical neck region and a carboxy terminal CRD.[85,86] Of the six known proteins of this group (MBP A and C, pulmonary surfactant apoproteins A and D, collectin CL-43 from bovine serum and bovine conglutinin), the MBPs have been most thoroughly investigated. The structural unit of the MBP is a trimer of 32 kDa subunits, formed by a triple helix of the collagenous portion of the subunit and is stabilized by the association of the α helices of the neck into a parallel triple-stranded coiled coil[87,88] (Figure 23). Two homologous, yet distinct forms of MBPs have been described, the serum type (MBP-A) and the liver type (MBP-C). The latter is characterized by an insertion of nine amino acids in the amino terminal cysteine-rich region. MBP-A circulates in the sera of higher animals as a hexamer of trimeric units of apparent MW of ~650 kDa, a partial model of which is shown in Figure 24. MBP-C is smaller and probably consists of two associated trimers.[89] The high-resolution three-dimensional structure of the CRD of MBP-A, the first of a C-type lectin to be elucidated, revealed that over 50% of the CRD is formed by loops and extended structures; the remainder comprises two short α-helices and five *β*-strands.[90]

Despite the great similarity in the architecture of the binding sites of the two MBP's, and their comparable affinity for monosaccharides, MBP-C binds branched mannooligosaccharides more strongly than does MBP-A. Binding studies, using mono- and oligosaccharides and synthetic cluster glycosides, led to the conclusion that MBP-C has two binding sites per subunit, one only for mannose, the other for both mannose and *N*-acetylglucosamine; the former appears to be extended, probably the size of a trimannoside. In contrast, MBP-A has only one site of the latter type.[91,92]

c. Selectins. This group consists of three members—E-selectin (MW 115 kDa), P-selectin (140 kDa), and L-selectin (90−110 kDa)—all highly asymmetric membrane-bound proteins.[3,93−97] They are so named because they mediate selective contact between cells. Each contains, in addition to the CRD located at the amino terminal part of the molecule, an adjoining epidermal growth factor (EGF)-like domain, several short repeating units related to complement-binding protein, a membrane-spanning region, and a cytoplasmic, carboxy terminal domain (Figure 21). The crystal structure of the CRD together with the EGF-like domain of E-selectin shows a very similar fold to that of the MBP-A.[98] The selectins interact specifically with sLe^x and its positional isomer sLe^a (3 and 4, respectively in Table 5), with both fucose and *N*-acetylneuraminic acid required for binding; sialic acid can be replaced by

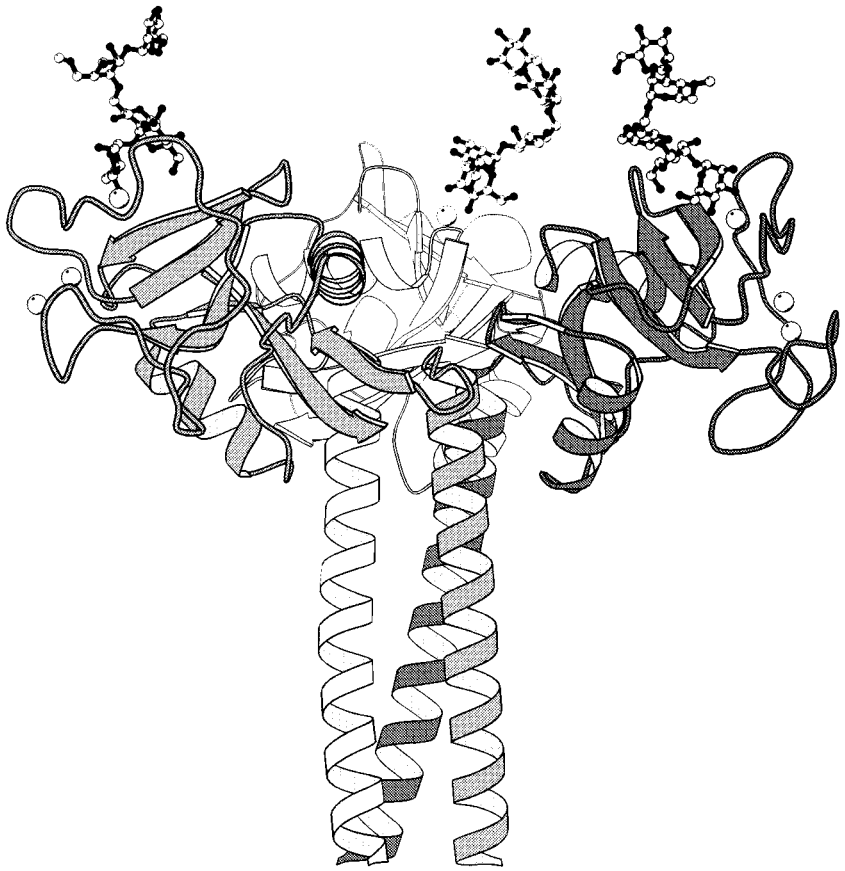


Figure 23. Three-dimensional structure of the CRD and the neck region of mannose binding protein A. Oligomannosides (top) are inserted at each of the three binding sites of the trimer, based on the known orientation of this sugar in the oligosaccharide-CRD struture. The Ca^{2+} ions are shown as spheres. (Reprinted with permission from ref 87. Copyright 1994 Current Biology Ltd.)

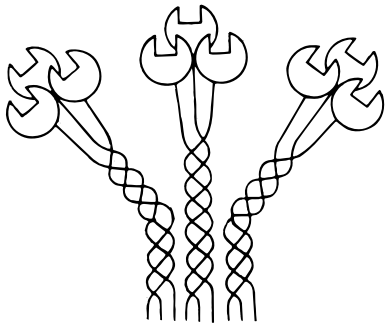


Figure 24. Oligomer form ("bouquet") of mannose binding protein. (Reprinted with permission from ref 3. Copyright 1993 Annual Review Inc. URL: http://www.AnnualReviews.org.)

another negatively charged group such as sulfate. Recognition of the carbohydrate ligands is possible only when they are present on particular glycoproteins, such as cell surface mucins, pointing to the role of the carrier molecule and carbohydrate presentation in the interaction with the lectins. In addition, selectins also bind oligonucleotides,[99] with affinities surpassing those of oligosaccharides by a factor of 10^4-10^5, providing another example of glycomimetics (cf. section II.A.1).

3. P-Type Lectins

The P-type CRD has been found only in two closely related lectins, the mannose 6-phosphate (Man-6-P) receptors.[100,101] One of these is of high molecular weight (~300 kDa), and does not require cations for activity; the other is of low molecular weight (~45 kDa) and is Ca^{2+}-dependent. Both are type I transmembrane glycoproteins. The extracellular domain of the high molecular weight receptor consists of 15 contiguous, homologous repeating units and contains two high-affinity binding sites, while that of the low molecular weight receptor is similar both in size and in sequence to the repeating units of the high molecular weight receptor and contains one binding site.

4. I-Type Lectins

I-type lectins are characterized by variable numbers of extracellular immunoglobulin-like domains and are thus members of the immunoglobulin superfamily.[102] By far the most important and best characterized lectins of this type are the sialoadhesins, a family of sialic acid specific type I membrane glycoproteins[102−105] (Figure 25). They include the macrophage receptor that mediates the adhesion of these cells to sheep erythrocytes (referred to simply as sialoadhesin), the lymphocyte surface antigen CD22 found only on B cells, CD33 present on early myeloid cells, and a myelin-associated glycoprotein, MAG. In all these lectins, the amino terminal, extracellular domain is similar to the variable region (V-type domain) of immunoglobulin G (IgG). The remaining domains, the number of which may vary

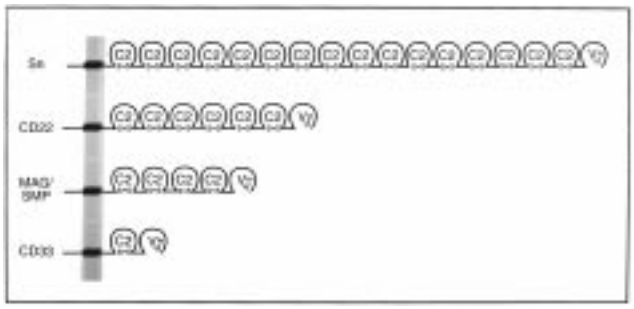


Figure 25. I-type lectins. (Reprinted from Kelm, S.; et al. *Curr. Biol.* **1994**, *4*, 965−972. Copyright 1994 Current Biology Ltd.)

from 1 to 16, are similar to the C2 segment of the constant region of IgG.

With the aid of a series of mutants of I-type lectins, from which various extracellular domains have been deleted, it was shown that in sialoadhesin the amino terminal, V-type domain is both necessary and sufficient for sialic acid dependent binding. In CD22, on the other hand, the adjacent C2-like domain is also required, apparently for correct folding of the protein.[106] A conserved arginine was implied by site-directed mutagenesis studies to play a key role in the affinity of the I-type lectins to glycoconjugates containing sialic acid.[107,108] CD22 recognizes specifically NeuAc(α2−6)Gal(β1−4)GlcNAc, known to occur in varying numbers on the *N*-linked oligosaccharides of many surface glycoproteins; the α2−6 linkage is an absolute requirement. In contrast, all other known I-type lectins bind structures containing *N*-acetylneuramnic acid that is α2−3 linked.

C. Macromolecular Assemblies

Lectins of this type are common in bacteria, usually in the form of fimbriae (or pili). These are filamentous, heteropolymeric organelles present on the surface of the bacteria, 3−7 nm in diameter and 100 to 200 nm in length, consisting of helically arranged subunits (pilins) of several different types, assembled in a well-defined order[109−111] (Figure 26). The bulk of the fimbrial filament (shaft) is made up of polymers of the major subunit, which thus plays a structural role. Only one of the subunits, usually a minor component of the fimbriae, possesses a carbohydrate combining site and is responsible for the binding activity and sugar specificity of the fimbriae, e.g., for mannose (in type 1 fimbriae) or galabiose, Gal(α1−4)Gal (in P fimbriae). In type 1 fimbriae, which are made up of hundreds of subunits of four different kinds, this subunit (MW 29−31 kDa) is present in small numbers at intervals along the fimbrial filament and at the distal tip. However only the latter subunit appears to be able to mediate mannose-sensitive adhesive interactions, whereas the subunits at the other positions are inaccessible to the ligand. In other types of fimbriae (e.g., type P) the carbohydrate-binding subunit (MW 36 kDa) is exclusively located at the tip. The combining sites of type 1 fimbriae of *E. coli* and *K. pneumoniae* correspond to the size of a trisaccharide and are probably in the form of a depression or pocket on the surface of the lectin. Several of the carbohydrate binding subunits have been sequenced, but in no case has the three-dimensional structure of any of them been solved.

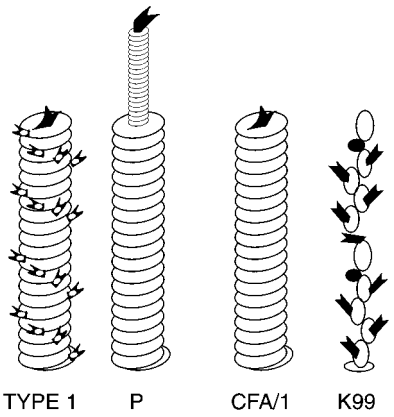


Figure 26. Schematic representation of different types of fimbriae of *E. coli*. The disks stand for the fimbrial subunits; the black shapes symbolize the carbohydrate binding sites. In the type 1 fimbriae, the symbols with white dots denote carbohydrate binding sites not available to the ligand. (Modified from ref 111.)

IV. Combining Sites

The combining sites of lectins are in the form of shallow depressions on the surface of the protein. Typically, only one or two edges or faces of the carbohydrate ligand are bound to the protein. This is in contrast to carbohydrate-binding bacterial periplasmic receptors, specific for, e.g., glucose or galactose, in which the ligand is buried in the interior of the protein.[112] In lectins, the combining sites appear to be preformed, since few conformational changes occur upon ligand binding. In general, the sites within a lectin family are similar, but quite different in different families, even if the specificity is the same, emphasizing the fact that nature finds different solutions to the problem of the design of combining sites for structurally similar ligands.[6]

Lectins combine with carbohydrates by a network of hydrogen bonds and hydrophobic interactions; coordination with metal ions may also play a role.[13−15,113] (Table 6). The hydrogen bonds are formed between carbohydrate hydroxyl groups and NH groups, hydroxyls, and oxygen atoms of the protein. When each of two adjacent hydroxyls of a monosaccharide interacts with a different atom of the same amino acid (e.g., the two oxygens of the carboxylate of glutamic or aspartic acid), they form bidentate hydrogen bonds.[114] Such bonds are quite common in protein−carbohydrate complexes. A different kind of hydrogen-bond characteristic for such complexes is the cooperative bond, in which the hydroxyl group acts simultaneously as donor and acceptor. van der Waals forces, although rather weak (often only a fraction of 1 kcal mol^{-1} for each pair of atoms), are frequently numerous and together may make a significant contribution to binding.

Even though carbohydrates are highly polar molecules, the steric disposition of hydroxyl groups creates hydrophobic patches on sugar surfaces that can form contacts with hydrophobic regions in the protein molecules. One common type of interaction

Table 6. Amino Acids and Metal Ions in the Monosaccharide Combining Sites of Plant and Animal Lectins[a]

types	H-bonding amino acids[b]	hydrophobic residues	metal ion	metal role
simple				
legumes	Asn, Asp	+	Ca^{2+}, Mn^{2+}	structural
cereals	Glu, Ser, Tyr	++	None	
amaryllidaceae	Asn, Asp, Gln, Tyr	+	None	
moraceae	Asn, Gly[c]	+	None	
euphorbiaceae	Asp, Gln			
galectins	Arg, Asn, His, Glu	+	None	
mosaic (multidomain)				
influenza virus hemagglutinin	Ser	++	None	
C-type	Asn, Glu		Ca^{2+}	coordinate ligand

[a] Additional side chain interactions occur when the lectins bind oligosaccharides [b] Only bonds with amino acid side chains are mentioned [c] This bond is to the N-terminal amino group

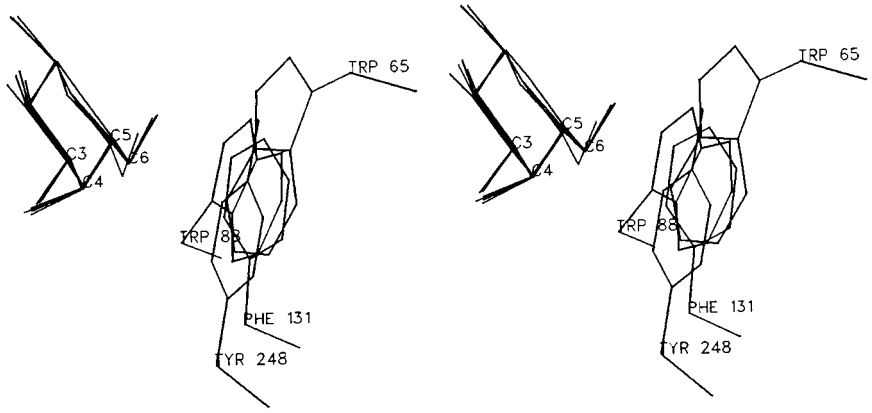


Figure 27. Stereodiagram showing the relative placement of galactose and binding site residues in different lectins: Trp65 (Galectin-1), Phe131 (ECorL), Tyr248 (ricin site 2) and Trp88 (LT, enterotoxin from *E. coli*). (Reprinted with permission from ref 13. Copyright 1995 Annual Reviews Inc. URL: http://www.AnnualReviews.org.)

is the stacking of a monosaccharide on the side chains of the aromatic amino acids such as phenylalanine, tyrosine, or tryptophan (Figure 27). In addition, the methyl group of the acetamide moiety of *N*-acetyl-amino sugars often interacts with aromatic residues in lectins (e.g., WGA and influenza virus hemagglutinin specific for the above sugars). Since most saccharides are uncharged, ionic (charge−charge) interactions do not commonly participate in the formation of their complexes with proteins. An exception is the heparin−antithrombin III complex, in which four basic amino acids of the protein form an elongated, positively charged binding site complementary to a specific oligosaccharide sequence of the polysaccharide with a unique sulfate substitution pattern.[115]

Contacts between the protein and its ligands are sometimes mediated by water bridges.[43,116] Water acts as a molecular "mortar", its small size and ability to act as both a hydrogen donor and hydrogen acceptor conferring ideal properties for this function. Tightly bound bridging water can be thought of as structural water, essentially an extension of the protein surface. Comparison of a series of sugars bound to a given lectin, or a series of related lectins bound to a given sugar, sometimes reveal common water molecules, suggesting that they are important elements in ligand recognition.

A. Simple Lectins

1. Legume

Legume lectins, irrespective of their specificity, bind ligands through the side chains of a constellation of three invariant combining site residues, an aspartic acid, an aparagine, and an aromatic amino acid[117,118] or leucine.[41] Replacement, by site-directed mutagenesis, of the aspartic acid or asparagine by another amino acid (e.g., alanine) in several of these lectins (e.g., ECorL,[118] pea lectin,[119] and GSII[120]) resulted in loss of sugar-binding ability. The key role of these amino acids in ligand binding has been similarly demonstrated for the homologous, mannose-specific animal lectin MR60/ERGIC 53 (cf. section III.A.1).[121] The aspartic acid and asparagine also participate in coordinating the calcium ion present in all members of this family, which explains the metal ion requirement for carbohydrate binding. Another characteristic of the combining site of the legume lectins is the presence of a rare *cis*-peptide bond between the critical asparagine just mentioned and the preceding amino acid, which is almost always alanine. This bond is necessary for the proper orientation of the asparagine.

The fact that the key amino acids involved in the binding of the carbohydrate are highly conserved in all legume lectins and have an identical spatial disposition raises the puzzling question of how discrimination between very similar monosaccharides,

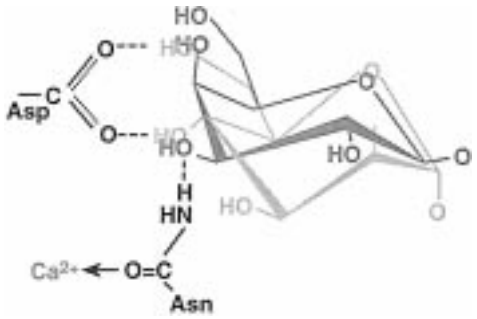


Figure 28. Hydrogen bonds between side chains of conserved amino acids that ligate mannose (light lines) to concanvalin A or *Lathyrus ochrus* lectin I and galactose (darker lines) to ECorL or GSIV. (Reprinted from Sharon, N. In *Lectin Blocking. New Strategies for the Prevention and Therapy of Tumor Metastasis and Infectious Diseases*; Beuth, J., Pulverer, G., Eds.; Verlag: Stuttgart, 1994; p 3. Copyright 1994 Gustav Fischer Verlag.)

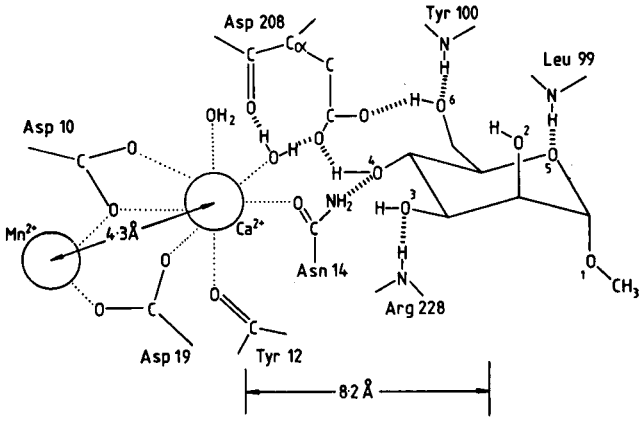


Figure 29. Methyl α-mannoside in the combining site of concanavalin A. (Reprinted with permission from ref 124. Copyright 1994 International Union of Crystallography.)

e.g., mannose (or glucose) and galactose is achieved. This is possible because the monosaccharide in the combining site of the mannose (and glucose)-specific lectins is oriented differently than in the galactose-specific ones (Figure 28). For instance, concanavalin A[122,123] and the lectins from lentil[124] and *Lathyrus ochrus* [125] bind glucose and mannose such that the Oδ1 and Oδ2 of asparagine accept hydrogen bonds from 6-OH and 4-OH, whereas Nδ2 of asparagine donates such a bond to the 4-OH of the sugar. In addition, the main-chain amide of glycine, which is conserved in all legume lectins with the exception of concanavalin A, forms a hydrogen bond with the 3-OH of the monosaccharide (Figure 29). On the other hand, in lectins specific for galactose, such as ECorL,[45] PNA,[47] and SBA,[48] Oδ1 and Oδ2 of asparagine accept hydrogen bonds from the 4-OH and 3-OH, respectively, while the Nδ2 of asparagine and the NH of glycine donate such bonds to the 3-OH (Figure 30). In this way, the same constellation of highly conserved residues in legume lectins provides the framework required for binding of diverse monosaccharides, while specificity apparently derives from variability in the structure of the amino acids that line other parts of the binding pocket (Figure 31). As expected, the mannose/glucose-specific lectins do not form bonds with the 2-OH which differs in configuration between these two monosaccharides.

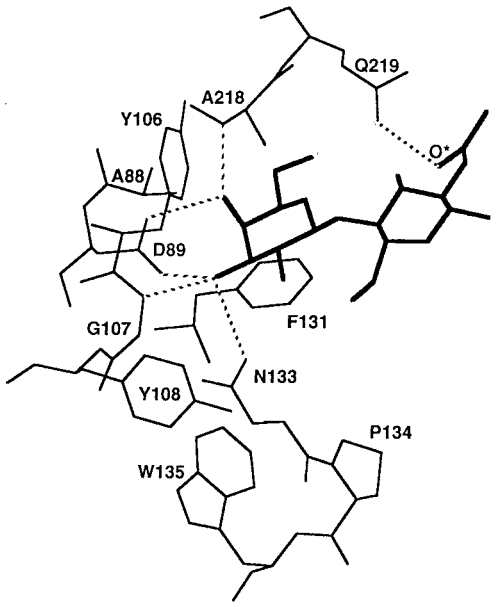


Figure 30. *N*-Acetyllactosamine in the combining site of *Erythrina corallodendron* lectin. (Based on data in ref 118.)

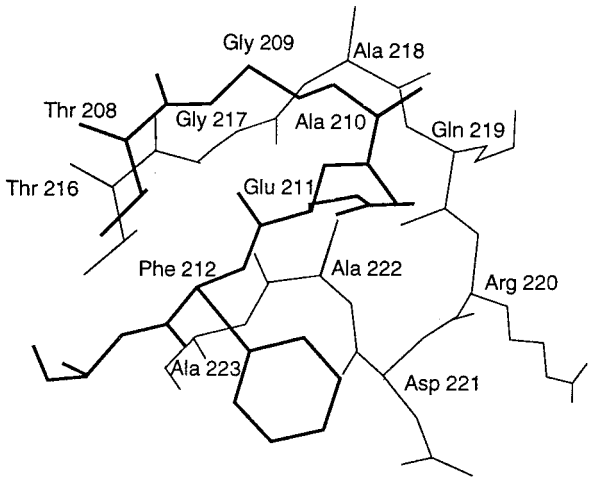


Figure 31. Superimposition of the variable regions of the combining sites of *Erythrina corallodendron* lectin (residues 216−223, thin lines) and *Lathyrus ochrus* lectin I (residues 208−212, thick lines). (Reprinted from Sharon, N. In *Lectin Blocking. New Strategies for the Prevention and Therapy of Tumor Metastasis and Infectious Diseases;* Beuth, J., Pulverer, G., Eds.; Verlag: Stuttgart, 1994, p 3. Copyright 1994 Gustav Fischer Verlag.)

The amino acids that form the monosaccharide combining sites of legume lectins are derived from the four distinct regions of the polypeptide chain that form loops, designated A, B, C, and D, connecting the β strands.[126] The conserved aspartic acid and glycine belong to loops A and B, respectively, whereas the asparagine and the hydrophobic residue are in loop C. Additional interactions, usually with backbone atoms, are provided by amino acids in loop D. There is a clear correlation between the lengths of loop D and the monosaccharide specificity of the legume lectins. Thus, in all mannose-specific lectins this loop is identical in size, consisting of 18 amino acids of variable sequence, and it is different in size from loop D found in lectins specific for Gal/GalNAc or for oligosaccharides only (Figure 32).

Lectin	Loop A	Loop B	Loop C	Loop D	Number of gaps				Sugar type
					A	B	C	D	
WBAI	APFP--RPHPADGLVF	GE-GGGYFG	VEFDTFRN----TWDP	GFSAATGDPSGKQRNATETHDILSW	2	1	4	0	GalNAc
ECorL	GPYT--RPLPADGLVF	AQ-GYGYLG	VEFDTFSN----PWDP	GLSGATG----AQRDAAETHDVYSW	2	1	4	4	GalNAc
DBL	APSK---ASFADGIAF	RR-NGGYLG	VEFDTLSNS---GWDP	GFSATTGLSEG----YIETHDVLSW	3	1	3	4	GalNAc
DB58	APNK---SSSADGIAF	KS-NSGFLG	VEFDTFSNT---DWDP	GFSATTGFFEG----YTETHDVLSW	3	1	3	4	GalNAc
CSII	APNPS--TAATDGLAF	QSA-GGYLG	VEFDTYYNS---AWDP	GFSATTGQTDN----YIETHDILSW	2	1	3	4	GalNAc
PHAE	VPNN---SGPADGLAF	KD-KGGLLG	VEFDTLYNV---HWDP	GFTATTGITKG----NVETNDILSW	3	1	3	4	Complex
PHAL	VPNN---AGPADGLAF	KD-KGGFLG	VEFDTLYNK---DWDP	GFSATTGINKG----NVETNDVLSW	3	1	3	4	Complex
PHAM	VPNN---AGPADGLAF	KD-KGGFLG	VEFDTLYNK---DWDP	GFSATTGITKG----NVETNDILSW	3	1	3	4	Complex
SBA	APDT---KRLADGLAF	QT-HAGYLG	VEFDTFRN----SWDP	GFSAATGLDIP-----GESHDVLSW	3	0	4	5	GalNAc
PNA	KD--IKDYDPADGIIF	GSIGGGTLG	VEFDTYSNS--EYNDP	GFSASGSL------GGRQIHLIRSW	2	1	2	6	Gal
LTA	IR--ELKYTPTDGLVF	GS-TGGFLG	VEFDSYHN----IWDP	GFSATTGN------PEREKHDIYSW	2	1	4	6	Fuc
UEAI	SANP---KAATDGLTF	RRA-GGYFG	VEFDTI-GSPVNFDDP	GFSGGTYI------GRQATHEVLNW	3	1	1	6	Fuc
UEAII	EPDE--KIDGVDGLAF	GS-SAGMFG	VEFDSYFGKTYNPWDP	GFSGGVGN------AAKFDHDVLSW	2	1	0	6	GlcNAc
LAAI	PPIQSRKADGVDGLAF	GS-SAGMFG	VEFDTYFGKAYNPWDP	GFSAGVGN------AAKFNHDILSW	0	1	0	6	GlcNAc
OVL	RENI---NRGGDGITF	KS-GGGYLG	VEFDTFSN----RWDP	GLSAATGDLV-------EQHRLYSW	3	1	4	7	Man/Glc
MTA	APYS---SNVADGLAF	IG-RAGFLG	VEIDTFHN----TWDP	GFSAATGAEF-------AEHDIRYW	3	1	4	7	Man/Glc
LSL	RPNSDS-QVVADGFTF	RG-DGGLLG	VEFDTFHNQ---PWDP	GLSASTATYY-------SAHEVYSW	1	1	3	7	Man/Glc
ConA	SPDS----HPADGIAF	GS-TGRLLG	VELDTYPNT--DIGDP	GLSASTGL-------YKETNTILSW	4	1	2	7	Man/Glc
DiocL	SPDH----EPADGITF	GS-GGRLLG	VELDSYPNT--DIGDP	GLSATTGL-------YKETNTILSW	4	1	2	7	Man/Glc
LOLI	APNS---YNVADGFTF	QT-GGGYLG	VEFDTFYNT---AWDP	GFSATTGAEF-------AAHEVLSW	3	1	3	7	Man/Glc
LenL	SPNG---YNVADGFTF	QT-GGGYLG	VEFDTFYNA---AWDP	GFSATTGAEF-------AAQEVHSW	3	1	3	7	Man/Glc
PSL	APNS---YNVADGFTF	QT-GGGYLG	VEFDTFYNA---AWDP	GFSATTGAEY-------AAHEVLSW	3	1	3	7	Man/Glc
Favin	APNG---YNVADGFTF	QT-GGGYLG	VEFDTFYNA---AWDP	GFSATTGAEY-------ATHEVLSW	3	1	3	7	Man/Glc
DlabL	TNYT---SRIADGLAF	SY-HGGFLG	VEFDTDYLN-PDYGDP	GLSASTGQ-------NIERNTVHSW	3	1	1	7	Man/Glc
GSIV	KNYG---APTADGLAF	KDY-GGFLG	VEFDTWINK--DWNDP	GFSAGVG--------YDEVTYILSW	3	1	2	8	GalNAc
SPL	IDVP---HITADGFAF	KDY-GGCLG	VEFDTWPNT--EWSDL	GFSGGTG--------FNETQYILSW	3	1	2	8	Gal
		** *	*	** **	*		*		

Figure 32. Correlation between the size of binding loops and monosaccharide specificity of legume lectins. The number of gaps present in the four loops is shown along with the specificity of the lectins on the right. The residues that are highly conserved have been indicated with asterisks (*) and the ligand binding residues are shown in bold. (Reprinted with permission from ref 126. Copyright 1997 Academic Press.)

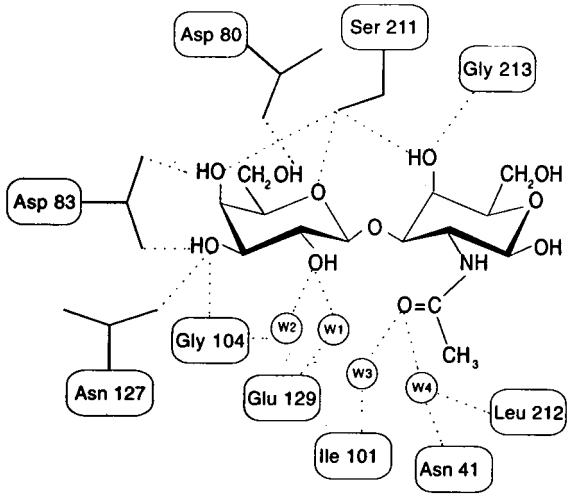


Figure 33. Gal(β1−3)GalNAc in the combining site of peanut agglutinin. The terminal galactose of the disaccharide forms, in addition to the commonly occurring bonds with the side chains of asparagine (Asn127) and aspartic acid (Asp83) and the main chain amide of glycine (Gly104), unique ones, namely between 6-OH and the side chain of Asp80 and between the ring oxygen and Ser211. The 4-OH of the *N*-acetylgalactosamine is hydrogen-bonded to Ser211 and Gly213. (Reprinted with permission from ref 128. Copyright 1997 Current Science.)

When lectins bind disaccharides, the nonreducing residue occupies the monosaccharide combining site, i.e., where the mannose of methyl α-mannoside binds, with additional contacts to the protein provided by the reducing residue. This is illustrated by ECorL in complex with *N*-acetyllactosamine (Figure 30)[118] or PNA in complex with Gal(β1−3)GalNAc (Figure 33).[127,128]

In the crystal of concanavalin A with the branched trisaccharide Man(α1−6)[Man(α1−3)]Man,[129] the α1−6-linked, nonreducing mannose of the trisaccharide occupies the monosaccharide combining site of the lectin, and forms essentially the same contacts (Figure 34). There are bonds also to the reducing mannose. The demonstration that the trisaccharide combines with an extended site on concanavalin A is in agreement with the results of titration microcalorimetry.[130,131] A single bridging water molecule is seen in the trisaccharide−concanavalin A complex.[129] In contrast, 20 water molecules are involved in the binding of the linear trisaccharide Man(α1−3)Man(β1−4)GlcNAc to *Lathyrus ochrus* lectin I, (LOL I), which is also mannose-specific.[132] In this complex, the α1−3-linked mannose at the nonreducing end of the trisaccharide occupies the monosaccharide combining site, but does not form the same contacts as does methyl α-mannoside in the same site[125] (see below). The only direct interactions between the trisaccharide and the protein are with this mannose, whereas those of the two remaining sugars are all mediated by long-chain water bridges. An example of such a bridge is the nine water molecules connecting the atoms of Manβ1−4 and of *N*-acetylglucosamine to LOL I over a distance of 13 Å.

The dibranched *N*-acetyllactosamine-type octasaccharide (entry 12 in Table 5) binds to LOL I with the α1−3 linked mannose occupying the monosaccharide binding site.[133] The complex is stabilized by a large number of hydrogen bonds, several via water, as well as by numerous van der Waals contacts. The above mannose and the GlcNAcβ1−2 of the α1−6-linked branch interact on each side of Phe123 and grip the aromatic ring as a clamp. A neighboring tyrosine is

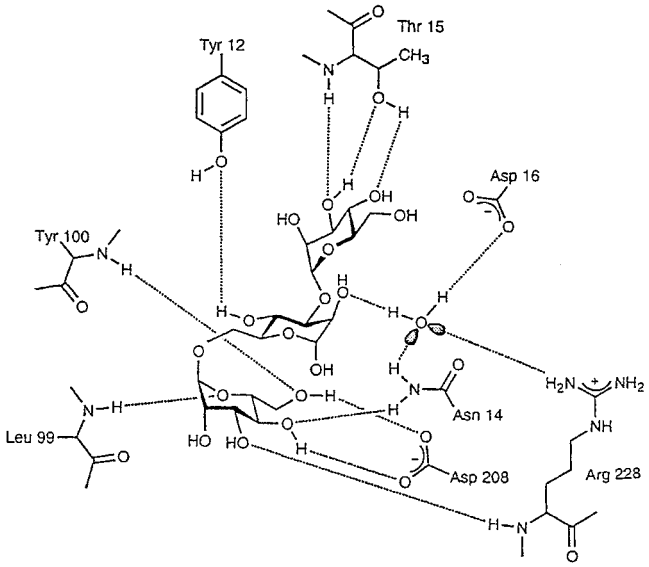


Figure 34. The extended trimannoside binding site of concanavalin A. In the crystal of the lectin with Manα6-(Manα3)Man the sugar is bound in the extended conformation, with a torsional angle af 180° about the C5−C6 bond of the α1−6 linkage. The α1−3-linked mannose contributes to the binding of the trisaccharide by hydrogen bonds between the 3-OH and 4-OH and the hydroxyl of Thr15, as well as between the 3-OH and the main chain nitrogen of the same amino acid. The reducing mannose of the trisaccharide is hydrogen bonded via its 4-OH to the hydroxyl of Tyr12. (Reprinted with permission from ref 129. Copyright 1996 American Society for Biochemistry and Molecular Biology.)

stacked against the terminal galactose of the α1−6 branch. As a result, the Gal(β1−4)GlcNAc of this branch fits in a partly hydrophobic cleft and displays numerous interactions with the lectin. The multiplicity of contacts that the octasaccharide forms with the protein explains why it exhibits an affinity to the lectin a 1000-fold higher than the monosaccharide.

Although mannose specific legume lectins do not bind free fucose, some of them (e.g., those of pea lectin and *L. ochrus*) bind fucose-containing glycopeptides of the type shown as entry 13 in Table 5, with the fucose α1−6 linked to the asparagine-bound *N*-acetylglucosamine, approximately 8−10-fold more tightly than their unfucosylated analogues.[21] The X-ray structure of cocrystals of LOL II, an isoform of LOL I, with such a fucosylated glycopeptide (or with the corresponding free oligosaccharide) has revealed that the α1−3-linked mannose resides in the monosaccharide binding site.[134] The fucose is hydrogen bonded to a shallow crevice on the distal side of Phe123, resulting in extended van der Waals contacts between the oligosaccharide and the lectin. Although concanavalin A has the same monosaccharide specificity as LOL II, it does not distinguish between oligosaccharides containing fucose and those that lack this sugar. Superposition of the main-chain carbon atoms of the two lectins revealed that the overall binding scheme between the oligosaccharide and LOL II can be preserved in concanavalin A without any steric clashes, except for the fucose binding site of the former lectin, which is occupied by a histidine in the latter.

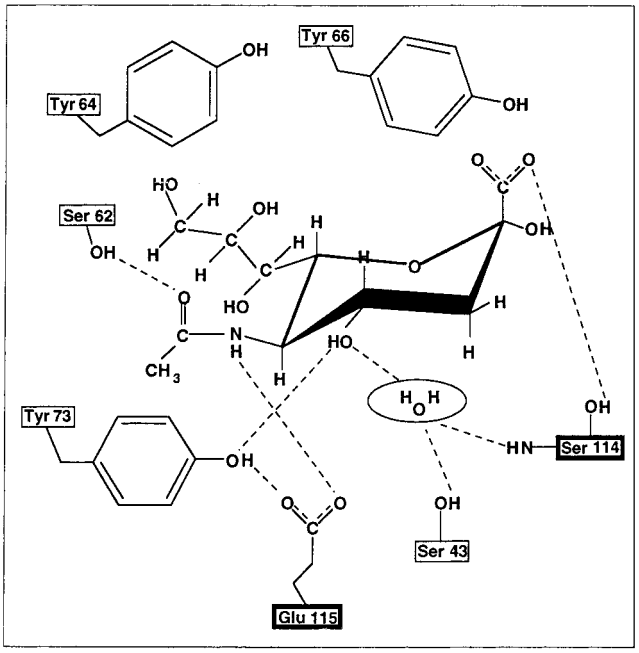


Figure 35. Sialyllactose in the combining site of wheat germ agglutinin. The carbonyl of the trisaccharide is hydrogen bonded to the hydroxyl of Ser62 and its amide to the carbonyl of Glu115. In addition, it forms five to seven van der Waals contacts with the phenyl ring of Tyr73. The adjacent ring hydroxyl (4-OH) is fixed by hydrogen bonds to the hydroxyl of the same tyrosine and an ordered water molecule. The carboxylate group of the *N*-acetylneuraminic acid is within hydrogen-bonding distance of the hydroxyl of Ser114. Several van der Waals contacts stabilize the orientation of the sugar ring through nonpolar stacking interactions with the aromatic side chain of Tyr66. A third aromatic side chain, that of Tyr64, interacts through nonpolar contacts with the glycerol tail of the *N*-acetylneuraminic acid. Only one water molecule appears to be involved in stabilizing the lectin−sugar complex. It is tetrahedrally coordinated by hydrogen bonds with the 4-OH of *N*-acetylneuraminic acid, the backbone amide of Ser114 and the hydroxyl of Ser43. (Figure modified by Joseph Crane from ref 135. Reprinted with permission from ref 117. Copyright 1993 Elsevier Trends Journals.)

2. Cereal

In the crystalline complex of WGA with sialyllactose, the sialic acid interacts with the lectin by a number of hydrogen bonds—none, however, via aspartic acid or asparagine, as found in the legume lectins—as well as by nonpolar contacts with aromatic amino acids[135] (Figure 35). The amino acids forming these contacts are not located in the same subunit, as usually found in other lectins, but belong to the two subunits of the lectin dimer. One subunit contributes an array of three aromatic amino acids (Tyr 64, Tyr 66, and Tyr73) and a serine (Ser 62), while the other provides two polar residues (Ser115 and Glu115). All the sialic acid ring substituents participate in interactions with WGA: the acetamide and carboxyl groups, as well as the hydroxyls attached to the pyranose ring and the glycerol side chain. However, the essential specificity determinants for this monosaccharide, as well as for *N*-acetylglucosamine (and *N*-acetylgalactosamine) which WGA also binds (cf. section II.A), are the *N*-acetyl group and the adjacent 3-OH. In this way a cluster

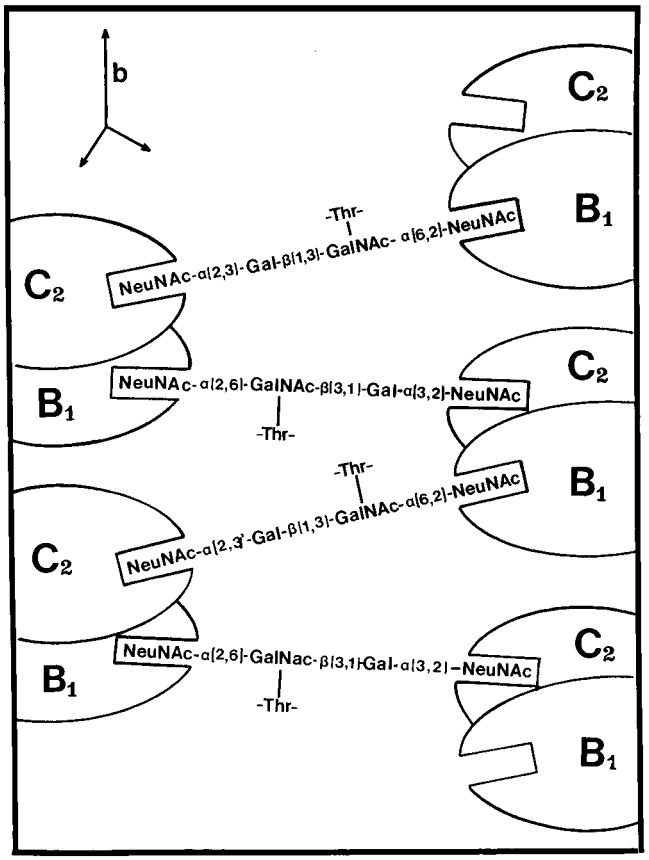


Figure 36. Schematic representation of wheat germ agglutinin cross-linked with a branched sialoglycopeptide from glycophorin. (Reprinted with permission from ref 136. Copyright 1992 American Society for Biochemistry and Molecular Biology.)

of three spatially close hydrogen bonds with the ligand and a hydrophobic contact (acetamide-CH_3 with the aromatic ring of Tyr73) is formed in the least exposed part of the combining site, where the conformation of the protein is most stable.

In the cocrystal of WGA with a branched sialoglycopeptide (entry 18 in Table 5) the carbohydrate, which possesses both α2−3- and α2−6-linked terminal *N*-acetylneuraminic acid moieties, cross-links two crystallographically related dimers (Figure 36). It binds to the protein so that the α2−6-linked sugar occupies the combining site in domain B of one dimer and the α2−3-linked sugar occupies the combining site of domain C of an opposing dimer.[136,137] This mode of binding, which is seen also with other lectins and branched ligands (cf. section IV.A.3), provides an insight into a possible mechanism of cross-linking that occurs when lectins interact with cell surface receptors (see section IV.D).

WGA contains four unique carbohydrate binding sites (section III.A.2). Crystallographic studies revealed that all four sites are functional, although two of them appear to have affinities too weak to be detectable in solution. Theoretical modeling, in conjunction with the crystallographic data, was employed to analyze and compare the binding interactions of *N*-acetylneuraminic acid and GlcNAc(β1−4)GlcNAc at each of the four sites.[51] The similiarity between the interactions is limited to those involving the three tyrosines that are quasi-conserved in all four sites and the fully conserved serine, all of which are part of a shallow aromatic pocket on the surface of one of the lectin subunits. In contrast, variability is observed in the interactions involving the contacting domain on the other subunit. Quantitative estimates for polar, nonpolar, and ionic interactions revealed that hydrogen bonding makes the largest contribution to complex stabilization, in agreement with thermodynamic data.[138]

3. Amaryllidaceae and Related

As mentioned (section III.A.4), the three-dimensional structure of only one of these lectins—that of the snowdrop (*Galanthus nivalis*)—has been elucidated, and of its 12 combining sites only those in subdomain I of each of the four subunits of the lectin and subdomain III of one of the subunits have been characterized in detail.[55] Every one of these combining sites contains five contact residues, Gln, Asp, Asn, Val, and Tyr, interspersed with three additional hydrophobic residues that are part of the hydrophobic core of the lectin (Figure 37). The five contact residues are invariant in all three subdomains, which led to the assumption that the same or very similar residues will form contacts with the ligand in the remaining sites. The 2-OH of the ligated mannose interacts with the side chains of the aspartic acid and asparagine. This accounts for the specificity of the lectin for mannose and its inability to bind glucose, a property that distinguishes lectins of this family from those of the legumes, such as concanavalin A or LOL I, that combine with both monosaccharides (cf. section IV.A.1).

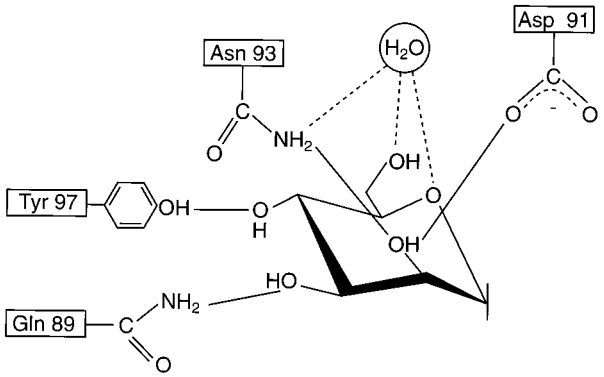


Figure 37. Mannose in the combining site of the first subdomain of *Galantis nivalis* lectin. Gln89 interacts with the 3-OH and the hydroxyl of Tyr97 with the 4-OH of the bound mannose, which together with the interactions of the 2-OH with the side chains of Asp91 and Asn93 provide four H-bonds in the complex. Additionally, Val95 makes hydrophobic contacts with C3 and C4. Three water molecules form a network of hydrogen bonds with the ring oxygen, the 6-OH and the nitrogen of the Asn93 amide. The binding is enhanced by participation of residues from neighboring subunits: His107 from a subunit in the same tetramer in the combining site of subdomain 1 and Leu48 from another tetramer in the site of subdomain 3. (Reprinted with permission from ref 55. Copyright 1995 Nature Publishing Co.)

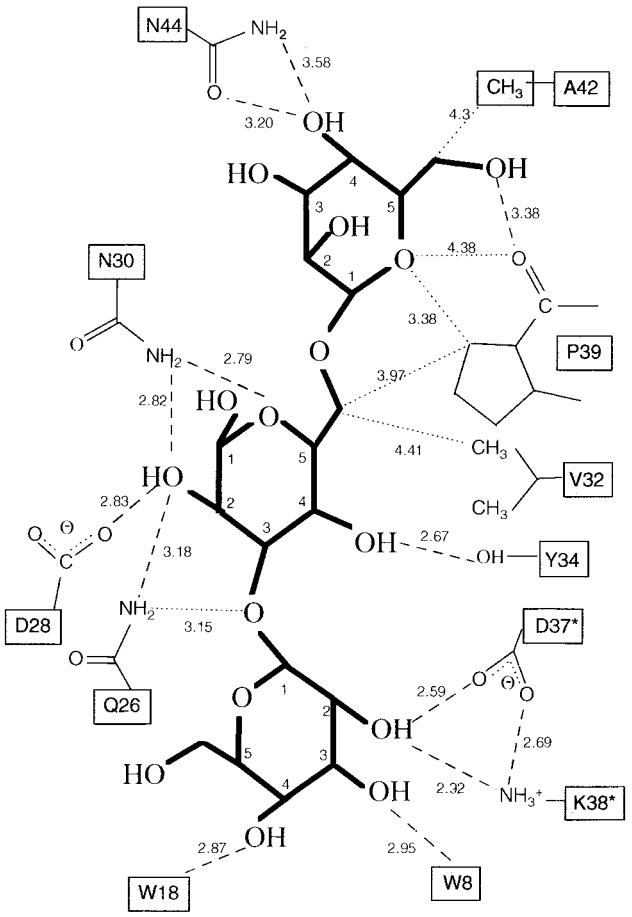


Figure 38. Manα3(Manα6)Man in the combining site of *Galantus nivalis* lectin. (Reprinted with permission from ref 56. Copyright 1996 Current Biology Ltd.)

A comparison with the legume lectins reveals that although in all cases an aspartic acid and an asparagine are involved in ligand binding, the bonding pattern is distinct and less complex in the snowdrop lectin. There are, for instance, no bidentate interactions of the aspartic acid in the combining site of the latter lectin with hydroxyls of the monosaccharide nor hydrophobic interactions of a tyrosine ring with the ligand. This may explain why the affinity of the snowdrop lectin for mannose is by an order of magnitude lower than that of the legume lectins for the same monosaccharide.

In the cocrystal of the snowdrop lectin with a branched pentamannose (entry 16 in Table 5), the oligosaccharide occurs in two distinct binding modes.[56] In one, the entire outer trimannose arm, Man(α1−6)[Man(α1−3)]Man(α1−6), combines with a single lectin subunit through an extended binding region of subdomain 3, which is made up of the conserved monosaccharide combining site and specific subsites for the terminal nonreducing residues, Man(α1−6) and Man(α1−3). In the second binding mode, the latter two mannose residues bind to subdomains 1 of two subunits belonging to different dimers, thus forming a bridge between them (Figure 38), analogous to the mode of binding of a branched sialoglycopeptide to WGA (cf. section IV.A.2).

4. Moraceae

A novel type of carbohydrate binding site has been observed in the crystal of jacalin in complex with methyl α-galactoside.[58] The ligand is substantially enclosed by the protein, with about 60 protein atoms at a distance of 4.0 Å or less. As in other lectins, the side chain of Tyr78 stacks against the hydrophobic face of the galactose. Quite unusually, the carbohydrate forms two hydrogen bonds with the amino group of Gly1 of the α-chain of the lectin (Figure 39). Furthermore, only a single side chain,

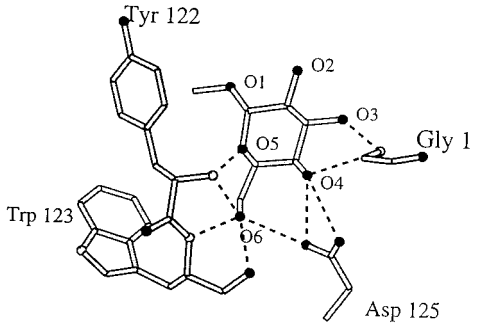


Figure 39. Combining site of jacalin with bound methyl α-galactoside. O1 to O6 mark the oxygen atoms of the monosaccharide. (Reprinted with permission from ref 58. Copyright 1996 Nature Publishing Co.)

that of Asp125, participates in interactions with the ligand. Model building shows that carbohydrates in which the 4-OH is not axial as in galactose, but equatorial as in glucose and mannose, can still form a hydrogen bond with Asp125, but not with Gly1. The specificity of the lectin for galactose is thus chiefly determined by the amino terminal residue, another unusual feature.

5. Galectins

In the crystal structures of galectin-2 from human spleen in complex with lactose[69] and of a galectin-1 from bovine spleen in complex with *N*-acetyllactosamine,[70] the amino acids forming the combining site are contained in four adjacent β-strands. These strands are contiguous in the primary sequences of the galectins and are conserved among the different members of this family. In both complexes, the side chains of three amino acids (histidine, asparagine, and arginine), invariant among all galectins sequenced, are hydrogen-bonded to the 4-OH of the terminal, nonreducing galactose, and a likewise conserved tryptophan is stacked against the sugar ring; the 6-OH is also hydrogen bonded with the protein, but neither is the 2-OH nor the 3-OH. In the galectin−lactose complex (Figure 40a) the glucose moiety contributes to binding by interacting with the protein via its 2-OH and 3-OH, while in the galectin−*N*-acetyllactosamine complex, only the 3-OH of *N*-acetylglucosamine is hydrogen bonded to the protein. For the structure of cocrystals of a galectin with larger oligosaccharides, see section IV.C.

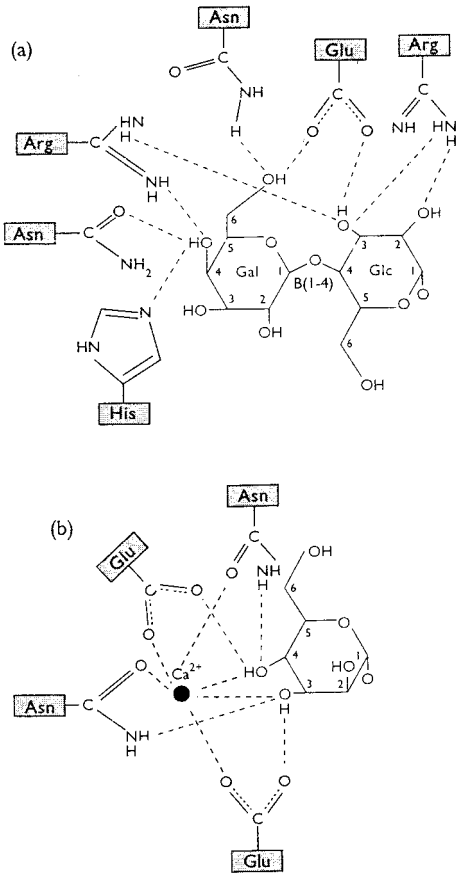


Figure 40. Combining sites of animal lectins: (a) bovine galectin with bound lactose; and (b) mannose binding protein C with bound mannose. Hydrogen bonds and Ca^{2+} coordination bonds are in dashed lines. (Reprinted with permission from ref 4. Copyright 1995 Portland Press Ltd.)

B. Mosaic Lectins

1. Viral Hemagglutinins

a. Influenza Virus. X-ray crystallographic analysis of the complex of the influenza virus hemagglutinin with *N*-acetylneuraminyl(α2−6)lactose placed the *N*-acetylneuraminic acid in the binding pocket with one side of the pyranose ring in tight contact to the protein and the other side facing the solvent.[139] It also permitted prediction of the potential hydrogen bonds and van der Waals contacts between the sialic acid atoms and amino acids in the binding site. The validity of these predictions was tested by determining (a) the ability of a series of synthetic analogues of *N*-acetylneuraminic acid to inhibit viral attachment to cells,[140] and (b) the dissociation constants for the binding of the analogues to the hemagglutinin by nuclear magnetic resonance spectroscopy.[73] It was concluded that the carboxyl of *N*-acetylneuraminic acid forms a bidentate bond with the hydroxyl of Ser136 and with the main chain amide of Asn137 of the hemagglutinin, both of which are necessary for ligand binding (Figure 41). These studies also provided additional proof for the critical importance of the hydrophobic contacts between the acetamide group of *N*-acetylneuraminic acid with the indole ring of Trp153. On the other hand, no evidence was

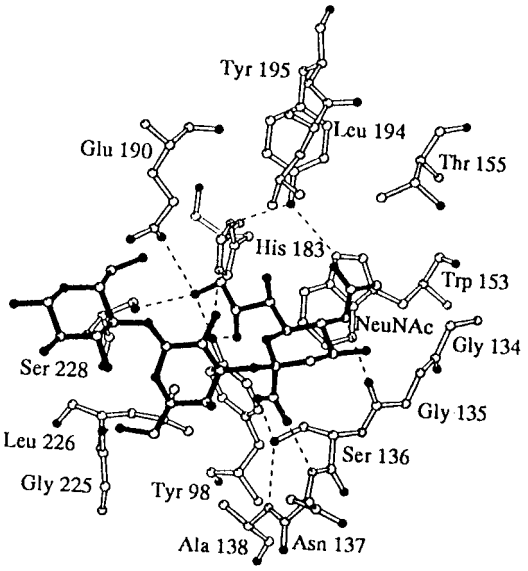


Figure 41. Complex of *N*-acetylneuraminyl(α2−3)lactose (solid line) with influenza virus hemagglutinin. (Reprinted with permission from ref 14. Copyright 1996 Annual Reviews Ltd. URL: http://www.AnnualReviews.org.)

obtained for the participation of the hydroxyl at C-9 in ligand binding, contrary to what has been concluded from the crystal structure.

Comparison of the primary sequences of hemagglutinins of the human virus with those of mutants showing decreased affinity for NeuAc(α2−6)Gal and increased affinity for NeuAc(α2−3)Gal, revealed that they differ in a single residue, Leu226 in the parental strains being replaced by glutamine in the mutants.[72] Avian isolates and their variants showing the reverse change in specificity (from α2−3 linked to α2−6 linked *N*-acetylneuraminic acid), again revealed a substitution only at position 226—from glutamine to leucine. This illustrates that replacement of a single amino acid can alter the sugar specificity of a lectin. Athough residue 226 is located in the carbohydrate binding site of the hemagglutinin, it is not in direct contact with the bound sugar, as shown by the crystallographic studies of the wild-type influenza virus hemagglutinin complexed with *N*-acetylneuraminyl(α2−6)lactose and of a mutant hemagglutinin complexed with *N*-acetylneuraminyl(α2−3)lactose. The suggestion was therefore made that the change in specificity is due to conformational differences between the mutant and the wild-type proteins.

The complexes of the hemagglutinin with *N*-acetylneuraminic acid analogues, having naphthyl or dansyl groups attached to the O-4 and O-6 positions, were also examined by high-resolution X-ray crystallography.[71,141] In each, the sialic acid occupied the same position in the combining site, while the substituents interacted with adjacent hydrophobic patches and polar residues, accounting for the high affinity (10−100-fold higher than that of *N*-acetylneuraminic acid) of the derivatives to the hemagglutinin.

b. Murine Polyoma Virus. Two types of strain of murine polyoma virus are known that differ in their ability to form tumors in mice and in their specificity for sialic acid oligosaccharides: highly tumorigenic

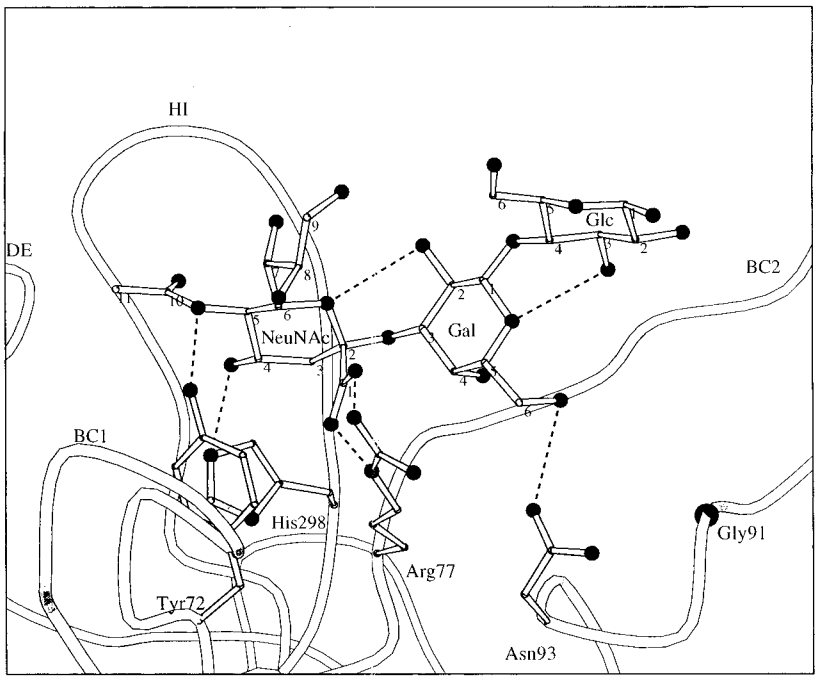


Figure 42. Combining site of polyoma virus. The side chains of Tyr72, Arg77, Asn93, and His298, as well as the trisaccharide NeuAcα2,3Galβ4Glc are shown as ball-and-stick models. The Cα-atom of Gly91 is marked with a black sphere. Hydrogen bonds are in broken lines. (Reprinted with permission from ref 143. Copyright 1994 Mcmillan Magazines Limited.)

(or "large plaque") specific for oligosaccharides terminating in NeuAc(α2−3)Gal, and poorly tumorigenic (or "small plaque") that also tolerate branched structures having a second, α2−6-linked, sialic acid, e.g., NeuAc(α2−3)Galβ3[NeuAc(α2−6)]GalNAc. The critical difference in the structure of the viral protein (VP1) which contains the carbohydrate binding site of these strains is in residue 91, which is glycine in the poorly tumorigenic strains and glutamic acid in the highly tumorigenic ones.[142] Low-resolution crystallographic studies of the intact poorly tumorigenic virus particle in complex with *N*-acetylneuraminyl-(α2−3)lactose[143] and with a branched sialohexasaccharide with both α2−3- and α2−6-linked *N*-acetylneuraminic acid[144] (entry 19 in Table 5) have shown that the combining site is in a shallow groove and that the α2−3 linked sialic acid and the adjoining galactose form contacts with the protein (Figure 42). The inability of the highly tumorigenic, "high plaque" strains to bind the branched ligand is due to electrostatic repulsion between the side chain of Glu91 and the carboxylate of NeuAc(α2−6). These results have recently been confirmed and extended in a high-resolution crystallographic study of a pentamer of VP1 from the "small plaque" virus in complex with the above pentasaccharide.[75]

2. C-Type Lectins

The three-dimensional structures of cocrystals of two C-type lectins with carbohydrate have been elucidated, namely that of the MBP-A with asparaginyloligomannose[145] (entry 17 in Table 5) and that of the highly homologous rat liver MBP (MBP-C) in complex with the methyl α-glycosides of mannose and *N*-acetylglucosamine and the methyl α- and β-glycosides of fucose.[89] In all these complexes the interactions between the protein and the sugar (a terminal nonreducing α-linked mannose in the case of the Asn-oligomannose complex with MBP-A) are the same. A calcium ion serves as the nucleus of the combining site, interacting with the two vicinal equatorial hydroxyls possessing the same stereochemical arrangement as the 3-OH and 4-OH of mannose (Figure 40b). This unusual role of a metal ion as a direct sugar ligand differs fundamentally from that in the legume lectins (cf. section III.A.1) and has not been observed elsewhere. Four of the five additional bonds that coordinate the metal ion to the protein are provided by the side chains of two glutamic acids and two asparagines that also form hydrogen bonds to the same (3 and 4) mannose hydroxyls, producing an intimately linked ternary complex of protein, Ca^{2+}, and sugar.

The four amino acids just mentioned are conserved in all C-type lectins specific for mannose, two of them in the sequence Glu-Pro-Asn (positions 185−187 in MBP-A). However, in lectins of this family, which recognize galactose instead of mannose, the glutamic acid is replaced by glutamine and the asparagine by aspartic acid, effectively reversing the position of the side chain amide and carboxylate groups. It was therefore assumed that the nature of the amino acids at these positions (whether Glu-Pro-Asn or Gln-Pro-Asp) is the primary determinant of specificity of the C-type lectins. This has indeed been proven by genetic engineering experiments, in which Glu185 and Asn187 in MBP-A were replaced by glutamine and aspartic acid, respectively, to yield a mutant protein designated QPD. The simple switch in

Table 7. Relative Affinities of Wild-Type MBP-A and Its Mutants by Competition Binding Assay[a]

lectin	K_a, mM		RIA[b]
	MeαMan	Meβ Gal	
wild type	0.34 ± 0.11[c]	ND	0.07 ± 0.03
QPD		ND	3.5 ± 0.7
QPDW	0.05 ± 0.01	0.4	4.6 ± 1.1
QPDWG	<0.007	0.5	43 ± 8
RHL−1[d]		1.0	110 ± 30

[a] Calculated from data in refs 147 and 148. [b] RIA, ratio of inhibitory activities of galactose vs mannose. [c] K_a for mannose, not MeαMan. [d] RHL-1, rat hepatic lectin 1, a galactose-specific C-lectin, included for comparison.

position of a single amide group altered the specificity of the lectin so that galactose became its preferred ligand (by a factor of 3 as compared to mannose), although with poor affinity.[146]

High affinity, comparable to that of rat hepatic lectin-1 (RHL-1), a paradigm of the galactose-specific C-type lectins, was obtained with a mutant of MBP-A in which His189 was replaced by tryptophan, in addition to the switch described above (mutant QPDW)[147] (Table 7). The amino acid in this position is variable in the mannose-specific C-type lectins, but is always either tryptophan or phenylalanine in the galactose specific ones. However mutant QPDW still exhibited an affinity for mannose which was only 5−8 times lower than that for galactose. Insertion into this mutant (between positions 191 and 192) of an additional, glycine-rich, segment of five amino acids found in galactose specific C-lectins yielded a mutant, designated QPDWG, with both the affinity and selectivity for galactose approaching that of RHL-1[148] (Table 7). These experiments indicate that the determinants for affinity and selectivity are somewhat distinct.

The crystal structures of a trimeric fragment of mutant QPDWG containing the neck and carboxy terminal CRD, both alone and in complex with the methyl β-glycosides of galactose and of *N*-acetylgalactosamine revealed that, as with mannose in wild-type MBPs, the 3-OH and 4-OH groups of the sugar directly coordinate Ca^{2+} and form hydrogen bonds with amino acids that also serve as Ca^{2+} ligands.[149] However the different stereochemistry of the 4-OH in mannose and galactose, combined with a fixed Ca^{2+} coordination geometry, leads to different orientations of the bound sugar in complex with the wild-type and mutant lectins. In the latter, the apolar patch, formed by the 3, 4, 5, and 6 carbons of galactose or *N*-acetylgalactosamine, packs against the side chain of Trp189. In contrast, the stacking of the monosaccharide against an aromatic amino acid, as commonly observed in complexes of galactose specific lectins with their ligands, is not seen when mannose is bound to the wild-type MBP. The glycine-rich loop in QPDWG serves as a rigid unit that holds Trp189 in a position for optimal interaction with galactose, but which is incompatible with mannose binding. The results of the crystallographic studies are consistent with earlier NMR measurements and mutagenesis experiments of the various MBP mutants.[147,148]

Comparison of the data from NMR measurements for mutant QPDWG and RHL-1, both in complex with methyl β-galactoside, revealed similar modes of ligand binding.[147] Thus, carbohydrate recognition by galactose-specific lectins the X-ray structure of which is not known can be deduced on the basis of the crystallographically well characterized MBP-A system. Use was made of this possibility to identify the amino acids that contribute to the preferential binding of *N*-acetylgalactosamine over galactose by certain C-type lectins. Thus, asialoglycoprotein receptors on both hepatocytes and macrophages recognize terminal galactose residues of desialylated glycoproteins. However, the hepatic receptor binds oligosaccharides with terminal *N*-acetylgalactosamine 60 times better than those with terminal galactose, whereas the macrophage receptor binds both types of ligand equally well. Studies with chimeric and mutant versions of the CRDs of the two receptors produced in bacteria have indicated that the substitution of only four amino acids in the macrophage receptor with the corresponding amino acids of the hepatic receptor is sufficient to endow it with an affinity for *N*-acetylgalactosamine comparable to that of the hepatic receptor.[150] Using the three-dimensional structure of the QPDWG mutant of MBP-A in complex with *N*-acetylgalactosamine as a scaffold, a model of the relative positions of the side chains of these four amino acids in the hepatic receptor has been created, suggesting the presence of a binding pocket for the acetamide of *N*-acetylgalactosamine.

Until recently, no structural information about the sugar binding sites in the selectins was available. Therefore, X-ray studies of unligated E-selectin, together with site-directed mutagenesis experiments and molecular modeling based on the three-dimensional structures of MBP−carbohydrate complexes, have been combined to suggest possible modes of interaction of E-selectin and P-selectin with their ligand, sLe^x. The first direct data on such interactions were obtained from studies of a mutant of MBP-A, in which residues 211−213 (Ala-Ser-His) were substituted with a stretch of three positively charged residues corresponding to Lys111−Lys113 in the selectins, confering on MBP-A the ability to recognize sLe^x in addition to mannose.[151] Examination, by X-ray, of complexes of this mutant with sialylated and sulfated derivatives of Le^x revealed that, as predicted, the 2-OH and 3-OH groups of fucose form the same set of hydrogen and cordination bonds with the protein as in MBP-C.[152] Surprisingly, the introduced lysines did not interact electrostatically with sialic acid. Instead, Lys211 was in direct contact with the 4-OH of the galactose of the ligand, and via a water molecule with its 6-OH.

3. P-Type Lectins

Although the two lectins belonging to this group—the cation-dependent and cation-independent mannose 6-phosphate receptors (cf. section II.B.3)—have been extensively studied and some modification and mutagenesis data are available[153] no

protein structure exists by which to interpret these results.

C. Energetics of Protein–Carbohydrate Interactions

For a long time the main tool for the study of the specificity of lectins and of their affinity for ligands was inhibition of hemagglutination or of oligosaccharide precipitation, a simple and rapid method that requires small amounts of material. Although it is not accurate enough for the determination of thermodynamic parameters, it is very useful, since the inhibitory activity of a carbohydrate correlates well with the affinity constant, K_a, measured by different physicochemical methods, such as equilibrium dialysis, spectrophotometry, fluorimetry, NMR, and microcalorimetry.[45,154−156] However, a complete thermodynamic profile of a binding interaction includes changes in the free energy, ΔG, in the heat of binding (or enthalphy), ΔH, and in the entropy of binding, ΔS, as well as heat capacity changes, ΔC_p (Table 8). These parameters, in combination with information from high-resolution X-ray structures, are essential for the understanding of carbohydrate-based biological recognition. ΔH and ΔS for a given complex establish the magnitude and sign of ΔG. Greater $-\Delta H$ values for the binding of an oligosaccharide relative to $-\Delta H$ values for those of monosaccharides indicate the existence of an extended combining site. By monitoring changes in the thermodynamic parameters of binding, resulting from the replacement

Table 8. Thermodynamic Parameters of Lectin−Carbohydrate Interactions[a]

lectin	carbohydrate	K_a, (10^{-3} M^{-1})	$-\Delta G$, kcal/mol	$-\Delta H$, kcal/mol	$-T\Delta S$, kcal/mol	ΔC_p, cal mol^{-1} deg^{-1}	ref
concanvalin A	MeαMan		5.3	6.6	1.3	−60	*b*
		8.2	5.3	8.2	2.9	−90	130,131
		7.0	5.2	6.8	1.6		*c*
	Manα6(Manα3)Man		7.2	9.8	2.6	−110	*b*
		490	7.8	14.4	6.6	−110	130,131
ECorL	MeαGal	1.4	4.3	5.2	0.85		*d*
	Me*β*Gal	0.7	3.9	4.3	0.47		*d*
		0.4	3.6	4.4	0.8		*e*
	GalNAc	1.3	4.3	5.5	1.2		*d*
		1.2	4.2	7.1	2.9		*e*
	Me*β*GalNAc	1.3	4.3	6.8	2.5		*e*
	Gal*β*4Glc	1.9	4.5	9.8	5.4		*d*
		1.6	4.4	6.3	1.9		*e*
	Gal*β*4GlcNAc	9.7	5.4	11.3	5.9		*d*
		42	5.0	10.9	5.9	+94	*e*
	MeαGalNDns	350	7.6	5.5	−2.1		*d*
galectin 1[f]	Gal*β*Glc	5.6	5.2	8.8	3.6		*g*
	Gal*β*4GlcNAc	32	6.2	10	3.9		*g*
galectin 1[h]	Gal*β*4GlcNAc	6.2	7.6	6.6	1.4	−90	*e*
lentil	MeαMan	0.8	3.9	4.1	0.2		*c*
MBP-A	MeαMan	1.0	3.8	4.7	0.9		91
	MeαGlcNAc	1.0	3.8	5.2	1.4		91
	NAcYD (G-ah-Man)$_2$[i]	1.0	3.8	5.2	1.4		91
MBP-C	MeαMan	1.0	3.8	5.1	1.3		91
	MeαGlcNAc	1.0	3.8	4.7	0.9		91
	NAcYD (G-ah-Man)$_2$[i]	7	5.9	7.4	1.5		91
pea	MeαMan	1.6	4.4	5.9	1.5		*c*
SBA	MeαGal	1.0	4.1	9.1	4.9		*j*
	Me*β*Gal	0.5	3.7	10.6	6.9	−94	*e*
	GalNAc	9	5.4	9.5	4.1		*e*
	MeαGalNAc	24	6.0	10.7	4.5		*j*
	Me*β*GalNAc	22	5.9	13.9	8.0	−100	*e*
	GalNDns	590	3.7	7.9	4.3		*j*
	Gal*β*4Glc	0.2	3.1	5.5	2.4		131
	Gal*β*4GlNAc	0.5	3.9	8.2	4.3		*j*
WGA	GlcNAc	0.4	3.7	6.1	2.4		138
	(GlcNAc)$_2$	5.3	5.1	15.6	10.5		138
	(GlcNAc)$_3$	11	5.5	19.4	13.9		138
winged bean	Gal	1.2	4.2	5.8	1.6		*k*
	MeαGal	6.6	5.2	5.6	0.4		*k*
	Me*β*Gal	1.0	4.1	4.7	0.6		*k*
	GalNAc	7.2	5.3	6.7	1.4		*k*

[a] All measurements were done by titration microcalorimetry, except for those in De Boeck et al. (ref *j*), where spectrophotometric methods were used. [b] Williams, B. A; Chervenak, M. C.; Toone, E., Jr. *J. Biol. Chem.* **1992**, *267*, 22907. [c] Schwarz, F. P.; Puri, K. D.; Bhat, R. G.; Surolia, A. *J. Biol. Chem.* **1993**, *268*, 7668. [d] Surolia, A.; Sharon, N.; Schwarz, F. P. *J. Biol. Chem.* **1966**, *271*, 17697. [e] Gupta, D.; Cho, M.; Cummings, R. D.; Brewer, C. F. *Biochemistry* **1996**, *35*, 15236. [f] From sheep spleen. [g] Ramkumar, R.; Surolia, A.; Podder, S. K. *Biochem. J.* **1995,** *308*, 237. [h] Recombinant lectin from Chinese hamster ovary cells. [i] Synthestic ligand with two mannose residues with an intersugar distance of 25 Å. [j] DeBoeck, H.; Lis, H.; van Tilbeurgh, H.; Sharon, N.; Loontiens, F. G. *J. Biol. Chem.* **1984**, *259*, 7067. [k] Schwarz, F. P.; Puri, K.; Surolia, A. *J. Biol. Chem.* **1991**, *266*, 24344.

of a particular hydroxyl in the ligand with hydrogen or fluor (as in monodeoxy or fluorodeoxy derivatives), the contribution of that hydroxyl can be assessed.[131,156] There is however no way to evaluate the contribution of individual hydrogen bonds. This is clearly illustrated by a recent study, in which thermodynamic parameters of binding of the trimannoside Man(α1−6)[Man(α-3)]Man to concanavalin A, as measured by titration microcalorimetry, were compared with those of the interaction of the lectin with a series of mono-, di-, and trideoxy analogues of the ligand.[131] The results were in agreement with the X-ray crystal structure of the concanavalin A−trisaccharide complex with respect to the hydroxy groups involved in binding (cf. section IV.A.1). The free energy and enthalpy contribution of the individual groups was however not linear, indicating that the differences measured are due not only to the loss of hydrogen bonds but also to differential contributions of other factors, such as protein and solvent effects.

Both the protein and the ligand in aqueous solutions are normally hydrogen-bonded to water molecules. In the process of complexation, these bonds are replaced by protein−ligand bonds and the released water returns to bulk solvent. The net binding energy represents the differences between the hydrogen bond energies of the protein and of the carbohydrate with water (solute−solvent interactions) and those of the protein and carbohydrate with each other (solute−solute interactions).

The calorimetric data reveal that protein−carbohydrate interactions are enthalpy driven and in almost all cases the enthalpy of binding is more negative than, or equal to, the free energy of binding (Table 8). A significant fraction (25−100%) of the enthalpy of complexation arises from solvent reorganization.[157] The data also show strong linear enthalpy−entropy compensation. Thus, the unfavorable loss in entropy resulting from changes in rotational degrees of freedom as torsion angles around the glycosidic bonds are frozen upon binding is compensated by an advantageous change in enthalpy due to the removal of ligated water. ΔC_p for lectin−carbohydrate binding, a term generally thought to reflect changes in solvent structure during binding, are small (<100 cal mol^{-1} deg^{-1}) and negative. A more detailed discussion of the energetics of protein-carbohydrate interactions can be found in ref 155.

D. Multivalent Binding

The most striking features of the monosaccharide−lectin interactions are that they are relatively weak (in the millimolar range), and may show relaxed specificity, when compared with the strict specificity of enzyme−substrate interactions, for example. Thus, as mentioned, a lectin may bind different sugars that have little in common except the orientation of a few hydroxyls, for instance in mannose and fucose, or a hydroxyl and an acetamide in *N*-acetylglucosamine and *N*-acetylneuraminic acid (see Figure 2; section II.A). Nevertheless, lectins exhibit both high affinity and exquisite specificity for oligosaccharides of cell surface glycoproteins and glycolipids, a prerequisite for their function as recognition molecules in biological processes. It has therefore been suggested that multiple protein−carbohydrate interactions cooperate in each recognition event to give the necessary functional affinity (or avidity) and specificity.[158−160] There are several possible ways, either alone or together, in which this is achieved (a) ligand multivalency; (b) an extended binding region capable of interaction with more than just a single monosaccharide residue of an oligosaccharide, as seen in e.g. concanvalin A and LOL I (cf. section IV.A.1); and (c) clustering of several identical binding sites by formation of protein oligomers. Such an oligomer can bind simultaneously to different, appropriately spaced, arms of a branched oligosaccharide, or to separate carbohydrate chains of the same glycoprotein.

A synthetic polymer carrying multiple mannose residues exhibited a 10^5-fold higher affinity for concanavalin A than methyl α-mannoside.[161] About the same increase in affinity (from ca. 2×10^{-5} to $0.5-2.6 \times 10^{-9}$) was observed with the human macrophage mannose receptor and a series of lysine-based cluster mannosides when the number of mannose residues per molecule increased from two to six.[162] On average, expansion of the cluster with the addition of an α-mannose group resulted in a 10-fold increase in affinity. Similarly, the affinity of the rat hepatic asialoglycoprotein receptor for aminotris-(hydroxymethyl)methane to which three lactose residues have been attached was 100-fold higher than for the derivative with a single lactose subsitiuent.[163] By varying the structure of the carrier and introducing flexible linkers of different length between the scaffold and the carbohydrate, trivalent lactose derivatives that were 1000-fold better ligands than lactose were generated.[159] The above studies emphasize the importance not only of the number of carbohydrate residues in the ligand but that of their spacing and orientation as well.

Ligand multivalency also affects the specificity of lectin−carbohydrate interactions. Thus, whereas concanavalin A binds methyl α-mannose with a 4-fold higher affinity than methyl α-glucose, it discriminates between polyvalent analogues of the corresponding monosaccharides with an up to 160-fold difference in affinity.[161] Recently, the effect of ligand clustering on the specificity of lectins was demonstrated with a solid-phase carbohydrate library of approximately 1300 related di- and trisaccharides attached to beads so that each bead contained clusters of a single carbohydrate species.[164] Out of this pool, *Bauhinia purpurea* lectin (that binds *N*-acetyllactosamine) was able to pick out the beads carrying two specific disaccharides, namely the *N*-*p*-nitrobenzoyl and isovaleroyl derivatives of lactosamine. In solution, these derivatives showed no higher affinity to the lectin than *N*-acetyllactosamine. Therefore, the amplified affinity and specificity of the lectin to the beads containing the two derivatives appear to result from their polyvalent presentation. Since carbohydrates on the beads are analogous to clusters on cell surfaces, the above results can be seen as illustrating the selectivity of lectins in biological systems.

The disposition of the individual binding sites of a polymeric lectin plays an important role in defining with which multivalent saccharide it can interact strongly. In the trimeric MBP-A, for example, the sites project in one direction and are widely spaced (cf. Figure 23), thus making branched oligosaccharides that can span the required distance between the binding sites optimal ligands for these lectins. In the rat MBP the distance between the combining sites in the trimer is 53 Å (ref 87), and in the human analogue it is 45 Å (ref 88). Such distances are far too great to allow contact with multiple terminal sugar residues of mannose-containing oligosaccharides of the type found in mammalian glycoproteins, a situation of great relevance for the biological activity of this class of lectin (see section V.D.2.b). The ability of different members of the plant legume lectin family to form tetramers through different types of dimer−dimer contacts, thereby projecting pairs of binding sites with different orientations, may represent another means of diversifying their preference for particular ligands.

The multivalency of both lectins and oligosaccharides provides them with a potential to form linear, as well as more complex arrays. The former are obtained, for example, with divalent oligosaccharides and dimeric lectins, in which equivalent binding sites are located at the two sites of the dimer. Such structures have been visualized by high-resolution X-ray studies of three crystal forms of galectin-1, a bovine heart muscle lectin, in complex with an octasaccharide and an asparaginyl-nonasaccharide, both dibranched (or biantennary), containing *N*-acetyllactosamine at the nonreducing end of the branches.[165,166] With each of these ligands, three crystal structures are formed—hexagonal, trigonal, and triclinic—arising from the selection of different low-energy oligosaccharide conformations present in solution; electron densities in each of the forms are identical for both the octasaccharide and the asparaginyl−nonasaccharide. The structures reveal infinite chains of lectin dimers cross-linked through the *N*-acetyllactosamine units of the oligosaccharides (Figure 43). In the hexagonal crystal form the ensemble of cross-linked molecules consists of roughly parallel motifs, in the triclinic form the lectin−saccharide motifs are bent with a large bending radius, and in the monoclinic crystal the chains are helicoidal, each turn of the helix containing between four and five galectin−oligosaccharide complexes. One or more of the binding modes observed may be representative of those that are formed when the galectins combine with sugars on cell surfaces. However, available techniques do not permit visualization of lectin-mediated cross-links on cells.

Lectins with more than two binding sites, as found on, e.g., tetrameric soybean agglutinin, allow for the formation with divalent oligosaccharides of cross-linked, three-dimensional lattices and precipitation of the lectin−oligosaccharide complex.[167] Homogeneous precipitates, with distinct lattice patterns, are formed even from solutions containing a single such lectin in the presence of a mixture of two oligosaccharides or, for that matter, a single oligosaccharide and two lectins that have the same specificity (Figure 44). It has been suggested that these findings point to the presence of long-range order and well-defined geometry in the cross-linked complexes and thus possibly a new source of specificity for lectins, namely the ability to selectively cross-link and aggregate glycoproteins in mixed systems. This could have important implications for the interaction of lectins with cells, where they are confronted with large, nearly planar arrays of oligosaccharides, and may also account for the biological activity of lectins, for which cross-linking and clustering of receptors is required.

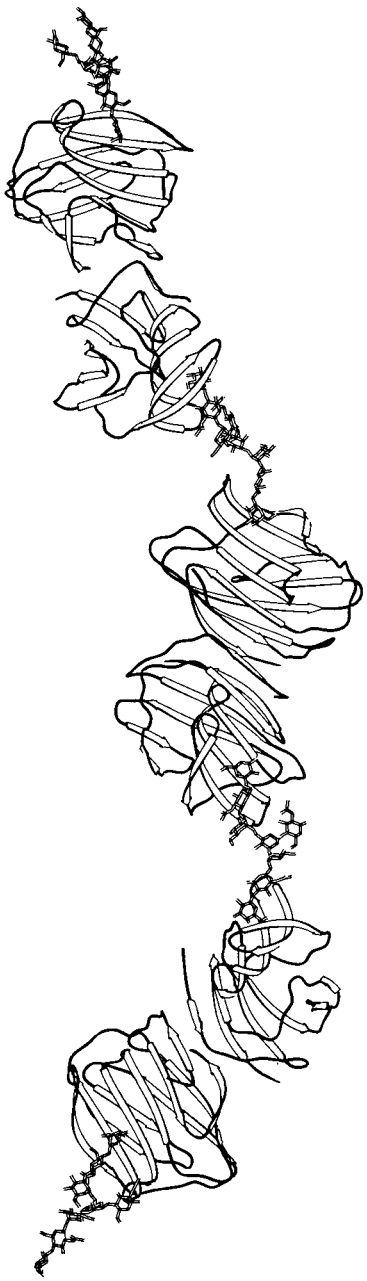


Figure 43. Ribbon model of bovine heart galectin-1 cross-linked by a biantennary octasaccharide (shown as stick model) as found in the hexagonal crystal form. (Reprinted with permission from Lobsanov, Y. D.; Rini, J. M. *Trends Glycosci. Glycotechnol.* **1997**, *9*, 145−154. Copyright 1997 Gakushin Co. Ltd.)

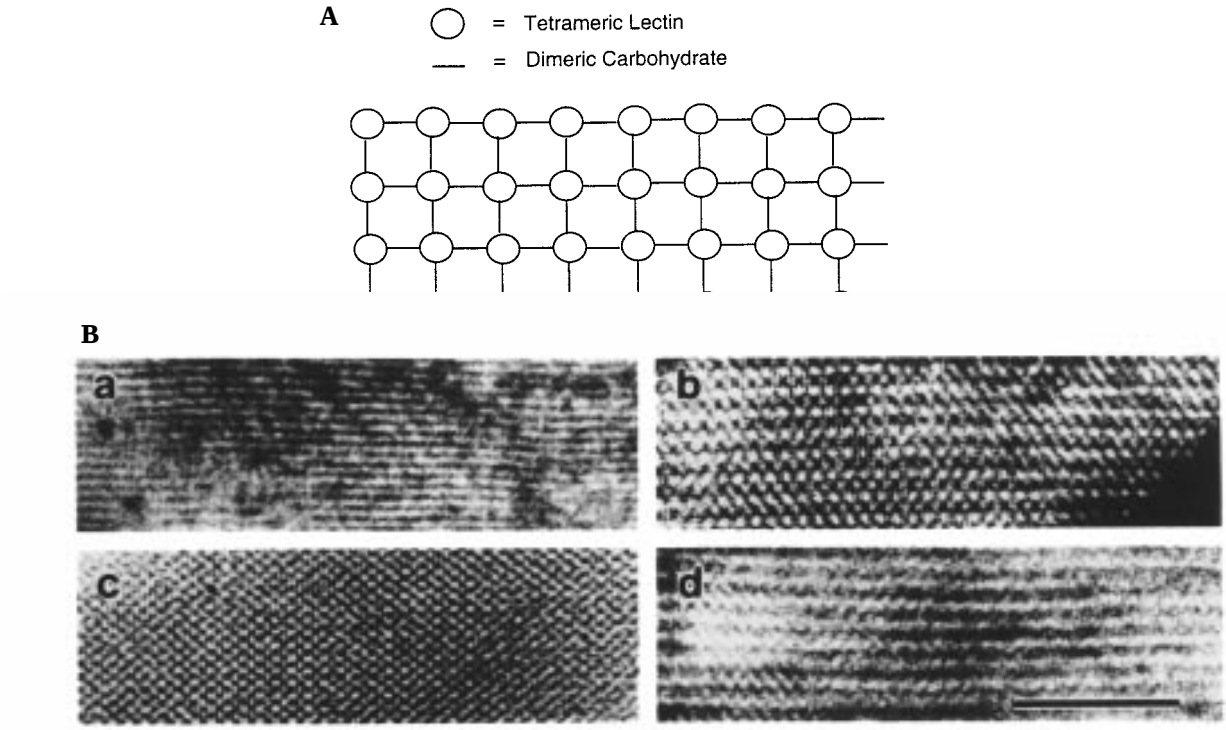


Figure 44. (A) Schematic diagram of a cross-linked complex between a tetravalent lectin and a divalent carbohydrate. The lectin is represented by a circle and the carbohydrate by a line (Reprinted with permission from from Brewer, C. F. *Trends Glycosci. Glycotechnol.* **1997**, *9*, 155−165. Copyright 1997 Gakushin Co. Ltd.) (B) Negative stain electron micrographs of the precipitates of soybean agglutnin with dfferent biantennary pentasaccharides: (a) pentasaccharide 8 in Table 5; (b) pentasaccharide 9 in Table 5; (c) pentasaccharide 10 in Table 5; and (d) pentasaccharide 11 in Table 5. (Reprinted from Gupta, D.; Bhattacharyya, L.; Fant, J.; Macaluso, F.; Sabesan, S.; Brewer, C. F. *Biochemistry* **1994**, *33*, 7495. Copyright 1994 American Chemical Society.)

Table 9. Carbohydrates and Lectins in Cell−Cell Recognition

process	sugars on	lectins on
infection	host cells	microorganisms
defense	phagocytes	microorganisms
	microorganisms	phagocytes
fertilization	eggs	(sperm)[a]
leukocyte traffic	leukocytes	endothelial cells
	endothelial cells	lymphocytes
metastasis	target organs	malignant cells
	malignant cells	(target organs)[a]

[a] Presumed, no experimental evidence available.

V. Functions

The major function of lectins appears to be in cell recognition (Table 9). The basis of this recognition is the molecular fit between pairs of complementary structures on the surfaces of the interacting cells, or between those on a cell surface and on a molecule in solution, one carrying encoded biological information and the other capable of deciphering the code. This concept represents an extension of the lock-and-key hypothesis introduced by Emil Fischer at the end of last century to explain the specificity of interactions between enzymes and their substrates.

A. Theoretical Considerations

The idea that lectins and carbohydrates are eminently suitable to act in cell recognition evolved with the demonstration that both classes of compound are commonly present on cell surfaces and with the realization that carbohydrates have an enormous potential for encoding biological information. This potential derives from the fact that monosaccharides are multifunctional compounds, usually with three or four free hydroxyls, and can thus combine covalently with each other at different positions in the molecule and also form branched oligomers and polymers. In addition, the linkages may vary in anomeric configuration (α or β). As a result, two identical monosaccharides can form 11 different disaccharides, whereas two amino acids can make only one dipeptide. With increase in size, the difference in the number of isomeric oligosaccharides and nucleotides or polypeptides increases markedly and quickly reaches several orders of magnitude. Four different nucleotides can form only 24 distinct tetranucleotides, but four different monosaccharide can make 3.55×10^4 unique tetrasaccharides. A recent calculation[168] of the number of reducing hexasaccharides that can be formed from six different monosaccharides gave the astronomic number of 1.95×10^{12}, in contrast with the just 4.66×10^4 different hexapeptides obtainable from six different amino acids. Theoretically, therefore, carbohydrates can carry per unit weight much more information than nucleic acids and proteins, linear polymers based on a single type of linkage. Although only a small fraction of the possible carbohydrate structures has as yet been found in nature, their number continues to grow at a fast rate and is now already in the thousands. Also, clear evidence for the role of carbohydrates as information molecules has become available in many

Table 10. Inhibitors of Sugar-Specific Adhesion Prevent Infection in Vivo[a]

organism	animal, site[b]	inhibitor
Escherichia coli type 1	mice, UT	MeαMan
	mice, GIT	Mannose
	mice, UT	Anti-Man antibody
Klebsiella pneumoniae type 1	rats, UT	MeαMan
Shigella flexnerii type 1	guinea pigs, eye	Mannose
Escherichia coli type P	mice	Globotetraose
	monkeys	Gal(α1−4)GalβOMe
Escherichia coli K99	calves, GIT	Glycopeptides of serum glycoproteins
Pseudomonas aeruginosa	human, ear	Gal+Man+Neu5Ac

[a] Based on refs 170−172. [b] UT, urinary tract; GIT, gastrointestinal tract.

systems.[12,32] Since lectins are geared to distinguish between different oligosaccharides, they are admirably suited as decoders of the carbohydrate-encoded information.

B. Microbial Lectins

Viral and microbial surface lectins mediate the adhesion of the organisms to host cells, a prerequisite for infection to occur.[109,110,169] This was first demonstrated in the 1940s for the influenza virus hemagglutinin, a lectin specific for *N*-acetylneuraminic acid, a sialic acid. Removal of this sugar from the cell membranes by sialidase abolishes viral binding and prevents infection, while enzymatic reattachment of *N*-acetylneuraminic acid or insertion of sialic acid-containing oligosaccharides (for example, in the form of glycolipids) into the membranes of sialidase-treated cells restores the ability of the cells to bind the virus and to be infected.

Lectin−carbohydrate interactions also function in bacterial infections. An interesting illustration for the role of bacterial lectins (occasionally referred to as sugar-specific adhesins) in recognition of host cells by the bacteria is provided by *E. coli* K99. This organism binds to glycolipids containing *N*-glycolylneuraminic acid but not to those containing *N*-acetylneuraminic acid, two sugars that differ in a single hydroxyl group, present in the acyl substituent on the 4-OH group of the former compound and absent in the latter. *N*-Glycolylneuraminic acid is found on intestinal cells of newborn piglets, but it disappears when the animals develop and grow. It is also not formed normally by humans. This explains why *E. coli* K99 can cause diarrhea (often lethal) in piglets, but not in adult pigs nor in humans.

Compelling evidence for the role of lectins in bacterial infection derives from experiments in which blocking the lectins by suitable sugars provided protection against such infection[170−172] (Table 10). Thus, mannose and methyl α-mannoside inhibited specifically infection of the urinary tract of mice and rats by different strains of type 1 fimbriated *E. coli* and *Klebsiella pneumoniae*, respectively, while *N*-glycolylneuraminic acid containing glycopeptides, administered orally, protected colostrum-deprived, newborn calves against lethal doses of enterotoxigenic *E. coli* K99. Similarly, *N*-acetylneuraminic acid considerably reduced colonization of lung, liver, and kidney by *Pseudomonas aeruginosa* injected intravenously to the animal. Further, introduction of galactose into the trachea of rabbits infected with *Bordetella pertussis* prevented colonization of the respiratory tract by the bacteria and blocked pulmonary edema. In a clinical trial in humans, patients with otitis externa (a painful swelling with secretion from the external auditory canal caused by *P. aeruginosa*) were treated locally with a solution containing a mixture of galactose, mannose, and *N*-acetylneuraminic acid. The results were fully comparable to those obtained with local antibiotic treatment. Very recently, it was shown that oral adminstration of NeuAc(α2−3)lactose to patients infected with *Helicobacter pylori*, the etiologic agent in the development of gastroduodenal ulcers and gastric cancers, significantly decreased the gastric bacterial load.[173] These findings illustrate the great potential of carbohydrates in the prevention of infections caused by bacteria that express surface lectins and provide a basis for the development of antiadhesion therapy of microbial infections. New classes of drugs such as antiadhesives are in great need, because of the increased occurrence of pathogenic organisms that are resistant to conventional antibiotics.

Attachment of a pathogen to a body site does not, in and of itself, initiate a disease. It must be coupled to a specific response that leads to infection. In viral infections, attachment of the virus by its hemagglutinin to *N*-acetylneuraminic acid residues on the surface of the target cells is followed by fusion of the viral and cellular membranes, allowing release of the viral genome into the cytoplasm and subsequent replication. Studies with P-fimbriated *E. coli* (as well as purified P-type fimbriae) and uroepithelial cells suggested that such adherence induces a two-way flow of biological cross-talk via the lectin bridge, affecting both partners: the target cell is activated, with resultant production of cytokines that engender acute inflammation and symptoms of disease, while in the bacterium the interaction leads to up-regulation of two signal transduction systems that allow responses to the changing environment.[174,175]

Bacteria, e.g., *Klebsiella pneumoniae*, can attach by their surface lectins also to mast cells and such attachment results in activation of the target cells and production of high levels of certain cytokines, in particular TNF-α. Activation can also be induced by the purified type 1 fimbriae, as well as by the isolated carbohydrate binding subunit of the fimbriae.[176−178] The cytokines released by the activated mast cells cause rapid recruitment of neutrophils into the site

of infection, resulting in early clearance of the bacteria. This is supported by the finding that mast cell deficient mice were significantly less efficient in clearing enterobacteria in either an intranasal or intraperitoneal infection with pathogenic *K. pneumoniae*.

Earlier it has been shown that some bacterial surface lectins may allow the specific attachment of the bacteria to human polymorphonuclear cells and human and mouse macrophages, an interaction which usually is mediated by serum constituents termed opsonins (mainly IgG antibodies and fragments of the C3 component of complement) as part of the immune reponse of the body to the invasion by bacteria. In both cases attachment leads to activation of the phagocytes and ingestion and killing of the bacteria. Phagocytosis resulting from lectin-mediated binding has been named lectinophagocytosis, in analogy to opsonophagocytosis, in which recognition is mediated by the opsonins.[179] It refers also to phagocytosis resulting from the interaction between bacteria or other infectious agents and macrophages, which is mediated by a lectin on the, latter cells that binds to carbohydrates on the surface of the infectious agent.

Lectinophagocytosis may function as a defense mechanism against microbial infections in vivo at sites low in serum, such as the renal medulla and peritoneal cavity, especially during dialysis, or in situations where opsonic activity is poor, e.g., in patients infected by microorganisms before the development of an immune response. Evidence for the possible occurrence of lectinophagocytosis in vivo comes from animal experiments. Injection of type 1 fimbriated *E. coli* into the peritoneal cavity of mice led to the activation of the peritoneal macrophages; no activation was observed in the presence of methyl α-mannoside or when the nonfimbriated bacteria were used.

C. Plant Lectins

Although known longer that lectins from other organisms, and also more extensively studied, the role of plant lectins is still not well understood. Of the many theories proposed, only two are currently under serious consideration. The first assumes that lectins function in the establishment of symbiosis between nitrogen-fixing bacteria, mainly rhizobia, and leguminous plants, a process of cardinal importance in both the nitrogen cycle of terrestrial life and in agriculture. However, it can account only for the role of lectins from one plant family. The other theory is more general and proposes that plant lectins are defense agents against different kinds of predatory invertebrates and higher animals, as well as against phytopathogenic fungi.[10,180]

The ability of legumes to associate specifically and form symbiosis with soil bacteria of the rhizobia family, thus making them independent from soil nitrogen supplies, has long been an intriguing phenomenon. When rhizobia encounter root hairs in the soil, several profound developmental events take place in the infected roots. The mechanism whereby most of these events occur is still obscure, but it is believed that invasion into the root cells requires a highly specific association between the bacteria and the root hair surface. Thus, rhizobia that infect and nodulate soybeans cannot nodulate garden peas or white clover, and vice versa. The idea that lectins are responsible for this association was advanced over 20 years ago. It was based on the finding that a lectin from a particular legume binds in a carbohydrate-specific manner to the corresponding rhizobial species and not to bacteria that are symbionts of other legumes. The suggestion was therefore made that rhizobial attachment to plant roots occurs by interaction between the rhizobial surface carbohydrates and lectins present in the roots of the legume plants. This became known as the "lectin recognition hypothesis". From the start, however, this hypothesis became the subject of controversy, mainly because of the lack of unequivocal evidence and many inconsistencies.[10] For most host−symbiont systems, there is no proof for the presence of lectins and their receptors on plant roots and bacteria, respectively, at precisely the right time and location. Furthermore, the correlation between the carbohydrate specificity of the host lectin and its ability to recognize nodulating bacteria specific for that host is not very strict. For instance, heterologous rhizobia adhere to pea roots equally well as does its symbiont, *R. leguminosum*, and sugars for which the pea lectin is specific do not inhibit the adhesion of this symbiont to pea root hairs. Also, several mutants of soybeans were found that lack the seed lectin, but they all are 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.[180,181] Transformation of clover plants with the gene of the pea lectin conferred upon their roots the ability to be nodulated by the pea specific rhizobia. In the roots of the transformed clover, pea lectin appeared to be present at sites similar to those on pea roots. Furthermore, transfection with a pea lectin gene in which a key combining site residue (Asn125) was mutated so that the lectin lost its ability to combine with sugars, did not render the clover roots susceptible to infection by the pea rhizobia.[182] In an extension of these studies, *Lotus* and alfalfa plants were transfected with the soybean agglutinin gene.[181] However, while the transgenic *Lotus* plants responded to *Bradyrhizobium japonicum*, the symbiont for soybean, the transgenic alfalfa did not.

The toxicity of various plant lectins for animals and their growth inhibitory effect on fungi are the basis for the assumption that they function in the defense of plants against phytopathogenic fungi and predatory animals. This assumption has received considerable attention during the last two decades, but it still lacks definite proof.[180−183] Much of the information about the toxic effects of plant lectins on animals comes from feeding experiments with PHA and accidental poisoning of humans by raw or insufficiently cooked beans.[184−186] Ingested PHA binds to the brush border cells of the intestine, where it is

rapidly endocytosed. Upon entering the cells, the lectin enhances their metabolic activity which eventually leads to hyperplasia and hyperthrophy of the small intestine. Moreover, ingestion of PHA or raw beans causes acute nausea followed by vomiting and diarrhea. The discomfort is so severe that experimental animals are very reluctant to consume a diet containing PHA, and in some instances they rather starve. The bark lectins of black locust (*Robinia pseudoacacia*) and elderberry (*Sambucus nigra*), among others, provoke similar toxic effects. These examples illustrate the potential of lectins in protection against predators. Indeed, since both bark lectins mentioned are abundant, elderberry and black locust are never attacked by rodents, deer, or other wildlife, whereas the bark of lectin-free plants, e.g., poplar, willow, and wild apple, is a favorite food for the same animals.

D. Animal Lectins

1. Galectins

The galectins are believed to function in cell adhesion. These lectins are found both inside the cytoplasm and the nucleus of different cells and occasionally also on the cell surface and outside the cell. Their expression is developmentally regulated, i.e., their synthesis in a given tissue takes place only during particular developmental or physiological stages.[187] They are therefore thought to be essential for the normal development and differentiation of all multicellular animals. For instance, galectin-1 is prominently expressed at early stages of embryonal development. However, "knock out" mice, i.e., genetically engineered animals in which the gene for galectin-1 has been inactivated, showed no apparent abnormality.[188] These results suggest that animals may maintain some compensation system in case one important gene is not functional. The elevated levels of galectin-3 present on the surface of metastatic murine and human cancer cells may be responsible for the adhesion of the cells to target organs, a step necessary for metastasis.[189,190] For instance, a good correlation was found between the amount of the lectin expressed on melanoma cells and the formation of pulmonory metastases after injection of the cells into syngeneic mice. Exposing highly metastatic cells to lactose or its derivatives before injecting them into the mice reduced the metastatic spread almost by half. Therefore, antiadhesive drugs may turn out to be antimetastatic.[171]

2. C-Type Lectins

a. Endocytic Lectins. The endocytic lectins, a class of C-type lectins, are membrane-bound receptors with different specificities (Table 11). In experimental animals, the mammalian hepatic asialoglycoprotein receptor, specific for galactose and *N*-acetylgalactosamine, was shown to facilitate the clearance from the circulation of glycoproteins with complex oligosaccharide units (e.g., ceruloplasmin and α_1-acid glycoprotein) from which the terminal sialic acid has been removed, exposing subterminal galactose. These findings were traditionally interpreted as representing a physiological mechanism for regulating the turnover of serum glycoproteins (and cells).[78] This is however far from certain.[191] Genetically engineered mice that lack the ability to synthesize the receptor do not exhibit increased levels of desialylated forms of predominant circulating glycoproteins.[192] It is possible that the natural ligands for the hepatic lectin are minor species of desialylated serum glycoproteins or perhaps glycoproteins that bear terminal *N*-acetylgalactosamine, for which it has a 60-fold higher affinity that for galactose.

Table 11. Clearance and Targeting of Glycoproteins

	endocytic lectin (receptor)	
glycoprotein	specificity	location
asialoglycoproteins	galactose	liver (hepatocytes)
diverse	fucose	liver (Kupffer cells)
hormones	SO_4-GalNAc	liver (Kupffer cells, endothelial cells)
lysosomal enzymes	Man−6-phosphate	ubiquitous
diverse	mannose	macrophages, liver (endothelial cells)

The mannose-specific receptor present on the surface of macrophages has been implicated in antimicrobial defense. It binds infectious organisms that expose mannose-containing glycans on their surface, leading to their ingestion and killing by lectinophagocytosis[3,83,179] (cf. section V.B). This type of defense, which does not depend on antibodies to the infectious agents, is known as innate immunity.

b. Collectins. A similar function, albeit by a different mechanism, is performed by the MBP's of mammalian serum and liver.[3,85,86] These lectins 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. As mentioned (cf. section IV.B), the elucidation of the spatial arrangement of the subunits in the lectin trimer provides a structural basis for understanding the fact that the mannose-binding protein can target the innate immune response to pathogens such as bacteria and yeast, while avoiding the endogenous sugars of the cell surfaces and soluble mannose-containing glycoproteins of the host. It is likely that the oligo- and polymannosides found in the outer walls of yeasts, as well as the *N*-acetylglucosamine-containing polymers in many bacterial walls, serve as targets for mannose-binding proteins because they present monosaccharides at regularly repeating intervals, suitable for interacting with widely spaced binding sites in the lectin.

The soluble MBPs also enhance phagocytosis of the invading organisms by acting as opsonins, thus bypassing the need for an antibody-binding step. Clinical evidence for their importance has come from the identification of a MBP deficiency syndrome. It is caused by a mutation of a single amino acid in the collagen-like domain of the lectin and results in recurrent, severe bacterial infections in infants.[3]

c. Selectins. The selectins, another family of C-type lectins, provide the best paradigm for the role of sugar−lectin interactions in biological recognition. The selectins mediate the adhesion of circulating

leukocytes to endothelial cells of blood vessels, a prerequisite for the exit of the former cells from the circulation and their migration into tissues. They thus control leukocyte trafficking to sites of inflammation and the migration (homing) of lymphocytes to specific lymphoid organs.[93−97] L-selectin, also known as "homing receptor", is found on all leukocytes and is involved in the recirculation of lymphocytes, directing them specifically to peripheral lymph nodes. The two other selectins, E and P, are expressed on endothelial cells only when these cells are activated by inflammatory mediators (e.g., inteleukin-2 and tumor necrosis factor) released from tissue cells in response to, e.g., wounding, infection, or ischemia. Genetically engineered mice which lack both E- and P-selectin are susceptible to severe and ultimately fatal bacterial infections, highlighting the role of the lectins in acute inflammatory responses.[193] The clinical relevance of selectin−carbohydrate interactions in such responses in humans is illustrated by the finding that the neutrophils of two patients with recurrent bacterial infections had a deficiency in sLe^x and were unable to bind to E-selectin.[194] The specific biochemical lesion responsible for these defects is believed to be a reflection of a general deficiency in the metabolism of fucose, which is an essential part of the selectin ligands sLe^x and sLe^a.

In cases of microbial infection, the selectin-mediated adhesion of leukocytes to activated endothelial cells is beneficial, since it is a major factor in clearance of the infectious agents. However in other situations, such adhesion may lead to the harmful accumulation of leukocytes, causing tissue damage, swelling, and pain, for instance the inflammation of rheumatoid arthritis or myocardial injury during reperfusion of an ischemic heart. Prevention of adverse inflammatory reactions by inhibition of leukocyte−endothelium interactions, another application of antiadhesion therapy, has become a major aim of the biomedical and pharmacological industry.[195] As shown in animal models, oligosaccharides recognized by the selectins protect against experimentally induced lung injury. Application of the selectin ligand sLe^x, or of anti-P- or anti-L-selectin antibodies, attenuates myocardial necrosis after myocardial ischemia and reperfusion. In addition to their involvement in inflammation, selectins may play a role in the spread of cancer cells from the main tumor throughout the body. Inhibition of the selectins, as of the galectins (cf. section V.D.1) may therefore be useful in the prevention of metastases in humans.

3. P-Type Lectins

The function of the mannose 6-phosphate receptors is well established: these lectins serve for targeting lysosomal enzymes to their subcellular compartment.[100,101,196] The targeting is mediated by the recognition between Man-6-P attached to the oligomannose unit(s) of such enzymes, and the above receptor(s). A defect in the synthesis of the Man-6-P marker results in I-cell disease (also called mucolipidosis II or MLII), an inherited lysosomal storage disease, characterized by a lack in the lysosome of those enzymes that normally carry the marker and resulting in intracellular accumulation of undigested glycoconjugates.[197] It is caused by a deficiency of GlcNAc-phosphotransferase, the first enzyme in the pathway of mannose phosphorylation. Therefore, even though the disease is transmitted by a single gene, some 20 enzymes are affected. The enzymes lacking the recognition marker do not reach their destination (the lysosomes) and are secreted into the extracellular milieu, which is one of the biochemical abnormalities of the affected cells. The importance of the two mannose 6-phosphate receptors in the routing of lysosomal enzymes was further demonstrated by "knock out" experiments with fibroblasts in which the genes for both receptors have been disrupted.[101,196] Such cells secrete their lysosomal enzymes and accumulate undigested material, similar to fibroblasts from patients with the I-cell disease.

4. I-Type Lectins

I-type lectins have been implicated in cell−cell interactions: sialoadhesin and CD22 in those of the immune system and MAG in the maintenance of myelin and in neuronal regeneration[105] (Figure 45).

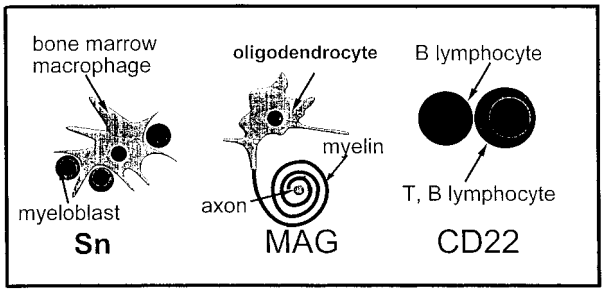


Figure 45. Sialic acid-dependent cellular interactions mediated by sialoadhesins. (Courtesy of Dr. S. Kelm, Biochemistry Institute, University of Kiel, Germany.)

5. Other Lectins

The recently discovered intracellular lectins calnexin, calreticulin, MR60/ERGIC-53, and VIP-36 play a role in the biosynthesis of glycoproteins as well as in their intracellular sorting, trafficking, and delivery to the cell surface.[40] Calnexin, a nonglycosylated membrane protein of 65.4 kDa molecular weight, is located intracellularly, in the endoplasmic reticulum, where biosynthesis of glycoproteins is initiated.[198,199] It binds transiently to terminal glucose on oligomannose units of newly formed glycoproteins and controls their folding and oligomer assembly, thus acting as a chaperone. Calreticulin, the soluble analogue of calnexin,[200] acts in a similar manner. MR60/ERGIC-53 is a mannose-specific lectin that shuttles between two intracellular compartments−the Golgi apparatus and the endoplasmic reticulum−through which glycoproteins pass at various stages of their biosynthesis. It has been postulated that this lectin carries nascent glycoproteins between these compartments.[40]

VI. Applications

Native lectins are used predominantly for applications that are based on precipitation and agglutina-

tion reactions or for mitogenic stimulation of lymphocytes (see below). For numerous purposes, however, lectin derivatives are required. Thus, lectins derivatized with fluorescent dyes, gold particles, or enzymes are employed as histochemical and cytochemical reagents for detection of glycoconjugates in tissue sections, on cells and subcellular organelles, and in investigations of intracellular pathways of protein glycosylation.[201] Lectin binding has been used to demonstrate that membrane receptors for hormones, growth factors, neurotransmitters and toxins are glycoconjugates. Immobilized lectins, such as those that are covalently bound to Sepharose, are indispensable for the purification and isolation by affinity chromatography of glycoproteins, glycopeptides, and oligosaccharides.[202−205]

Mouse and human cortical (immature) thymocytes differ markedly from the medullary (mature) ones in their surface carbohydrates, as evidenced by the fact that the former are bound and agglutinated by peanut agglutinin (PNA^+ cells), whereas the latter are not (PNA^-).[206] Separation with peanut agglutinin provides facile access to the individual thymocyte subpopulations and makes it possible to examine in vitro their developmental and functional relationship.

Selective agglutination by SBA permits separation of B and T mouse splenocytes. The main application of this lectin is for purging human bone marrow for transplantation.[207] It is employed routinely for transplantations into children born with severe combined immune deficiency ("bubble children", since they are highly susceptible to microbial infections and have to be kept all the time in a plastic bubble) with close to 70% success. SBA purging is also used experimentally in bone marrow transplantation of leukemic patients, as an alternative to other accepted techniques for T cell depletion, such as monoclonal antibodies.

Another clinical application of lectins is in blood typing.[208] Thus, the lectins from *Lotus tetragonolobus* and *Ulex europaeus*, both specific for fucose, are employed to identify blood type O cells, and for the identification of secretors of blood group substances. The lectin from *Dolichos biflorus* is used to distinguish between A_1 and A_2 subgroups and that from *Vicia graminea*, specific for blood type N, to differentiate between M and N cells. In addition PNA, specific for Gal(β1−3)GalNAc (see Table 1), is employed in the detection of "polyagglutination" (or "polyagglutinability"), a condition accompanying certain bacterial and viral infections, in which human erythrocytes become agglutinable by antibodies normally present in the sera of nearly all adults. If not diagnosed in time, it may lead to serious complications and death.

Certain lectins are potent mitogens, activating lymphocytes and inducing them to divide; PHA and concanavalin A, for example, stimulate T lymphocytes, while pokeweed mitogen (PWM) stimulates both T and B cells.[209,210] The mitogenic lectins are polyclonal activators, in that they activate lymphocytes irrespective of their antigenic specificity. Prior to the advent of monoclonal antibodies to cell surface antigens, lectins were the major tool for studies of the mechanism of cell activation. Mitogenic stimulation by lectins provides an easy and simple means to assess the immunocompetence of patients suffering from a diversity of diseases, including AIDS, and to monitor the effects of various immunosuppressive and immunotherapeutic manipulations. It has been used to examine the effect of stress, both physical and psychological, on the immune system, e.g., sport, weightlessness in space, bereavement. It is also employed for the preparation of chromosome maps for different purposes, such as karyotyping, sex determination, and detection of chromosome defects, since chromosomes are easily visualized in the stimulated cells.

VII. Acknowledgments

Our sincere thanks to Dr. Miriam Eisenstein from the Department of Structural Biology for the preparation of the computer drawings.

VIII. References

(1) (1) Liener, I. E., Sharon, N., Goldstein, I. J., Eds.; *The Lectins: Properties, Functions and Applications in Biology and Medicine*; Academic Press, Inc.: Orlando, 1986; 600 p.
(2) Sharon, N.; Lis, H. *Lectins*; Chapman and Hall: London, 1989; 127 pp.
(3) Drickamer, K.; Taylor, M. E. *Annu. Rev. Cell Biol.* **1993**, *9*, 237.
(4) Sharon, N.; Lis, H. *Essays Biochem.* **1995**, *30*, 59.
(5) Kilpatrick, D. C., van Driessche, E., Bøg-Hansen T. C., Eds. *Lectin Reviews*; Sigma Chemical Co.: St. Louis, MO., 1991; Vol. 1, 218 pp.
(6) Sharon, N.; Lis, H. *Trends Biochem. Sci.* **1987**, *12*, 488−491.
(7) Kocourek, J. In *The Lectins: Properties, Functions and Applications in Biology and Medicine*; Liener, I. E., Sharon, N., Goldstein, I. J., Eds.; Academic Press, Inc.: Orlando, 1986; p 1.
(8) Goldstein, I. J.; Winter, H. C.; Poretz, R. D. In *Glycoproteins II*; Vliegenthart, J. F. G., Montreuil, J., Schachter, H., Eds.; Elsevier Science B.V.: Amsterdam, 1997; p 403.
(9) Lis, H.; Sharon, N. In *The Lectins: Properties, Functions and Applications in Biology and Medicine*; Liener, I. E., Sharon, N., Goldstein, I. J., Eds.; Academic Press, Inc.: Orlando, 1986; p 293.
(10) Sharon, N.; Lis, H. *Science* **1989**, *246*, 227.
(11) Sharon, N.; Lis, H. *Sci. Am.* **1993**, *268* (1), 82.
(12) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683.
(13) Rini, J. M. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 551.
(14) Weis W.; Drickamer K. *Annu. Rev. Biochem.* **1996**, *65*, 441.
(15) Cambillau, C. In *Glycoproteins I*; Montreuil, J., Schachter, H., Vliegenthart, J. F. G., Eds.; Elsevier Science B.V.: Amsterdam, 1995; p 29.
(16) Gabius, H.-J. *Eur. J. Biol.* **1997**, *243*, 543.
(17) Sharon, N.; Lis, H. In *Glycoproteins II*; Vliegenthart, J. F. G., Montreuil, J., Schachter, H., Eds.; Elsevier Science B.V.: Amsterdam, 1997; p 475.
(18) Loris, R.; Bouckaert, J.; Hamelryck, T.; Wyns, L. *Biochim. Biophys. Acta* **1998**, in press
(19) Burnette, W. N. *Structure* **1994**, *2*, 151.
(20) Emsley, J.; White, H. E.; O'Hara, B. P.; Oliva, G.; Srinivasan, N.; Tickle, I. J.; Blundell, T.; Pepys, M. B.; Wood, S. P. *Nature* **1994**, *367*, 338.
(21) Goldstein, I. J.; Poretz, R. D. In *The Lectins: Properties, Functions and Applications in Biology and Medicine*; Liener, I. E., Sharon, N., Goldstein, I. J., Eds.; Academic Press, Inc.: Orlando, 1986; p 35.
(22) Lis, H.; Sharon, N. In *Biology of Carbohydrates*; Ginsburg, V., Robbins, P. W., Eds.; John Wiley & Sons: New York, 1984; Vol. 2. p 1.
(23) Oldenberg, K. R.; Loganathan, D.; Goldstein, I. J.; Schultz, P. G.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5393.
(24) Scott, J. K.; Loganathan, D.; Easley, R. B.; Gong, X.; Goldstein, I. J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5398.
(25) Kaur, K. J.; Khurana, S.; Salunke, D. M. *J. Biol. Chem.* **1997**, *272*, 5539.
(26) Carver, J. P. *Pure Appl. Chem.* **1993**, *65*, 763.
(27) Bush, A. C. *Curr. Opin. Struct. Biol.* **1992**, *2*, 655.
(28) Homans, S. W. In *Molecular Glycobiology*; Fukuda, M., Hindsgaul, O., Eds.; Oxford University Press: Oxford, 1994; p 230.

(29) Homans, S. W.; Rutheford, T. *Biochem. Soc. Trans.* **1993**, *21*, 449.
(30) Imberty, A.; Bourne, Y.; Cambillau, C, Rougé, P.; Perez, S. *Adv. Biophys. Chem.* **1993**, *3*, 61.
(31) Rice, K. G.; Wu, P. Brand, L.; Lee, Y. C. *Curr. Opin. Struct. Biol.* **1993**, *3*, 669.
(32) Lis, H.; Sharon, N. *Eur. J. Biochem.* **1993**, *218*, 1.
(33) Sharon, N.; Lis, H. In *Glycosciences-Status and Perspectives*; Gabius, H., Gabius, S., Eds.; Chapman & Hall; Weinheim, 1966; p 134.
(34) Sharon, N.; Lis, H. *FASEB J.* **1990**, *4*, 3198.
(35) Konami, Y.; Yamamoto, K.; Osawa, T.; Irimura, T. *Glycoconjugate J.* **1995**, *12*, 128.
(36) Pusztai, A. *Plant Lectins*; University Press: Cambridge, 1991; 263 pp.
(37) Emmerich, C.; Helliwell, J. R.; Redshaw, M.; Naismith, J. H.; Harrop, S. J.; Raftery, J.; Kalb (Gilboa), J. A.; Yariv, J.; Dauter, Z.; Wilson, K. S. *Acta Crystallogr.* **1994**, *D50*, 749.
(38) Min, W.; Jones, D. H. *Nature Struct. Biol.* **1994**, *1*, 502.
(39) Fiedler, K.; Simons, K. *Cell* **1994**, *77*, 625.
(40) Roche, A.-C.; Monsigny, M. *Chemtracts Biochem. Mol. Biol.* **1996**, *6*, 188.
(41) Imberty, A.; Casset, F.; Gegg, C. V.; Etzler, M. E.; Perez, S. *Glycoconjugate J.* **1994**, *11*, 400.
(42) Srinivasan, N.; Rufino, S. D.; Pepys, M. B.; Wood, S.; Blundell, T. L. *Chemtracts Biochem. Mol. Biol.* **1996**, *6*, 149.
(43) Loris, R.; Stas, P. P. G.; Wyns, L. *J. Biol Chem.* **1994**, *269*, 26722.
(44) Bouckaert, J.; Loris, R.; Poortmans, F.; Wyns, L. *Proteins* **1995**, *23*, 510.
(45) Shaanan, B.; Lis, H.; Sharon, N. *Science* **1991**, *254*, 862.
(46) Delbaere, L. T. J.; Vandonselaar, M.; Prasad, L.; Quail, J. W.; Nikrad, P. V.; Pearlstone, J. R.; Carpenter, M. R.; Smillie, L. B.; Spohr, U.; Lemieux, U. *Can. J. Chem.* **1990**, *68*, 1116.
(47) Banerjee, R.; Das, K.; Ravishankar, R.; Suguna, K.; Surolia, A.; Vijayan, M. J. *Mol. Biol.* **1996**, *259*, 281.
(48) Dessen, A.; Gupta, D.; Sabesan, S.; Brewer, C. F.; Sacchettini, J. C. *Biochemistry* **1995**, *34*, 4933.
(49) Hamelryck, T. W.; Dao-Thi, M.-H.; Poortmans, F.; Chrispeels, M. J.; Wyns, L.; Loris, R. *J. Biol. Chem.* **1996**, *271*, 20479.
(50) Wright, C. S. *J. Mol. Biol.* **1987**, *194*, 501.
(51) Wright, C. S.; Kellog, G. E. *Protein Sci.* **1996**, *5*, 1466.
(52) Van Damme, E. J. M.; Goldstein, I. J.; Peumans, W. J. *Phytochemistry* **1991**, *30*, 509.
(53) Van Damme, E. J. M.; Balzarini, J.; Smeets, K.; van Leuven, F.; Peumans, W. J. *Glycoconjugate J.* **1994**, *11*, 321.
(54) Chervenak, M. C.; Toone, E. J. *Biochemistry* **1995**, *34*, 5685.
(55) Hester, G.; Kaku, H.; Goldstein, I. J.; Wright, C. S. *Nature Struct. Biol.* **1995**, *2*, 472.
(56) Wright, C. S.; Hester, G. *Structure* **1996**, *4*, 1339.
(57) Kabir, S.; Daar, A. S. *Immunol. Invest.* **1994**, *23*, 167.
(58) Sankaranarayanan, R.; Sekar, K.; Banerjee, R.; Sharma, V.; Surolia, A.; Vijayan, M. *Nature Struct. Biol.* **1996**, *3*, 596.
(59) Rutenber, E.; Katzin, B. J.; Collins, E. J.; Mlsna, D.; Ernst, S. E.; Ready, M. P.; Robertus, J. D. *Proteins* **1991**, *10*, 240.
(60) Rutenber, E.; Robertus, J. D. *Proteins* **1991**, *10*, 260.
(61) Sweeney, E. C.; Palmer, R. A.; Tonevitsky. A. G.; Temiakov, D. E. *Proteins* **1997**, *28*, 586.
(62) Barondes, S. H.; Cooper, D. N. W.; Gitt, M. A.; Leffler, H. *J. Biol. Chem.* **1994**, *269*, 20807.
(63) Kasai, K.; Hirabayashi, J. *J. Biochem.* **1996**, *119*, 1.
(64) Leffler, H. *Trends Glycosci. Glycotech.* **1997**, *9*, 9.
(65) Ahmed, H.; Pohl, J.; Fink, N. E.; Strobel, F.; Vasta, G. R. *J. Biol. Chem.* **1996**, *271*, 33083.
(66) Wagner-Hülsmann, C.; Bachinski, N.; Diehl-Seifert, B.; Blumbach, B.; Steffen, R.; Pancer, Z.; Müller, E. G. *Glycobiology* **1996**, *6*, 785.
(67) Teichberg, V. I.; Silman, I.; Beitsch, D. D.; Resheff, G. *Proc. Natl. Accad. Sci. U.S.A.* **1975**, *72*, 1383.
(68) Abbott, W. M.; Feizi, T. *J. Biol. Chem.* **1991**, *266*, 5552.
(69) Lobsanov, Y. D.; Gitt, M. A.; Leffler, H.; Barondes, S. H.; Rini, J. M. *J. Biol. Chem.* **1993**, *268*, 27034.
(70) Liao, D.-I.; Kapadia, G.; Ahmed, H.; Vasta, G. R.; Herzberg, O. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1428.
(71) Srinivasan, N.; White, H.; Emsley, J.; Wood, S. P.; Pepys, M. B.; Blundell, T. L. *Structure* **1994**, *2*, 1017.
(72) Wiley, D. C.; Skehel, J. J. *Annu. Rev. Biochem.* **1987**, *56*, 365.
(73) Sauter, N. K.; Hanson, J. E.; Glick, G. D.; Brown, J. H.; Crowther, R. L.; Park, S.-J.; Skehel, J. J.; Wiley, D. C. *Biochemistry* **1992**, *31*, 9609.
(74) Rayment, I.; Baker, T. S.; Caspar, D. L. D.; Murkami, W. T. *Nature* **1982**, *295*, 110.
(75) Stehle, T.; Harrison, S. C. *EMBO J.* **1997**, *16*, 5139.
(76) Drickamer, K. *J. Biol. Chem.* **1988**, *263*, 9557.
(77) Drickamer, K. In *Molecular Glycobiology*; Fukuda, M., Hindsgaul, O., Eds.; Oxford University Press: Oxford, 1994; p 53.
(78) Ashwell, G., Harford, J. *Annu. Rev. Biochem.* **1982**, *51*, 531.
(79) Spiess, M. *Biochemistry* **1990**, *29*, 10008.
(80) Stockert, R. J. *Physiol. Rev.* **1995**, *75*, 591.
(81) Stahl, P. D. *Curr. Opin. Immunol.* **1992**, *4*, 49.
(82) Taylor, M. E.; Conary, J. T.; Lennartz, M. R.; Stahl, P. D.; Drickamer, K. *J. Biol. Chem.* **1990**, *265*, 12156.
(83) Ezekowitz, R. A. B.; Sastry, K.; Bailly, P.; Warner, A. *J. Exp. Med.* **1990**, *172*, 1785.
(84) Drickamer, K. *Curr. Opin. Struct. Biol.* **1995**, *5*, 612.
(85) Epstein, J.; Eichbaum, Q.; Sheriff, S.; Ezekowitz, R. A. B. *Curr. Opin. Immunol.* **1996**, *8*, 29.
(86) Holmskov, U.; Malhotra, R.; Sim, R. B.; Jensenius, J. C. *Immunol. Today* **1994**, *15*, 67.
(87) Weis, W. I.; Drickamer, K. *Structure* **1994**, *2*, 1227.
(88) Sheriff, S.; Chang, C. Y.; Ezekowitz, R. A. B. *Nature Struct. Biol.* **1994**, *1*, 789.
(89) Ng, K. K.-S.; Drickamer, K.; Weis, W. I. *J. Biol. Chem.* **1996**, *271*, 663.
(90) Weis, W. I.; Kahn, R.; Fourme R.; Drickamer, K.; Hendrickson, W. A. *Science* **1991**, *254*, 1608.
(91) Quesenberry, M. S.; Lee, R. T.; Lee Y. C. *Biochemistry* **1997**, *36*, 2724.
(92) Lee, R. T.; Lee, Y. C. *Glycoconjugate J.* **1997**, *14*, 357.
(93) Rosen, D. D.; Bertozzi, C. R. *Curr. Opin. Cell Biol.* **1994**, *6*, 663.
(94) Lasky, L. A. *Annu. Rev. Biochem.* **1995**, *64*, 113.
(95) Springer, T. A. *Annu. Rev. Physiol.* **1995**, *57*, 827.
(96) McEver, R. P.; Moore, K. L.; Cummings, R. D. *J. Biol. Chem.* **1995**, *270*, 11025.
(97) Nelson, R. M.; Venot, A.; Bevilacqua, M. P.; Linhardt, R. J. Stamenkovic, I. *Annu. Rev. Cell Dev. Biol.* **1995**, *11*, 601.
(98) Graves, B. J.; Crowdther, R. L.; Chandran, C. Rumberger, J. M.; Li, S.; Huang, K.-S.; Presky, D. H.; Familletti, P. C.; Wolitzky, B. A.; Burns, D. K. *Nature* **1994**, *367*, 532.
(99) O'Connell, D.; Koenig, A.; Jennings, S.; Hicke, B.; Han, H. L.; Fitzwater, T.; Chang, Y. F.; Varki, N.; Parma, D.; Varki, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5883.
(100) Kornfeld, S. *Annu. Rev. Biochem.* **1992**, *61*, 307.
(101) Ludwig, T.; Le Borgne, R.; Hoflack, B. *Trends Cell Biol.* **1995**, *5*, 202.
(102) Powell, L. D.; Varki, A. *J. Biol. Chem.* **1995**, *270*, 14243.
(103) Varki, A. *Curr. Opin. Cell Biol.* **1992**, *4*, 257.
(104) Kelm, S.; Schauer, R.; Crocker, P. R. *Glycoconjugate J.* **1996**, *13*, 913.
(105) Crocker, P. R.; Kelm, S.; Hartnell, A.; Freeman, S., Nath, D.; Vinson, M.; Mucklow, S. *Biochem. Soc. Trans.* **1996**, *24*, 150.
(106) Nath, D.; Van der Merwe, P. A.; Kelm, S.; Bradfield, P.; Crocker, P. R. *J. Biol. Chem.* **1995**, *270*, 26184.
(107) Vinson, M.; van der Merwe, P. A.; Kelm, S.; May, A.; Jones, E. Y. Crocker, P. R. *J. Biol. Chem.* **1996**, *271*, 9267.
(108) Van der Merwe, P. A.; Crocker, P. R.; Vinson, M.; Barclay, A. N.; Schauer, R.; Kelm, S. *J. Biol. Chem.* **1996**, *271*, 9273.
(109) Ofek, I.; Sharon, N. *Curr. Top. Microbiol. Immunol.* **1990**, *152*, 91.
(110) Ofek, I.; Doyle, R. J. *Bacterial Adhesion to Cells and Tissues*; Chapman and Hall: London 1994; 578 pp.
(111) Gaastra, W.; Svennerholm, A.-M. *Trends Microbiol.* **1996**, *4*, 444.
(112) Quiocho, F. A. *Pure Appl. Chem.* **1989**, *61*, 1293.
(113) Sharon, N.; Lis, H. *Chem. Brit.* **1990**, *26*, 679.
(114) Quiocho, F. A. *Biochem. Soc. Trans.* **1993**, *21*, 442.
(115) Grootenhuis, P. D. J.; van Boeckel, C. A. A. *J. Am. Chem. Soc.* **1991**, *113*, 2743.
(116) Lemieux, R. U. *Acc. Chem. Res.* **1996**, *29*, 373.
(117) Sharon, N. *Trends Biochem. Sci.* **1993**, *18*, 221.
(118) Adar, R.; Sharon, N. *Eur. J. Biochem.* **1996**, *239*, 668.
(119) van Eijsden, R. R.; De Pater, B. S.; Kijne, J. W. *Glycoconjugate J.* **1994**, *11*, 375.
(120) Zhu, K.; Bressan, R. A.; Hasegawa, P. M.; Murdock, L. L. *FEBS Lett.* **1996**, *390*, 271
(121) Itin, C.; Roche, A.-C.; Monsigny, M.; Hauri, H.-P. *Mol. Biol. Cell* **1996**, *7*, 483.
(122) Derevenda, Z.; Yariv, J.; Helliwell, J. R.; Kalb, A. J.; Dodson, E. J.; Papiz, M. Z.; Wan, T.; Campbell, J. *EMBO J.* **1989**, *8*, 2189.
(123) Naismith, J. H.; Emmerich, C.; Habash, J.; Harop, S. J.; Helliwell, J. R.; Hunter, W. N.; Raftery, J.; Kalb (Gilboa), A. J.; Yariv, J. *Acta Crystallogr.* **1994**, *D50*, 847.
(124) Loris, R.; Casset, F.; Bouckaert, J.; Pletinckx, J.; Dao-Thi, M.-H.; Poortmans, F.; Imberty, A.; Perez, S.; Wyns, L. *Glyconjugate J.* **1994**, *11*, 507.
(125) Bourne, Y.; Roussel, A.; Frey, M. Rougé, P.; Fonntecilla-Camps, J.-P.; Cambillau, C. *Proteins* **1990**, *8*, 365.
(126) Sharma, V.; Surolia, A. *J. Mol. Biol.* **1997**, *267*, 433.
(127) Sharma, V.; Vijayan, M.; Surolia, A. *J. Biol. Chem.* **1996**, *271*, 21209.
(128) Ravishankar, R.; Ravindran, N.; Suguna, A.; Surolia, A.; Vijayan, M. *Curr. Science* **1997**, *72*, 855.
(129) Naismith, J. H.; Field, R. A. *J. Biol. Chem.* **1996**, *271*, 972.
(130) Mandal, D. K.; Kishore, N. Brewer, C. F. *Biochemistry* **1994**, *33*, 3, 1149.
(131) Gupta, D.; Dam, T. K.; Oscarson, S.; Brewer, C. F. *J. Biol. Chem.* **1997**, *272*, 6388.
(132) Bourne, Y.; Rougé, P. Cambillau, C. *J. Biol. Chem.* **1990**, *265*, 18161.

(133) Bourne, Y.; Rougé, P.; Cambillau, C. *J. Biol. Chem.* **1992**, *267*, 197.
(134) Bourne, Y.; Mazurier, J.; Legrand, D.; Rougé, P.; Montreuil, J.; Spik, G.; Cambillau, C. *Structure* **1994**, *2*, 209.
(135) Wright, C. S. *J. Mol. Biol.* **1990**, *215*, 635.
(136) Wright, C. S. *J. Biol. Chem.* **1992**, *267*, 14345.
(137) Wright, C. S.; Jaeger, J. *J. Mol. Biol.* **1993**, *232*, 620.
(138) Bains, G.; Lee, R. T.; Lee, Y. C.; Freire, E. *Biochemistry* **1992**, *31*, 12624.
(139) Weis, W. I.; Brown, J. H.; Cusack, S.; Paulson, J. C.; Skehel, J. J.; Wiley, D. C. *Nature* **1988**, *333*, 426.
(140) Kelm, S.; Paulson, J. C.; Rose, U.; Brossmer, R.; Schmid, W.; Bandgar, B. P.; Schreiner, E.; Hartmann, M.; Zbiral, E. *Eur. J. Biochem.* **1992**, *205*, 147.
(141) Watowich, S. J.; Skehel, J. J.; Wiley, D. C. *Structure* **1994**, *2*, 719.
(142) Freund, R.; Garcea, R. L.; Sahli, R.; Benjamin, T. L. *J. Virol.* **1991**, *65*, 350.
(143) Stehle, T.; Yan, Y.; Benjamin, T. L.; Harrison, S. C. *Nature* **1994**, *369*, 160.
(144) Stehle, T.; Harrison, S. C. *Structure* **1996**, *4*, 183.
(145) Weis, W. I.; Drickamer, K.; Hendrickson, W. A. *Nature* **1992**, *360*, 127.
(146) Drickamer, K. *Nature* **1992**, *360*, 183.
(147) Iobst, S. T.; Wormald, M. R.; Weis, W. I.; Dwek, R. A.; Drickamer, K. *J. Biol. Chem.* **1994**, *269*, 15505.
(148) Iobst, S. T.; Drickamer, K. *J. Biol. Chem.* **1994**, *269*, 15512.
(149) Kolatkar, A.; Weis, W. I. *J. Biol. Chem.* **1996**, *271*, 6679.
(150) Iobst, S. T.; Drickamer, K. *J. Biol. Chem.* **1996**, *271*, 6686.
(151) Blanck, O.; Iobst, S. T.; Gabel, C.; Drickamer, K. *J. Biol. Chem.* **1996**, *271*, 7289.
(152) Ng, K. K.-S.; Weis, W. I. *Biochemistry* **1997**, *36*, 979.
(153) Dahms, N. M. *Biochem. Soc. Trans.* **1996**, *24*, 136.
(154) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L.-N. *Anal. Biochem.* **1989**, *179*, 131.
(155) Toone, E. J. *Curr. Opin. Struct. Biol.* **1994**, *4*, 719.
(156) Schwarz, F. P.; Misquith, S.; Surolia, A. *Biochem. J.* **1996**, *316*, 123.
(157) Chervenak, M. C.; Toone, E. J. *J. Am. Chem. Soc.* **1994**, *116*, 10533.
(158) Drickamer, K. *Struct. Biol.* **1995**, *2*, 437.
(159) Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *28*, 321.
(160) Kiessling, L. L.; Pohl, N. L. *Chem. Biol.* **1996**, *3*, 71.
(161) Mortell, K. H.; Weatherman, R. V.; Kiessling, L. *J. Am. Chem. Soc.* **1996**, *118*, 2297.
(162) Biessen, E. A. L.; Noorman, F.; van Teijlingen, M. E.; Kuiper, J.; Barrett-Bergshoeff, M.; Bijsterbosch, M. K.; Rijken, D. C.; van Berkel, T. J. C. *J. Biol. Chem.* **1996**, *271*, 28024.
(163) Lee, R. T.; Lin, P.; Lee, Y. C. *Biochemistry* **1984**, *23*, 4255.
(164) Liang, R.; Yan, L.; Loebach, J.; Ge, M.; Uozumi, Y.; Sekanina, K.; Horan, N.; Gildersleeve, J.; Thompson, C.; Smith, A.; Biswas, K.; Still, W. C.; Kahne, D. *Science* **1966**, *274*, 1520.
(165) Bourne, Y.; Bolgiano, B.; Liao, D.-I.; Strecker, G.; Cantau, P.; Herzberg, O.; Feizi, T.; Cambillau, C. *Nature Struct. Biol.* **1994**, *1*, 863.
(166) Sharon, N. *Nature Struct. Biol.* **1994**, *1*, 843.
(167) Brewer, C. F. *Chemtracts Biochem. Mol. Biol.* **1996**, *6*, 165.
(168) Laine, R. A. *Glycobiology* **1994**, *4*, 759.
(169) McCoy, J. J.; Mann, B. J.; Petri, W. A. Jr. *Infect. Immun.* **1994**, *62*, 3045.
(170) Kahane, I., Ofek, I., Eds. *Towards Anti-Adhesion Therapy for Microbial Diseases*; Plenum Press; New York, 1996; 288 pp (*Adv. Exp. Med., 408*).
(171) Beuth, J., Pulverer, G., Eds. *Lectin Blocking. New Strategies for the Prevention and Therapy of Tumor Metastasis and Infectious Diseases*; Gustav Fischer Verlag; Stuttgart, 1994; 171 pp.
(172) Zopf, D.; Roth, S. *Lancet* **1996**, *347*, 1017.
(173) Simon, P. M.; Goode, P. L.; Mobasseri, A.; Zopf, D. *Infect. Immun.* **1997**, *65*, 750.
(174) Svensson, M.; Lindstedt, R.; Radin, N. S.; Svanborg, C. *Infect. Immun.* **1994**, *62*, 4404.
(175) Normark, S. *Abstracts of Microbial Pathogenesis and Immune Response II*, October 25, 1995; New York Academy of Sciences: New York, 1995; abstract S9.
(176) Echtenacher, B.; Mannel, D. N.; Hultner, L. *Nature* **1996**, *381*, 75.
(177) Malaviya, R.; Ikeda, T.; Ross, E.: Abraham, S. N. *Nature* **1996**, *381*, 77.
(178) Abraham, S. N.; Malaviya, R. *Infect. Immun.* **1997**, *65*, 3501.
(179) Ofek, I.; Goldhar, J.; Keisari, Y.; Sharon, N. *Annu. Rev. Microbiol.* **1995**, *49*, 239.
(180) Kijne, J. W. *Chemtracts Biochem. Mol. Biol.* **1996**, *6*, 180.
(181) Hirsch, A. M.; Brill, L. M.; Lim, P. O.; Scambray, J.; van Rihn, P. *Symbiosis* **1995**, *19*, 155.
(182) van Eijsden, R. R.; Diaz, C. D.; de Patter, B. S.; Kijne J. W. *Plant Mol. Biol.* **1995**, *29*, 431.
(183) Etzler, M. E. In *Glycoconjugates*; Allen, H. J., Kisailus, E. C., Eds.; Marcel Dekker: New York, 1992; p 521.
(184) Peumans, W. J.; van Damme E. J. M. *Plant Physiol.* **1995**, *109*, 347.
(185) Pusztai, A., Bardocz S. Eds. *Lectins: Biomedical Perspectives*. Taylor & Francis: London, 1995; 331 pp.
(186) Pusztai, A.; Bardocz, S. *Trends Glycosci. Glycotechnol.* **1996**, *8*, 149.
(187) Colnot, C.; Ripoche, M.-A.; Fowlis, D.; Cannon, V.; Scaerou, F.; Cooper, D. N. W.; Poirier, F. *Trends Glycosci. Glycotechnol.* **1997**, *9*, 31.
(188) Poirier, F.; Robertson, E. *Development* **1993**, *119*, 1229.
(189) Ohannesian, D. W.; Lotan, R. In *Glycosciences: Status and Perspectives*; Gabius, H.-J., Gabius, S., Eds.; Chapman & Hall: Weinheim, 1997; p 459.
(190) Akahani, S.; Inohara, H.; Mangia-Makker, P.; Raz, A. *Trends Glycosci. Glycotechnol.* **1997**, *9*, 69.
(191) Ashwell, G. In *Lectin Blocking. New Strategies for the Prevention and Therapy of Tumor Metastasis and Infectious Diseases*; Beuth, J., Pulverer, G., Eds.; Gustav Fischer: Stuttgart, 1994; p 26.
(192) Ishibashi, S.; Hammer, R. E.; Herz, J. *J. Biol. Chem.* **1994**, *269*, 27803.
(193) Frenette, P. S.; Mayadas, T. N.; Rayburn, H.; Hynes, R. O.; Wagner, D. D. *Cell* **1996**, *84*, 563.
(194) Etzioni, A.; Phillips, L. M.; Paulson, J. C.; et al. In *Cell adhesion and human disease*; Marsh, J., Goode, J. A., Eds.; Ciba Foundation Symposium 189; Wiley: Chichester, 1995; p 51.
(195) Parekh, R. B.; Edge, C. J. *Trends Biotechnol.* **1994**, *12*, 339.
(196) Munier-Lehmann, H.; Mauxion, F.; Hoflack, B. *Biochem. Soc. Trans.* **1996**, *24*, 133.
(197) Neufeld, E. F. *Annu. Rev. Biochem.* **1991**, *60*, 257.
(198) Hammond, C.; Helenius, A. *Curr. Opin. Cell Biol.* **1995**, 7, 523.
(199) Williams, D. B. *Glycoconjugate J.* **1995**, *12*, iii−iv.
(200) Spiro, R. G.; Zhu, Q.; Bhoyroo, V.; Soeling, H. D. *J. Biol. Chem.* **1996**, *271*, 11588.
(201) Rhodes, J. M., Milton, J. D., Eds.; *Lectin Methods and Protocols*; Humana Press Inc.: Totowa, NJ, 1998; 616 pp.
(202) Carlson, S. R. In *Glycobiology, a Practical Approach*; Fukuda, M., Kobata, A., Eds.; IRL Press: Oxford, 1994; p 1.
(203) Debray, H.; Montreuil, J. *Adv. Lectin Res.* **1991**, *4*, 51.
(204) Hasselbeck, A.; Hosel, W. In *Glycoprotein Analysis in Biomedicine*. Hounsell, E. L., Ed. *Methods Mol. Biol.* **1993**, *14*, 161.
(205) Yamamoto, K.; Tsuji, T.; Osawa, T. In *Glycoprotein Analysis in Biomedicine*. Hounsell, E. L., Ed. *Methods Mol. Biol.* **1993**, *14*, 17.
(206) Sharon, N. *Adv. Immunol.* **1983**, *34*, 213.
(207) Aversa, F.; Tabillo, A.; Terenzi, A.; Velardi, A.; et al. *Blood* **1994**, *84*, 3948.
(208) Kilpatrick, D. C.; Green, C. *Adv. Lectin Res.* **1992**, *5*, 51.
(209) Borrebaeck, C. A. K.; Carlsson, R. *Adv. Lectin Res.* **1989**, *2*, 10.
(210) Di Sabato, G.; Hall, J. M.; Thompson, L. *Methods Enzymol.* **1987**, *150*, 3.

CR940413G