Lectin–carbohydrate interactions: different folds, common recognition principles

Sharona Elgavish and Boaz Shaanan

Lectins can be found in many organisms and are involved in a multitude of cellular processes that depend on specific recognition of complex carbohydrates. The stereochemical principles underlying the recognition process have been the subject of extensive biochemical and structural studies. When examined from the viewpoint of the bound sugar, the structural information accumulated so far on lectins and other proteins that are specific to galactose and glucose (or mannose), provides suggestive evidence for distinct ligand-dependent distribution of hydrogen-bond partners in the combining site.

LECTINS ARE PROTEINS that bind carbohydrates with considerable specificity. They are found in a variety of organisms and are involved in numerous cellular processes, such as host–pathogen interactions, targeting of proteins within cells and cell–cell interactions[1]. High-resolution studies on bacterial sugar-transport proteins have provided a detailed description of the elementary interactions sustaining carbohydrate recognition by proteins[2]. Recent studies on lectin–carbohydrate complexes from various sources (reviewed in Refs 3, 4) have further extended our understanding of the structural basis of lectin–carbohydrate interactions. The thermodynamics of the recognition process has also been studied extensively (reviewed in Ref. 5).

In this review, we focus on the stereochemistry of lectin–carbohydrate interactions at the primary binding site[3]. In particular, we examine the capacity of lectins to discriminate between galactose (Gal) and glucose (Glc) or mannose (Man), in an attempt to underline structural features common to different families of lectins.

Classification of carbohydrate-binding proteins and lectins

Traditionally, lectins have been classified according to sequence similarity and tertiary structure[6]. The reader is referred to a recent review for an extensive survey of the variety of lectin folds[3]. For the purpose of detecting the possible common principles underlying carbohydrate recognition, it is convenient to classify carbohydrate-binding proteins according to topological features of the combining site.

Accordingly, carbohydrate-binding proteins have been divided into two major groups[3]. The proteins constituting group I, such as the bacterial periplasmic transport proteins[2] and enzymes, have a buried binding site and engulf the ligand fully upon binding. Proteins belonging to group II, however, have a shallow binding site, mostly in the form of a depression on the protein surface. In this versatile group, we find the three classical lectin families – legume lectins, C-type lectins and galectins – as well as other plant lectins, viral proteins (e.g. influenza hemagglutinin), toxins (e.g. cholera toxin[7]), anti-carbohydrate antibodies[8] and pentraxins[3,4,9].

The key interactions involved in carbohydrate recognition

Carbohydrates interact with lectins through hydrogen bonds, metal coordination, van der Waals and hydrophobic interactions. The availability of large numbers of hydroxyl groups on sugars renders them obvious partners in complex networks of hydrogen bonds, usually formed by cooperative hydrogen bonds in which the hydroxyl serves both as a donor and an acceptor[2,4]. Sidechain atoms of Asp and Asn, and the main-chain amide hydrogens and carbonyl oxygen, commonly participate in such hydrogen bonds, with a less frequent occurrence of sidechains of other amino acids[4]. The protein–carbohydrate hydrogen bonds are both direct and water mediated[2,4,10]. Divalent cations, such as Ca^{2+} and Mn^{2+}, are involved in carbohydrate recognition either indirectly, by shaping the combining site, as in the legume lectins[4,11], or through direct binding of the carbohydrate to Ca^{2+} as in the C-type lectins[4,12].

Despite the overall hydrophilic character of carbohydrates, hydrophobic interactions play a major role in their recognition by lectins. Particularly noticeable is the interaction between aromatic residues and galactose (Gal) in the combining sites of Gal-specific lectins, which was attributed to an interaction between an extended patch of partially charged aliphatic protons on the B face of the hexose ring and the partial negative charge on the π-electrons of the aromatic system[4,13,14].

Increased selectivity through multiple binding

A most-intriguing question is how the specificity of lectins towards their natural targets, usually branched carbohydrates, is elicited despite the fact that their affinity toward monosaccharides is relatively low (in the mM range). The common view is that the selectivity towards a particular target is augmented by several orders of magnitude through multiple binding, by mechanisms of additional binding in sub-sites (or extended sites) and subunit multivalency[3,4].

In sub-site binding, one monosaccharide, usually the terminal one, is bound at the primary binding site of the lectin, with additional monosaccharides along the carbohydrate chain bound to secondary sub-sites on the lectin. This kind of selectivity enhancement is demonstrated in the binding of the Glc/Man-specific *Lathyrus ochrus* lectin (LOL I) to a series of mannose-containing oligosaccharides[15], and that of cholera toxin (CT) to GM1 ganglioside through the terminal sialic acid and galactose[7]. Subunit multivalency is exhibited when several subunits of the same lectin bind to different extensions of a branched carbohydrate as in the case of the asialoglycoprotein receptor (ASGPR)[16], or to separate carbohydrate chains as in the case of the trimeric mannose-binding protein (MBP)[17,18] and the pentameric cholera toxin[7].

The primary specificity

While the ultimate selectivity of lectins is probably attained by the multivalency

S. Elgavish and **B. Shaanan** are in the Department of Biological Chemistry and The Wolfson Centre for Applied Structural Biology, Institute of Life Sciences, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel.
Email: boaz@sugar.ls.huji.ac.il

 0968–0004/97/$17.00 PII: S0968-0004(97)01146-8

mechanisms, a major discriminatory factor in carbohydrate recognition is the primary monosaccharide specificity, which is usually also indicative of the particular branched carbohydrate to be recognized by the lectin. Thus, the Glc/Man-specific legume lectins (e.g. LOL I and Con A) bind complex carbohydrates bearing Glc or Man, whereas the Gal-specific legume lectins (e.g. *Erythrina corallodendron* EcorL[11]) require the presence of Gal (or its amino-acetylated derivative) for binding (Gal and Glc/Man have different stereochemistry at the C-4 positions and are therefore called C-4 epimers; Gal has an axial 4-OH, whereas Glc/Man has an equatorial 4-OH). Likewise, the C-type lectins, MBP and human ASGPR, which are specific for Man and Gal, respectively, bind complex carbohydrates terminated with the corresponding monosaccharides[19]. It should be noted, however, that the primary specificity does not always serve as a clear indication for the overall selectivity. In the case of CT, both Gal and sialic acid are essential for GM1 binding[7], but only monosaccharide specificity towards Gal has been shown[20]. Similarly, Gal is essential for the Lewis b blood group oligosaccharide binding to the *Griffonia simplicifolia* lectin (GS4), but the lectin does not show any detectable monosaccharide-binding capacity[3].

Analysis of the structural origins of the primary specificity, based on high-resolution structures of protein–carbohydrate complexes representing various families and folds, suggests that although the key interactions responsible for carbohydrate recognition are common, each family has evolved a unique stereochemistry at the principal combining site in order to discriminate between ligands[3,4]. However, as described below, a closer inspection of the structural basis of the Gal vs. Glc/Man discrimination in several lectin families reveals that the distribution of atoms serving as hydrogen-bond partners for the carbohydrate at the primary site, points to an intriguing ligand-dependent stereochemistry of the hydrogen-bonding pattern around the 4-OH. In combination with the preferential disposition of the aromatic residues[3,4,21], this stereochemistry could play a key role in eliciting the primary specificity.

Distribution of hydrogen-bonding partners in Gal- and Glc/Man-specific lectins

With the structural information currently available on protein–carbohydrate complexes, a search for common recognition features can be conducted with a reasonable degree of fidelity only on the group of lectins and proteins that bind Gal or Glc/Man at the primary site. The diversity of folds of the proteins included in this group, together with the relatively simple structure of Gal and Glc/Man, support the notion that the potentially small number of such recognition factors[3,4] would be well represented within this group.

To detect the potential discriminating factors in the Gal- and Glc/Man-binding proteins, it is convenient to examine the ligand environments in the primary combining sites as they are 'seen' by the carbohydrates and within a common framework. Thus, we superimposed the hexose rings of Gal- and Glc/Man-binding lectins and group-I proteins, for which coordinates of the complexes with these carbohydrate were available at the Brookhaven Protein Data Bank (PDB; Table I), and applied the resulting transformations to the atoms that hydrogen-bond the carbohydrates at the primary site, as well as to the aromatic residues implicated in ligand stacking[3]. The results are depicted in two orthogonal views (Fig. 1). The bacterial Gal/Glc-binding protein[22], in which all hydroxyls are extensively hydrogen bonded (Table I), serves as a reference point in the analysis (Fig. 1).

It is noticeable that the carbohydrate ligands span, to a good approximation, dome-shaped loci around the sugar hydroxyls (Fig. 1). Ligands common to neighbouring sugar hydroxyls (e.g. 3-OH and 4-OH donors of the b c-terial toxins or Ca^{2+} in MBP-A and -C; Table I) and ligands emanating from the same sidechain (e.g. OD1 and OD2 of Asp89 in EcorL; Table I) are found in, or proximal to, the tangential regions of the respective domes (Fig. 2). Such spatial distribution conforms with the expected stereochemistry of the hydrogen bond[23] and the distributions observed around sidechains of polar amino acids involved in hydrogen bonds in protein structures[24].

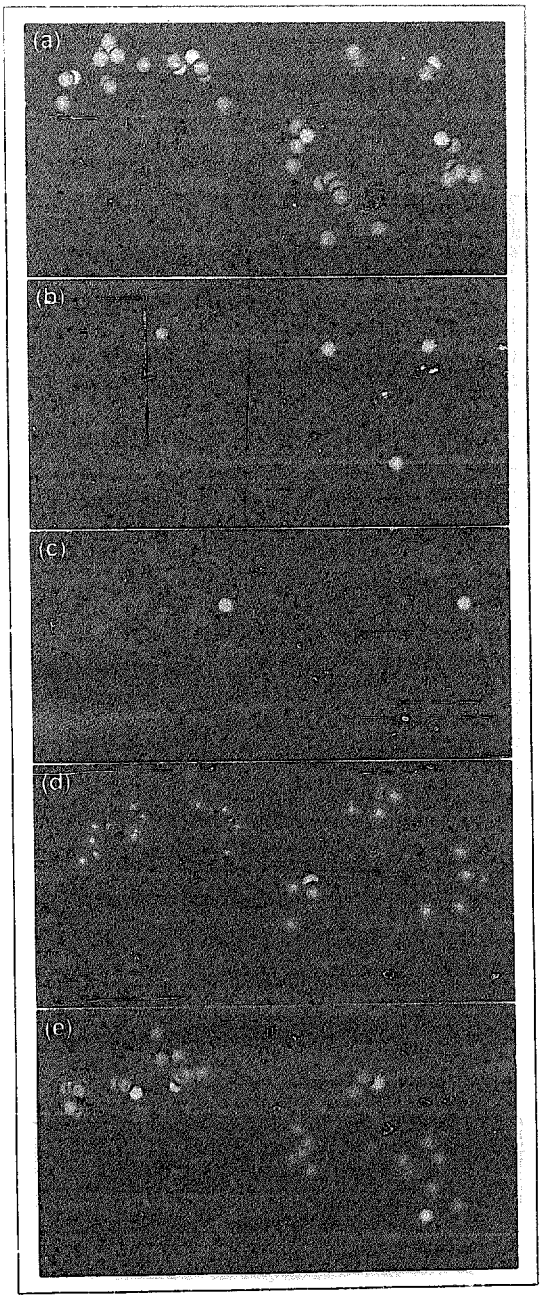

Figure 1

Side view (left) and view after a 90° rotation around horizontal axis (right) of the distribution of hydrogen-bond partners in Gal- and Glc/Man-binding proteins listed in Table I. Boundaries of dome-shaped distribution are marked by white lines. Carbohydrate is shown in blue and white spheres are used for Glc/Gal-binding protein (GBP), which serves as a reference (see text). Green spheres represent average position of distribution. O1 is omitted for clarity. In all other figures, the colour scheme for the distributions is as in this figure. **(a)** Distribution around O2; **(b)** O3; **(c)** axial O4; **(d)** equatorial O4; and **(e)** distribution around O6 for the dihedral angle ω, C4-C5-C6-O6, with mean value of 180°. The distribution for ω ~60° is similar.

Table I. Galactose- and glucose/mannose-binding proteins with known three-dimensional structure

Protein[a]	PDB code	Aromatic residue involved in stacking	Ligand[b]	Positions of sugar hydroxyls[c] 01	02	03	04	05	06
GS4	1led	Trp133	Gal		wat382	Asp89 OD1 (A) Gly107 N (D) Asn135 ND2 (D)	Asp89 OD2 (A)		
EcorL	1ite	Phe131	Gal			Asp89 OD2 (A) Gly107 N (D) Asn133 ND2 (D)	Asp89 OD1 (A) Ala218 N (D)		wat836 Gln219 NE2 (D)
SBA	1sba	Phe128	Gal			Asp88 OD2 (A) Ala105 O (A) Asn130 ND2 (D)	Asp88 OD1 (A) Leu214 N (D)		Asp215 OD1 (A)
Peanut lectin	2pel	Tyr125	Gal		wat179 wat437	Asp83 OD1 (A) Gly104 N (D) Asn127 ND2 (D)	Asp83 OD2 (A) Ser211 OG (D)		Asp80 OD2 (A)
ConA	5cna	Tyr12	Man			Arg228 N (D)	Asp208 OD1 (A) Asn14 ND2 (D)		Asp208 OD2 (A) Leu99 N (D) Tyr100 N (D)
Pea lectin	1rin	Phe123	Man			Gly99 N (D)	Asp81 OD1 (A) Asn125 ND2 (D)		Asp81 OD2 (A) Glu218 N (D) Ala217 N (D)
LoL I	1loa	Phe123	Glc			Gly99 N (D)	Asp81 OD1 (A) Asn125 ND2 (D)	Ala210 N (D)	Asp81 OD2 (A) Gly209 N (D) Glu211 N (D)
Ricin	2aai	Trp37	Gal			Lys40 NZ (D) Asn46 ND2 (D)	Asp22 OD1 (A)		Gln35 NE2 (D)
GNA	1msa		Man		Asp91 OD2 (A) Asn93 ND2 (D)	Gln89 NE2 (D)	Tyr97 OH (A)		wat45
Galectin-1	1slt	Trp68	Gal			wat25 (D)	His44 NE2 (A) Asn46 OD1 (A) Arg48 NH2 (D) wat25 (D)		Glu71 OE2 (A) Asn61 ND2 (D)
Galectin-2	1hlc	Tyr65	Gal				His45 NE2 (A) Asn47 OD1 (A) Arg49 NH1 (D)		Glu68 OE1 (A) Asn58 ND2 (D)
MBP-A	2msb		Man			Glu185 OE2 (A) Asn187 ND2 (D) Ca^{2+} (D)	Glu193 OE2 (A) Asn205 ND2 (D) Ca^{2+} (D)		
QPDWG	1afa	Trp189	Gal		Asn210 ND2 (D)	Glu198 OE2 (A) Asn210 ND2 (D) Ca^{2+} (D)	Asp187 OD2 (A) Gln185 NE2 (D) Ca^{2+} (D)		
MBP-C	1rdl		Man			Glu198 OE2 (A) Asn210 ND2 (D) Ca^{2+} (D)	Glu190 OE2 (A) Asn192 ND2 (D) Ca^{2+} (D)		
LT	1lta	Trp88	Gal		wat279 Asn90 ND2 (D)	Asn90 OD1 (A) Lys91 NZ (D)	Glu51 OE2 (A) Lys91 NZ (D)		wat291 wat92 Gln61 NE2 (D)
LT-I	1lti	Trp88	Gal			Asn90 OD1 (A) Lys91 NZ (D)	Glu51 OE2 (A) Lys91 NZ (D)		Gln56 O (A) Gln61 NE2 (D)
CT	1chb	Trp88	Gal		Asn90 ND2 (D) wat115 wat17 Asn90 ND2 (D)	Asn90 OD1 (A) Lys91 NZ (D)	Glu51 OE2 (A) Lys91 NZ (D)		wat207 wat206
GBP[d]	1glg	Trp183	Gal	Asp154 OD2 (A) Arg158 NH2 (D) Asn256 ND2 (D)	Asp236 OD1 (A) Arg158 NH1 (D)	Asp236 OD2 (A) Asn211 ND2 (D) wat313 (D)	Asp14 OD1 (A)	Asn91 ND2 (D)	Asn91 OD1 (A) His152 NE2 (D)
GBP	2gbp	Trp183	Glc	Asp154 OD1 (A) Arg158 NH2 (D)	Asp236 OD2 (A) Arg158 NH1 (D)	Asp236 OD1 (A) Asn211 ND2 (D) wat313 (D)	Asp14 OD2 (A)	Asn91 ND2 (D)	Asn91 OD1 (A) His152 NE2 (D)
GPB	2gpb		Glc	Asn256 ND2 (D) wat279	Glu672 OE1 (A) wat458 (D) Asn284 ND2 (D)	Glu672 OE1 (A) Ala673 N (D) Ser674 N (D) wat c21 (A) Arg349 NH2 (D)	Asn484 OD1 (A) Gly675 N (D) wat522 (D) Glu739 OE2 (A) Arg349 NH1 (D)		Asn484 OD1 (A) His377 ND1 (D)
Chitobiase	1qbb	Trp737	Glc			Asp319 OD2 (A)			
CelC A[e]	1cel	Trp212	Glc		Met318 O (A) Trp313 NE1 (D)			Gln16 NE2 (D)	Gln16 NE2 (D) wat610 (A)
CelC B[f]	1cel	Tyr200	Glc	Gln140 OE1 (A)	Glu280 OE1 (A) His90 NE2 (D)	His90 NE2 (D)			Trp212 NE1 (D) His207 NE2 (D)

[a]Full names of proteins: GS4, *Griffonia simplicifolia* lectin IV (Lewis b tetrasaccharide[4]); EcorL, *Erythrina corallodendron* (Galβ 1–4Glc); SBA, Soybean agglutinin (pentasaccharide[4]); Pea lectin (trimannoside[4]); ConA, Concanavalin A (Manα1-Me); LoL I, *Lathyrus orchus* isolectin I (Manα1-Me); Galectin-1 (Galβ 1–4NAc); Galectin-2 (Galβ 1–4Glc); GNA, Snowdrop lectin (Manα 1-Me); Ricin (Galβ 1–4Glc); MBP-A, Mannose-binding protein-A (high Man octasaccharide[20]); QPDWG, Gal-binding mutant of mannose-binding protein-A (Galβ 1-Me); MBP-C, Mannose-binding protein-A (Manα1-Me); LT, Heat-labile enterotoxin (Gal); LT-I, Heat-labile enterotoxin (Galβ 1–3GalNAc); CT, Cholera toxin (ganglioside GM1 pentasaccharide); GBP, Bacterial periplasmic glucose/galactose-transport protein (Glc/Gal); GPB, Glycogen phosphorylase B (Glc); Chitobiase (di-GlcNAc); CelC, mutant of *Clostridium thermocellum* endoglucanase [cellohexose (Glc_6)]. [b]Carbohydrate in primary binding site. [c]Donor (D)/Acceptor (A) assignment as in original articles (see Ref. 4) or by chemical consideration. Where no D/A assignment is given it is to be considered as undefined. [d]The bacterial periplasmic glucose/galactose-transport protein, as well as the enzymes glycogen phosphorylase, chitobiase and endoglucanase represent group I proteins. [e]Refers to glucose 350 in the PDB file 1cel.pdb. [f]Refers to glucose 351 in the PDB file 1cel.pdb

Distinct distribution of hydrogen-bond donors and acceptors around 4-OH

Most intriguingly, two distinctly different spatial distributions become apparent upon division of the ligands according to donors and acceptors (Fig. 3). Whereas donors and acceptors of the equatorial 3-OH (and similarly those of 2-OH or 6-OH; data on the distribution around 1-OH are scarce) are evenly scattered across the dome (Fig. 3a), the donors and acceptors of the axial and equatorial 4-OH tend to cluster in two segregated regions of their respective domes (Fig. 3b,c). If the distributions of hydrogen-bond partners around the 4-OH are viewed down the C-OH bond in what is known as Newman projection, the donors of the axial 4-OH and the acceptors of the equatorial 4-OH span the segment between *gauche*$^-$ and *trans* conformations around the 4-OH–C-4 bond, whereas the axial 4-OH acceptors and equatorial 4-OH donors span the *trans-gauche*$^+$ segment (Fig. 4). Furthermore, the acceptors of the two 4-OH epimers converge onto neighbouring zones, while the donors occupy diametrically opposing ones (Fig. 5). Interestingly, the Ca^{2+} ion, the primary affinity determinant for carbohydrate binding by C-type lectins[4,12], straddles the line of demarcation between the donors and acceptors regions of both 4-OH epimers (Fig. 3b,c) and is in *trans* to the C-5 (Fig. 4). The aromatic residues interacting with Gal (Fig. 5), which have been shown to be nearly coplanar when aligned as described above[3,21], are located in close proximity to the area occupied by the equatorial 4-OH donors.

Implications of the donors/acceptors distribution around 4-OH

The analysis presented in this review is based largely on the three-dimensional structures of Gal- and Glc/Man-specific lectins. However, several carbohydrate-binding proteins from group I with a similar specificity[3] were also included in the analysis in order to detect potential ligand-dependent, rather than protein-dependent, discriminatory factors. It seems as if such ligand-dependent factors do indeed exist, although, as could be expected, subdivision according to protein families[25] can also be discerned.

The donors. The unique clustering mode of donors and acceptors around the 4-OH, as opposed to the even distributions around other hydroxyls (Fig. 3), emphasizes the importance of the stereochemistry around the sugar C-4 position for Gal vs. Glc/Man specificity[4]. Moreover, the stereochemical constraints around C-4 appear to be more stringent than the mutual spatial disposition of the 3-OH and 4-OH, which has been implicated as a discriminatory factor[4,25]. The 3-OH donors and acceptors can, in principle, be presented to the carbohydrate from any point around the locus (dome) of 3-OH ligands (Fig. 3a). By contrast, the 4-OH donors and acceptors are presented only from four segregated zones (Figs 3b,c, 4, 5) according to the C-4 epimer. With minor variations, the stereochemical requirements imposed on 4-OH ligands seem to be also generally valid across protein families.

The fact that the clusters of both the axial and equatorial 4-OH acceptors are located close in space while the clusters of donors are far apart (Fig. 5), amplifies the significance of the location of the 4-OH donors. Different protein families have devised different means to position the donor correctly with respect to the preferred ligand. For example, axial 4-OH donors in the Gal-specific legume lectins are located in a loop, which has long been recognized as an important discriminatory factor[11] (in EcorL, this loop spans residues 217–226 and has recently been named as loop D of the legume lectins[26]), but equatorial 4-OH donors in the Glc/Man-specific members of this family project from a loop facing the opposite side of the hexose ring (the loop spanning residues 131–141 in EcorL, or loop C according to the above naming convention[26]). In the C-type lectins, the donors are correctly positioned with respect to Man or Gal by mutating the sequence EPN in MBP to QPD in ASGPR, which leads to rearrangement of donors and acceptors in the combining site[19,21,27] (Fig. 6).

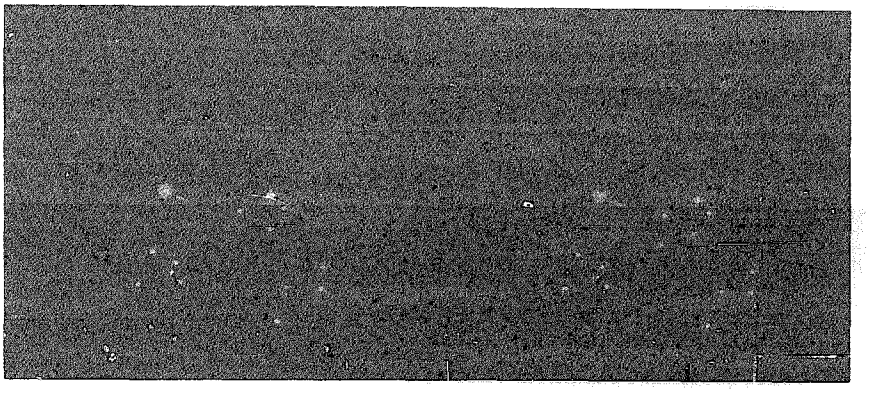

Figure 2
Spatial relations (stereo image) between domes of hydrogen-bonds partners. Distribution around O3 (blue sphere) is shown in purple, partners of axial O4 (yellow sphere) are shown in red and partners of equatorial O4 (orange sphere) are shown in grey.

The acceptors. Interestingly, in both legume lectins and C-type lectins, the

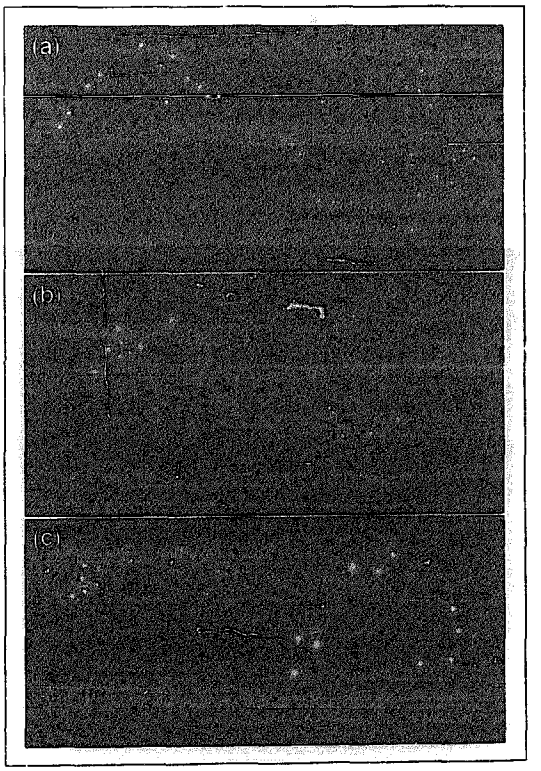

Figure 3
Segregation of 4-OH donors/acceptors. Distributions are displayed as in Fig. 1, showing separation to donors (pink) and acceptors. Large brown spheres represent Ca^{2+} atoms. **(a)** O3 (acceptors in purple); **(b)** axial O4 (acceptors in red); and **(c)** equatorial O4 (acceptors in grey). Dotted line marks donor/acceptors boundary.

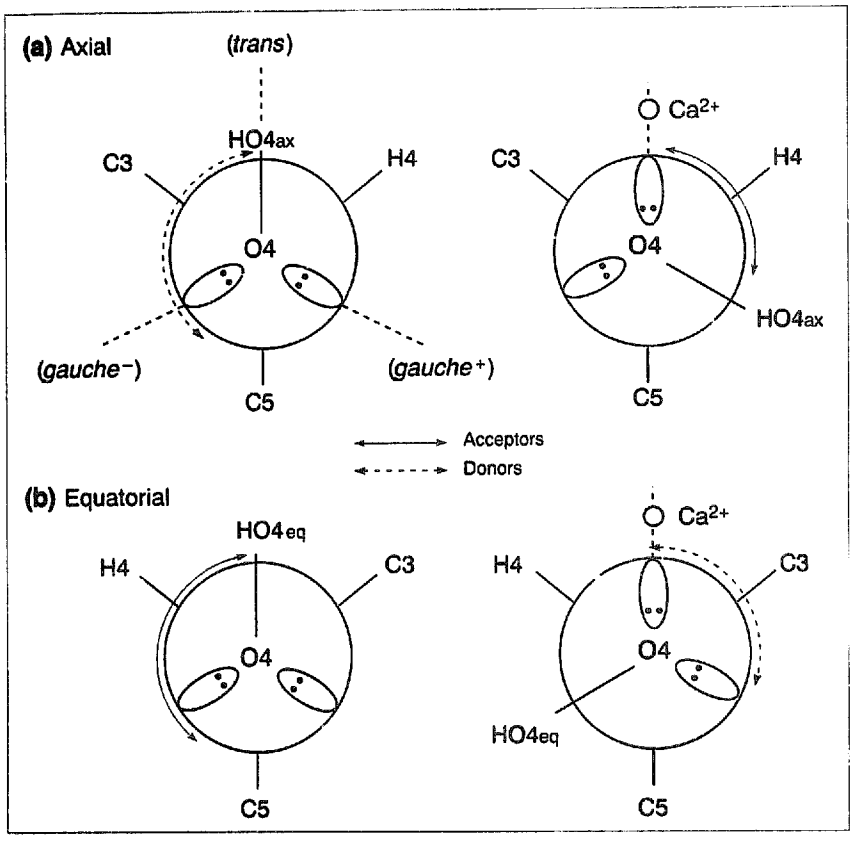

Figure 4

Newman projections around C4–O4 bond of the carbohydrate. Projections correspond to views in Fig. 3b,c (right) and show the range of orientations occupied by **(a)** axial and **(b)** equatorial donors and acceptors. The lone electron pairs of O4 are depicted as ellipses.

only lectin families in which specificity for carbohydrate with the two 4-OH epimers has been found, the contiguous acceptors regions serve as the pivot point around which the location of donors interchange according to the C-4 stereochemistry of the entering ligand (Fig. 5). The acceptor regions of the axial and equatorial 4-OH are proximal because the acceptors either project from the same sidechain in the legume lectins (Table I), or are tethered by the Ca^{2+} in the C-type lectins[4].

In the bacterial Glc/Gal-binding protein[22], which, unlike the legume and C-type lectins, does not discriminate between Gal and Glc, the above epimer-dependent interchange of the 4-OH donor disposition is not observed. Both Glc and Gal are bound in exactly the same orientation, the only difference in hydrogen bonding being the interchange of Asp14 OD2 for OD1, respectively[22]. Furthermore, none of the 4-OH epimers has a protein atom serving as a donor (the function of the water molecule w313 as a donor is questionable[22]).

The aromatic residues. The disposition of the aromatic residues interacting with the carbohydrate is also coupled with that of the donors (Fig. 5). The overlapping aromatic groups of Gal-specific proteins[3,21] (represented in Fig. 5 by Trp88 of CT; Table I) and the respective donors, face opposite sides of the hexose ring and are thus far apart (Fig. 5). However, the donors in Glc/Man-binding proteins occupy the space usually taken by the aromatics of Gal-specific proteins. Consequently, the aromatic groups involved in ligand binding to Glc/Man-specific proteins (Table I) tend to be grouped in regions proximal to the less-hydrophobic B face of the carbohydrate of Glc/Man but, at the same time, avoid steric clash with the donors (Fig. 5). We have noticed that the aromatic residues present in Glc/Man-binding proteins are clustered in two regions, one characteristic of the legume lectins (represented in Fig. 5 by Tyr12 of Con A; Table I) and the other characteristic of proteins in group I (represented in Fig. 5 by Trp183 of GBP; Table I). In both cases, the contribution of the aromatic stacking to the binding energy is probably lower than in the Gal-binding proteins, where the disposition of the aromatic groups with respect to the ligand is optimal[3,4,21]. The ligand-dependent coupling between the disposition of the aromatic groups and donors, together with the convergence of the acceptors of both 4-OH epimers to proximal regions with respect to the bound carbohydrate, are interesting observations that deserve further investigation.

Contributions to primary selection of ligand. Although it is difficult to uncouple the various contributions to the primary specificity, it seems as if a crucial level of selection, which favours the 'correct' carbohydrate in the combining site, is achieved by matching the C-4 epimer (equatorial/axial 4-OH) with the given constellation of 4-OH donors/acceptors (Figs 3b,c, 5, 6). When a match is established, the binding energy is enhanced by contributions from other ligand- and protein-dependent factors, such as additional hydrogen bonds, van der Waals interactions and aromatic stacking.

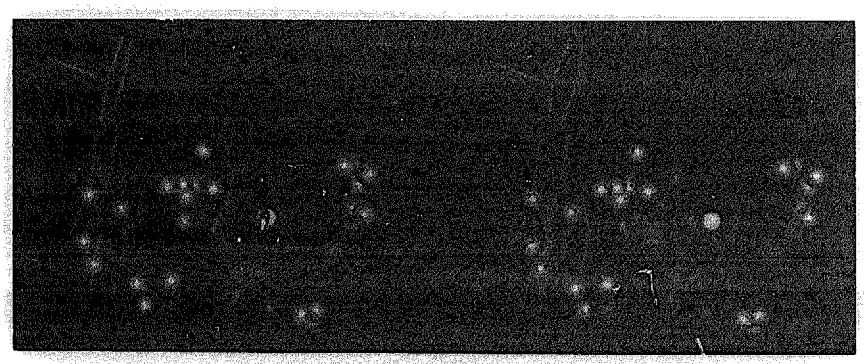

Figure 5

Side view (stereo image) of segregated donors (pink) and acceptors of axial O4 (yellow sphere, acceptors shown in red) and of equatorial O4 (orange sphere, acceptors in grey). Large brown spheres represent Ca^{2+} atoms. Trp88 of cholera toxin (CT; in green) represents the location of aromatic residues in Gal-binding proteins. The aromatic residues of Glc/Man proteins are represented by Tyr12 of Con A (cyan) for legume lectins, and by Trp183 of the Glc/Gal-binding protein (GBP; orange) for group I proteins (see text; Table I).

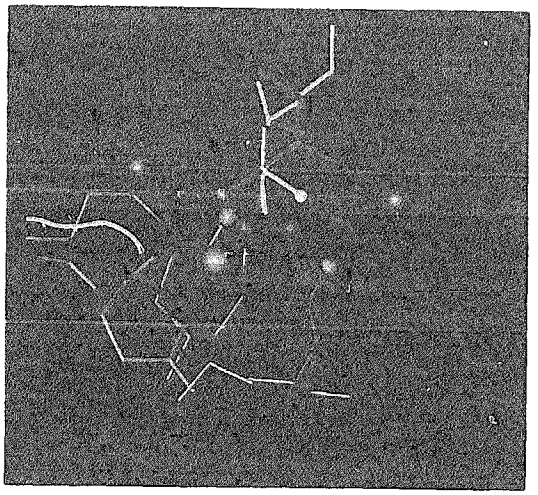

Figure 6

Superposition of the combining sites of the complexes of the mannose-binding protein (MBP-A; grey bonds) with Man (green) and of the QPDWG mutant (red bonds) with Gal (yellow). Models were superimposed on atoms of hexose rings. O3 donor/acceptors are shown as spheres in the same colour as their bonds, large spheres represent Ca^{2+} atoms, O4 acceptors are in blue and donors are in pink for both complexes. Hydrogen bonds (dotted lines) to 4-OH and bonds between Ca^{2+} and 4-OH (dashed lines) are shown for both complexes.

Overall, this dissection of the ligand discrimination seems to be consistent with the results of the elegant biochemical and structural studies of Drickamer, Weis and their colleagues[19,21,27], in which Gal specificity has been introduced into MBP by a series of crucial mutations and the structural consequences have been inspected by crystallographic determination of the complexes. The EPN→QPD conversion mentioned above is sufficient to elicit Gal binding, which reflects the effect of rearranging the donors/acceptors according to the consensus of axial 4-OH constraints (Figs 5, 6). However, full discrimination against Man is achieved only upon further mutagenesis and the introduction of Trp and a Gly-rich loop, which probably ensure optimization of the aromatic stacking with respect to Gal[19,21,27]. Further studies on the C-type lectins system would undoubtedly provide more insight into the factors controlling their carbohydrate specificity and would hopefully inspire studies of similar quality on other lectin families.

Concluding remarks

By analysing the currently known structures of complexes of Gal- and Glc/Man-binding proteins from the point of view of the ligand in the combining site, we can present strongly suggestive evidence that a particular organization of the protein hydrogen-bond donors and acceptors is a necessary condition for specific carbohydrate binding at the primary combining site. The requirements for the disposition of donors/acceptors necessary to recognize the axial 4-OH of Gal are clearly distinguishable from the requirements for binding the equatorial 4-OH of Glc and Man. While satisfying these basic epimeric-dependent hydrogen-bonding constraints around C-4 may be crucial for the initial selection of primary ligand, additional ligand-dependent hydrogen bonds, van der Waals and hydrophobic interactions are obviously essential for highly specific binding to take place. It would be interesting to watch whether the stereochemistry at the combining sites of complexes of Gal- and Glc/Man-binding proteins determined in the future conforms with the view presented in this article.

Considerations related to the availability of information led us to concentrate on the structural aspects of the discrimination between Gal and Glc/Man. With the accumulation of structural information on additional protein–carbohydrate complexes, similar analyses may unravel distinct trends in the distribution of hydrogen-bond partners pertinent to the specificity towards other ligands.

Acknowledgements

We are grateful to W. Weis, A. Kolatkar and K. Ng for helpful comments on the manuscript, and to R. Gabbay for help with the figures. Because of the limited number of references permitted, some references to articles describing the complexes included in the analysis had to be omitted. We apologize to the authors of these articles.

References

1 Sharon, N. (1993) *Trends Biochem. Sci.* 18, 221–226
2 Quiocho, F. A. (1989) *Pure Appl. Chem.* 51, 1293–1306
3 Rini, J. M. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 551–577
4 Weis, W. I. and Drickamer, K. (1996) *Annu. Rev. Biochem.* 65, 441–473
5 Toone, E. J. (1994) *Curr. Opin. Struct. Biol.* 4, 719–728
6 Drickamer, K. (1988) *J. Biol. Chem.* 263, 9557–9560
7 Merritt, E. A. *et al.* (1994) *Protein Sci.* 3, 166–175
8 Bundle, D. R. and Young, N. M. (1992) *Curr. Opin. Struct. Biol.* 2, 666–673
9 Emsley, J. *et al.* (1994) *Nature* 367, 338–344
10 Bourne, Y. and Cambillau, C. (1993) *Topics Mol. Biol.* 17, 321–337
11 Shaanan, B., Lis, H. and Sharon, N. (1991) *Science* 254, 862–866
12 Weis, W. I., Drickamer, K. and Hendrickson, W. A. (1992) *Nature* 360, 127–134
13 Burley, S. K. and Petsko, G. A. (1985) *Science* 229, 23–28
14 Levitt, M. and Perutz, M. F. (1988) *J. Mol. Biol.* 201,751–754
15 Bourne, Y., Rougé, P. and Cambillau, C. (1992) *J. Biol. Chem.* 267, 197–203
16 Rice, K. G. *et al.* (1990) *J. Biol. Chem.* 265, 18429–18434
17 Sheriff, S., Chang, C. Y. Y. and Ezekowitz, R. A. B. (1994) *Nat. Struct. Biol.* 1, 789–794
18 Weis, W. I. and Drickamer, K. (1994) *Structure* 2, 1227–1240
19 Drickamer, K. (1992) *Nature* 360, 183–186
20 Merritt, E. A. *et al.* (1994) *Mol. Microbiol.* 13, 745–753
21 Kolatkar, A. R. and Weis, W. I. (1996) *J. Biol. Chem.* 271, 6679–6685
22 Vyas, M. N., Vyas, N. K. and Quiocho, F. A. (1994) *Biochemistry* 33, 4762–4768
23 Creighton, T. C. (1993) *Proteins Structures and Molecular Properties* (2nd edn), W. H. Freeman
24 Ippolito, J. A., Alexander, R. S. and Christianson, D. W. (1990) *J. Mol. Biol.* 215, 458–471
25 Drickamer, K. (1997) *Structure* 5, 465–468
26 Sharma, V. and Surolia, A. (1997) *J. Mol. Biol.* 267, 433–445
27 Iobst, S. T. and Drickamer, K. (1994) *J. Biol. Chem.* 269, 15512–15519