Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structures of the complexes of peanut lectin with methyl-β-galactose and *N*-acetyllactosamine and a comparative study of carbohydrate binding in Gal/GalNAc-specific legume lectins

R. Ravishankar, K. Suguna, A. Surolia and M. Vijayan*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: mv@mbu.iisc.ernet.in

Received 3 March 1999
Accepted 10 May 1999

PDB References: peanut agglutinin–*N*-acetyllactosamine complex, 1ciw; peanut agglutinin–methyl-β-galactose complex, 1qf3.

The crystal structures of complexes of peanut lectin with methyl-β-galactose and *N*-acetyllactosamine have been determined at 2.8 and 2.7 Å, respectively. These, and the complexes involving lactose and the T-antigenic disaccharide reported previously, permit a detailed characterization of peanut-lectin–carbohydrate association and the role of water molecules therein. The water molecules in the combining site are substantially conserved in the four complexes. The role of interacting sugar hydroxyl groups, when absent, are often mimicked by ordered water molecules not only at the primary combining site, but also at the site of the second sugar ring. The similarity of peanut-lectin–sugar interactions with those in other galactose/*N*-acetylgalactosamine-specific lectins also extend to a substantial degree to water bridges. The comparative study provides a structural explanation for the exclusive specificity of peanut lectin for galactose at the monosaccharide level, compared with that of the other lectins for galactose as well as *N*-acetylgalactosamine. The complexes also provide a qualitative structural rationale for differences in the strengths of binding of peanut lectin to different sugars.

1. Introduction

Protein–carbohydrate interactions form the basis of many biological recognitive processes. Lectins are multivalent proteins which specifically bind to different cell-surface carbohydrates and have therefore received considerable attention in recent years (Sharon & Lis, 1989*a*,*b*; Chrispeels & Raikhel, 1991; Rini, 1995). Crystallographic studies on lectins from widely different sources and their carbohydrate complexes have provided valuable information on the geometric features of protein–carbohydrate interactions (Weis & Drickamer, 1996; Drickamer, 1997; Elgavish & Shaanan, 1997; Loris *et al.*, 1998). Lectins from the seeds of leguminous plants constitute the most thoroughly studied family. Crystal structures of the complexes of a number of legume lectins, including those from *Canavalia ensiformis*, *Erythrina corallodendron*, pea, *Griffonia simplicifolia*, *Lathrus ochrus*, lentil, *Vicia faba*, peanut, soybean and winged bean, with a variety of saccharides are now known (Hardman & Ainsworth, 1972; Becker *et al.*, 1975; Reeke *et al.*, 1975; Einspahr *et al.*, 1986; Reeke & Becker, 1986; Bourne, Abergel *et al.*, 1990; Bourne, Roussel *et al.*, 1990; Bourne, Rouge *et al.*, 1990; Bourne *et al.*, 1992, 1994; Delbaere *et al.*, 1993; Rini *et al.*, 1993; Banerjee *et al.*, 1996; Ravishankar *et al.*, 1997; Naismith *et al.*, 1994; Naismith & Field, 1996; Loris *et al.*, 1993, 1996; Shaanan *et al.*, 1991; Dessen *et al.*, 1995; Prabu *et al.*, 1998; Elgavish & Shaanan, 1998). All of them have essentially the same tertiary

fold, but exhibit a variety of quaternary structures and carbohydrate specificities.

Peanut (*Arachis hypogaea*) agglutinin (PNA)[1] is a non-glycosylated tetrameric lectin, M_r = 110 kDa, of identical subunits. It is specific to galactose at the monosaccharide level and binds preferentially to the tumour-associated Thomsen–Friedenreich antigen (T-antigen; Galβ1–3GalNAc). The X-ray analysis at 2.25 Å resolution of the complex of the protein with lactose revealed that the tetrameric molecule has an unusual 'open' quaternary structure, without the expected 222 or fourfold symmetry (Banerjee *et al.*, 1994, 1996). The structure also provides a clear picture of PNA–lactose interactions. Indeed, it formed the basis of the structure-based redesign of the carbohydrate-binding region (Sharma *et al.*, 1996, 1998). The structure of its complex with T-antigen at 2.5 Å resolution was subsequently determined (Ravishankar *et al.*, 1997, 1999). PNA binds T-antigen 20 times more strongly than it does lactose (Swamy *et al.*, 1991; Pereira *et al.*, 1976). Yet the only additional interactions T-antigen has with PNA are two water bridges; this is an interesting instance where specificity is generated apparently through water-mediated interactions. PNA also recognizes (with much lower affinity) *N*-acetyllactosamine (Galβ1-4GlcNAc), which contributes to the residual binding of the lectin at the cell surface, as the disaccharide occurs at the termini of saccharide chains of many glycoproteins (Langklide & Ørntoft, 1995). The structure of the complex of the lectin with this disaccharide is reported here. Also reported is the structure of the complex involving a monosaccharide (methyl-β-galactose). The four PNA–carbohydrate complexes, with a tetramer in the asymmetric unit in each case, including the two reported here, provide 16 crystallographically independent copies of the lectin's combining site. The structures thus provide a reliable means of delineating the salient features of the combining site and the water structure associated with it.

The legume lectins of known structure can be broadly classified into two groups in terms of specificity at the primary site: mannose/glucose specific and galactose (or *N*-acetylgalactosamine) specific. Most of the early crystallographic studies have been concerned with the first group. However, structures of a few galactose-specific lectins such as *E. corallodendron* (EcorL; Shaanan *et al.*, 1991; Elgavish & Shaanan, 1998), lectin IV from *G. simplicifolia* (GS4; Delbaere *et al.*, 1990, 1993), soybean agglutinin (SBA; Dessen *et al.*, 1995; Olsen *et al.*, 1997), basic winged-bean agglutinin (WBAI; Prabu *et al.*, 1998) and peanut lectin (Banerjee *et al.*, 1996; Ravishankar *et al.*, 1997, 1998) and/or their carbohydrate complexes have become available in recent years. A detailed comparative study of the protein–carbohydrate interactions in these structures and those in the complexes reported here should provide further insights into the geometrical basis of the galactose (*N*-acetylgalactosamine) specificity of these proteins and the water structure associated with the combining sites in them.

[1] Abbreviations: PNA, peanut agglutinin; GS4, lectin IV from *Griffonia simplicifolia*; EcorL, *Erythrina corallodendron* lectin; SBA, soybean agglutinin; WBAI, basic winged-bean agglutinin; r.m.s., root-mean-square; PTant, PNA–T-antigen complex; PLac, PNA–lactose complex; PLacNAc, PNA–*N*-acetyllactosamine complex; PMeGal, PNA–methyl-β-galactose complex; Gal, galactose; Glc, glucose; GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; LacNAc, *N*-acetyllactosamine; MeGal, methyl-β-galactose.

Table 1
Summary of data collection and refinement statistics.

Data collection.

	PLacNAc	PMeGal
Space group	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å)	129.38, 126.92, 76.07	128.95, 126.68, 77.28
Resolution (Å)	2.7	2.8
No. of observations	185971	573611
No. of unique reflections	34005	31250
Data completeness (%)	97.75	96.5
$R_{\rm merge}$ (%)	12.4	12.8

Refinement parameters.

	PLacNAc	PMeGal
Resolution limits (Å)	10–2.7	10–2.8
$R_{\rm cryst}$ ($F > 0\sigma F$) (%)	19.4	19.5
$R_{\rm free}$ (%)	24.4	24.7
No. of protein atoms	6976	6976
No. of solvent molecules	427	373
No. of carbohydrate atoms	104	52
Other ions	4 Mg^{2+}, 4 Ca^{2+}	4 Mg^{2+}, 4 Ca^{2+}

R.m.s. deviations from standard values.

	PLacNAc	PMeGal
Bond lengths (Å)	0.017	0.008
Bond angles (°)	2.1	1.5
Dihedral angles (°)	28	27

2. Materials and methods

2.1. Crystallization

The protein was prepared by affinity chromatography on cross-linked arabinogalactan (Majumdar & Surolia, 1978). The crystallization conditions for the complexes are similar and crystals were grown from a hanging drop of 5 mg ml^{-1} protein in 0.05 *M* sodium phosphate buffer pH 7.0 containing 0.2 *M* sodium chloride, 0.02% sodium azide, 1.5–10 m*M* of sugar and 12%(*w*/*v*) PEG 8000, equilibrated against 40%(*w*/*v*) PEG 8000 in the same buffer. All the crystals were isomorphous to the T-antigen (Ravishankar *et al.*, 1997) and lactose (Banerjee *et al.*, 1996) complexes (hereafter referred to as PTant and PLac, respectively). Data were collected at room temperature on a MAR Research imaging-plate system mounted on a Rigaku rotating-anode generator. Data for the *N*-acetyllactosamine and methyl-β-galactose complexes (hereafter referred to as PLacNAc and PMeGal, respectively) were processed using *XDS* (Kabsch, 1988) and *DENZO*/*SCALEPACK* (Otwinowski, 1993), respectively. The data-collection statistics are summarized in Table 1.

2.2. Structure refinement

The same refinement protocol was used for the two complexes. The structure of the PNA–lactose complex was used as the starting model, after removing water and sugar molecules. Refinement was carried out using *X-PLOR* (Brünger, 1992). Standard Powell minimization was carried out throughout with non-crystallographic symmetry imposed. The NCS restraints were progressively relaxed in the final cycle of refinement. Both $2F_o - F_c$ and $F_o - F_c$ maps calculated after a few cycles of refinement showed clear density for the carbohydrates. The carbohydrates were fitted into the map at this stage. All maps were inspected using *FRODO* (Jones, 1978). Only unrestrained group temperature factors were refined. Water molecules were added in steps from difference maps contoured at 2.5σ and $2F_o - F_c$ maps contoured at 0.8σ. Extensive use was made of omit-type maps (Bhat & Cohen, 1984; Vijayan, 1980) to inspect the model and water positions. The refinement was monitored at every stage using the $R_{\rm free}$ value. Engh and Huber parameters (Engh & Huber, 1991) were used throughout the refinement. The refinement statistics are summarized in Table 1.

2.3. Accuracy, conformation angles, displacement parameters and hydrogen bonds

The final models for both complexes had over 90% of the residues in the most-favoured region of the Ramachandran plot (Ramachandran *et al.*, 1963). No residues were in the disallowed region. The stereochemistry and geometry of the models were checked using *PROCHECK* (Laskowski *et al.*, 1993). The Luzzati plot (Luzzati, 1952) calculated at the end of the refinement indicated average coordinate errors of 0.28 and 0.26 Å in the atomic positions in PLacNAc and PMeGal, respectively. The average B values are 15.8 and 21.9 Å^2 for main-chain and side-chain atoms, respectively, in PMeGal. The corresponding values for PLacNAc are 25.4 and 25.9 Å^2, respectively. The average values for the sugar and water in PMeGal are 31.9 and 34.4 Å^2, respectively. The values in PLacNAc are 34.6 and 141.3 Å^2 for the Glc and GlcNAc moieties, respectively, and 43.8 Å^2 for the water. Hydrogen bonds were calculated using the program *CONTACT* in the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994). Acceptor-to-donor distances less than 3.6 Å were treated as hydrogen bonds if the angle at the H atom was greater than 120° in N—H···O contacts and the C—O···O angle was greater than 90° in the O···O contacts. A distance cutoff of 3.6 Å was only used in the case where water O atoms are involved.

3. Results and discussion

3.1. Tertiary and quaternary structure

The tertiary and quaternary structures of PNA in its complexes with methyl-β-galactose and *N*-acetyllactosamine are the same as those observed in its complexes with lactose (Banerjee *et al.*, 1996) and T-antigen (Ravishankar *et al.*, 1997). The striking feature of the tetrameric molecule remains its unusual quaternary structure which, contrary to expectations (Monod *et al.*, 1965; Klotz *et al.*, 1975), does not exhibit 222 or fourfold symmetry. The two dimers in the tetramer are related to each other by a non-crystallographic twofold symmetry, as are the monomers in each dimer. However, these twofold axes do not intersect, nor are they mutually perpendicular. Each subunit has a jelly-roll fold characteristic of legume lectins (Fig. 1). It consists of a six-stranded nearly flat β-sheet, a curved seven-stranded sheet, a small five-stranded sheet which plays an important role in connecting the two larger sheets, and a number of loops of differing length and conformation which account for about 54% of the polypeptide chain. There are two hydrophobic cores in the structure. The first one occurs between the six- and seven-stranded β-sheets. The loops, which connect the β-strands in the curved sheet and the curved sheet itself, enclose the second hydrophobic core (Banerjee *et al.*, 1996). There are 236 residues in each subunit. In the crystal structures, only the first 232 residues are seen clearly in the electron-density maps and the last four residues are presumably disordered. As illustrated in Fig. 1, the carbohydrate-binding site, one per subunit, occurs at one edge of the subunit. The root-mean-square (r.m.s) deviations between the C$^\alpha$ positions when pairs of subunits are superposed range between 0.28 and 0.29 Å in the PLacNAc complex, and between 0.06 and 0.07 Å in PMeGal. The r.m.s. deviation of the C$^\alpha$ atoms in all the four complexes range between 0.27

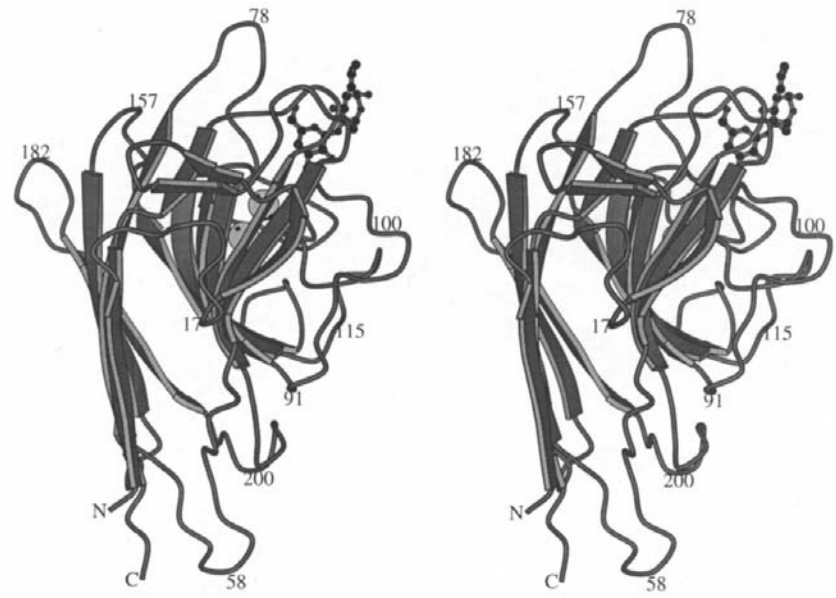


Figure 1
A stereo representation of a subunit of peanut lectin with *N*-acetyllactosamine shown as a ball-and-stick model at one edge. The figure was prepared using *MOLSCRIPT* (Kraulis, 1991).

and 0.59 Å when the respective subunits are superposed onto subunit 1 of the PTant complex. The metal-binding sites, including the coordination details, are the same in the four complexes.

3.2. PNA–carbohydrate interactions

The saccharide-binding region is made up of residues in the four loops 91–106, 125–135, 75–83 and 211–216. There is well defined electron density for the saccharides in both the crystal structures, as illustrated in Fig. 2, and this permits an unambiguous description of protein–carbohydrate interactions in the respective complexes (Fig. 3; Table 2). A constellation of four invariant hydrogen bonds occur in carbohydrate binding in all galactose-specific lectins. They are, in the PNA numbering scheme, Asp83 OD1–galactose O3, Asp83 OD2–galactose O4, Gly104 N–galactose O3 and Asn127 ND2–galactose O3. In addition, there are other invariant interactions, which include a stacking interaction with an aromatic side chain (Tyr125 in the case of PNA) and the proximity of an Ala residue seen in all the legume lectin–carbohydrate complexes.

In addition to the interactions referred to above, found in all legume lectins with galactose specificity, Asp80 OD2 is hydrogen bonded to galactose O6 in all four PNA–carbohydrate complexes. The same is true of the hydrogen bonds involving Ser211 OG. Of the four carbohydrate-binding loops (Young & Oomen, 1992), the fourth loop (211–216 in PNA) of variable length and composition is implicated in determining the specificity of individual lectins for different sugars (Sharma & Surolia 1997). Ser211 belongs to this loop. The water bridges involving W1 and W2 also occur in all four complexes (in both PLacNAc and PMeGal, galactose O2 is connected to Glu129 OE2 through W1 and to Gly104 N through W2; galactose O1 is connected to Ser211 OG and Gly213 N through W8 in PMeGal). The residues within 4 Å of the sugar (Asp80, Ala82, Asp83, Gly103, Gly014, Tyr125, Asn127, Ser211, Leu212 and Gly214 in PLacNAc; Asp80, Ala82, Asp83, Gly103, Gly104, Tyr125, Asn127, Ser211 and Gly214 in PMeGal) make about 40 van der Waals contacts with the galactose moiety in each of them. In fact, the position of this moiety is remarkably similar in all four complexes. Atomic deviations in the galactose moiety, when the 16 subunits are superposed, vary beween 0.1 and 0.2 Å. The four loops which constitute the carbohydrate-binding region also remain invariant in the four complexes. A second sugar ring exists in three of the four complexes. The protein–sugar interactions are the same in PLac and PLacNAc, as the acetamido group points to the solution. Lactose and LacNAc have the same orientation and conformation in the complexes (Fig. 4 and Table 2). Ser211 OG and Gly213 N, both from the variable fourth loop, interact with Glc O3. The number of residues within 4 Å of the sugar increases by one when the second sugar ring is added. The number of atoms in these residues making a contact within this distance limit increases from approximately 40 to 60. PNA binds T-antigen 20 times more strongly than it binds lactose. Yet the interactions between the galactose moiety and the lectin are similar in PLac (and PLacNAc) and PTant (Banerjee *et al.*, 1996; Ravishankar *et al.*, 1997, 1999). The same is true of the nonpolar contacts. The linkage and the conformation of the disaccharides are such that O4 of the second ring of T-antigen occupies the same position as O3 as in the second ring of the other two disaccharides (Fig. 4). O4 in T-antigen is therefore hydrogen bonded to Ser211 OG and Gly213 N instead of O3 as in the other two disaccharides. The major additional interactions T-antigen makes with PNA are two water bridges involving the carbonyl O atom of the acetamido group. One water molecule (W3) connects the O atom to Ile101 O, while the other (W4) connects it to Asn41 ND2 and to Leu212 N. These two water molecules also exist in the other complexes.

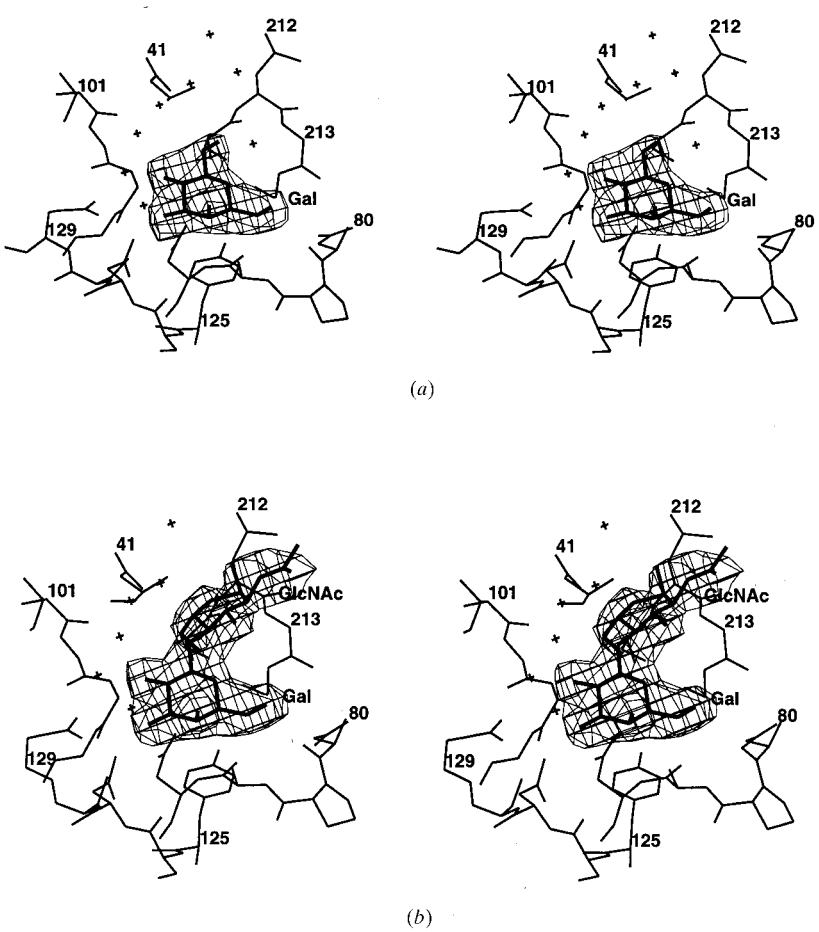


Figure 2
Omit map, shown in stereo, (*a*) contoured at 2.5σ for methyl-β-galactose and (*b*) contoured at 2.0σ for *N*-acetyllactosamine. The surrounding peptide stretches and water molecules are also shown. The figure was prepared using *FRODO* (Jones, 1978).

However, in these cases they interact only with the protein atoms, as there is no sugar atom in the immediate vicinity. Thus, the high specificity of PNA for T-antigen in comparison with that for other disaccharides appears to be generated by two water bridges.

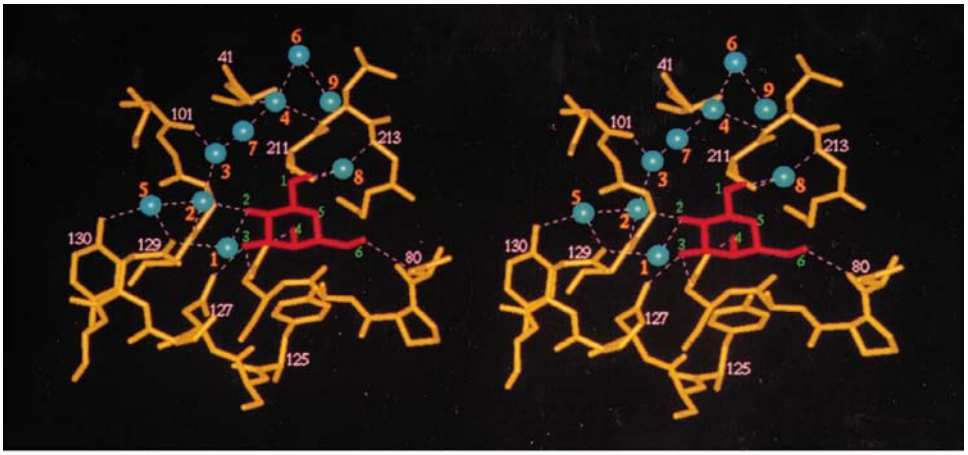
3.3. Water structure at the combining site

The role of water molecules in the carbohydrate-binding sites of lectins has received considerable attention in recent years (Quiocho, 1989; Quiocho *et al.*, 1989; Toone, 1994; Elgavish & Shaanan, 1998; Loris *et al.*, 1998). The generation of T-antigen specificity apparently through water bridges (Ravishankar *et al.*, 1997, 1999) is a dramatic illustration of this role. The fact that the two water molecules exist in the combining site even when the lectin is not bound to T-antigen suggests that at least some of the water molecules could be legitimately considered as an extension of the combining site. Therefore, a detailed study of the water molecules in the combining site appeared worthwhile.

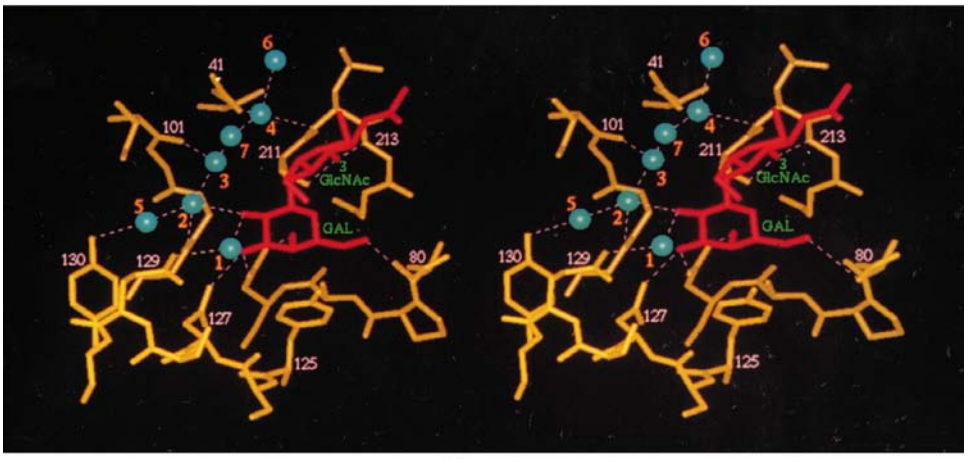
The remaining 15 subunits, along with the hydration shells, were superposed onto subunit A of the T-antigen complex. All water molecules within 5 Å from any carbohydrate atom was considered to belong to the combining site. A superposition of such water molecules in the 16 subunits is given in Fig. 5. Most water molecules appeared in clusters, indicating the relative invariance of water positions in the combining site. A cluster was defined as a set of water molecules from different subunits, in which all water O atoms are within a sphere of 1.5 Å radius with the average position as the centre. These clusters, named W1–W9, are listed in Table 3.

Water clusters W1–W6 are not displaced by sugar molecules in any of the four complexes and, with four subunits in each complex, a maximum of 16 water molecules could occur in each cluster. The full potential is realised in W2 and W4, while the number is 15, 14 and 11 in W1, W3 and W5, respectively. The occupancy is rather low in W6, but the corresponding water molecule occurs in at least two subunits in each complex. The acetamido group in the second sugar ring displaces W7 in the T-antigen complex. It occurs in at least two subunits in the remaining three complexes. W8 and W9 are displaced by the second ring and therefore occur only in PMeGal. W8 occurs in all four subunits and W9 in two.

W1 and W2 are involved in bridging Gal O2 to Glu129 OE1 and Gly104 N in all the complexes. W5 is hydrogen bonded to W2 on the one hand and Glu129 OE1 and Tyr130 OH on the other. Thus W1, W2 and W5 form a network which bridges galactose and peanut lectin at the primary binding site. W3 and W4 are also hydrogen bonded to the protein: W3 to Ile101 O and W4 to Asn41 ND2 and Leu212 N. In T-antigen, both hydrogen bond to GalNAc O7. The water bridges thus generated lead to the specificity of PNA for T-antigen at the disaccharide level, as mentioned earlier. W6 is hydrogen bonded to W4, but not to any protein or sugar atom. In the three complexes other than PTant, where it does not exist, W7 bridges W3 and W4. W7 again does not interact with any protein or sugar atom. W8 and W9 occur only in PMeGal. The former interacts with Ser211 OG,



(*a*)



(*b*)

Figure 3
Stereo diagram of the protein–carbohydrate interactions in (*a*) PMeGal complex and (*b*) LacNAc complex. The sugar (red) is labelled in green. The interacting residues (yellow) are numbered in white. Hydrogen bonds are represented by white dotted lines. The water molecules (cyan-coloured spheres) are numbered in orange. See Table 2 for details. This and the subsequent figures were prepared using *Insight*II (Biosym Technologies, 1993)

Gly213 N and Gal O1, forming a water bridge between the protein and the glycosidic O atom, while the latter is hydrogen bonded to W6 only.

Interestingly, after superposition the average position of W7 is at a distance of 1.1 Å from GalNAc O7 of T-antigen in its complexes. Thus, when T-antigen is not present in the combining site, the position corresponding to GalNAc O7 is occupied by a water molecule. As mentioned previously, O3 in the second sugar ring in PLac and PLacNAc occupy nearly the same position as O4 in the second ring in PTant. The distance between the average positions of O3 and O4 after superposition is as low as 0.9 Å. The average distance of W8 from the average positions of O3 and O4 in the second ring are 0.5 and 0.8 Å, respectively. Thus, in the absence of the second ring, a water molecule occupies the position of the hydroxyl group in the complex. W9 has a similar relationship with O1 in the second ring. O1 is attached to the anomeric carbon. The moderate resolution of the PLacNAc structure and the location of O1 at the terminus of the second GlcNAc ring meant that the configuration at the anomeric carbon could not be ascertained with absolute certainty. In PTant, O1 with a β configuration appeared to fit better, although both configurations were compatible with the electron-density map. W9 is at an average distance of 1.0 Å from the refined position of O1 in the β configuration in this structure. The corresponding average distance from the refined position of O1 in the α configuration is 1.25 Å. The average distances of W9 from O1 in the α configuration in PLac and PLacNAc are 0.59 and 1.36 Å, respectively, while its distance from O1 in the β configuration in one of the PLac subunits is 2.3 Å. Thus, irrespective of the configuration at the anomeric carbon, W9 in PMeGal is close to O1 in the second ring in the other three complexes. The occurrence of ordered water molecules in the unliganded form at positions corresponding to hydroxyl groups in the liganded form is also found in other carbohydrate-binding proteins. In particular, it has been shown that O3 and O4 of galactose at the primary binding site are replaced by water molecules in unliganded EcorL and GS4 (Elgavish & Shaanan, 1998; Delbaere *et al.*, 1990, 1993). The material presented above demonstrates that the water molecules in the carbohydrate-binding region mimic the ligand to a substantial extent not only at the primary site, but also in the regions adjacent to it.

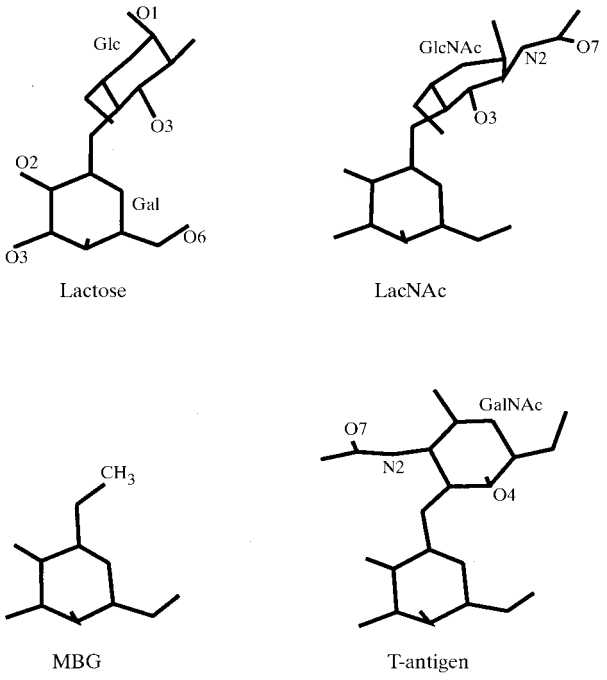


Figure 4
Lactose, *N*-acetyllactosamine, methyl-β-galactose and the T-antigenic disaccharide, shown with the galactose moiety oriented in the same manner. The saccharide conformations are as they occur in the respective complexes with PNA.

Table 2
Protein–sugar hydrogen bonds.

N-acetyllactosamine complex.

Protein atom	Sugar atom	Distance (Å) in subunit			
		1	2	3	4
Asp83 OD1	Galactose O3	2.6	2.5	2.8	2.8
Gly104 N	Galactose O3	3.2	3.4	3.0	2.9
Asn127 ND2	Galactose O3	2.9	2.8	2.8	2.7
Asp83 OD2	Galactose O4	2.7	2.8	3.0	2.8
Ser211 OG	Galactose O4	2.7	2.8	2.9	3.0
Ser211 OG	Galactose O5	3.0	3.3	3.6	3.3
Asp80 OD2	Galactose O6	3.2	3.5	3.3	3.4
Ser211 OG	GlcNAc O3	3.1	3.3	3.6	3.2
Gly213 N	GlcNAc O3	2.7	2.9	3.3	2.7

Methyl-β-galactose complex.

Protein atom	Sugar atom	Distance (Å) in subunit			
		1	2	3	4
Asp83 OD1	Galactose O3	2.6	2.7	2.7	2.6
Gly104 N	Galactose O3	3.2	3.0	2.9	2.9
Asn127 ND2	Galactose O3	2.8	2.9	2.9	2.8
Asp83 OD2	Galactose O4	2.7	2.7	2.8	2.7
Ser211 OG	Galactose O4	2.8	2.7	2.8	2.7
Ser211 OG	Galactose O5	3.2	3.5	3.6	3.3
Asp80 OD2	Galactose O6	3.0	3.2	3.3	3.0

3.4. Comparison with the water structure in the combining sites of other Gal/GalNAc-specific lectins

Protein–carbohydrate interactions at the primary site are remarkably similar in Gal/GalNAc-specific lectins, although differences in detail exist (Banerjee *et al.*, 1996; Prabu *et al.*, 1998; Elgavish & Shaanan, 1998; Loris *et al.*, 1998). It is, however, of interest to examine whether this similarity also extends to the water structure. The water molecules associated with the combining sites of PNA, W1 and W2 are of particular importance, as they bridge O2 of galactose to Glu129 OE1 and Gly104 N, respectively. A water molecule equivalent to W1, with a similar role, exists in the lactose, *N*-acetylgalactosamine and *N*-acetyllactosamine complexes of EcorL (Elgavish & Shaanan, 1998), SBA (Dessen *et al.*, 1995; Olsen *et al.*, 1997) and GS4 (Delbaere *et al.*, 1990, 1993), although an asparaginyl residue instead of a glutamyl residue interacts with the water

molecule. This asparaginyl residue is invariant in all legume lectins and is involved in a direct protein–carbohydrate hydrogen bond. A similar water site also exists in the methyl-α-galactose complex with WBAI, except that it is a little over 4 Å from galactose O2. An equivalent of W2, with a similar role, exists in WBAI–methyl-α-galactose and SBA. It occurs only in the lactose complex of EcorL. The position corresponding to W2 is occupied by a sugar atom in the complexes of GS4 (Delbaere *et al.*, 1993). An equivalent water, however, exists in sugar-free GS4. Thus, the similarity of the PNA–sugar interactions with those in other Gal/GalNAc-specific lectins also extend to a substantial degree to water bridges.

3.5. Protein–carbohydrate interactions and sugar specificity

Structural data on protein–carbohydrate interactions and the corresponding thermodynamic data are notoriously difficult to correlate on a one-to-one basis (Toone, 1994), partly because the interatomic contacts seen in structures are more related to binding enthalpies than to free energies. The data, however, do provide a qualitative rationale for the observed specificities. Among the Gal/GalNAc-specific lectins, PNA is the only one which does not bind GalNAc at all (Swamy *et al.*, 1991; Young *et al.*, 1984). WBAI, EcorL and SBA bind Gal and GalNAc, although the affinities are not the same for the sugars in each case (De Boeck *et al.*, 1984; Schwarz *et al.*, 1991; Surolia *et al.*, 1996; Young *et al.*, 1984). GS4 does not bind monosaccharides, but Gal and GalNAc can occur at the primary site. In order to facilitate comparison among them, the GalNAc moiety in the EcorL–GalNAc complex was built into the PNA binding site by superposition of the galactose ring in the EcorL–GalNAc complex and the PMeGal complex. Models of the GalNAc complex of WBAI, SBA and GS4 were similarly constructed. The water-mediated interaction between the protein and the sugar involving W2 is disrupted by the acetamido group in all the models. The short contacts between the group and the water molecule is most severe in PNA. In all the structures except PNA, however, there is the possibility of a compensating hydrogen bond between the O atom of the acetamido group and the main-chain N atom of a glycyl residue (107 in EcorL, 105 in WBAI, 106 in SBA and 107 in GS4). The nature of the combining site near the acetamido group in PNA is also different from that in the other four lectins. An invariant aspariginyl residue (Asn127 in PNA) and a glycyl residue are close to the C atoms of the acetamido group in all the five lectins. The other residue close to the C atoms in PNA is Glu129. Tyr106, Tyr108 and Trp 135 are close to the C atoms in EcorL. Similar residues in WBAI, SBA and GS4 are Tyr106 and Trp130, Tyr107 and Trp132, and Phe108 and Trp138, respectively. Thus, favourable hydrophobic interactions with the acetamido group exist in EcorL, WBAI, SBA and GS4, but not in PNA.

PNA understandably has nearly the same affinity for galactose and methyl-β-galactose (Swamy *et al.*, 1991; Young *et al.*, 1984). It has higher affinity for lactose on account of the additional hydrogen bonds the second ring makes with the lectin (Banerjee *et al.*, 1996). The lectin binds the T-antigenic disaccharide still more strongly, primarily on account of the additional water bridges (Ravishankar *et al.*, 1997, 1999). The reason for its lower affinity for *N*-acetyllactosamine compared with that for lactose is not immediately obvious, as the protein–carbohydrate interactions in PLac and PLacNAc are the same. The acetamido group in PLacNAc points to the solution and the *N*-acetylglucosamine moiety has high displacement parameters. In fact, the average atomic *B* factors for the atoms in this moiety is as high as 141 Å^2, compared with 44.8 Å^2 for the

Table 3
Clusters in the binding site of PNA.

W1 to W6 can occur in all four structures while W7 can occur in all but PTant. W8 and W9 can only occur in PMeGal.

Cluster	Atoms interacting with	Expected number	Number observed	Max deviation from average (Å)
W1	E129 OE1, Gal O2	16	15	1.5
W2	G104 N, Gal O2	16	16	1.4
W3	I101 O, GalNAc O7	16	14	1.1
W4	N41 ND2, L212 N, GalNAc O7	16	16	1.1
W5	Y130 OH, W2	16	11	1.0
W6	W4, W9	16	9	1.5
W7	W3, W4	12	7	1.0
W8	S211 OG, G213 N, MeGal O1	4	4	0.8
W9	W6	4	2	0.7

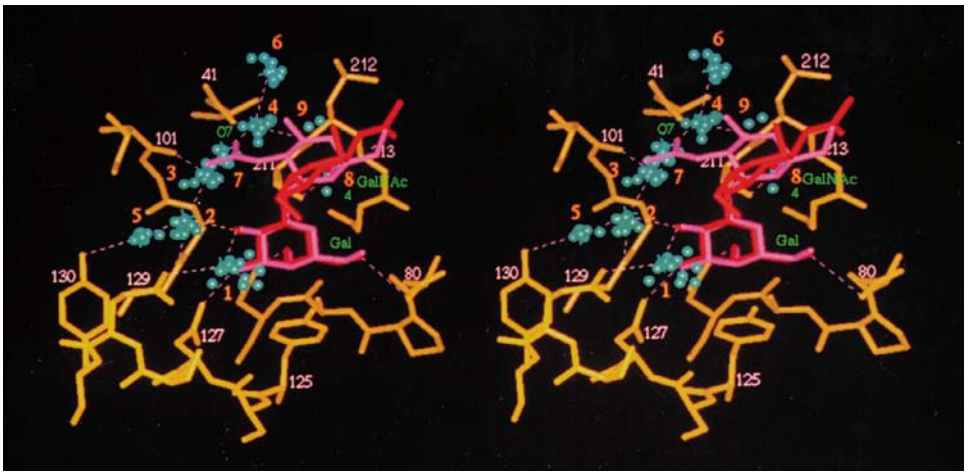


Figure 5
Stereo diagram of the water clusters (cyan), numbered in orange, in the binding site after superposition of the 16 subunits in the four complexes of PNA. The numbering in green is for the T-antigenic disaccharide (pink). The *N*-acetyllactosamine is coloured red. The surrounding peptide stretches (yellow) are numbered in white. Hydrogen bonds are represented by white dotted lines.

glucose moiety in PLac. The average value for the remaining sugar atoms in PLacNAc and PLac are 34.6 and 28.8 Å^2, respectively. Thus, the exposed nature of the acetamido group and its consequent mobility appears to percolate to the sugar rings, rendering the binding of PNA to LacNAc weaker.

The data were collected at the National Area-Detector Facility, supported by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT). Facilities at the Supercomputer Education and Research Centre and the Interactive Graphics-based Facility and Distributed Information Centre (both supported by DBT) were used in the work. The work was funded by the DST.

References

Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1996). *J. Mol. Biol.* **259**, 281–296.

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A. & Vijayan, M. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 227–231.

Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1513–1524.

Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.

Biosym Technologies (1993). *InsightII User Guide.* Biosym Technologies, San Diego, CA, USA.

Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rouge, P. & Fontecilla-Camps, J.-C. (1990). *J. Mol. Biol.* **214**, 571–584.

Bourne, Y., Mazurier, J., Legrand, D., Rouge, P., Montreuil, J., Spik, G. & Cambillau, C. (1994). *Structure*, **2**, 209–219.

Bourne, Y., Rouge, P. & Cambillau, C. (1990). *J. Biol. Chem.* **265**, 18161–18165.

Bourne, Y., Rouge, P. & Cambillau, C. (1992). *J. Biol. Chem.* **267**, 197–203.

Bourne, Y., Roussel, A., Frey, M., Rouge, P., Fontecilla-Camps, J-C. & Cambillau, C. (1990). *Proteins Struct. Funct. Genet.* **8**, 365- 376.

Brünger, A. T. (1992). *X-PLOR Version* 3.1. *A System for X-ray Crystallography and NMR.* Yale University, Connecticut, USA.

Chrispeels, M. J. & Raikhel, N. V. (1991). *Plant Cell*, **3**, 1–9.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

De Boeck, H., Lis, H., van Tilbeurgh, H., Sharon, N. & Loontiens, F. G. (1984). *J. Biol. Chem.* **259**, 7067–7074.

Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Pearlstone, J. W., Carpenter, M. R., Smillie, L. B., Nikrad, P. V., Spohr, U. & Lemieux, R. U. (1990). *Can. J. Chem.* **68**, 116–121.

Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S. & Dauter, Z. (1993). *J. Mol. Biol.* **230**, 950–965.

Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F. & Sacchettini, J. C. (1995). *Biochemistry*, **34**, 4933–4942.

Drickamer, K. (1997). *Structure*, **5**, 465–468.

Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E. & Suddath, F. L. (1986). *J. Biol. Chem.* **261**, 16518–16527.

Elgavish, S. & Shaanan, B. (1997). *Trends Biochem. Sci.* **12**, 462–467.

Elgavish, S. & Shaanan, B. (1998). *J. Mol. Biol.* **277**, 917–932.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Hardman, K. D. & Ainsworth, C. F. (1972). *Biochemistry*, **11**, 4910–4919.

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.

Klotz, I. M., Darnall, D. W. & Langerman, N. R. (1975). *The Proteins*, Vol. 3, edited by H. Neurath & R. L. Hill, pp. 293–411. New York: Academic Press

Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.

Langklide, N. C. & Ørntoft, T. F. (1995). *Glycoconj. J.* **12**,567.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). *Biochim. Biophys. Acta*, **1383**, 9–36.

Loris, R., Maes, D., Poortmans, F., Wyns, L. & Bouckaert, J. (1996). *J. Biol. Chem.* **271**, 30614–30618.

Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. & Wyns, L. (1993). *Biochemistry*, **32**, 8772–8781.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Majumdar, T. & Surolia, A. (1978). *Prep. Biochem.* **8**, 119–131.

Monod, J., Wyman, J. & Changeux, J.-P. (1965). *J. Mol. Biol.* **12**, 88–118.

Naismith, J. H., Emmerich, C., Habash, J., Harrop, S., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb (Gilboa), A. J. & Yariv, J. (1994). *Acta Cryst.* D**50**, 847–858.

Naismith, J. H. & Field, R. A. (1996). *J. Biol. Chem.* **271**, 972–976.

Olsen, L. R., Dessen, A., Gupta, D., Sabesan, S., Sacchettini, J. C. & Brewer, C. F. (1997). *Biochemistry*, **36**, 15073–15080.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Pereira, M. E. A., Kabat, E. A., Lotan, R. & Sharon, N. (1976). *Carbohydr. Res.* **51**, 107–118.

Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. & Suguna, K. (1998). *J. Mol. Biol.* **276**, 787–796.

Quiocho, F. A. (1989). *Pure Appl. Chem.* **61**, 1293–1306.

Quiocho, F. A., Wilson, D. K. & Vyas, N. K. (1989). *Nature (London)*, **340**, 404–407.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.

Ravishankar, R., Ravindran, M., Suguna, K., Surolia A. & Vijayan, M. (1997). *Curr. Sci.* **72,** 855–861.

Ravishankar, R., Ravindran, M., Suguna, K., Surolia, A. & Vijayan, M. (1999). *Curr. Sci*. **76**, 1393.

Ravishankar, R., Surolia, A., Vijayan, M., Lim, S. & Kishi, Y. (1998). *J. Am. Chem. Soc.* **120**, 11297–11303.

Reeke, G. N. Jr & Becker, J. W. (1986). *Science*, **234**, 1108–1111.

Reeke, G. N. Jr, Becker, J. W. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1525–1547.

Rini, J. M. (1995). *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–557.

Rini, J. M., Hardman, K. D., Einspahr, H., Suddath, F. L. & Carver, J. P. (1993). *J. Biol. Chem.* **268**, 10126–10132.

Schwarz, F. P., Puri, K. & Surolia, A. (1991). *J. Biol. Chem.* **266**, 24344–24350.

Shaanan, B., Lis, H. & Sharon, N. (1991). *Science*, **254**, 862–866.

Sharma, V., Srinivas, V. R., Adhikari, P., Vijayan, M. & Surolia, A. (1998). *Glycobiology*, **10**, 1007–1012.

Sharma, V. & Surolia, A. (1997). *J. Mol. Biol.* **267**, 433–445.

Sharma, V., Vijayan, M. & Surolia, A. (1996). *J. Biol. Chem.* **271**, 21209–21213.

Sharon, N. & Lis, H. (1989*a*). *Lectins*. London: Chapman & Hall.

Sharon, N. & Lis, H. (1989*b*). *Science*, **246**, 227–246.

Surolia, A., Sharon, N. & Schwarz, F. P. (1996). *J. Biol. Chem.* **271**, 17697–17703.

Swamy, M. J., Gupta, D., Mahanta, S. K. & Surolia, A. (1991). *Carbohydr. Res.* **213**, 59–67.

Toone, E. J. (1994). *Curr. Opin. Struct. Biol.* **4**, 719–728.

Vijayan, M. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 19.01–19.26. Bangalore: Indian Academy of Sciences.

Weis, W. I. & Drickamer, K. (1996). *Annu. Rev. Biochem.* **65**, 441–473.

Young, N. M & Oomen, R. P. (1992). *J. Mol. Biol.* **228**, 924–934.

Young, N. M., Watson, D. C. & Williams, R. E. (1984). *Biochem. J.* **222**, 41–48.