Structural Basis of the Carbohydrate Specificities of Jacalin: An X-ray and Modeling Study

A. Arockia Jeyaprakash[1], S. Katiyar[1], C. P. Swaminathan[1], K. Sekar[2] A. Surolia[1] and M. Vijayan[1]*

[1]*Molecular Biophysics Unit UGC Centre of Advanced Study, Indian Institute of Science, Bangalore 560 012 India*

[2]*Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India*

**Corresponding author*

The structures of the complexes of tetrameric jacalin with Gal, Me-α-GalNAc, Me-α-T-antigen, GalNAcβ1-3Gal-α-*O*-Me and Galα1-6Glc (mellibiose) show that the sugar-binding site of jacalin has three components: the primary site, secondary site A, and secondary site B. In these structures and in the two structures reported earlier, Gal or GalNAc occupy the primary site with the anomeric carbon pointing towards secondary site A. The α-substituents, when present, interact, primarily hydrophobically, with secondary site A which has variable geometry. O–H···π and C–H···π hydrogen bonds involving this site also exist. On the other hand, β-substitution leads to severe steric clashes. Therefore, in complexes involving β-linked disaccharides, the reducing sugar binds at the primary site with the non-reducing end located at secondary site B. The interactions at secondary site B are primarily through water bridges. Thus, the nature of the linkage determines the mode of the association of the sugar with jacalin. The interactions observed in the crystal structures and modeling based on them provide a satisfactory qualitative explanation of the available thermodynamic data on jacalin–carbohydrate interactions. They also lead to fresh insights into the nature of the binding of glycoproteins by jacalin.

Keywords: Moraceae lectin; carbohydrate specificity; O/C–H···π interactions; water bridges; glycoproteins

Introduction

Lectins are characterized by their ability to specifically bind diverse sugars. They exert their biological effects on cell–cell recognition, host–pathogen interactions, malignancy, cellular signaling and differentiation, and immune response through binding to appropriate carbohydrates.[1–5] Although this group of proteins exists in all forms of life, the most thoroughly studied lectins are those extracted from plants.[6–8] On the basis of molecular structure, plant lectins belong to five different families, one of which is characterized by the β-prism fold†. Those with β-prism fold include jacalin, *Maclura pomifera* agglutinin, artocarpin and hetuba. The first lectin to be shown to have this fold is jacalin, one of the two lectins from jackfruit (*Artocarpus integrifolia*) seeds.[9] The other lectin, artocarpin, also assumes the β-prism I fold. Both the lectins are tetrameric and have 57% sequence identity.[10] However, artocarpin is mannose-specific at the monosaccharide level while jacalin is essentially galactose-specific, although it is known to bind methyl-α-mannose as well.[11] Artocarpin is non-glycosylated, whereas jacalin is a glycoprotein containing approximately 7–10% carbohydrate. Each subunit of artocarpin is composed of a single polypeptide chain, while the jacalin subunit consists of a long α-chain and a short β-chain. The two chains are produced by post-translational proteolysis.[12–14] The new amino terminus, that of the α-chain, has been shown to be important for the higher affinity of jacalin for galactose at the primary binding site.

Jacalin exhibits interesting biological properties.[15] For instance, it selectively binds IgA_1 and other glycoproteins such as carcinoma-related mucins. It is also selectively mitogenic for human $CD4^+$ T-cells, leading to its use in AIDS

Abbreviation used: XRD1, X-ray diffraction 1.

E-mail address of the corresponding author: mv@mbu.iisc.ernet.in

research.[16–23] It has also been used as a tool to investigate IgA_1 nephropathy, to identify O-linked glycoproteins and in histochemistry. Each subunit of the tetrameric lectin, M_r 66,000, binds one sugar molecule. The details of the protein–sugar interactions at the primary site were originally defined from the crystal structure of the jacalin–methyl-α-galactose (Me-α-Gal) complex.[9] At the disaccharide level, jacalin is specific to the tumor-associated sugar Galβ1-3GalNAc, generally known as T-antigen.[24,25] The structure of the complex of this disaccharide with jacalin confirmed the details of interactions at the primary site defined earlier and also provided information on the nature of the interactions of the second residue with a secondary binding site on the lectin. The structure also provided a plausible model of the binding of the lectin to T-antigen O-linked to seryl or threonyl residues.[26] Here, we report the crystal structures of the complexes of jacalin with galactose (Gal), methyl-α-*N*-acetyl galactosamine (Me-α-GalNAc), Galβ1-3GalNAc-α-*O*-Me (Me-α-T-antigen), GalNAcβ1-3Gal-α-*O*-Me and Galα1-6Glc (mellibiose). These structures along with the structures of the complexes with Me-α-Gal and T-antigen reported earlier, provide a comprehensive picture of jacalin–carbohydrate interactions. In particular, an additional, flexible secondary binding site has been identified. Detailed analysis of the complexes reveals how the nature of the linkage governs the use of the primary and the two secondary binding sites. The structures provide insights into the binding of glycoproteins by jacalin. The interactions observed in them lead to a qualitative explanation of the available thermodynamic data on jacalin–carbohydrate interactions.

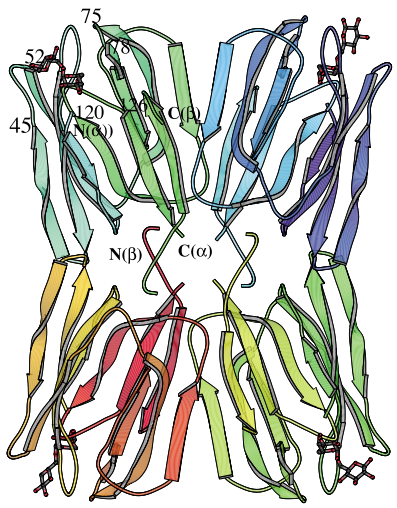
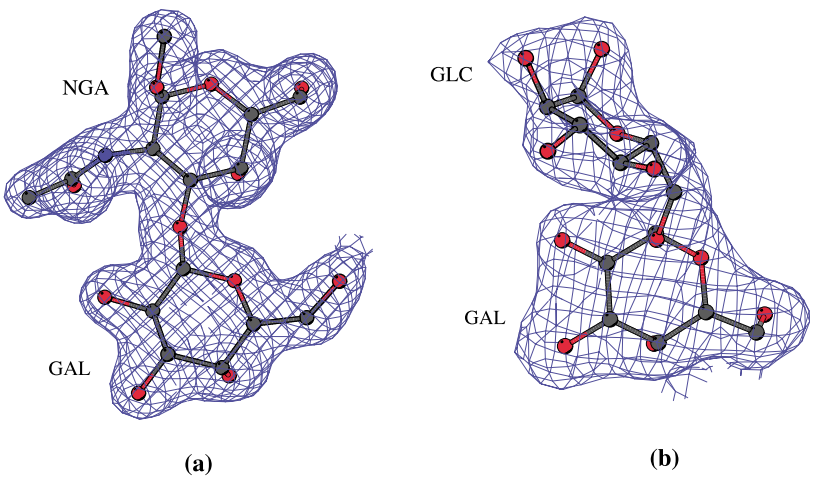


Figure 1. Structure of tetrameric jacalin with bound melibiose. The loops involved in sugar binding are numbered in one subunit. The termini of the two chains are also indicated.

Results and Discussion

Overall features

Jacalin is a two-chain lectin, consisting of a heavy chain (α) of 133 amino acid residues and a light chain (β) of 20 amino acid residues.[13,14] Each subunit has a 3-fold symmetric β-prism fold made up of three four-stranded β sheets (Figure 1). The 3-fold symmetry is not, however, reflected in the sequence. Two sheets have Greek-key topology. The third also is Greek key-like, but with a break in the outer loop caused by the post-translational cleavage which gives rise to the α and the β chains. The strands in the sheets and the sheets themselves are connected by loops. The loops at one end of the prism constitute the carbohydrate-binding site. The subunits associate in the tetramer with 222 symmetry. The complexes of jacalin with Me-α-Gal and T-antigen have already been reported by us. The high-resolution structure of the T-antigen complex revealed the presence of *cis* peptides and the heterogeneity in the amino acid composition among the subunits. The high-resolution structures of Gal and Me-α-T-antigen complexes, reported here, also confirm the above observations. The r.m.s. deviation value of C^{α} positions when pairs of subunits from the same complex are superposed



(a) **(b)**

Figure 2. Electron densities in the $2F_o - F_c$ map contoured at 1σ for (a) Galβ1-3GalNAc-α-*O*-Me (Me-α-T-antigen). (b) Galα1-6Glc (melibiose).

Table 1. Modes of sugar binding in different complexes

Complex	Primary site	Secondary site A	Secondary site B
Gal	Gal	–	–
Me-α-Gal	Gal	Me	–
Me-α-GalNAc	GalNAc	Me	–
Galβ1-3GalNAc	GalNAc	–	Gal
Galβ1-3GalNAc-α-*O*-Me	GalNAc	Me	Gal
GalNAcβ1-3Gal-α-*O*-Me	Gal	Me	GalNAc
Galα1-6Glc	Gal	Glc	–

varies between 0.03 Å and 0.27 Å. The structure of the lectin is essentially the same in all the complexes. The subunits in one structure superpose on those in the other with r.m.s. deviation values ranging from 0.08 Å to 0.38 Å.

Jacalin–carbohydrate interactions

The carbohydrate-combining site of jacalin is made up of loops 46–52, 76–82, 122–125, which connect the inner strands of Greek keys II, I and III, respectively, and the N terminus of the α-chain (Figure 1). The well-defined electron density for the bound carbohydrates (see Figure 2 for representative densities) in all the subunits permits a detailed characterization of their interactions with the lectin. A thorough examination of the protein–ligand interactions in the complexes shows that the binding site is composed of three subsites (Table 1, Figure 3). The primary binding site is composed of the side-chains of Phe47, Tyr78 and Asp125, backbone nitrogen and oxygen atoms of Tyr122 and Trp123 and the free amino terminus of the α-chain. Secondary site A involves the side-chains of Tyr78, Tyr122 and Trp123. Thus, the primary binding site and the secondary binding site A have the side-chain of Tyr78 in common. Secondary site B is composed of backbone nitrogen and oxygen atoms of Val79, Ser119 OG and the carboxy-terminal region of the β-chain. The primary binding site of the lectin is occupied by Gal/GalNAc/Man in the monosaccharide complexes with the anomeric carbon atom pointing towards secondary site A. An α-substituent, even if it is a sugar residue, does not disturb this arrangement. However, β-substitution leads to unacceptable steric clashes with the protein. Therefore, in β-substituted disaccharides, the reducing sugar binds at the primary site with the non-reducing end located at secondary site B (Figure 4). This situation obtains in the disaccharides, other than melibiose, considered here and in T-antigen considered earlier.

Interactions at the primary site

The interactions of a given sugar at the primary site are the same in all the subunits (Table 2). As in the case of the Me-α-Gal complex,[9] the terminal amino group of the α-chain hydrogen bonds to O3 and O4, Tyr122 N to O5 and O6, the NH and O of Trp123 to O6, and the side-chain oxygen atoms of Asp125 to O4 and O6, in all the complexes. Also, the side-chain of Tyr78 stacks on the B face of Gal in all of them. The high-resolution structure of the Gal complex also reveals an equal preference for α as well as β-D-galactose in three subunits as observed in two subunits in the T-antigen complex.[26] In the Gal complex, the β anomer is stabilized through an O–H···π interaction involving Tyr122 (Figure 5(a)). Me-α-GalNAc in its complex has two additional interactions with the lectin. They are between the methyl group of the sugar and the π electrons of the aromatic side-chain Tyr122 (discussed later), and between the carbonyl group of the sugar acetamido group

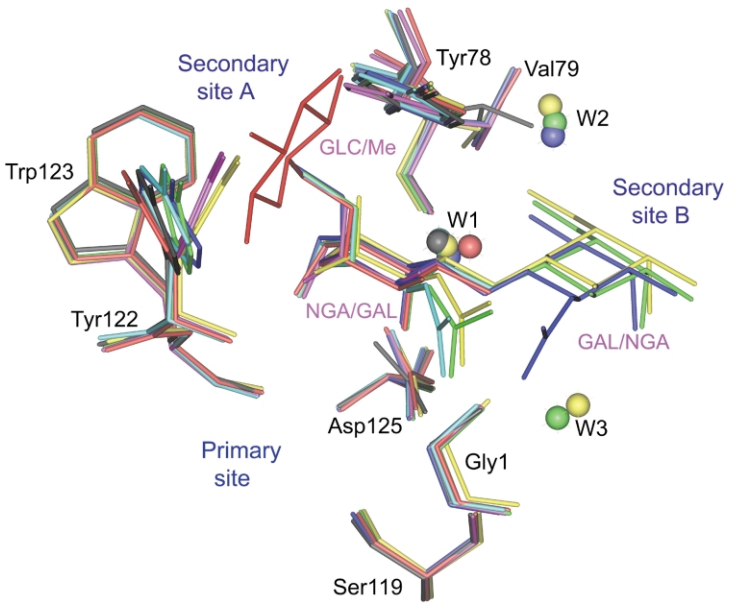
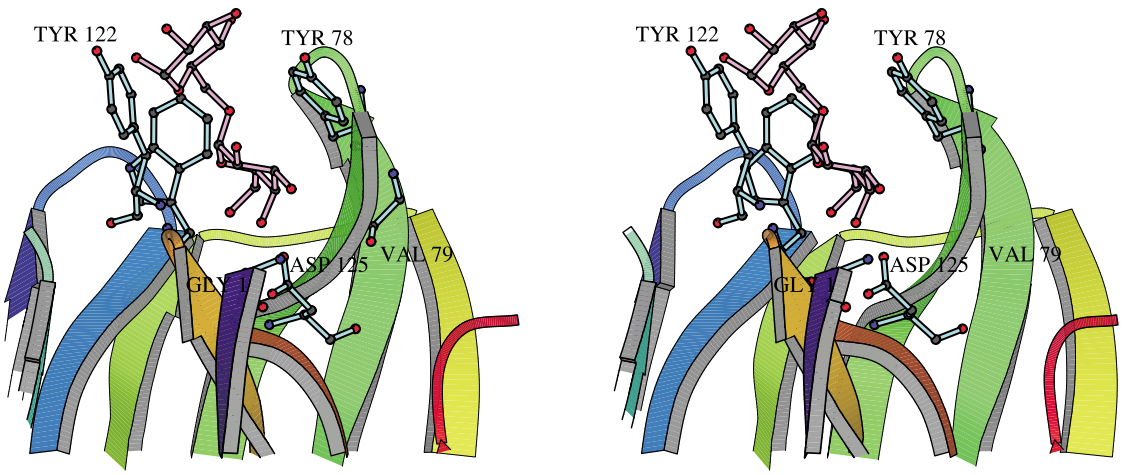
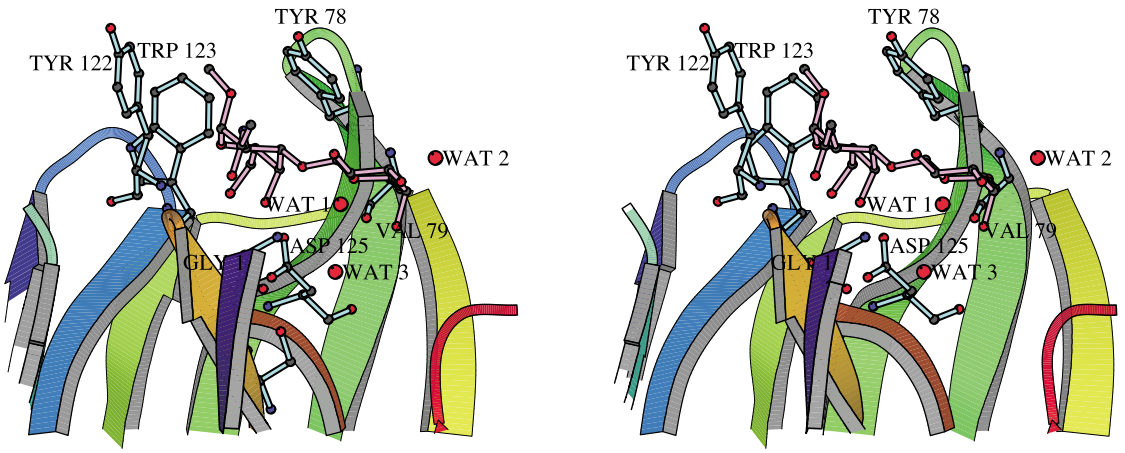


Figure 3. Composite view of the sugar-binding site of jacalin. The protein residues which are involved in sugar binding are shown along with the bound carbohydrate. Phe47 is not shown in the Figure for the clarity of the picture. The colours used are pink, black, cyan, red, yellow, blue and green, respectively, for the complexes in the order in which they are listed in Table 1.



(a)



(b)

Figure 4. Stereoviews of the binding site of jacalin with bound (a) Galα1-6Glc (melibiose) and (b) Galβ1-3GalNAc-α-*O*-Me (Me-α-T-antigen).

and the amino terminus of the α-chain. The interactions of melibiose and Me-α-T-antigen at the primary site are identical with those of Gal and Me-α-GalNAc, respectively. All the interactions observed in the Me-α-Gal complex are also present in the complex of GalNAcβ1-3Gal-α-*O*-Me. The latter has an additional interaction between O7 of GalNAc and the amino terminus of the α-chain. This is the only case where both the residues in a disaccharide interact with the primary site.

Interactions at secondary site A

The secondary binding site, involving the side-chains of Tyr78, Tyr122 and Trp123, is utilized by methyl groups or sugar residues α-linked to Gal/GalNAc located on the primary site. In the absence of an α-linked group, as in the complexes with Gal and T-antigen, χ^{21} ($C^{\alpha}-C^{\beta}-C^{\gamma}-C^{\delta}$) of Tyr122 is in the range of $-8°$ to $-46°$. Tyr122 is now involved in an O–H$\cdots\pi$ interaction which stabilizes the β anomer (Figure 5(a)). In the presence of an α-linked group, the side-chain of Tyr122 moves to accommodate the group (Figures 3 and 5(b)). In particular, χ^{21} now is in the range of 16–38°. This creates a groove involving the side-chains of Tyr78 and Tyr122 and the α-linked group nestles among the side-chain of Tyr78, Tyr122 and Trp123. This involves an increased burial of apolar accessible surface area (Table 3). The methyl group also

Table 2. Jacalin–carbohydrate interactions in the complexes with (1) Gal, (2) Me-α-GalNAc, (3) Galβ1,3GalNAcαMe, (4) GalNAcβ1,3GalαMe, (5) Galα1,6Glc

Interaction	1	2	3	4	5
Direct interactions at the primary site					
Gal/GalNAc O3···N Gly1	2.95	2.96	3.01	2.85	2.96
Gal/GalNAc O4···N Gly1	2.88	2.89	2.90	2.92	2.85
Gal/GalNAc O4···OD1 Asp125	2.79	2.63	2.65	2.70	2.80
Gal/GalNAc O4···OD2 Asp125	3.04	3.02	3.04	3.03	3.00
Gal/GalNAc O5···N Tyr122	3.08	3.10	3.07	3.07	3.04
Gal/GalNAc O6···N Tyr122	3.02	3.09	2.95	3.13	2.91
Gal/GalNAc O6···N Trp123	3.00	3.00	2.89	2.95	2.90
Gal/GalNAc O6···O Trp123	3.11	3.11	3.05	3.07	2.95
Gal/GalNAc O6···OD1 Asp125	2.76	2.82	2.81	2.75	2.83
GalNAc O7···N Gly1	–	3.32	3.21	3.21	–
Water bridges involving the primary site					
GalNAc N2···W···OH Tyr78	–	–	3.03,2.99	–	–
GalNAc O7···W···N Gly1	–	–	2.87,3.23	–	–
Direct interactions at secondary site A					
Glc O1···OH Tyr122	–	–	–	–	3.00
Glc O4···OH Tyr78	–	–	–	–	2.95
Water bridges involving secondary site B					
Gal/GalNAc O4···W···OG Ser119	–	–	3.06,3.34	–	–
Gal/GalNAc O6···W···O Val79	–	–	2.56,2.77	3.03,2.82	–
Gal/GalNAc O6···W···OD1 Asp125	–	–	2.56,2.75	3.03,2.70	–
Gal/GalNAc O6···W···N Val79	–	–	2.61,3.06	3.00,3.01	–

The distances (Å) are averaged over subunits. Two distances are given in the case of water bridges. They correspond to those from the sugar and the protein atoms to the central water molecule. Secondary site A is involved only in direct interactions, while only water bridges occur at secondary site B.

makes a C–H···π interaction with Tyr122 with a C–M distance in the range of 3.2–3.8 Å and ω within 4° and 15°. These values match reasonably well with the geometry required for a good C–H···π interaction.[27,28] Naturally, the movement caused by the α-linked group is the maximum in the case of Galα1-6Glc (melibiose) in which the substituent is a bulky sugar residue. In the complex with this disaccharide, the hydroxyl groups of Tyr78 and Tyr122 also form hydrogen bonds with O1 and O4 of Glc.

Interactions at secondary site B

In the crystal structures of the β-linked disaccharide complexes, the residue at secondary site B exhibits considerable variability in its interactions with the lectin. Except in the GalNAcβ1-3Gal-α-*O*-Me complex, the residue directly interacts with the combining site in none of the subunits. In the Me-α-T-antigen complex, Gal, the residue at secondary site B, interacts with the lectin through three water bridges (Figure 4(b), Table 2). One of these is invariant in all structures. It bridges O6 of Gal/GalNAc to Asp125 OD1 and Val79 O. So this water can be thought as an extention of the protein surface. Another water bridge between O6 of Gal/GalNAc and Val79 N is present in all complexes involving β-linked disaccharides, except in one subunit of the T-antigen complex. The sugar binding at this subsite is also accompanied by the burial of Ala17, the penultimate residue at the carboxy terminus of the β chain. Of the total non-polar surface area buried on complexation, 13–16% belongs to this residue in complexes involving T-antigen and Me-α-T-antigen. The corresponding values are 4–6% in the GalNAcβ1-3Gal-α-*O*-Me complex.

Correlation with binding data

The crystal structures of the complexes presented here and those reported earlier provide a reasonable qualitative explanation for the vast amount of available thermodynamic data on jacalin–sugar interactions, especially the enthalpy component of it. Table 3 lists the affinity of jacalin to different sugars. The relation between the nature of the glycosidic linkage and the occupancy of the primary and the secondary sites, established through crystallographic studies presented here and earlier, gives reasonable indications of the mode of binding of even those sugars whose complexes have not been studied crystallographically. These indications and those established experimentally are also given in the Table. The order of binding affinity of jacalin at the monosaccharide level to Gal derivatives is as follows: Me-α-GalNAc > Me-α-Gal > GalNAc > Gal > > > > > Me-β-GalNAc = Me-β-Gal. The same at the disaccharide level is: Galβ1-3GalNAc-α-*O*-Me > GalNAcβ1-3Gal-α-*O*-Me > Galβ1-3GalNAc > Glcβ1-3GalNAc > GlcNAcβ1-3Gal >

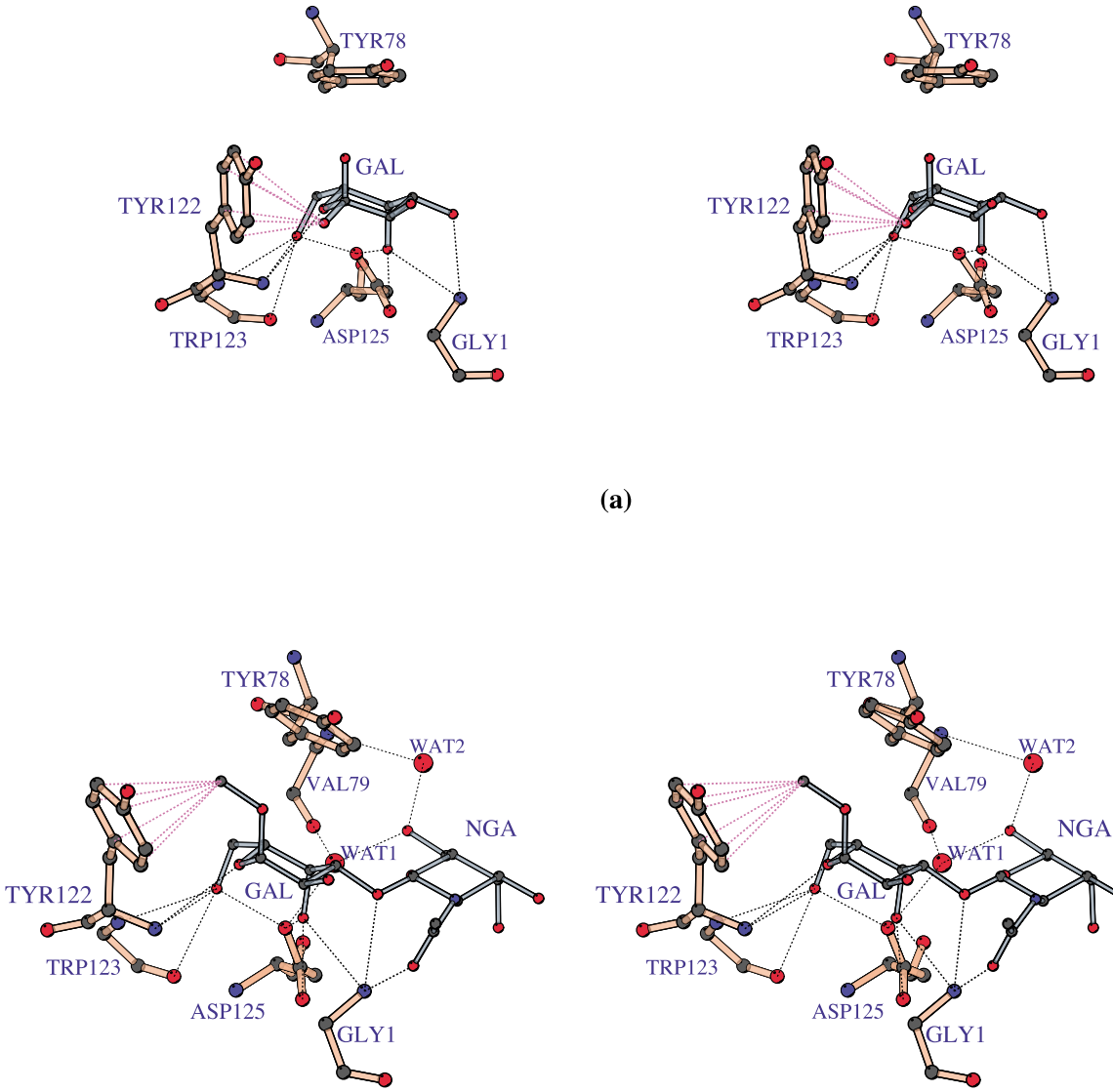


Figure 5. Stereoviews of jacalin–carbohydrate interactions involving (a) Gal and (b) GalNAcβ1-3Gal-α-*O*-Me. The O–H···π and C–H···π interactions are also shown.

Galα1-3Gal-α-*O*-Me = Galα1-6Glc.[24,25,29] The structure of the Me-α-GalNAc complex reveals an additional hydrogen bond between the carbonyl oxygen of the acetamido group and the N terminus of the α-chain, which is responsible for the higher affinity for GalNAc compared to Gal. The favourable steric interaction of the methyl group at secondary site A with the aromatic residues Tyr78, Tyr122 and Trp123 observed in the methylated sugar complexes explains the better affinity of Me-α-Gal over Gal. The simultaneous presence of the hydrogen bond between the acetamido group and the free amino terminus, and the interaction of the methyl group at secondary site A explains the better affinity of Me-α-GalNAc over the individual affinities of Me-α-Gal, GalNAc and Gal. As indicated earlier, the architecture of secondary site A is such that it can accommodate only the α-linked methyl group. Any β-substitution will lead to severe steric clash with Tyr122 resulting in, for instance, the poor affinity of the lectin for Me-β-GalNAc and Me-β-Gal.

In the structure of the T-antigen complex reported earlier, the primary site is occupied by GalNAc. The structure also revealed the role of water-mediated interactions at the secondary site B in enhancing the affinity of T-antigen compared to that of the corresponding monosaccharide. GalNAcβ1-3Gal-α-*O*-Me interacts with all the three sites. The primary site is occupied by Gal. Therefore, the interaction involving the acetamido group present in T-antigen does not exist in its complex. Instead, the acetamido group of GalNAc has a direct hydrogen bond with the protein. In

Table 3. Information pertaining to jacalin–carbohydrate interactions in complexes with different sugars

		Location in binding site			
	Binding constants ($10^{-3} \times K_a$)	Primary	Sec A	Sec B	Remarks
Gal	1.22(−38.0)	Gal	–	–	350(176)
Me-α-Gal	40.00(−55.0)	Gal	Me	–	390(264)
Me-β-Gal	0.20(−42.0)	Gal	–	–	Me has steric clash with Tyr122
GalNAc	3.31(−55.0)	GalNAc	–	–	–
Me-α-GalNAc	73.28(−63.7)	GalNAc	Me	–	430(256)
Me-β-GalNAc	0.55(n/a)	GalNAc	–	–	Me has steric clash with Tyr122
Galβ1-3GalNAc	122.00(−105.0)	GalNAc	–	Gal	477(237)
Galβ1-3GalNAc-α-*O*-Me	806.00(−95.6)	GalNAc	Me	Gal	540(321)
Galβ1-3GalNAc-β-*O*-Me	0.40(−36.9)	GalNAc	–	Gal	Me has steric clash with Tyr122
Galβ1-3GlcNAc	0.04(n/a)	GlcNAc	–	Gal	O4 of GlcNAc cannot make hydrogen bond with Gly1N
GalNAcβ1-3Gal-α-*O*-Me	302.00(−68.8)	Gal	Me	GalNAc	523(316)
Glcβ1-3GalNAc	18.76(−42.2)	GalNAc	–	Glc	–
GlcNAcβ1-3Gal	11.74(n/a)	Gal	–	GlcNAc	–
Galα1-6Glc	5.31(−38.0)	Gal	Glc	–	468(262)
Galα1-3Gal-α-*O*-Me	5.75(−36.8)	Gal	Gal	–	Me is exposed to solvent
Galα1-4Gal	0.84(n/a)	Gal	–	–	Reducing Gal has steric clash with Tyr78
Blood group A Tri	0.27(n/a)	GalNAc	Gal	–	Steric clash between fucose and the side-chains of Trp123 and Tyr122
Blood group B Tri	0.24(n/a)	Gal	Gal	–	Steric clash between fucose and the side-chains of Trp123 and Tyr122

Sugars in crystallographically characterized complexes are given in bold. The values of enthalpy (kJ mol^{-1}), wherever available, are given in parentheses by the side of the binding constants. Surface areas ($Å^2$) buried on complexation, with the hydrophobic components in parentheses, are listed in the last column in the case of complexes of known structure.

addition, the α-linked methyl group interacts with secondary site A, an interaction that does not exist in the T-antigen complex. That explains the higher affinity of GalNAcβ1-3Gal-α-*O*-Me compared to that of T-antigen. Me-α-T-antigen also interacts with all the three binding sites. Its interaction at secondary site A is similar to that of GalNAcβ1-3Gal-α-*O*-Me. At the primary site, the reducing sugar of Me-α-T-antigen and the non-reducing sugar of the other disaccharide have an additional hydrogen bond in their respective complexes. Also, the former has an additional water bridge between the nitrogen atom of the acetyl arm and the OH of Tyr78. Thus, the affinity of jacalin to Me-α-T-antigen is higher than that of GalNAcβ1-3Gal-α-*O*-Me. The only available crystal structure of an α-linked disaccharide is that of melibiose (Galα1-6Glc) and this structure has two additional direct hydrogen bonds at the secondary site A between O1 and O4 of Glc with the hydroxyl groups of Tyr122 and Tyr78.

The other disaccharides that are known to bind jacalin but whose interactions with the lectin have not been studied crystallographically, include Glcβ1-3GalNAc and GlcNAcβ1-3Gal. Since we now know that only the reducing sugar moiety of the β-linked disaccharides can occupy the primary binding site, one would expect GalNAc and Gal to occupy the primary binding site in the respective complexes, which will result in Glc and GlcNAc occupying secondary site B. The binding of GalNAc at the primary binding site in the Glcβ1-3GalNAc complex explains the better affinity of this disaccharide compared to that of GlcNAcβ1-3Gal. Galβ1-3GalNAc-β-*O*-Me binds very poorly to jacalin due to the presence of a methyl group linked to the β-anomer of the GalNAc. Galβ1-3GlcNAc also binds very weakly to jacalin. In the light of the above discussion, GlcNAc of Galβ1-3GlcNAc will occupy the primary site, which will result in the loss of the crucial hydrogen bond between O4 of GlcNAc and the amino terminus at the primary binding site. This could be the reason for the poor binding of Galβ1-3GlcNAc. Galα1-3Gal-α-*O*-Me is the other α-linked disaccharide, which binds equally as well as melibiose. Simple modeling revealed that the occupation of the primary site by Gal-α-*O*-Me would lead to unacceptable steric clashes of Gal with the protein. On the other hand, Gal can comfortably occupy the primary site with Gal-α-*O*-Me at secondary site A. However, the methyl group is exposed to solvent, resulting in no additional interaction compared to melibiose. Although Galα1-4Gal is very similar to melibiose, it binds very poorly to jacalin. Modeling of this sugar at the binding site revealed severe steric clash at the secondary site A between the reducing Gal and Tyr78, a stacking residue important for the binding of Gal at the primary site. Interestingly, the blood group trisaccharides A (GalNAcα1-3Galα1-2Fuc) and B (Galα1-3Galα1-2Fuc) bind very poorly to jacalin, even though they contain the disaccharides GalNAcα1-3Gal and Galα1-3Gal, which can bind jacalin with reasonable affinity. The energy minimized model of Galα1-3Gal, reveals unfavorable steric clashes between the fucose attached to the reducing Gal through an α1-2 linkage and the side-chains of Trp123 and Tyr122, thus explaining the low affinity of jacalin for blood group substances A and B.

Structural insights into the binding of glycoproteins

Jacalin has received considerable attention on account of its ability to bind a large number of glycoproteins including IgA_1, IgD, $CD4^+$, hemopexin, plasminogen, chorionic gonadotropin, bovine protein Z, bovine coagulation factor X, fetuin and asialofetuin. All of them contain the T-antigenic disaccharide core O-linked to the protein, often in addition to N-linked glycans. The ability of jacalin to bind IgA subclass IgA_1 is used in clinical monitoring in IgA nephropathy.[30,31] Jacalin does not bind to glycoproteins containing only N-linked oligosaccharides such as ovalbumin, transferrin and α1-acid glycoprotein. But a recent observation of jacalin binding to arcelin1, a glycoprotein with high mannose and complex oligosaccharides necessitates a relook at the binding of jacalin to glycoproteins.[11,32]

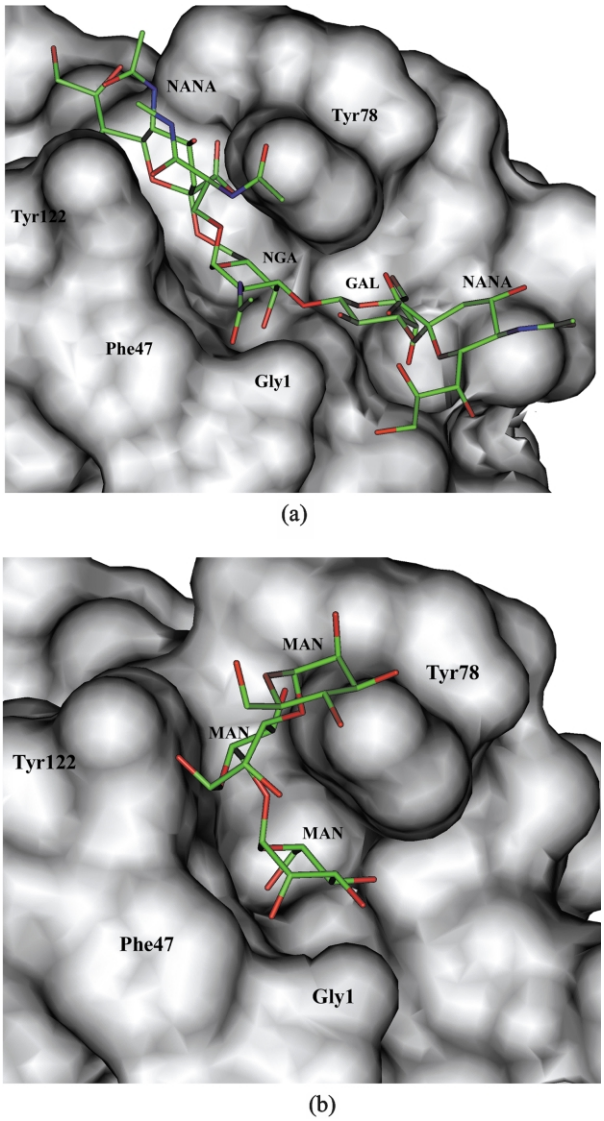


Figure 6. Surface representation of the sugar-binding site with modeled (a) disialyl-T-antigen O-linked to *N*-acetyl *N*-methyl amide of serine and (b) mannotriose (Manα1-3Manα1-6Man).

In general, the mucin-type O-glycans (which contain T-antigen core) can be classified into five major types. They are GalNac-α-*O*-Ser/Thr (Tn-antigen), Galβ1-3GalNAc-α-*O*-Ser/Thr (T-antigen), NeuNAcα2-3Galβ1-3GalNAc (monosialyl T-antigen), NeuNAcα2-3Galβ1-3(NeuNAcα2-6)GalNAc (disialyl T-antigen) and NeuNAcα2-3Galβ1-3(NeuNAcα2-3Galβ1-4GlcNAcβ1-6)GalNAc (disialyl hexasaccharide). The ability of jacalin to bind mono- and disialylated T-antigen, but not the hexasaccharide has been demonstrated.[33] Modeling and energy minimization calculations earlier revealed a feasible mode for the binding of T-antigen-α-*O*-Ser/Thr to jacalin.[26] The addition of a sialic acid α2-3 linked to Gal, as in monosialylated T-antigen, did not yield any additional interaction of the sugar with the lectin while retaining all the interactions found in the T-antigen complex. But the addition of sialic acid α2-6 linked to GalNAc necessitates some modifications in the way GalNAc binds to jacalin. O6 of GalNAc is fully buried in the T-antigen complex and it makes hydrogen bonds with Asp125 OD1 and the main-chain oxygen and nitrogen atoms of Tyr122 and Trp123. Energy minimization of this disialylated T-antigen at the jacalin combining site led to the disruption of this hydrogen bonding network, while, OG of Ser76 and the main-chain oxygen atom of Val79 O could now make favourable hydrogen bonds with sialic acid. In addition, the side-chain of Tyr122 stacks on the sialic acid linked to GalNAc (Figure 6(a)). These compensatory interactions therefore allow binding of the disialylated tetrasaccharide. Attempts to dock the disialylated hexasaccharide led to severe steric clash between Trp123 and GlcNAc β1-6 linked to GalNAc, which explains the inability of jacalin to bind the hexasaccharide.

The structures of N-glycans can be categorized into three major types; namely, high-mannose, complex, and hybrid. All of them have a common pentasaccharide core (Man_3 $GlcNAc_2$). The structures of jacalin with Me-α-mannose[11] and melibiose together provide sufficient information to model the binding of mannotriose (Manα1-3Manα1-6Man) to jacalin. Manα1-6Man of mannotriose could occupy the combining site of jacalin as Galα1-6Glc does, with glycosidic torsion angles ϕ and ψ of 79° and 101° (as seen in the crystal structure of the melibiose complex). These conformational angles correspond to one (67(±14)° and 109(±14)°) among the three distinct comformers seen in 69 crystal structures.[34] The glycosidic torsion angles of α1-3 linkage were modeled using the single conformer observed in 84 of the 91 relevant crystal structures,[34] where ϕ and ψ are 73(±11)° and −112(±23)°. This led to a model where only Manα1-6Man of the mannotriose interacts with the lectin, while the other mannose is exposed to solvent (Figure 6(b)). Admittedly, this is only one of the possible models, but it is the most likely. Interestingly, O2 of mannose bound at

the primary site is fully buried and hence any β-linked sugar at this atom as in complex-type N-glycans will lead to severe steric clashes. This could be the reason for the restricted ability of jacalin to bind N-linked glycans. Although, the binding of mannotriose is not sterically unfavorable, the affinity of N-glycans was shown to be very weak compared to that of O-glycans. The absence of the crucial hydrogen bond between the O4 of Man with the free amino terminus of the α-chain, unlike in the case of Gal, explains the poor affinity for mannose itself[11,32] and possibly for the N-glycans. This suggests the possibility of favorable contribution from the non-specific protein–protein interactions in the binding of glycoproteins containing high-mannose and hybrid N-glycans.

Materials and Methods

Crystallization

Jacalin was extracted and purified from crude jackfruit seeds on cross-linked guar gum columns using salinated phosphate buffer as reported elsewhere.[35] Crystals of jacalin complexed with Me-α-T-antigen and Me-α-GalNAc were grown using the vapour-diffusion technique by equilibrating a 10 μl drop of 10 mg ml^{-1} protein in the presence of 20 times molar excess (12 mM) of Me-α-T-antigen and 30 times molar excess (18 mM) of Me-α-GalNAc, respectively, in 0.02 M phosphate buffer (pH 7.3), containing 0.1 M NaCl and 0.025% (w/v) sodium azide and 8–10% PEG4000 against a reservoir solution of 40% PEG4000, in the same buffer. In the presence of Me-α-T-antigen, crystals of size 0.8 mm × 0.6 mm × 0.6 mm grew in two weeks. Crystals of size 0.4 mm × 0.4 mm × 0.2 mm grew in the presence of Me-α-GalNAc in a week. Efforts to crystallize the protein in the presence of GalNAcβ1-3Gal-α-*O*-Me failed. So, the crystals of Me-α-GalNAc complex were soaked in the mother liquor containing 18 mM GalNAcβ1-3Gal-α-*O*-Me and allowed to equilibrate for 48 hours. The crystals of jacalin complexed with Gal were grown by equilibrating 10 μl of 15 mg ml^{-1} protein in the presence of 20 times molar excess (18 mM) of Gal in 0.1 M sodium acetate trihydrate buffer (pH 7.3), containing 0.2 M ammonium sulfate and 5–6% PEG4000 against a reservoir solution of 25% PEG4000 in the same buffer. Big crystals of size 1.2 mm × 0.6 mm × 0.6 mm grew in a week. These crystals were soaked in 36 mM solution of Galα1-6Glc and allowed to equilibrate for 48 hours to prepare the complex of the disaccharide.

Data collection and processing

Data from the crystals of the Gal and the Me-α-T-antigen complexes were collected on the X-ray diffraction 1 (XRD1) beam line of wavelength 1.0 Å at Elettra Synchrotron Light Laboratory, Trieste, Italy, using a Mar research MAR345 imaging plate. Data from the other crystals were collected at home using a Mar research MAR300 imaging plate mounted on a Rigaku RU-200 X-ray generator. The data were processed using DENZO and SCALEPACK of the HKLsuite of programs.[36] The processed data were truncated using the program TRUNCATE of CCP4.[37] Data collection statistics along with the cell parameters are given in Table 4. The Matthews coefficient[38] indicated the presence of one tetramer in the asymmetric unit in all the crystals except in those of Me-α-T-antigen, in which only one subunit is present in the asymmetric unit.

Structure solution and refinement

The structures of different sugar complexes were solved using the molecular replacement program AMoRe.[39] The structure of jacalin complexed with T-antigen (PDB code 1M26), which was solved earlier in this laboratory, was used as the starting model. Unique solutions were obtained with CC and *R*-factors in the range of 0.58–0.69 and 0.29–0.38, respectively. All the structures were refined in a similar manner. To start with, 40 cycles of rigid body refinement followed by 100 cycles of positional refinement using CNS[40] with "mlf" target were carried out. Clear unambiguous density for the sugars appeared in the $F_o - F_c$ and $2F_o - F_c$ maps. The coordinates of the sugars were generated using the web-based program SWEET† and were fitted into the electron density using FRODO.[41] Subsequent 30 cycles of positional refinement revealed clear densities for ordered solvent molecules and the additional density for the β-anomer of Gal in three of the four subunits. At this stage, the β-anomer of Gal was modeled and water oxygen atoms were added successively to the model, using peaks with heights greater than 2.5σ in $F_o - F_c$ maps and 0.8σ in $2F_o - F_c$ maps. Omit maps were used in the course of refinement to remove model bias. Bulk solvent corrections and anisotropic *B*-factor corrections were used throughout the refinement. Iterative cycles of model building and refinement were carried out till *R* and R_{free} converged. *B*-Factors of the atoms were refined individually for the structures solved at resolutions of 2.5 Å or better. In other cases, group *B*-factor refinement was carried out. The final values of *R* and R_{free} and other relevant refinement statistics are given in Table 4. The model was checked using PROCHECK.[42]

Analysis and modeling

Molecular superpositions were performed using program ALIGN.[43] Program NACCESS‡ was employed for calculating accessible surface areas. Possible hydrogen bonds were identified using program HBPLUS.[44] Contacts involving oxygen and nitrogen atoms with distances less than 3.6 Å and with donor–hydrogen–acceptor angle greater than 90° were treated as hydrogen bonds. Binding of various sugars to jacalin were modeled using INSIGHT II. The refined coordinates of the Me-α-T-antigen complex structure were used for the modeling. Disialyl-T-antigen O-linked to the *N*-acetyl *N*-methyl amide of serine was modeled and optimized using the BIOPOLYMER module of INSIGHT II. The T-antigen coordinates of the model glycopeptide were superposed on the crystal structure coordinates of the T-antigen before minimization. A distance-dependent dielectric constant was used throughout the minimization. During minimization, the loop residues 73–76 and the side-chain of Tyr122 and Trp123 only were allowed to move. The sugar residue at the primary site

†

Table 4. Data collection and refinement statistics

	Gal	Me-α-GalNAc	Galβ1,3GalNAcα-*O*-Me	GalNAcβ1,3Galα-*O*-Me	Galα1-6Glc
Space group	$P\,2_12_12_1$	$P\,6_522$	$I\,222$	$P\,6_522$	$P\,2_12_12_1$
Unit cell dimensions					
a (Å)	80.44	129.81	43.29	129.45	80.46
b (Å)	99.55	129.81	100.56	129.45	99.83
c (Å)	105.85	158.53	102.4	158.35	105.65
Z	4	12	2	12	4
Resolution (Å)	1.7	2.8	1.6	2.8	2.4
Last shell (Å)	1.76–1.70	2.90–2.80	1.63–1.60	2.90–2.80	2.49–2.40
No. of observations	350,856	113,368	139,555	81,752	106,303
No. of unique reflections	91,056(8911)	19,864(1944)	27,418(2452)	19,332(1860)	32,331(3044)
Reflections with $I = 0$	2995(394)	370(92)	2540(502)	436(88)	1617(299)
Completeness (%)	96.8(95.8)	99.3(99.6)	91.5(83.0)	97.4(97.0)	95.2(91.1)
R merge[a] (%)	5.5(44.7)	13.5(49.0)	8.6(16.3)	11.9(46.4)	13.8(40.0)
Multiplicity	3.9	5.7	5.1	4.2	3.3
Protein atoms	4611	4588	1144	4588	4611
Sugar atoms	45	64	27	97	92
Solvent atoms	489	109	123	120	255
R-factor[b] (%)	18.5	19.4	18.9	19.2	18.0
R_{free}[b] (%)	19.4	21.8	20.5	23.2	21.4
Resolution range (Å)	20.0–1.70	20.0–2.8	30.0–1.60	20.0–2.8	20.0–2.4
No. of reflections	91,005	19,485	27,409	18,869	32,307
RMS deviations from ideal values					
Bond length (Å)	0.005	0.019	0.005	0.012	0.006
Bond angle (deg.)	1.3	1.7	1.3	1.4	1.3
Dihedral angle (deg.)	26.5	26.9	26.4	26.7	26.5
Improper (deg.)	0.78	1.03	0.80	0.80	0.81
Residues (%) in Ramachandran plot					
Core region	89.1	87.7	88.1	87.3	85.1
Additionally allowed region	10.7	12.3	11.9	12.7	14.6
Generously allowed region	0.2	0.0	0.0	0.0	0.3
Disallowed region	0.0	0.0	0.0	0.0	0.0

[a] $R_{merge} = \sum|I_i - \langle I\rangle|/\sum\langle I\rangle$. The values within the parentheses refer to the last shell.
[b] $R = \sum||F_o| - |F_c||\sum|F_o|$; R_{free} is calculated the same way but for a subset of reflections that is not used in the refinement.

was kept rigid to preserve the invariant interactions. The whole conformation space of the glycopeptide was searched in combination with conjugate gradient minimization and torsion force constraint using DISCOVER. Figures 1, 2, 4, and 5 were prepared using BOBSCRIPT,[45] while Figures 3 and 6 were prepared using INSIGHT II.

Atomic coordinates

The atomic coordinates and the structure factors of all the jacalin–sugar complexes were deposited in the Protein Data Bank (PDB codes 1UGW, 1UH0, 1UGX, 1UH1 and IUGY).

Acknowledgements

Two data sets used in the present work were collected at Sincrotrone Trieste under the framework of the Indo-Italian Programme of Cooperation in Science and Technology, 2001. The remaining data sets were collected at the X-ray Facility for Structural Biology, at the Institute, supported by the Department of Science & Technology (DST) and the Department of Biotechnology (DBT) of the Government of India. Computations were performed at the Supercomputer Education and Research Centre of the Institute, and the Bioinformatics Centre and the Graphics facility, both supported by DBT. Financial support from DST is acknowledged.

References

1. Sharon, N. & Lis, H. (1989). Lectins as cell recognition molecules. *Science*, **246**, 227–246.
2. Lis, H. & Sharon, N. (1998). Lectins: carbohydrate specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637–674.
3. Drickamer, K. (1999). C-type lectin-like domains. *Curr. Opin. Struct. Biol.* **9**, 585–590.
4. Rini, J. (1999). New animal lectin structures. *Curr. Opin. Struct. Biol.* **9**, 578–584.
5. Vijayan, M. & Chandra, N. R. (1999). Lectins. *Curr. Opin. Struct. Biol.* **9**, 707–714.
6. Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). Legume lectin structure. *Biochim. Biophys. Acta*, **1383**, 9–36.
7. Rudiger, H. & Gabius, H. J. (2001). Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconj. J.* **18**, 589–613.
8. Barre, A., Bourner, Y., Van Damme, E. J., Peumans, W. J. & Rouge, P. (2001). Mannose-binding plant

lectins: different structural scaffolds for a common sugar-recognition process. *Biochimie*, **83**, 645–651.
9. Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. & Vijayan, M. (1996). A novel mode of carbohydrate recognition in jacalin, a *Moraceae* plant lectin with a beta-prism fold. *Nature Struct. Biol.* **3**, 596–603.
10. Pratap, J. V., Jeyaprakash, A. A., Rani, P. G., Sekar, K., Surolia, A. & Vijayan, M. (2002). Crystal structures of artocarpin, a *Moraceae* lectin with mannose specificity, and its complex with methyl-α-D-mannose: implications to the generation of carbohydrate specificity. *J. Mol. Biol.* **317**, 237–247.
11. Bourne, Y., Astoul, C. H., Zamboni, V., Peumans, W. J., Menu-Bouaouiche, L., van Damme, E. J. M. *et al.* (2002). Structural basis for the unusual carbohydrate-binding specificity of jacalin towards galactose and mannose. *Biochem. J.* **364**, 173–180.
12. Aucouturier, P., Mihaesco, E., Mihaesco, C. & Preud'homme, J. L. (1987). Characterization of jacalin, the human IgA and IgD binding lectin from jackfruit. *Mol. Immunol.* **24**, 503–511.
13. Mahanta, S. K., Sanker, S., Rao, N. V. S. A. V. P., Swamy, M. J. & Surolia, A. (1992). Primary structure of a Thomsen-Friedenreich-antigen-specific lectin, jacalin [*Artocarpus integrifolia* (jackfruit) agglutinin]. Evidence for the presence of an internal repeat. *Biochem. J.* **284**, 95–101.
14. Yang, H. & Czapla, T. H. (1993). Isolation and characterization of cDNA clones encoding jacalin isolectins. *J. Biol. Chem.* **268**, 5905–5910.
15. Kabir, S. (1998). Jacalin: a jackfruit (*Artocarpus heterophyllus*) seed-derived lectin of versatile applications in immunobiological research. *J. Immunol. Methods*, **212**, 193–211.
16. Bunn-Moreno, M. M. & Campos-Neto, A. (1981). Lectin(s) extracted from seeds of *Artocarpus integrifolia* (jackfruit): potent and selective stimulator(s) of distinct human T and B cell functions. *J. Immunol.* **127**, 427.
17. Pineau, N., Brugier, J. C., Breux, J. P., Becq-Giraudon, B., Descamps, J. M., Aucouturier, P. & Preud'homme, J. L. (1989). Stimulation of peripheral blood lymphocytes of HIV-infected patients by jacalin, a lectin mitogenic for human $CD4^+$ lymphocytes. *AIDS*, **3**, 659–663.
18. Pineau, N., Aucouturier, P., Brugier, J. C. & Preud'homme, J. L. (1990). Jacalin: a lectin mitogenic for human CD4 T lymphocytes. *Clin. Exptl Immunol.* **80**, 420–425.
19. Corbeau, P., Haran, M., Binz, H. & Devanux, C. (1994). Jacalin, a lectin with anti-HIV properties, and HIV-I gp120 envelope protein interact with distinct regions of the CD4 molecule. *Mol. Immunol.* **31**, 569–575.
20. Corbeau, P., Pasquali, J. L. & Devaux, C. (1995). Jacalin, a lectin interacting with O-linked sugars and mediating protection of $CD4^+$ cells against HIV-I, binds to the external envelope glycoprotein gp120. *Immunol. Letters*, **47**, 141–143.
21. Lafont, V., Nicolas, M., Dornand, J., Liautard, J. P. & Favero, J. (1993). Perturbation of *in vitro* HIV pathogenic effects by peptides showing sequence similarities with the C2 conserved domain of gp120. *Immunol. Letters*, **37**, 249–250.
22. Lafont, V., Dornand, J., d'Angeac, A. D., Monier, S., Alcover, A. & Favero, J. (1994). Jacalin, a lectin that inhibits *in vitro* HIV-1 infection, induces intracellular calcium increase *via* CD4 in cells lacking the CD3/TcR complex. *J. Leukoc. Biol.* **56**, 521–524.
23. Lafont, V., Dornand, J., Covassin, L., Liautard, J. P. & Favero, J. (1996). The lectin jacalin triggers CD4-mediated lymphocyte signaling by binding CD4 through a protein–protein interaction. *J. Leukoc. Biol.* **59**, 691–696.
24. Sastry, M. V., Banarjee, P., Patanjali, S. R., Swamy, M. J., Swarnalatha, G. V. & Surolia, A. (1986). Analysis of saccharide binding to *Artocarpus integrifolia* lectin reveals specific recognition of T-antigen (β-D-Gal(1–3)D-GalNAc). *J. Biol. Chem.* **261**, 11726–11733.
25. Mahanta, S. K., Sastry, M. V. & Surolia, A. (1990). Topography of the combining region of a Thomsen-Friedenreich-antigen-specific lectin jacalin (*Artocarpus integrifolia* agglutinin). A thermodynamic and circular-dichroism spectroscopic study. *Biochem. J.* **265**, 831–840.
26. Jeyaprakash, A. A., Rani, P. G., Reddy, G. B., Banumathi, S., Betzel, C., Sekar, K. *et al.* (2002). Crystal structure of the jacalin–T-antigen complex and a comparative study of lectin–T-antigen complexes. *J. Mol. Biol.* **321**, 637–645.
27. Steiner, T. & Koellner, G. (2001). Hydrogen bonds with π-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **305**, 535–557.
28. Brandl, M., Weiss, M. S., Jabs, A., Suhnel, J. & Hilgenfeld, R. (2001). C–H···interactions in proteins. *J. Mol. Biol.* **307**, 357–377.
29. Wu, M., Wu, J. H., Lin, L. & Liu, J. (2003). Binding profile of *Artocarpus integrifolia* agglutinin (jacalin). *Life Sci.* **72**, 2285–2302.
30. Allen, A. C. (1999). Structural features of IgA molecules which contribute to IgA nephropathy. *J. Nephrol.* **12**, 59–65.
31. Allen, A. C. (1999). Methodological approaches to the analysis of IgA1 O-glycosylation in IgA nephropathy. *J. Nephrol.* **12**, 76–84.
32. Astoul, C. H., Peumans, W. J., van Damme, E. J. M., Barre, A., Bourne, Y. & Rouge, P. (2002). The size, shape and specificity of the sugar-binding site of jacalin-related lectins is profoundly affected by the proteolytic cleavage of the subunits. *Biochem. J.* **367**, 817–824.
33. Maemura, K. & Fukuda, M. (1992). Poly-*N*-acetyllactosaminyl O-glycans attached to leukosialin. *J. Biol. Chem.* **267**, 24379–24386.
34. Petrescu, A. J., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. (1999). A statistical analysis of *N*- and *O*-glycan linkage conformations from crystallographic data. *Glycobiology*, **9**, 343–352.
35. Kumar, S. G., Appukuttan, P. S. & Basu, D. (1993). α-D-galactose-specific lectin from jackfruit (*Artocarpus integrifolia*) seed. *J. Biosci.* **4**, 257–261.
36. Otwinowsky, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Macromolecular Crystallography, Part A, Methods of Enzymology* (Carter, C. W., Jr & Sweet, R. M., eds), vol. 276, pp. 307–326, Academic Press, New York.
37. Collaborative Computational Project, No. 4 (1994). *Acta Crystallog. sect. D*, **50**, 760–763.
38. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
39. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 445–449.
40. Brunger, A. T., Adams, P. D. & Rice, L. M. (1998).

Recent developments for the efficient crystallographic refinement of macromolecular structures. *Curr. Opin. Struct. Biol.* **8**, 606–611.
41. Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallog.* **11**, 268–272.
42. Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067.
43. Cohen, G. E. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallog.* **30**, 1160–1161.
44. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
45. Esnouf, R. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132–134.