Crystal Structure of the Jacalin−T-antigen Complex and a Comparative Study of Lectin−T-antigen Complexes

A. Arockia Jeyaprakash[1], P. Geetha Rani[1], G. Banuprakash Reddy[1] S. Banumathi[2], C. Betzel[2], K. Sekar[3], A. Surolia[1] and M. Vijayan[1]*

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

[2]*Universitdtsklinikum Hamburg-Eppendorf, Institut fur Medizinische Biochemie und Molekularbiologie Hamburg, Germany*

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

Thomsen−Friedenreich antigen (Galβ1−3GalNAc), generally known as T-antigen, is expressed in more than 85% of human carcinomas. Therefore, proteins which specifically bind T-antigen have potential diagnostic value. Jacalin, a lectin from jack fruit (*Artocarpus integrifolia*) seeds, is a tetramer of molecular mass 66 kDa. It is one of the very few proteins which are known to bind T-antigen. The crystal structure of the jacalin−T-antigen complex has been determined at 1.62 Å resolution. The interactions of the disaccharide at the binding site are predominantly through the GalNAc moiety, with Gal interacting only through water molecules. They include a hydrogen bond between the anomeric oxygen of GalNAc and the π electrons of an aromatic side-chain. Several intermolecular interactions involving the bound carbohydrate contribute to the stability of the crystal structure. The present structure, along with that of the Me-α-Gal complex, provides a reasonable qualitative explanation for the known affinities of jacalin to different carbohydrate ligands and a plausible model of the binding of the lectin to T-antigen O-linked to seryl or threonyl residues. Including the present one, the structures of five lectin−T-antigen complexes are available. GalNAc occupies the primary binding site in three of them, while Gal occupies the site in two. The choice appears to be related to the ability of the lectin to bind sialylated sugars. In either case, most of the lectin−disaccharide interactions are at the primary binding site. The conformation of T-antigen in the five complexes is nearly the same.

Keywords: Moraceae lectin; T-antigen; carbohydrate specificity; O−H···π interaction; water-bridges

*Corresponding author

Introduction

Thomsen−Friedenreich antigen (Galβ1−3GalNAc), generally known as T-antigen, is a disaccharide of non-oncofetal origin with a well documented link to malignancy in man.[1] It is generally expressed in more than 85% of human carcinomas such as colon, breast, bladder, buccal cavity and prostate, as well as on poorly differentiated cells.[2−5] Therefore, proteins which specifically bind T-antigen have potential diagnostic

value. Peanut (*Arachis hypogaea*) agglutinin (PNA) was the first lectin to be shown to have T-antigen specificity. The crystal structures of PNA in complex with several sugars, including T-antigen, have been reported from this laboratory.[6−9] The other proteins complexed with T-antigen that have been analysed by X-ray so far are amaranthin (*Amaranthus caudatus* agglutinin),[10] *Maclura pomifera* agglutinin (MPA)[11] and heat-labile enterotoxin (LT) from *Escherichia coli*.[12] Another important protein which specifically binds T-antigen is jacalin,[13] one of the two lectins from jackfruit (*Artocarpus integrifolia*) seeds. Jacalin, like the lectins other than PNA mentioned above, binds also to the sialylated derivatives of T-antigen and Tn-antigen (GalNAc-Ser/Thr),[14] which are termed cryptic T- and Tn-antigens, respectively.

Furthermore, jacalin, again unlike PNA, selectively binds to the α-linked forms of T-antigen.[15,16]

Jacalin, which has received considerable attention on account of its interesting biological properties,[17] has been used extensively for the isolation of IgA[18-20] and other glycoproteins including carcinoma related mucins.[21] It is selectively mitogenic for human CD4+ T-cells and is used in AIDS research.[22-28] It is a tetrameric Gal/GalNAc specific lectin, M_r 66,000, with a subunit containing two chains generated by post-translational proteolysis.[29] The β prism I fold in a lectin was first identified in the structure of jacalin.[13] The crystal structure of the jacalin–Me-α-Gal complex also indicated that its specificity for Gal is generated by a post-translational modification, the first instance where such a strategy has been established for the generation of ligand specificity. Extensive thermodynamic and kinetic studies showed that the lectin binds the T-antigenic disaccharide with 3-, 36-, 100- and 610-fold more affinity than it binds Me-α-Gal, GalNAc, Gal and Me-β-Gal, respectively. Other disaccharides such as lactose, N-acetyllactosamine and Galβ1–3GlcNAc have at least 3000 times poorer affinity compared to T-antigen for the lectin.[15,16] The anomeric preference of the lectin for Galβ1–3GalNAc-α-Me over the corresponding β-substituted disaccharide, has already been referred to.

The crystal structure of a complex of jacalin with T-antigen at 1.62 Å resolution and the insights it provides into the structural basis of the carbohydrate specificities of the lectin, are presented here. *Inter alia*, the structure also provides the first instance of anomer selection through an O–H···π interaction when a carbohydrate binds to a protein. The present work also highlights the role of the T-antigen disaccharide in mediating crystal packing. Among the crystal structures of the protein–T-antigen complexes reported so far, the present one has been determined at the highest resolution. Therefore, this structure has been used as the basis for a comparative study of protein–T-antigen complexes.

Results and Discussion

Tertiary and quaternary structure

Jacalin is a tetramer with 222 symmetry and a subunit consisting of a heavy chain (α) of 133 amino acid residues and a light chain (β) of 20 amino acid residues (Figure 1). Each subunit has a three-fold symmetric β-prism fold made up of three four-stranded β sheets. The structure of the lectin is essentially the same in the present complex and in the complex with Me-α-Gal.[13] A notable feature revealed by the T-antigen complex is the presence of three X-Pro *cis* peptide bonds. They were not identified in the Me-α-Gal complex, presumably on account of the limited resolution of the structure. The *cis* peptides involve α61P, α95P
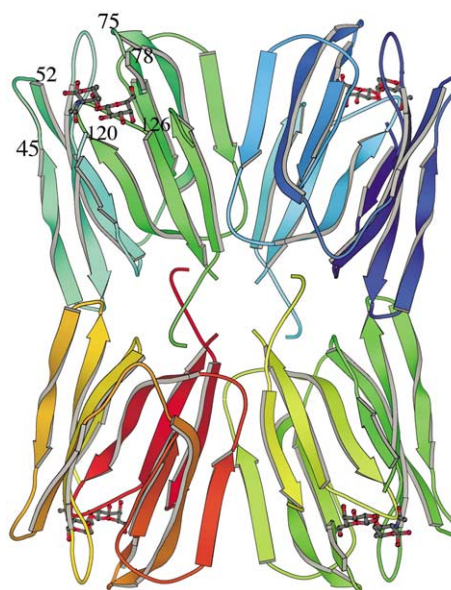

Figure 1. Structure of the jacalin–T-antigen complex. The loops involved in sugar binding are numbered in one of the four subunits.

and β14P and have reasonable geometry.[30] The present structure also corroborates the reported microheterogeneity in the jacalin sequences.[17]

Jacalin–T-antigen interactions

The carbohydrate combining site of jacalin is made up of loops 46–52, 76–82, 122–125, and the N terminus of the α-chain (Figure 1). The well-defined electron density for the bound T-antigen in all the subunits (Figure 2(a)) permits a detailed characterisation of its interactions with the lectin (Figure 2).

Interaction at the primary site

The primary carbohydrate binding site of the lectin is occupied by the GalNAc moiety of the disaccharide. The interactions of this residue with the lectin are the same in all the subunits (Table 1). Jacalin–carbohydrate interactions also appear to involve a strong electrostatic component. The surface of the combining site contains a positive patch formed by Gly121N, Tyr122N and Trp123N. The GalNAc/Gal moiety has a positive patch on the B face and a negative patch on the opposite face. The aromatic ring containing π electrons of Tyr78 stacks on the B face, as in other Gal/GalNAc specific lectins, while the positive patch on the combining site interacts with the negative patch on the opposite face.

The major difference between the Me-α-Gal complex analysed earlier and the present non-methylated T-antigen complex is in the orientation of the aromatic ring of Tyr122. The torsion angle χ^{21} (C^α–C^β–C^γ–C^δ) for this residue in the T-antigen complex varies between -24 and $-38°$ while it is
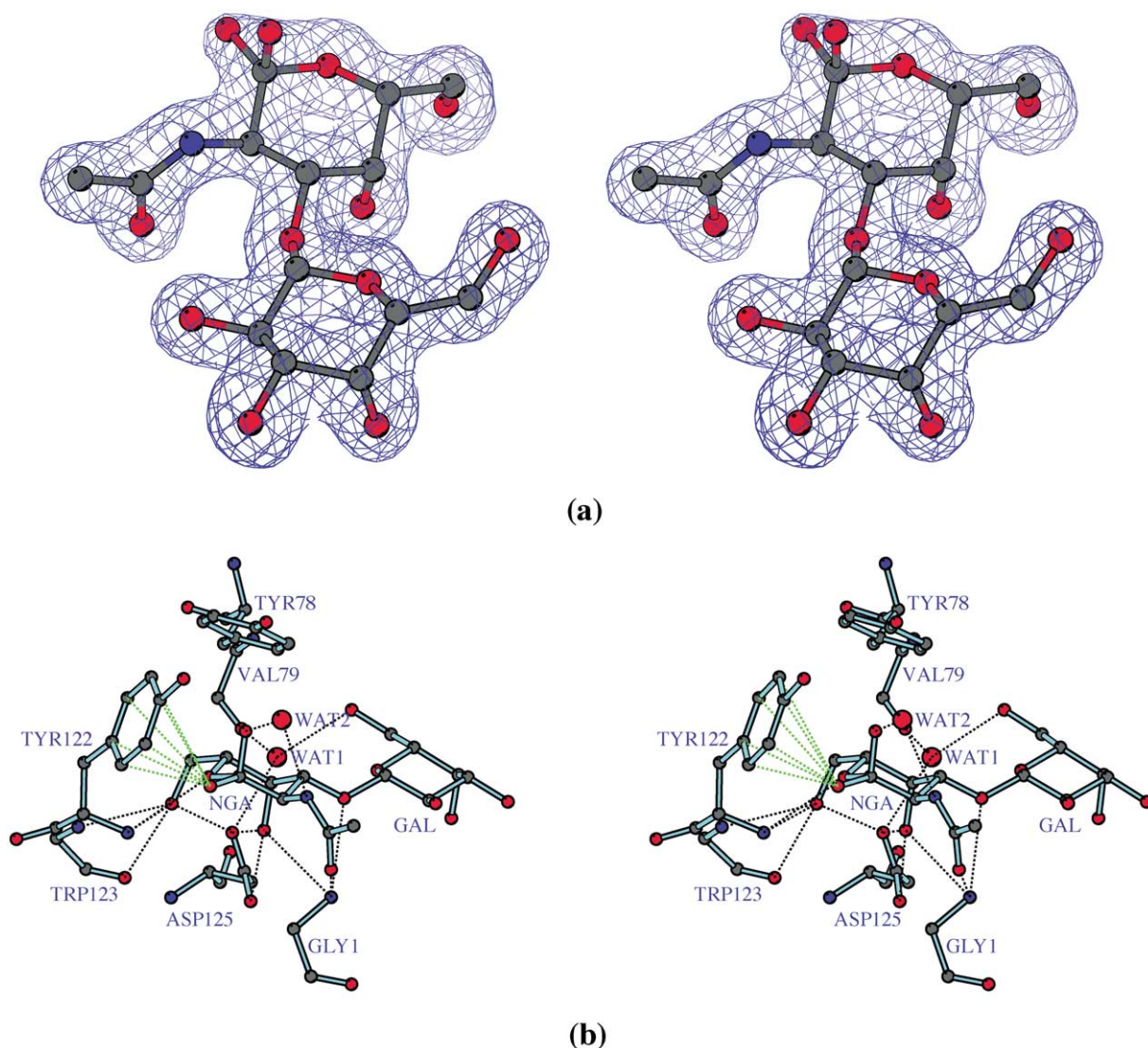
(a)

(b)

Figure 2. (a) Stereo view of the electron density ($2F_o - F_c$ map contoured at 1σ) for the bound sugar in the B subunit. (b) The lectin–T-antigen hydrogen bonding interactions. The O–H$\cdots\pi$ interaction involving the β-anomer exists in subunits B and C. Wat2 is hydrogen bonded to Tyr122 OH in subunit A and to Tyr78 OH in the remaining three subunits.

Table 1. Jacalin–T-antigen interactions which occur in all the four subunits

Sugar atom	Distance (Å)	Water	Distance (Å)	Protein atom
NGA O1	3.09	W	3.18	NGA N2
NGA O3	2.95	–	–	Gly1 N
NGA O4	2.97	–	–	Gly1 N
	2.69	–	–	Asp125 OD1
	2.99	–	–	Asp125 OD2
NGA O5	3.00	–	–	Tyr122 N
NGA O6	2.95	–	–	Tyr122 N
	2.91	–	–	Trp123 N
	3.06	–	–	Trp123 O
	2.82	–	–	Asp125 OD1
NGA O7	3.03	–	–	Gly1 N
Gal O6	2.63	W	2.86	Val79 O
	2.63	W	2.73	Asp125 OD1

The distances are averaged over the subunits.

in the 10–30° range in the Me-α-Gal complex. Consequently the groove formed by the side-chains of Tyr78 and Tyr122 opens up in the latter (Figure 3). The methyl group in the Me-α-Gal complex nestles among the side-chains of the aromatic residues Phe47, Tyr78, Tyr122, Trp123.

Anomer stabilisation at the primary site using O–H$\cdots\pi$ interactions

The present structure provides a unique case where an O–H$\cdots\pi$ hydrogen bond is used for anomer stabilisation. In two subunits (designated as A and D), O1 of GalNAc is wholly axial. In the remaining two subunits (B and C), O1 occupies the axial and equatorial positions with occupancies of 0.55 and 0.45, respectively. In all the four subunits the axial O1 has a water bridge with Tyr78
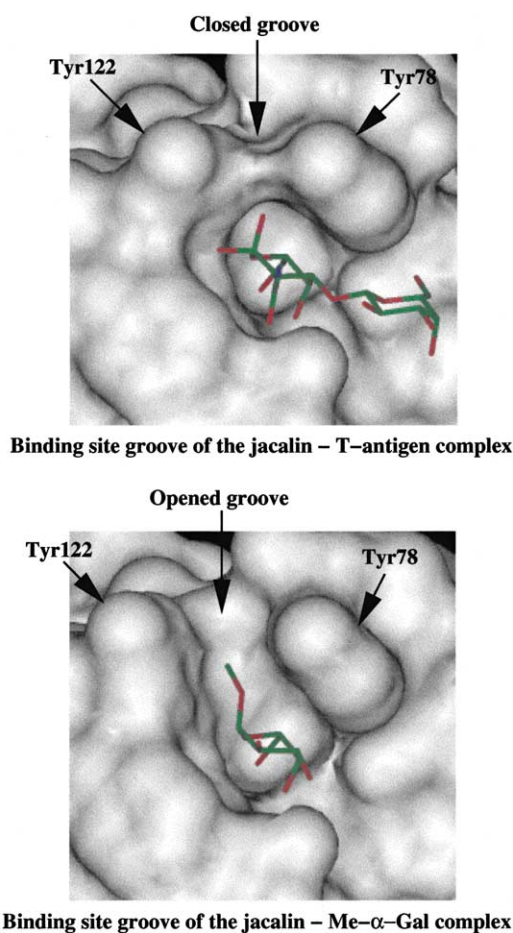
Closed groove

Tyr122   Tyr78

Binding site groove of the jacalin – T–antigen complex

Opened groove

Tyr122   Tyr78

Binding site groove of the jacalin – Me–α–Gal complex

Figure 3. Comparison of the sugar binding sites in the jacalin–T-antigen and jacalin–Me-α-Gal complexes.

OH or Tyr122 OH. In subunits B and C, the one at the equatorial position is hydrogen bonded (Figure 2(b)) to the π electron cloud of Tyr122 side-chain with an ω of 4 and 2°, and an O−M distance of 3.1 Å in both the cases. In subunits A and D the oxygen atom at the equatorial position, and hence the O−H···π hydrogen bond, do not occur presumably because of the additional stabilisation of the axial position by an inter molecular water bridge with Thr102 OG of a neighbouring molecule and Lys117 NZ of another neighbouring molecule, respectively.

Interactions involving the Gal residue

In none of the subunits does the Gal residue directly interact with the combining site. The only interaction present in all subunits is the bridges Gal O6 forms to Val79 O and Asp125 OD1 through a water molecule. This water molecule, present in the Me-α-Gal complex also, is believed to be important for correctly positioning the side-chain of Asp125 as well.[13] A water bridge of O6 with Val79N exists in three subunits. One involving O4 and Gly1N is present in two. A few more water bridges are present, but none of them occur in more than one subunit. The only β chain residue

involved in sugar binding is Ala17. In total 13–16% of the non-polar surface area buried on complexation belongs to this residue and the burial is entirely caused by Gal.

Interactions involving Gal and crystal packing

The Gal residue of T-antigen is involved in extensive intermolecular contacts. This is rather unusual in protein complexes involving monovalent ligands. The most extensive contacts are between tetramers related by the 2_1 screw axis. In fact Gal O2−4 of T-antigen associated with subunit B have five direct interactions with Asn74 and Ser76 of subunit D of the adjacent tetramer related by the 2_1 screw, resulting in a chain of linked tetrameric molecules along *b* (Figure 4). This arrangement is further stabilised by a couple of water bridges between the T-antigen molecule belonging to subunit D of one tetramer and the side-chains of Asp5 and Lys117 of the adjacent one. The chains along *c* are crosslinked by Gal O6 belonging to subunit A and Tyr71 OH of a molecule related by a translation. This connectivity between adjacent chains is strengthened by water bridges between the tetramers involving GalNAc O1 and Gal O2 belonging to subunit A of one molecule and Thr102N, O and OG of subunit C of the molecule related by a translation.

Structural rationale for carbohydrate specificity

The present structure, along with that of the Me-α-Gal complex provides a reasonable qualitative explanation of the known affinities of jacalin to different carbohydrate ligands.[15,16] At the monosaccharide level, Me-α-Gal and GalNAc binds with higher affinity to jacalin than Gal does. All of them appear to bind at the primary site. The structure of the Me-α-Gal complex clearly brings out the favourable steric interactions of the methyl group in the ligand with aromatic residues Phe47, Tyr78, Tyr122 and Trp123, thus explaining the higher affinity achieved by α-methyl substitution. The T-antigen complex shows that GalNAc at the primary site has an additional hydrogen bond through the oxygen atom in the acetamido group to the terminal amino group of the α-chain, in addition to those involving the Gal moiety. Thus the binding affinity of GalNAc to jacalin is higher than that of Gal. Understandably, the affinity of Me-α-GalNAc is higher than the individual affinities of Me-α-Gal and GalNAc. Simple modelling shows that methyl substitution at the β anomeric site would lead to severe steric clashes with Tyr122, leading to the lower affinity of the Me-β-Gal than Gal. At the disaccharide level, lactose (Galβ1-4Glc), *N*-acetyllactosamine (Galβ1-4GlcNAc) and Galβ1−3GlcNAc are very poor ligands of jacalin. These disaccharides cannot bind to jacalin as T-antigen does, with the second residue at the primary binding site as this site is designed for galactose and not for glucose/
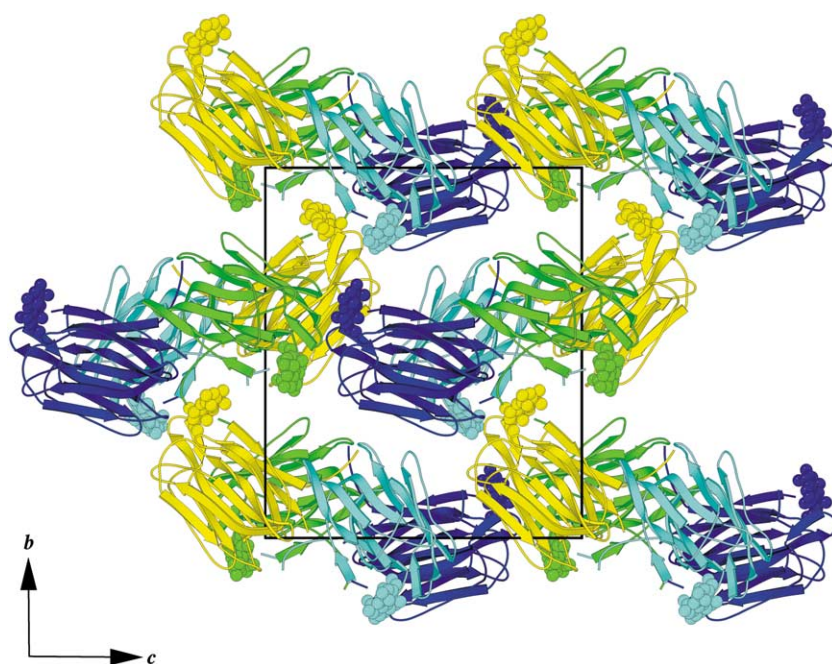
Figure 4. Involvement of T-antigen in crystal packing. Subunits A, B, C and D along with the bound sugar are shown in blue, cyan, green and yellow, respectively. See the text for details.

mannose. Attempts to dock them with the first residue at the primary site led to unacceptable steric clashes. Understandably, Galβ1–3GalNAc-α-Me (α-methylated T-antigen) has higher affinity for jacalin than T-antigen has, on account of the favourable steric interactions of the methyl group with the four aromatic residues, referred to earlier. Again, for reasons mentioned in relation to Me-β-Gal, Galβ1–3GalNAc-β-Me has much lower affinity for the lectin than T-antigen.

In a diagnostic context, jacalin would interact with T-antigen O-linked to a seryl or a threonyl residue. In order to explore the feasibility of such interactions with the GalNAc residue at the primary site, modelling studies were carried out with the α-anomer of bound T-antigen O-linked to N-acetyl-N-methylamides of serine and threonine. The aromatic ring of Tyr122 was oriented as in the Me-α-Gal complex, as the position of the methyl group corresponds to that of C_β in an O-linked amino acid residue. In both the cases, the protein and the sugar atoms (as in subunit C which is not involved in intermolecular interactions) were held constant while ϕ, ψ and χ^1 of the O-linked amino acid residues were varied and minimised using INSIGHT II as described in Materials and Methods. The minimisation led to models with acceptable torsion angles ($\phi = -145°$, $\psi = 144°$ and $\chi^1 = -81°$ in the seryl derivative and $\phi = -180°$, $\psi = 108°$ and $\chi^1 = -91°$ in the threonyl derivative) and good additional interactions with the protein. A C_β–H···π interaction involving the aromatic side-chain of Tyr122 exists in both the cases. Also observed are interactions of the main chain oxygen and nitrogen atoms in the seryl and threonyl derivatives with Tyr78 OH and Tyr122 OH. These interactions are presumably responsible for the higher affinity of Ser–T-antigen to the lectin compared to that of T-antigen itself.[16]

Comparison with MPA–T-antigen interactions

MPA is structurally very close to jacalin; so are MPA–T-antigen interactions to jacalin–T-antigen interactions.[11] However, the two sets of interactions exhibit some interesting differences. Only the α-anomer of O1 has been located in the MPA complex while both anomers exist in two subunits in jacalin. In jacalin, in all subunits, one water molecule connects Gal O6 to Val79 O and Asp125 OD1. The first of these bridges exists in the MPA complex also. However, the interaction with Asp125 is more tenuous with one more water molecule (altogether two water molecules) between Gal O6 and Asp125 OD1. Furthermore, unlike in the case of the jacalin complex, the β chain is not involved in interactions with Gal in the MPA complex. The weakening of the water bridge is presumably caused by a small outward movement of the sugar molecule in MPA in relation to its position in jacalin (Figure 5). This movement appears to follow that of the aromatic ring of Tyr78, which stacks against the GalNAc ring. In MPA, Tyr78 hydroxyls of two symmetry-related molecules are involved in a hydrogen bond. Furthermore, the phenolic ring also stacks against a Gal ring belonging to a neighbouring molecule. These intermolecular interactions apparently lead to the movement of the side-chain aromatic ring.

As in the case of jacalin, the crystal packing of MPA also involves protein–carbohydrate interactions. The interactions are, however, less numerous in MPA. Inter-molecular protein–carbohydrate interactions lead to a chain of linked tetramers in MPA also. However, adjacent tetramers are related by a translation in MPA whereas they are related by a 2_1 screw in jacalin.
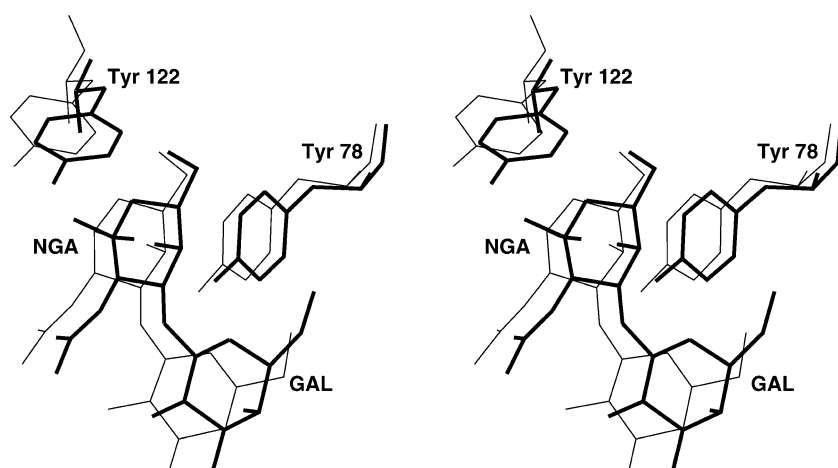
Figure 5. Stereo view of the positions of the bound sugar along with aromatic binding site residues of jacalin (thick line) on MPA (thin line) when the two lectins are superposed using all α carbon positions.

Comparison of lectin–T-antigen complexes

With the present structure, the crystal structures of the T-antigen complexes of five lectins, namely, PNA, LT, MPA, amaranthin and jacalin, are now available. These lectins, between them, represent four different structural families of lectins, three of plant origin and one of bacterial origin. Their combining sites are substantially different. However, in terms of their binding to T-antigen, they can be broadly classified into two categories. The Gal residue occupies the primary binding site in PNA and LT while the GalNAc residue occupies the site in the remaining three. This appears to be related to their ability to bind sialylated sugars (masked sugars). In sialylated sugars, N-acetyl neuraminic acid is linked to Gal of T-antigen through a α(2–3) linkage (NeuNAcα2–3 Galβ1–3GalNAc). Gal bound at the primary site in PNA and LT complexes cannot accommodate N-acetyl neuraminic acid without disrupting its interactions with the lectin. On the contrary, Gal is almost fully exposed to solvent in jacalin and MPA and at least O3 of Gal is exposed to solvent in amaranthin. The ability to discriminate between α and β substituents at the GalNAc end also depends on whether the residue is at the primary site. In jacalin, MPA and amaranthin, where the residue is at the primary site, steric factors favour α substitution. The residue is almost entirely exposed to solvent and there is no α, β discrimination in PNA and LT.

In all the complexes except that of amaranthin, most of the lectin–sugar interactions take place at the primary site. The number of such direct hydrogen bonds at the primary site varies between two and nine. In amaranthin the number of hydrogen bonds at the primary site is the lowest at two. There are, in addition, water bridges in PNA, LT and amaranthin. Stacking interactions involving the sugar ring at the primary site are also observed in all except the amaranthin complex. In PNA, the second sugar (GalNAc) has only just one direct hydrogen bond with the protein. However, there are bridges involving two water molecules between the acetamido group and the protein. This is the only additional interaction present in the PNA–T-antigen complex when compared to the PNA–lactose complex, and is believed to be responsible for the 20-fold increased affinity of T-antigen for PNA compared to the affinity of lactose.[6,7] Indeed the PNA–T-antigen complex provides a telling illustration of how carbohydrate specificity can be generated by water bridges. In the LT–T-antigen complex, the second sugar residue does not interact at all with the protein molecule to which the first residue is bound.

Among the three lectins to which T-antigen binds with GalNAc at the primary site, direct protein–Gal interactions exist only in amaranthin. Only water-bridges connect the Gal residue to the protein in jacalin and MPA. The residue is, however, involved in intermolecular interactions, the biological implications of which are not yet clear.

Conformation of T-antigen

The glycosidic torsion angles φ and ψ of the T-antigenic disaccharide (Galβ1–3GalNAc) bound to the lectins are listed in Table 2. The values of φ vary between 24 and 57° and ψ between −9 and

Table 2. Torsion angles (°) of the disaccharide T-antigen in different complexes

Subunit	φ	ψ
JAC A	29	−14
JAC B	34	−9
JAC C	28	−14
JAC D	24	−15
MPA	38	−15
PNA A	41	−25
PNA B	38	−27
PNA C	42	−41
PNA D	42	−32
AMA A	27	−38
AMA B	32	−35
LT	57	−14

The angles φ and ψ are defined as (H1–C1–O3′–C3′) and (C1–O3′–C3′–H3′), respectively.

−41°. Molecular mechanics and dynamics studies, supported by NMR investigations of free T-antigen, indicated a global valley of minimum energy around a φ value of 40° and the ψ values in the range of −60 to 60°.[31] These values are compatible with those found in the crystal structures of lectin–T-antigen complexes. The latter, however, access a narrower range of ψ values. A recently reported NMR spectroscopic study on the MPA–T-antigen complex suggests two different conformations with φ/ψ combinations of 45/ − 65° and −60/ − 18° for the bound sugar in solution.[32] Neither of the conformations is favoured in free state or observed in crystal structures. Simple modelling shows that in the jacalin–T-antigen complex φ/ψ values 45/ − 65° lead to a severe steric clash of Gal with the side-chain of the binding site residue Asp125 and a conserved water (W289), which is of crucial importance for sugar binding to the lectin. T-antigen with φ/ψ values −60/ − 18° has no steric clash with any of the binding site residues; also Gal can make two direct interactions with the lectin. The difference in glycosidic torsion angles observed in solution and in the crystals has been attributed to crystal packing in MPA.[32] However, the fact that nearly the same conformation is observed in all the lectin complexes, in spite of widely different environments, appears to suggest that it corresponds to the intrinsic conformational propensity of T-antigen.

Materials and Methods

Crystallisation

Crystals of jacalin purified as described earlier[33] complexed with T-antigen were grown using the vapour diffusion technique by equilibrating a 10 μl drop of 10 mg ml⁻¹ protein in the presence of 30 times molar excess of T-antigen in 0.02 M phosphate buffer (pH 7.3), containing 0.1 M NaCl, 0.025% (w/v) sodium azide and 8–10% (w/v) PEG4000 against a reservoir solution of 40% PEG4000 in the same buffer. Crystals of size 0.6 mm × 0.6 mm × 0.4 mm grew in two weeks.

Data collection and processing

The data were collected on the EMBL beam line X31 at a wavelength 1.1 Å at DESY, Germany, using a Mar research MAR345 imaging plate. The collected data were processed using programs DENZO and SCALE-PACK of the HKL suite of programs.[34] The processed data were converted to structure factors using the program TRUNCATE of CCP4.[35] Data collection statistics along with the cell parameters are given in the Table 3. Matthews coefficient[36] indicated the presence of one tetramer in the asymmetric unit.

Structure solution and refinement

The structure of jacalin complexed with T-antigen was solved using the molecular replacement program AmoRe,[37] with the jacalin–Me-α-Gal complex (PDB code 1JAC) as the starting model. The correct solution

Table 3. Data collection statistics

Space group	$P2_1$
Cell parameters	
a (Å)	58.92
b (Å)	78.00
c (Å)	67.91
$β$ (°)	100.73
Resolution (Å)	1.62
Last shell	1.65–1.62
No. of observations	188,335
No. of unique reflections	74,262 (3714)
No. of reflections with $I = 0$	1366 (216)
Data completeness (%)	96.7 (95.8)
Multiplicity	2.5 (1.8)
R_{merge} (%)	7.3 (28.9)
Z	2
Solvent content (%)	47

Values in parenthesis refer to the last resolution shell.

had a correlation coefficient (CC) of 0.68 and R-factor of 0.32. A total of 40 cycles of rigid body refinement followed by 100 cycles of positional refinement using CNS[38] with "mlf" target brought down the R and R_{free} to 0.269 and 0.285, respectively. Clear unambiguous density for the T-antigenic disaccharide appeared in the $F_o − F_c$ and $2F_o − F_c$ maps computed at this stage. The coordinates of the saccharide in the PNA–T-antigen complex (PDB code 1PEL) were used to fit the model in the electron density. Water oxygen atoms were added successively to the model, using peaks with heights greater than 2.5 sigma in $F_o − F_c$ maps and 0.8 sigma in $2F_o − F_c$ maps. Omit maps were used in the course of refinement to remove model bias. Bulk solvent corrections were used throughout the refinement. Refinement converged at R and R_{free} of 0.189 and 0.205, respectively. The model was checked using PROCHECK.[39] The refinement statistics are given in Table 4.

Table 4. Summary of refinement statistics

Resolution range (Å)	20–1.62
Number of reflections	
Total	74,262
Used in refinement	74,262
Working set	70,475
Test set	3787
Number of non-hydrogen atoms	
Protein	4592
Sugar	104
Solvent	497
R-factor (%)	18.9
R_{free}-factor (%)	20.5
Average of the temperature factor (Å²)	
Protein atoms	
Main chain	21.8
Side-chain	24.5
Sugar atoms	27.1
Solvent atoms	41.5
RMS deviations from ideal values	
Bond lengths (Å)	0.01
Bond angles (°)	2.3
Dihedral angles (°)	26.8
Improper angles (°)	2.69
Ramachandran plot	
Residues in most favoured regions (%)	89.5
Residues in additional allowed regions (%)	10.5

Analysis of the structure and modelling

Possible hydrogen bonds were identified using the program HBPLUS. Contacts involving oxygen and nitrogen atoms with distances less than 3.6 Å and with donar–hydrogen–aceptor angle greater than 90° were treated as hydrogen bonds. Molecular superpositions were performed using the program ALIGN.[40] The program NACCESS[41] was employed for calculating accessible surface areas. Electrostatic potentials were calculated using DELPHI-INSIGHT II (Biosym Technologies, 1992). T-antigen O-linked to the *N*-acetyl-*N*-methylamides of serine and threonine were modelled and optimised using the BIOPOLYMER module of INSIGHT II. The whole conformation space was searched in combination with conjugate gradient minimisation and torsion force constraint using DISCOVER. Figures 1, 2 and 4 were prepared using BOBSCRIPT[42] and Figures 3 and 5 were generated using INSIGHT II and FRODO,[43] respectively.

Atomic coordinates

The atomic coordinates and the structure factors were deposited in the Protein Data Bank (PDB code 1M26).

References

1. Springer, G. F. (1984). T and Tn, general carcinoma auto antigens. *Science*, **224**, 1198–1206.
2. Thatcher, N., Hashmi, K., Chang, J., Swindell, R. & Crowther, D. (1980). Anti-T antibody in malignant melanoma patients. Influence of response and survival following chemotherapy-changes in serum levels following C parvum, BCG immunization. *Cancer*, **46**, 1378–1382.
3. Vos, G. H. & Brain, P. (1981). Heterophile antibodies, immunoglobulin levels, and the evaluation on anti-T activity in cancer patients and controls. *S. Afr. Med. J.* **60**, 133–136.
4. Bray, J., Maclean, G. D., Dusel, F. J. & McPherson, T. A. (1982). Decreased levels of circulating lytic anti-T in the serum of patients with metastatic gastro-intestinal cancer: a correlation with disease burdan. *Clin. Expt. Immunol.* **47**, 176–182.
5. Springer, G. F., Desai, P. R., Wise, W., Carlstedt, S. C., Tegtmeyer, H., Stein, R. & Scanlon, E. F. (1990). Pan-carcinoma T and Tn epitopes: autoimmunogens and diagnostic markers that reveal incipient carcinomas and help establish prognosis. *Immunol. Ser.* **53**, 587–612. Review..
6. Banerjee, R., Das, K., Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1996). Conformation, protein–carbohydrate interactions and a novel sub-unit association in the refined structure of peanut lectin–lactose complex. *J. Mol. Biol.* **259**, 281–296.
7. Ravishankar, R., Ravindran, M., Suguna, K., Surolia, A. & Vijayan, M. (1997). Crystal structure of the peanut lectin–T-antigen complex. Carbohydrate specificity generated by water bridges. *Curr. Sci.* **72**, 855–861.
8. Ravishankar, R., Surolia, A., Vijayan, M., Lim, S. & Kishi, Y. (1998). Preferred conformation of *C*-lactose at the free and peanut lectin bound states. *J. Am. Chem. Soc.* **120**, 11297–11303.
9. Ravishankar, R., Suguna, K., Surolia, A. & Vijayan, M. (1999). 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. *Acta Crystallog. sect. D*, **55**, 1375–1382.
10. Transue, T. R., Smith, A. K., Mo, H., Goldstein, I. J. & Saper, M. A. (1997). Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nature Struct. Biol.* **4**, 779–783.
11. Lee, X., Thompson, A., Zhang, Z., Ton-that, H., Biesterfeldt, J., Ogata, C. *et al.* (1998). Structure of the complex of *Maclura pomifera* agglutinin and the T-antigen disaccharide, Galβ1,3GalNAc. *J. Biol. Chem.* **273**, 6312–6318.
12. Akker, F. V. D., Steensma, E. & Hol, W. G. (1996). Tumor marker disaccharide D-Gal-β1, 3-GalNAc complexed to heat-labile enterotoxin from *Escherichia coli. Protein Sci.* **5**, 1184–1188.
13. 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.
14. Allen, A. C. (1999). Structural features of IgA molecules which contribute to IgA nephropathy. *J. Nephrol.* **12**, 59–65.
15. Sastry, M. V., Banerjee, 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.
16. 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.
17. Kabir, S. (1998). Jacalin: a jackfruit (*Artocarpus heterophyllus*) seed-derived lectin of versatile applications in immunobiological research. *J. Immunol. Methods*, **212**, 193–211. Review..
18. Chatterjee, B. P., Vaith, P., Chatterjee, S., Karduck, D. & Uhlenbruck, G. (1979). Comparative studies of new marker lectins for alkali-labile and alkali-stable carbohydrate chains in glyco-proteins. *Int. J. Biochem.* **10**, 321–327.
19. Pereira, M. E., Loures, M. A., Villalta, F. & Andrade, A. F. (1980). Lectin receptors as markers for *Trypanosoma cruzi*. Developmental stages and a study of the interaction of wheat germ agglutinin with sialic acid residues on epimastigote cells. *J. Expt. Med.* **152**, 1375–1392.
20. 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.

21. Higuchi, T., Xin, P., Buckley, M. S., Erickson, D. R. & Bhavanandan, V. P. (2000). Characterization of the rabbit homolog of human MUC1 glycoprotein isolated from bladder by affinity chromatography on immobilized jacalin. *Glycobiology,* **10**, 659–667.

22. 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.

23. Pineau, N., Aucouturier, P., Brugier, J. C. & Preud'homme, J. L. (1990). Jacalin: a lectin mitogenic for human CD4 T lymphocytes. *Clin. Expt. Immunol.* **80**, 420–425.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. Yang, H. & Czapla, T. H. (1993). Isolation and characterization of cDNA clones encoding jacalin isolectins. *J. Biol. Chem.* **268**, 5905–5910.

30. Pal, D. & Chakrabarti, P. (1999). *Cis* peptide bonds in proteins: residues involved, their conformations, interactions and locations. *J. Mol. Biol.* **294**, 271–288.

31. Gilleron, M., Siebert, H. C., Kaltner, H., von der Lieth, C. W., Kozar, T., Halkes, K. M. *et al.* (1998). Conformer selection and differential restriction of ligand mobility by a plant lectin—conformational behaviour of Galβ1–3GlcNAcβ1-R, Galβ1–3Gal-NAcβ1-R and Galβ1-2Galβ1-R′ in the free state and complexed with galactoside-specific mistletoe lectin as revealed by random-walk and conformational-clustering molecular-mechanics calculations, molecular-dynamics simulations and nuclear Overhauser experiments. *Eur. J. Biochem.* **252**, 416–427.

32. Weimar, T., Bukowski, R. & Young, N. M. (2000). The conformation of the T-antigen disaccharide bound to *Maclura pomifera* agglutinin in aqueous solution. *J. Biol. Chem.* **275**, 37006–37010.

33. Kumar, S. G., Appukuttan, P. S. & Basu, D. (1993). α-D-Galactose-specific lectin from jackfruit (*Artocarpus integrifolia*) seed. *J. Biosci.* **4**, 257–261.

34. Otwinowsky, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Macromoleuclar Crystallography, Part A, Methods in Enzymology* (Carter, C. W. Jr & Sweet, R. M., eds), vol. 276, pp. 307–326, Academic Press, New York.

35. Collaborative Computational Project, No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.

36. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.

37. Navaza, J. (1994). AmoRe—an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.

38. 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.

39. 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.

40. Cohen, G. E. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallog.* **30**, 1160–1161.

41. Hubbard, S. J. & Thronton, J. M. (1993). NACCESS, Computer Program, Department of Biochemistry and Molecular Biology, University College, London

42. Esnouf, R. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132–134.

43. Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallog.* **11**, 268–272.