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. V. Pratap[1], A. Arockia Jeyaprakash[1], P. Geetha Rani[1], K. Sekar[2] A. Surolia[1] and M. Vijayan[1]*

[1]*Molecular Biophysics Unit and*

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

**Corresponding author*

The seeds of jack fruit (*Artocarpus integrifolia*) contain two tetrameric lectins, jacalin and artocarpin. Jacalin was the first lectin found to exhibit the β-prism I fold, which is characteristic of the *Moraceae* plant lectin family. Jacalin contains two polypeptide chains produced by a post-translational proteolysis which has been shown to be crucial for generating its specificity for galactose. Artocarpin is a single chain protein with considerable sequence similarity with jacalin. It, however, exhibits many properties different from those of jacalin. In particular, it is specific to mannose. The structures of two crystal forms, form I and form II, of the native lectin have been determined at 2.4 and 2.5 Å resolution, respectively. The structure of the lectin complexed with methyl-α-mannose, has also been determined at 2.9 Å resolution. The structure is similar to jacalin, although differences exist in details. The crystal structures and detailed modelling studies indicate that the following differences between the carbohydrate binding sites of artocarpin and jacalin are responsible for the difference in the specificities of the two lectins. Firstly, artocarpin does not contain, unlike jacalin, an N terminus generated by post-translational proteolysis. Secondly, there is no aromatic residue in the binding site of artocarpin whereas there are four in that of jacalin. A comparison with similar lectins of known structures or sequences, suggests that, in general, stacking interactions with aromatic residues are important for the binding of galactose while such interactions are usually absent in the carbohydrate binding sites of mannose-specific lectins with the β-prism I fold.

Keywords: β-prism I fold; *Moraceae* lectin; carbohydrate specificity; post-translational modification; stacking interactions

Introduction

Lectins are carbohydrate binding proteins which mediate various biological processes such as cell-cell communication, host-pathogen interactions, targeting of cells, cancer metastasis and differentiation, through recognising and binding specifically to diverse sugar structures. Although originally isolated from plants, they occur in animals, bacteria and viruses as well.[1,2] On account of their important biological properties and their use in research and medicine, structural studies on lectins have gathered added momentum in recent years.[3–8] Plant lectins account for about half the lectins of known three-dimensional structure†. Among them legume lectins constitute the most thoroughly studied family.

In addition to legume lectins, there are four other structural families of plant lectins. They involve hevein domains, β-trefoil, β-prism I and β-prism II folds.[6,9] Of these, the β-prism I fold in a lectin was first characterised in jacalin, one of the two lectins from jack fruit seeds (*Artocarpus integrifolia*). It was surmised that this fold was characteristic of *Moraceae* plant lectins. Each subunit of the

† http://www.cermav.cnrs.fr/database/lectine

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

tetrameric lectin, M_r 66,000 Da, essentially consists of two Greek key motifs and one Greek key-like motif. The lectin consists of two chains generated by post-translational modification involving proteolysis. The larger chain is 133 amino acid residues long and the smaller one has 20 residues. The lectin is galactose-specific at the monosaccharide level. It binds with high specificity to oligosaccharides α-linked to the tumour-associated T-antigenic disaccharide Galβ1,3GalNAc. The structure of the complex of jacalin with methyl-α-galactose clearly demonstrated that the specificity of the lectin to galactose is generated by the post-translational modification referred to earlier.[10] That from *Maclura pomifera* is another *Moraceae* lectin, which is, like jacalin, specific to galactose and is made up of two chains. Therefore, a framework for describing the carbohydrate specificity of galactose binding *Moraceae* lectins exists.

Artocarpin, the second lectin from jack fruit seeds, is, unlike jacalin, specific to mannose at the monosaccharide level. It has a high affinity for the hepta saccharide from horse radish peroxidase.[11–13] It possesses a potent and selective mitogenic effect on distinct T and B cells.[14] Like jacalin, artocarpin is tetrameric with a comparable molecular weight. However, unlike jacalin, its subunit is made up of a single polypeptide chain. Jacalin is glycosylated while artocarpin is not. Despite these and other differences between the two lectins, they were demonstrated to be homologous using partial sequence and X-ray rotation function studies.[15] One of the motivations for undertaking detailed structure analysis of artocarpin and its complex with methyl-α-mannose has been to explore the extent of this homology and the departures from it caused by sequence variation and the post-translational modification in jacalin. Artocarpin also turns out to be the first mannose-specific *Moraceae* lectin to be X-ray analysed. The present study is mainly intended to elucidate the precise structural basis for the difference in the carbohydrate specificities of jacalin and artocarpin, despite their similar three-dimensional structures. In the event, as discussed later, the study also provided useful general insights into the difference between galactose and mannose binding sites in lectins.

Results and Discussion

Three-dimensional structure

Native artocarpin crystallised in two different forms, both in the space group $P2_1$. The asymmetric unit of one of the forms (form I) contains one tetramer, while in the other form (form II) two tetramers are present in the asymmetric unit. Artocarpin complexed with methyl-α-D-mannose crystallised in the space group $P6_1$ with one tetramer in the asymmetric unit, thus accounting for 16 crystallographically independent subunits. Each subunit in the complexed form binds one sugar molecule. Each subunit contains 149 amino acid residues. The secondary and tertiary structures of artocarpin are similar to that of jacalin and other lectins having the β-prism I fold. Figure 1(a), (b) and (c) shows the schematic hydrogen-bonding pattern and the monomeric and tetrameric structures, respectively, of artocarpin. The 16 independent subunits have the same structure and so do the four tetramers. Unless otherwise specified, subunit A of form II is used for the discussion of the structure. Likewise, subunit A of the complex is used for discussion on sugar binding. Each subunit contains three four-stranded anti-parallel β-sheets, stacked like the faces of a prism. Two of the three β-sheets form Greek key motifs with topology (1,1,−3).[16] The third sheet also forms a Greek key motif, except that it is not made of a contiguous set of residues. The outer strands of this sheet come from the N terminus, while the inner strands are from the C terminus. The residues in the three Greek keys are:3-22 and 127-149 (Greek key 1); 24-69 (Greek key 2) and 77-124 (Greek key 3). The subunit possesses an internal 3-fold symmetry, which is not reflected in its sequence.

The artocarpin tetramer has 222 symmetry, as can be readily seen from Figure 1(c). Most of the hydrogen bonds between subunits A and B (C and D) come from the fourth strand of Greek key 3 of the first subunit and the first strand of Greek key 1 of the second subunit. Similarly, the inter subunit hydrogen bonds between subunits A and C (B and D) involve mainly the N-terminal arm of the two subunits and the loop region connecting Greek keys 1 and 2 (residues 22-25). There are no hydrogen bonds between subunits A and D (B and C). The surface area buried when pairs of subunits associate to form dimers, were calculated using the method of Lee and Richards[17] with a probe radius of 1.4 Å. The total surface area buried is consistently higher (≈25 %) in the A-B type of interface than that buried in the A-C type of interface. The surface area buried at the A-D interface is negligible. This, in combination with the fact that the number of hydrogen bonds are also higher in the A-B interface than in the A-C interface, appears to suggest that artocarpin is perhaps a dimer of a dimer.

Hydration

The final refined models of the two native forms contain 431 and 874 water molecules, respectively. The corresponding number in the complexed form is 146. A comparison of the hydration of the chemically equivalent but crystallographically independent subunits in the asymmetric unit was carried out for the 12 native subunits. The four ligand bound subunits were not used as the structure was solved at a comparatively lower resolution. On application of the criteria used earlier in this laboratory,[18,19] only one water molecule was found to be invariant in all the 12 subunits. This water molecule bridges the main-chain amide group of Ile126 with the carboxyl group of Leu76. A further

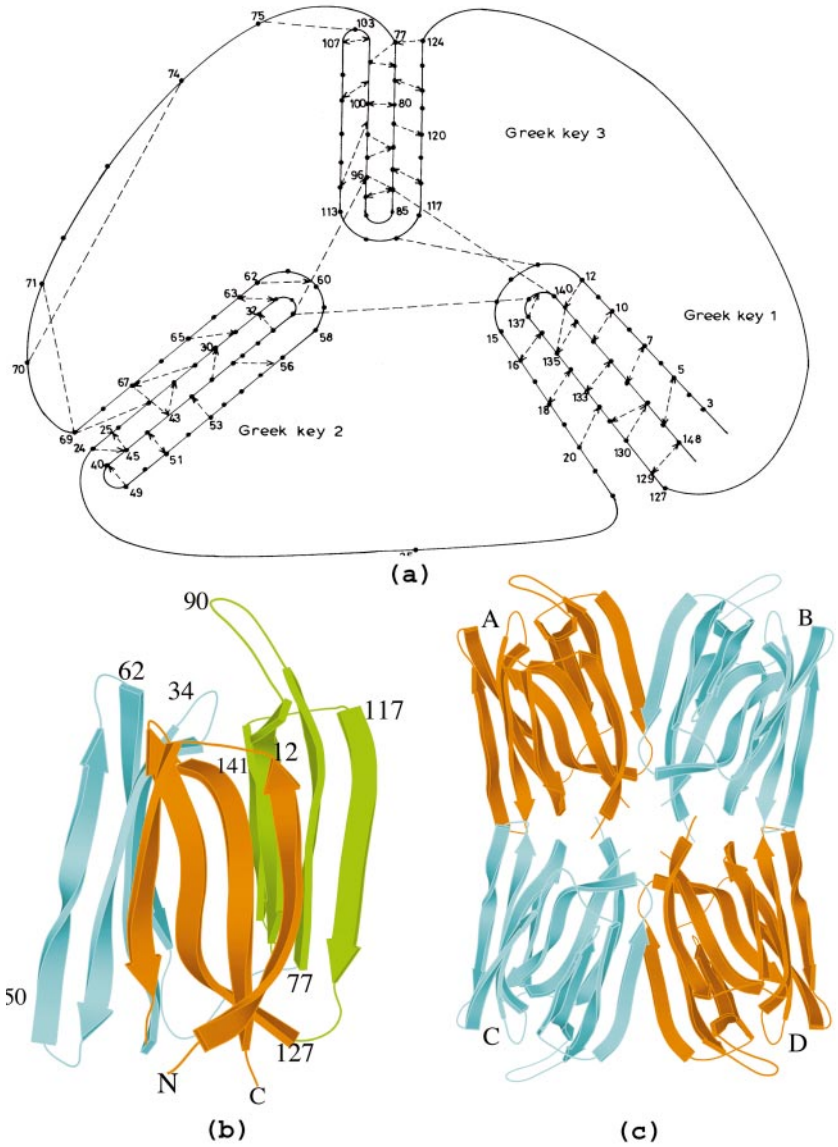

Figure 1. Structure of artocarpin. (a) Schematic representation showing hydrogen bonds. (b) The subunit with the three Greek keys coloured differently. (c) Quaternary structure with the four subunits coloured differently. Figures 1(b) 4 were prepared using BOBSCRIPT.[44]

five water molecules were found in any eight of the 12 subunits. Of these six water molecules, one interacts with the side-chain carboxylate of OD1 Asp19. The remaining five water molecules form a cluster near the bottom of the subunit and form a series of hydrogen bonds connecting the residues Ile26, Leu76 and Ile129. These three residues, are at the bottom of the subunit and have been implicated in the stabilisation of the subunit in jacalin. These water molecules thus appear to have a role in the stabilisation of the subunit. Figure 2 shows a close up view of these water molecules and their interactions.

Comparison with proteins of similar structure

The subunit of artocarpin is similar to that in jacalin, *Maclura pomifera* agglutinin (MPA),[20] *Helianthus tuberosus* lectin (heltuba),[7] domain II of δ-endotoxin[21] and the vitelline membrane outer layer protein.[22] A comparison of these structures shows that the structures belonging to the lectin family (jacalin, MPA, heltuba) have much more similarity amongst themselves than with those belonging to other types of proteins, as shown by the values of the relevant root mean square (r.m.s.) deviations. Also, among the lectins, the similarity is higher within the *Moraceae* family (with jacalin and MPA), even when their carbohydrate specificities are different. The subunit in all the proteins basically consists of three sub-domains, approximately 40 residues long, which appears to suggest that nature uses segments of this length for the generation of carbohydrate binding protein through gene fusion and multiplilication.[10] Recent reports have also shown that there are several lectins wherein 40 residue stretches are involved in carbohydrate binding.[6]

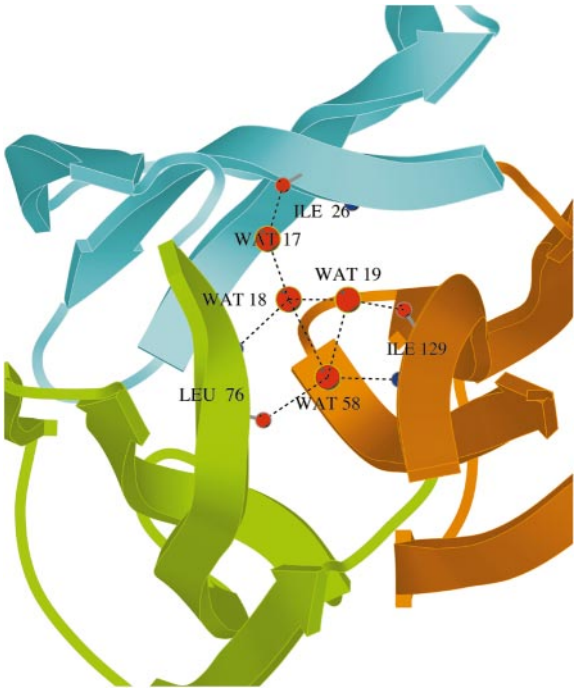

Figure 2. Close up view, approximately down the 3-fold axis, of the five invariant water molecules bridging the main-chain amide and oxygen atoms at the bottom of the subunit. See the text for details.

Lectin-carbohydrate interactions

Methyl-α-mannose has well defined density in all the four subunits (Figure 3). The lectin sugar interactions, which are the same in all the subunits,are also illustrated in the Figure. Sugar binding does not lead to any substantial change in the lectin structure. The native and the sugar-bound subunits superpose with an r.m.s. deviation of 0.30 Å in C^{α} positions, a value comparable to that obtained when subunits in the same tetramer are superposed on one another. The binding site is also unaffected by sugar binding. The r.m.s. deviation of C^{α} positions in the loop (residues 14-16, 91-93 and 137-141) constituting the site is as low as 0.25 Å. The average movement of the five atoms that hydrogen bond to the sugar is also about 0.30 Å. Thus the combining site in artocarpin is pre-formed to a remarkable extent.

The hydrogen bonds that stabilise the artocarpin-methyl-α-mannose complex are best discussed in relation to those in the methyl-α-galactose complex of jacalin and those in the dimannoside complexes of heltuba, the only other mannose-specific lectin of known structure with a β-prism I fold. The lectin sugar hydrogen bonds at the primary site in the three complexes are listed in Table 1. These are the same in the two complexes of heltuba and only those in one of them are listed. Superpositions of the combining sites in the jacalin and heltuba complexes on the artocarpin complex are illustrated in Figure 4. In all these complexes, the sugar is substituted at O1, which is not therefore readily available for interactions. O2 points to the solution in all the three and is not involved in protein-sugar interactions. All the other sugar oxygen atoms form hydrogen bonds with the protein in every complex. The hydrogen bonded interactions and the geometry of the sugar with respect to the lectin are very similar in artocarpin and heltuba. This is understandable as both the lectins are mannose-specific. However, the two lectins differ in their specificities for higher mannosides. Among dimannoses, heltuba binds Manα1-2Man more strongly than it does Manα1-3Man,[7] while the reverse is true with artocarpin.[11] In the crystal structures of the complexes of heltuba with Manα1-2Man and Manα1-3Man, the second mannose makes one hydrogen bond with the side-chain of His91 in the former and with the side-chain of Asp136 in the latter. However, Manα1-2Man has numerous van der Waals contacts with Met92, which makes it a stronger ligand. In artocarpin, His91 is replaced by Thr91, while Met92 is replaced by Pro92. Simple modelling of the two dimannosides into the carbohydrate binding site of artocarpin followed by rotation about the two glycosidic bonds showed that in the case of Manα1-3Man, O4 of the second mannose could make hydrogen bonds with the main-chain amide and oxygen atoms of the residues Leu89 and

Table 1. Protein-carbohydrate hydrogen bonds (lengths in Å) in artocarpin

Sugar atom	Protein atom	Distances (Å) in		
		Artocarpin	Jacalin	Heltuba
O3	Gly15 N	3.1	2.8	2.8
O4	Gly15 N	3.1	3.2	3.3
O4	Asp141 OD1	3.0	2.8	2.6
O5	Asp138 N	3.0	3.0	2.9
O6	Asp138 N	3.3	3.3	2.8
O6	Leu139 N	3.1	3.0	2.9
O6	Asp141 OD1	3.5	2.8	3.6
O6	Asp141 OD2	2.6	-	2.7

The corresponding values in heltuba (PDB code :1C3 M) and jacalin (PDB code 1JAC) are also given. Residue numbering is as in artocarpin.

(Gal O4 in jacalin has an additional interaction with OD2 of A125 Asp and Gal O6 has with the main-chain oxygen of A123 Trp.)

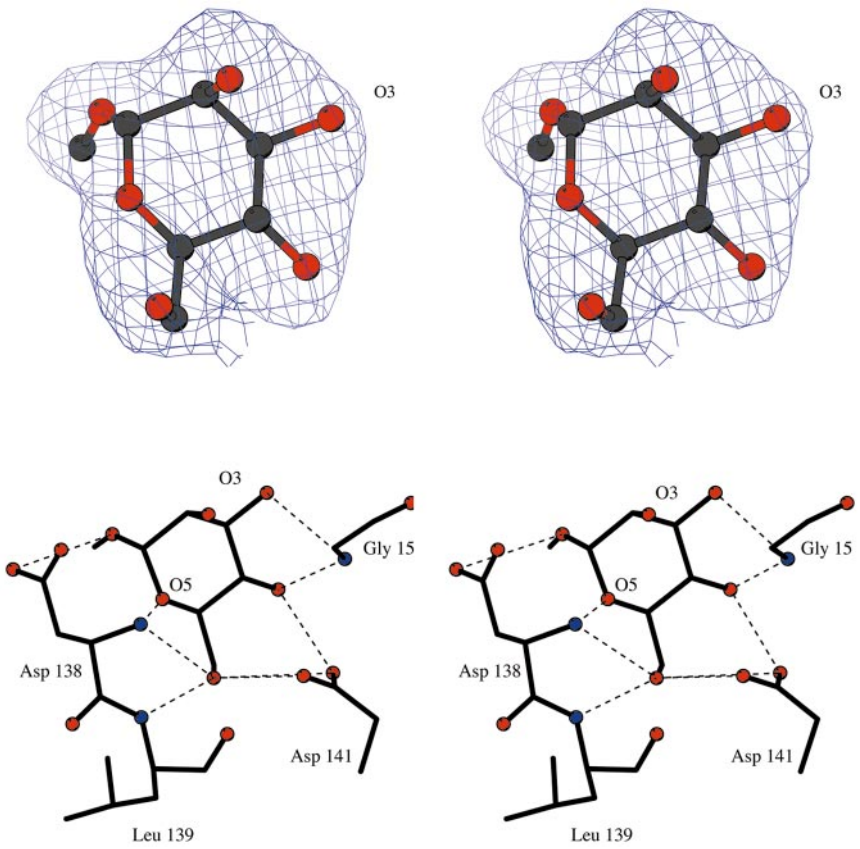

Figure 3. (a) Stereo view of the sugar molecule in the A subunit with the $2|F_o| - |F_c|$ map contoured at 1σ and (b) hydrogen bonds observed between the protein and sugar.

Ala90, while O2 could make hydrogen bonds with the side-chain oxygen atoms of Asp138. In the case of Manα1-2Man, O4 of the second ring points into the solution and O2, being the linker oxygen in the disaccharide, is not available for hydrogen bonding. O3 forms a hydrogen bond with one of the side-chain oxygen atoms of Asp138, while O6 interacts with carbonyl oxygen of Ala90. There is no significant van der Waal interaction between the second ring and the protein in both the models. Hence, the higher affinity of artocarpin for Manα1-3Man can be qualitatively explained based on the number of hydrogen bonds that the second ring makes with the protein. Any firm conclusion, however, should await the experimental determination of the structures of the two complexes.

The protein-sugar hydrogen bonds are almost the same in the jacalin-methyl-α-galactose and artocarpin-methyl-α-mannose complexes in spite of the different specificities of the two lectins. This is achieved by a set of small movements which were explored by superposing, to start with, C^{α} positions in the subunits of the two lectins after excluding the residues in the sugar binding sites (residues A1, A2, A121, A122, A123, A124 and A125 in jacalin and their equivalents in artocarpin). The deviations in the C^{α} positions in these seven common residues in the binding site range from 0.6 Å to 2.4 Å with an average value of 1.2 Å. The seven C^{α}s in the binding site of artocarpin were now superposed on those in jacalin. This involved a rotation of 18° (about an axis passing through the point (33.6, 54.7, 35.6) with direction cosines (−0.49, 0.72, −0.49)) and a translation of 0.82 Å. The r.m.s. deviations in the C^{α} positions reduced to 0.35 Å. Among the atoms that hydrogen bond to the sugar, the main-chain nitrogen atoms of 138 and 139 (artocarpin numbering) superposed with less than 0.2 Å while Gly15 N and the side-chain oxygen atoms of Asp141 exhibited deviations ranging from 0.7 to 1.5 Å. The superposition between the two sugars could be obtained by the rotation of bound mannose by 10° (about an axis passing through the point (33.7, 65.5, 71.2) with direction cosines (−0.66, 0.11, −0.72)) and a translation of 0.78 Å. All the ring atoms now superposed with deviations of less than 0.1 Å. As expected, O2 and O4, which have different orientations in galactose and mannose, exhibited deviations well over 2 Å. Thus the conservation of the lectin-sugar hydrogen bonds in the two complexes is achieved through semi-independent rigid body motions of the binding site and the sugar, and small displacements of a few atoms in the binding site. The delineation of these movements, however, does not provide a straight forward structural rationale for the difference in the specificities of jacalin and artocarpin.

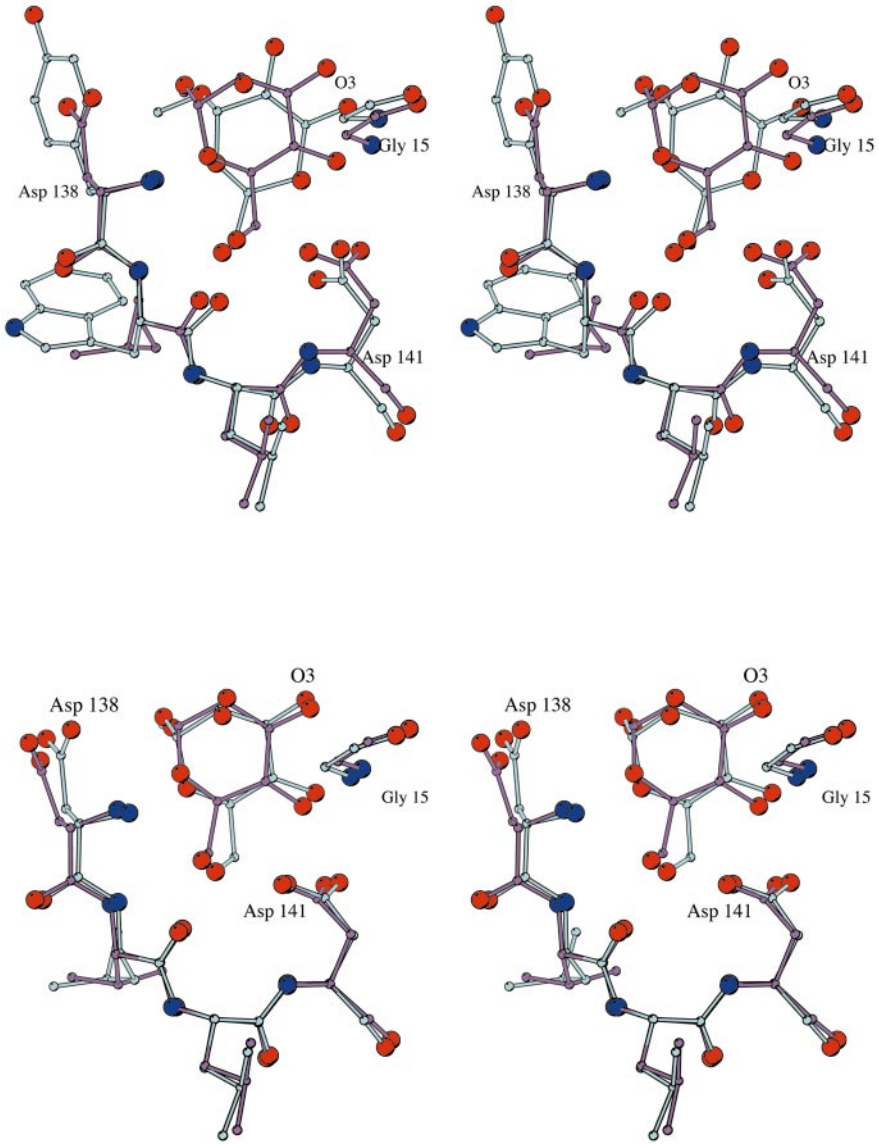

Figure 4. Stereo view of the superposition of the carbohydrate binding regions of artocarpin (cyan) on jacalin (top) and heltuba (bottom). The residue numbers correspond to artocarpin.

Simple modelling involving the replacement of an axial hydroxyl by an equatorial hydroxyl at C4 (the configurations at C2 is of no consequence), showed that jacalin can easily accommodate mannose in the binding site. Likewise, artocarpin can accommodate galactose as well. No steric clash occurs in either case when the configuration at C4 is changed. In both cases the change results in the abolition of the hydrogen bond of O4 with Asp125 or 141 OD2. In the case of jacalin, the change leads to the disruption of one more, and crucially important, hydrogen bond of O4 with the terminal amino group generated by post-translational modification. Thus, a preference of jacalin for galactose as against mannose, is understandable, but the observed degree of the preference is unexplainable. The structural basis for the discrimination between mannose and galactose in artocarpin appeared still more tenuous. The same is true about the mannose specificity of heltuba. Based on modelling studies,[7,23] steric clashes of one type or another have been invoked to explain the inability of heltuba and artocarpin to bind galactose. As indicated earlier, simple geometrical considerations based on the structure of the artocarpin-methyl-α-mannose complex do not substantiate this suggestion.

Interaction energies

As the crystal structures could not readily provide an explanation for the overwhelming preference of jacalin for galactose in relation to that for mannose and that of artocarpin for mannose over that for galactose, energy minimisations were performed on models, based on the crystal structures of the complexes of jacalin and artocarpin with galactose as well as mannose. That observed in the crystal structure, after removal of the methyl group

attached to O1, was used as the starting model of the jacalin-galactose complex. The model of the artocarpin-mannose complex was similarly obtained from the crystal structure of the complex of artocarpin with methyl-α-mannose. The starting models for the jacalin-mannose and artocarpin-galactose complexes were obtained by appropriately changing the positions of sugar hydroxyls O2 and O4 in their respective complexes.

Energy minimisation of the models of the four complexes were performed as described in Materials and Methods. The lectin-ligand interactions in the minimised models are by and large similar to those observed in the crystal structures. The main difference between the two jacalin complexes is that while O4 of mannose has no interactions with jacalin, that of galactose has several, particularly with the post-translationally generated N terminus. Presumably on account of the interactions involving O4, the interaction energy in the jacalin complexes favours galactose by 12 kcal/mol (Table 2). The apolar surfaces buried on complexation and shape complementarity[26] (Table 2) are nearly the same in the two complexes. Thus, the carbohydrate specificity of jacalin for galactose at the primary binding site does indeed appear to be primarily a result of the post-translational modification. The explanation provided by the minimised artocarpin complexes, however, is not as clear cut as in the case of jacalin. The apolar surface areas buried upon complexation again show only marginal differences between the two artocarpin complexes (Table 2). The interaction energy favours mannose over galactose in conformity with the experimental observation, but only by 2 kcal/mol. The hydrogen bonds in the minimised models of the artocarpin complexes do not indicate any preference for one or the other of the sugars. Thus the energy minimisation studies also fail to provide a convincing rationale for the specificity of artocarpin in a straightforward manner.

Aromatic residues and galactose specificity

A comparison of the sequences of lectins of known structure having the β-prism I fold, revealed that in the galactose binding proteins (jacalin, MPA), the carbohydrate binding region has four aromatic residues (Phe47, Tyr78, Tyr122 and Trp123), while there are none in the case of mannose binding proteins (artocarpin, KM+, heltuba). Corresponding to Phe47 in jacalin, there is a single residue deletion in artocarpin. The loop in which Tyr78 occurs has a four residue insertion in artocarpin and is replaced by a proline. Moreover, none of the residues in this longer loop are aromatic. Tyr122 and Trp123 are now an Asp and a Leu, respectively. It is well known that galactose binding is almost always accompanied by a stacking interaction with an aromatic residue against the B face of the sugar.[27,28] This stacking interaction involves an extended patch of partially positively charged aliphatic protons on the B-face of the ring and the π electrons of the aromatic residue. The work of Drickamer,[29] Iobst & Drickamer[30] and Kolatkar & Weis,[31] in engineering galactose specificity to a mannose binding protein amongst C-type lectins, showed the importance of the stacking interaction for Gal binding, although it did not confer specificity. (Specificity in that case was achieved by the insertion of a short Gly-rich loop, which made the aromatic residue adopt a conformation that sterically hindered mannose.) The equatorial hydroxyl at the fourth position, as in Glc/Man, reduces the accessibility to the B-face and hence stacking is not seen in Glc/Man-specific lectins.

In the light of the above observations, it appears that the galactose specificity for jacalin can be understood as generated from the mannose-specific artocarpin, a putative precursor of jacalin, by a two-step process: mutation of key residues in the vicinity of the sugar binding pocket to aromatic residues and the cleavage of a short loop, so as to create a positively charged N terminus which interacts specifically with O4 at the axial position.

In general, there are two significant differences between the mannose binding lectins and the galactose binding ones belonging to the β-prism I fold structural family in relation to carbohydrate specificity. The first of them is that the galactose binding proteins are two-chain molecules, generated by a post-translational modification, while the mannose binding proteins are single-chain molecules. The second difference between the two subgroups is the presence/absence of aromatic residues involved in stacking interactions with the carbohydrate in Gal/Man binding proteins. Although sequence (and structural) information for only two proteins belonging to the galactose

Table 2. Surface area buried on complexation, shape complementarity and interaction energies in the energy minimised structures

System	Surface area buried (Å^2)	Shape complementarity	Interaction energy (kcal/mol)
Jacalin-galactose	185 (128)	0.89	−33
Jacalin-mannose	198 (119)	0.86	−21
Artocarpin-galactose	175 (106)	0.77	−34
Artocarpin-mannose	201 (106)	0.70	−36

The apolar surface area buried is given in parenthesis.

binding sub-class are known, the sequences of six mannose binding proteins of this family are available. The latter occurs in taxonomically unrelated families including the *Moraceae* lectins (artocarpin, KM+), *Convolvulaceae* (calsepa), the *Gramineae* lectins (jacalin related lectins from barley and wheat), *Asteraceae* (heltuba) and *Musaceae* (banlec) lectins.. A comparison of the sequences of all these proteins[7] indicates the absence of aromatic residues near the carbohydrate molecule in all of them. One is therefore tempted to accord a more significant role for stacking interactions than previously thought of, in the generation of carbohydrate specificity.

Materials and Methods

Crystal structure determination

Artocarpin in the native form was purified and crystallised as reported earlier.[13] The complexed protein was crystallised using 40% (w/v) methoxy PEG 350 as the precipitant using the hanging drop method. A typical drop contained 10 µl of 20 mg/ml protein at pH 7.4 in phosphate buffer with 0.15 M NaCl and 1.5 µl of 40% methoxy PEG 350 equilibrated against a reservoir solution of 1 ml 40% methoxy PEG 350. Intensity data from form I were collected on a Siemens-Nicolet area detector, while those from form II and the complexed protein were collected on a MAR imaging plate system. XENGEN[32] and XDS[33] were used for data processing for forms I and II, respectively, while DENZO/SCALEPACK[34] was used for the complexed protein. Details of data processing are given in Table 3.

The structures of the two native crystal forms were solved with molecular replacement techniques using AMoRe[35] with the jacalin tetramer as the search model. X-PLOR[36] was used to refine the structures. Initially, refinement of form I was undertaken, as it has only one tetramer in the asymmetric unit. Rigid body refinement with the non-glycine residues in the jacalin model replaced by alanine, followed by 100 cycles of positional refinement using data from the 20-3 Å resolution shell, resulted in R and R_{free} of 37.2% and 44.1%, respectively. Electron density maps were calculated at this stage and the partial sequence information available[15] was used for model building followed by subsequent refinement. 50

Table 3. Data collection and refinement statistics

	Form I	Form II	Complex
Crystal size (mm)	0.5 × 0.5 × 0.2	0.5 × 0.5 × 0.2	1.5 × 0.3 × 0.3
Radiation used	CuK_α	CuK_α	CuK_α
Space group	$P2_1$	$P2_1$	$P6_1$
Unit cell dimensions			
a (Å)	69.88	87.69	129.20
b (Å)	73.74	72.19	129.20
c (Å)	60.64	92.63	78.61
β (°)	95.1	101.2	
Z	2	4	6
Resolution (Å)	2.5	2.4	2.9
Last shell (Å)	2.65-2.5	2.5-2.4	3.0-2.9
No. of observations	54,998	133,210	31,710
No. of unique reflections	17,578	42,342	15,219
Reflections with $I = 0$	0	5583	219
Completeness (%)	86.0 (39.9)	96.1 (56.0)	91.3 (83.7)
R_{merge} (%)[a]	9.0 (40.1)	8.9 (37.9)	15.7 (41.6)
Multiplicity	2.4 (2.6)	3.1 (3.2)	2.1 (1.8)
Protein atoms	4505	9053	4422
Sugar atoms	-	-	52
Solvent atoms	431	874	146
R factor[b]	19.9	19.1	22.2
R_{free}[c]	26.2	25.8	25.9
Resolution range (Å)	20-2.5	20-2.4	20.0-2.9
No. of reflections	17,578	36,759	15,000
RMS deviations from ideal values[d]			
Bond length (Å)	0.01	0.01	0.01
Bond angle (deg.)	1.7	1.5	3.5
Dihedral angle (deg.)	28.9	25.1	26.6
Improper (deg.)	1.6	1.5	2.0
Residues(%) in Ramachandran plot[e]			
Core region	84.0	86.0	81.8
Additionally allowed region	15.2	13.4	16.9
Generously allowed region	0.8	0.6	1.3
Disallowed region	0.0	0.0	0.0

[a] $R_{merge} = \Sigma|I_i - \langle I\rangle|/\Sigma\langle I\rangle$. The values within the parentheses refer to the last shell.
[b] $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$; R_{free} calculated the same way but for a subset of reflections that is not used in the refinement. No σ cutoff was applied.
[c] 1036 and 1441 reflections were set for calculating R_{free} in form I and II, respectively.
[d] Deviations from ideal geometry parameters as defined by Engh & Huber.[42]
[e] As calculated by PROCHECK.[43]

of the 87 known side-chains were fitted progressively. Refinement was then extended to 2.5 Å in steps. R and R_{free} were 33.1 and 44.4 %. The linker region connecting the C terminus of the β-chain to the N terminus of the α-chain of jacalin was identified from the subsequent symmetry-averaged map constructed using RAVE_SGI.[37] A two-residue deletion was also identified in one of the loops near the carbohydrate binding region. The sequence information corresponding to the linker region as the two-residue deletion region were not available at that time and hence, for confirmation, the structure analysis of the second form was also undertaken. The structure analyses of both crystal forms were then pursued simultaneously and independently. In the absence of further sequence information, a combination of the sequence of jacalin, with which artocarpin shares a sequence similarity of approximately 50 %, and the electron density maps were used and the structures refined. In places where the density was ambiguous, alanine was retained. In this manner, about 110 side-chains were progressively fitted into the electron density. The R and R_{free} at this stage were 29.3 % and 40.1 % (form I) and 28.2 % and 38.1 % (form II).

Still, a couple of loops near the carbohydrate binding region were not clearly defined. A map obtained after density modification using DM[38] in the CCP4 suite of programs[39] was used for further model building. The density modification involved solvent flattening, use of Sayre's equation, histogram matching and NCS averaging. A deletion and a couple of insertions could be identified from the map. At this stage (R and R_{free} 23.0 and 29.5 % for form I, and 23.0 and 29.1 % for form II), the sequence of KM+, a lectin from the south American jack fruit seeds became available.[23] This protein is similar to artocarpin in its physicochemical properties. It is a tetrameric, mannose-specific lectin with the subunit made of a single polypeptide chain of 147 residues. A comparison of this sequence with that of the sequence fitted for artocarpin showed that the two proteins share a high degree of sequence identity. Figure 5 shows a comparison of KM+ sequence with that of artocarpin. The insertions and deletions observed in the structure of artocarpin were observed in the sequence of KM+ as well. The insertion after jacalin Ala79, was four residues long in KM+, while we had added two residues from the electron density map. The extra residues also came up during subsequent refinement of the model. Also, unlike in jacalin, the N-terminal alanine was found to be acetylated. A total of 431 solvent oxygen atoms were identified in form I. The corresponding number in form II was 874. A 1/8th omit map[40] was computed during the final stages to remove model bias. The final R and R_{free} converged to 19.9 % and 26.2 % for form I and 19.1 % and 25.8 % for form II. The refinement statistics are given in Table 3.

The structure of artocarpin complexed with methyl-α-mannose was solved using AMoRe,[36] with tetramer 1 of form II as the search model. CNS[41] was used for the refinement of this structure. Conventional $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps were calculated after one cycle of refinement and the sugar molecules in the four subunits were modelled into the electron density. Solvent molecules were added during the next cycle of refinement. The final model has 146 solvent molecules and refined to a R value of 22.2 % and an R_{free} of 25.9 %. Refinement statistics are given in Table 3.

```
KM+   : ASQTITVGPW  GGPGGNGWDD  GSYTGIRQIE
ARTO  : ASQTITVGSW  GGPGGNGWDE  GSYTGIRQIE30

KM+   : LSYKEAIGSF  SVIYDLNGEP  FSGPKHTSKL
ARTO  : LSYDEAIGSF  SVIYDLNGDP  FSGPKHTSKL60

KM+   : PYKNVKIELR  FPDEFLESVS  GYTAPFSALA
ARTO  : PYANVKIALA  FPDEFLESVS  GYTAPFSALA90

KM+   : TPTPVVRSLT  FKTNKGRTFG  PYGDEEGTYF
ARTO  : TPTPVVRSLT  FKTNKGRTFG  PYGDEEGTYF120

KM+   : NLPIENGLIV  GFKGRTFDLL  DAIGVHMAL
ARTO  : NLPIENGLIV  GFKGRTGDLL  DAIGIHMSL149
```

Figure 5. Comparison of the sequence of KM+ with that of artocarpin finally arrived at. Identical residues are highlighted.

Energy minimisation of protein-ligand complexes

The final refined coordinates of subunit A of the complexed structure were used for the energy minimisation of the artocarpin complexes. Subunit A of jacalin (PDB code 1JAC) was used for the minimisation involving the jacalin complexes. Minimisation, restricted to residues near the sugar binding region, was performed using INSIGHTII (Biosym Inc.). Residues within 13 Å from the sugar ring oxygen (O5) were allowed to move. These residues include 11-18, 35-39, 57-62, 84-97 and 134-143 and the ligand in artocarpin. The corresponding residues in the case of jacalin are A1-A4, A17-A27, A44-A51, A71-A83, A119-A127 and B15-B18. Hydrogen atoms were added based on geometric considerations, using the HBUILDER module. The minimisation region was solvated using the SOAK module with a 5 Å water shell. A dielectric constant of 1.0 was used throughout the minimisation. A 12 Å cutoff was used for caluclating nonbonded interactions. A combination of steepest descent and conjugate gradient method was used for minimisation. During the steepest descent method, the heavy atoms were tethered initially with a force constant of 100 kcal/mol Å^2, which was subsequently reduced to 50 and 25 before removing it altogether. The steepest descent refinement was followed by 10,000 steps of conjugate gradient refinement. The final coordinate sets thus obtained were then analysed for hydrogen bonding pattern, van der Waals contacts and buried surface area. The interaction energy between the sugar and the protein atoms were calculated for the initial and final structures. The solvent molecules were excluded from this calculation. The interaction energy is the sum of the electrostatic and van der waals interaction terms between the two sets of atoms.

Atomic coordinates

The atomic coordinates and the structure factors have been deposited in the PDB. The codes for the atomic coordinates are 1J4S and 1J4T for the two forms of the native crystal and 1J4U for the complex crystal.

Acknowledgements

The intensity data were collected at the X-ray Facility for Structural Biology at the Institute, supported by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT). Computations and model building were done at the Supercomputer Education and Research Centre, and also at the Interactive Graphics Facility and the Distributed Information Centre, supported by the DBT. We thank Drs A. Stephen Suresh and R. Sankaranarayanan for their help during the early stages of the work. We thank the DST for financial support.

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. Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). Legume lectin structure. *Biochim. Biophys. Acta,* **1383**, 9-36.
4. Drickamer, K. (1999). C-type lectin-like domains. *Curr. Opin. Struct. Biol.* **9**, 585-590.
5. Rini, J. (1999). New animal lectin structures. *Curr. Opin. Struct. Biol.* **9**, 578-584.
6. Vijayan, M. & Nagasuma, R. C. (1999). Lectins. *Curr. Opin. Struct. Biol.* **9**, 707-714.
7. Bourne, Y., Zamboni, V., Barre, A., Peumans, W. J., van Damme, E. J. & Rouge, P. (1999). *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. *Struct. Fold. Des.* **7**, 1473-1482.
8. Imberty, A., Gautier, C., Lescar, J., Perez, S., Wyns, L. & Loris, R. (2000). An unusual carbohydrate binding site revealed by the structures of the *Maackia amurensis* lectins complexed with sialic-acid containing oligosaccharides. *J. Biol. Chem.* **275**, 17541-17548.
9. Wright, C. S. (1989). Comparison of the refined crystal structures of two wheat germ isolectins. *J. Mol. Biol.* **209**, 475-487.
10. 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 beta-prism fold. *Nature Struct. Biol.* **3**, 596-603.
11. Misquith, S., Rani, P. G. & Surolia, A. (1994). Carbohydrate binding specificity of the B-cell maturation mitogen from *Artocarpus integrifolia* seeds. *J. Biol. Chem.* **269**, 30393-30401.
12. Rani, P. G., Bachhawat, K., Misquith, S. & Surolia, A. (1999). Thermodynamic studies of saccharide binding to artocarpin, a B-cell mitogen, reveals the extended nature of its interaction with mannotriose [3,6-Di-O-(α-D-mannopyranosyl)-D-mannose]. *J. Biol. Chem.* **274**, 29694-29698.
13. Rani, P. G., Bachhawat, K., Reddy, G. B., Oscarson, S. & Surolia, A. (2000). Isothermal titration calorimetric studies on the binding of deoxytrimannose derivatives with artocarpin: implications for a deep-seated combining site in lectins. *Biochemistry,* **39**, 10755-10760.
14. De Miranda-Santos, I. K. F., Mengel, J. O., Jr, Bunn-Moreno, M. M. & Campos-Neto, A. (1991). Activation of T and B cells by crude extract of *Artocarpus integrifolia* is mediated by a lectin distinct from jacalin. *J. Immunol. Methods,* **140**, 197-203.
15. Suresh, A. S., Rani, P. G., Pratap, J. V., Sankaranarayanan, R., Surolia, A. & Vijayan, M. (1997). Homology between jacalin and artocarpin from jack fruit (*Artocarpus integrifolia*) seeds. Partial sequence and preliminary crystallographic studies of artocarpin. *Acta Crystallog. sect. D,* **53**, 469-471.
16. Richardson, J. S. (1981). The anatomy and taxonomy of protein structures. *Advan. Protein Chem.* **34**, 167-339.
17. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.
18. Nagendra, H. G., Sudarsanakumar, C. & Vijayan, M. (1996). An X-ray analysis of native monoclinic lysozyme. A case study on the reliability of refined protein structures and a comparison with the low humidity form in relation to mobility and enzyme action. *Acta Crystallog. sect. D,* **52**, 1067-1074.
19. Sadasivan, C., Nagendra, H. G. & Vijayan, M. (1998). Plasticity, hydration and accessibility in Ribonuclease A. The structure of a new crystal form and its low humidity variant. *Acta Crystallog. sect. D,* **54**, 1343-1352.
20. 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.
21. Li, J., Carrol, J. & Ellar, D. J. (1991). Crystal structure of insecticidal delta endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. *Nature,* **353**, 815-821.
22. Shimuzu, T., Vasylyev, D. G., Kido, S., Doi, Y. & Morikawa, K. (1994). Crystal structure of the vitelline membrane outer layer protein I (VMOI): a folding motif with homologous Greek key structures related by an internal 3-fold symmetry. *EMBO J.* **13**, 1003-1010.
23. Rosa, J. C., De Oliviera, P. S. L., Garratt, R., Beltramini, L., Resing, K., Roque-Barreira, M-C. & Greene, L. J. (1999). KM+, a mannose-binding lectin from *Artocarpus integrifolia*: amino acid sequence, predicted tertiary structure, carbohydrate recognition and analysis of β-prism fold. *Protein Sci.* **8**, 13-24.
24. Bhat, T. N., Sasisekharan, V. & Vijayan, M. (1979). An analysis of side-chain conformations in proteins. *Int. J. Peptide Protein Res.* **13**, 170-184.
25. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, J. C. (1999). Asparagine and glutamine rotamers: B-factor cutoff and correction of amide flips yield distinct clustering. *Proc. Natl Acad. Sci. USA,* **96**, 400-405.
26. Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946-950.
27. Weis, W. I. & Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441-473.
28. Elgavish, S. & Shaanan, B. (1998). Structures of *Erythrina corallodendron* lectin and of its complexes with mono and disaccharides. *J. Mol. Biol.* **277**, 917-932.
29. Drickamer, K. (1992). Engineering galactose-binding activity into a C-type mannose binding protein. *Nature,* **360**, 183-186.
30. Iobst, S. T. & Drickamer, K. (1994). Binding of sugar ligands to Ca^{2+}-dependent animal lectins II. Gener-

ation of high affinity galactose binding by site directed mutagenesis. *J. Biol. Chem.* **269**, 15512-15519.
31. Kolatkar, A. R. & Weis, W. I. (1996). Structural basis of galactose recognition by C-type animal lectins. *J. Biol. Chem.* **271**, 6679-6685.
32. Kabsch, W. (1988). Evaluation of single crystal X-ray diffraction data from a position sensitive detector. *J. Appl. Crystallog.* **21**, 916-924.
33. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallog.* **26**, 795-800.
34. Otwinowsky, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology Macromolecular Crystallography, Part A* (Carter, C. W., Jr & Sweet, R. M., eds), vol. 276, pp. 307-326, Academic Press.
35. Navaza, J. (1994). AMoRe - an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 445-449.
36. Brunger, A. T. (1992). *X-PLOR Version 3.1. A System for Crystallography and NMR*, Yale University, New Haven, USA, CT.
37. Kleywegt, G. J. & Jones, T. A. (1995). Halloween ... masks and bones. In *From First Map to Final Model* (Bailey, S., Hubbard, R. & Waller, D. A., eds), pp. 59-66, SERC Daresbury Laboratory, Warrington, UK.
38. Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, vol. 31, pp. 34-38.
39. SERC Daresbury Laboratory WarringtonW A4 4AD England (1994). *Acta Crystallog. sect. D,* **50**, 760-763.
40. Vijayan, M. (1980). On the Fourier refinement of protein structures. *Acta Crystallog. sect. A,* **36**, 295-298.
41. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P. & Grosse-Kunstleve, R. W., *et al.* (1998). Crystallography and NMR systems (CNS): a new software system for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905-921.
42. Engh, R. A. & Huber, R. (1991). Accurate bond angle parameters of X-ray protein structure refinement. *Acta Crystallog. sect. A,* **47**, 392-400.
43. Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
44. Esnouf, R. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134.