Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Two orthorhombic crystal structures of a galactose-specific lectin from *Artocarpus hirsuta* in complex with methyl-α-D-galactose

K. N. Rao, C. G. Suresh,* U. V. Katre, S. M. Gaikwad and M. I. Khan

Division of Biochemical Sciences, National Chemical Laboratory, Pune-8, India

Correspondence e-mail: suresh@ems.ncl.res.in

Received 2 December 2003
Accepted 4 June 2004

Based on their carbohydrate specificity, the jacalin family of lectins can be divided into two groups: galactose-specific and mannose-specific. The former are cytoplasmic proteins, whereas the latter are localized in the storage vacuoles of cells. It has been proposed that the post-translational modification in some of the lectins that splits their polypeptide chains into two may be crucial for galactose specificity. The mannose-specific members of the family are single-chain proteins that lack the above modification. Although the galactose-specific and the mannose-specific jacalin-type lectins differ in their sequences, they share a common fold: the β-prism I fold, which is characteristic of Moraceae plant lectins. Here, two crystal structures of a jacalin-related lectin from *Artocarpus hirsuta*, which is specific for galactose, in complex with methyl-α-D-galactose are reported. The lectin crystallized in two orthorhombic forms and one hexagonal form under similar conditions. The crystals had an unusually high solvent content. The structure was solved using the molecular-replacement method using the jacalin structure as a search model. The two orthorhombic forms were refined using data to 2.5 and 3.0 Å resolution, respectively. The structures of the *A. hirsuta* lectin and jacalin are identical. In orthorhombic form I the crystal packing provides three different micro-environments for sugar binding in the same crystal. The observed difference in the specificity for oligosaccharides between the *A. hirsuta* lectin and jacalin could only be explained based on differences in the molecular associations in the packing and variation of the C-terminal length of the β-chain. The observed insecticidal activity of *A. hirsuta* lectin may arise from its similar fold to domain II of the unrelated δ-endotoxin from *Bacillus thuringiensis*.

1. Introduction

Molecular recognition, interaction and binding play central roles in biological activity. Protein–carbohydrate recognition in particular triggers many biological processes, mediates cell–cell recognition, which is vital for the initiation and control of many cellular functions, and causes host–pathogen interactions through the specific recognition of carbohydrates present on the cell surface. Lectins, the proteins involved in carbohydrate recognition, are defined as

> proteins possessing at least one non-catalytic domain which binds reversibly to a specific mono- or oligosaccharide

(Peumans & Van Damme, 1999). The crystal structures of lectin–carbohydrate complexes have turned out to be excellent model systems for the study of protein–carbohydrate interactions.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the last resolution shell.

	Form I	Form II
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)		
a	92.4	89.9
b	98.7	121.9
c	164.9	131.6
Z	4	4
Resolution range (Å)	15.0–2.5 (2.6–2.5)	20.0–3.0 (3.5–3.0)
No. observations	169292	131164
Completeness (%)	96.2 (98.2)	98.9 (99.5)
R_{sym} (%)	7.2 (26.7)	9.6 (20.2)
Average $I/\sigma(I)$	11.5 (3.1)	19.2 (7.2)
Matthews coefficient (V_M; Å^3 Da^{-1})	6.5	6.0
Solvent content (%)	80.0	79.0
No. unique reflections	46666	26013
No. reflections for R_{free}	2358	1316
R value (%)	19.7	19.3
R_{free} value (%)	23.5	23.7
No. protein atoms	4568	4572
No. sugar atoms	52	52
No. water molecules	110	67
Average *B* factors (Å^2)		
Protein	44.3	33.5
α-Chain	44.4	33.4
β-Chain	43.2	35.2
Sugar atoms	44.3	30.7
Water molecules	50.4	45.4
R.m.s. deviations		
Bond lengths (Å)	0.009	0.008
Bond angles in distances (Å)	2.2	2.2
Residues in allowed regions (%)	85.4	84.2
Residues in additionally allowed regions (%)	14.6	15.8
Overall average *G* factor	−0.12	−0.13

Over 230 structures of lectins and their complexes with sugars are listed in the 3D Lectine Database (http://www.cermav.cnrs.fr/lectines/). More than half of the structures are from plants; the rest are from animals, bacteria and viruses. Although a variety of folds have been detected in lectin structures, they can be classified into a finite number of unique folds (around 17). Of these, the legume lectin fold and the C-type lectin fold are the most frequently found (Bouckaert *et al.*, 1999; Drickamer, 1999; Lis & Sharon, 1998; Rini, 1999; Vijayan & Chandra, 1999). Depending on their size, the carbohydrate-recognition domains (CRDs) are classified into two types (Vijayan & Chandra, 1999): the 120–250-residue CRD of legume lectins, jacalins, galectins, C-type lectins, influenza virus hemagglutinin *etc*. and the 40–50-residue CRD of bulb and cereal lectins. Presently, based on their sequences, the plant lectins are classified into seven distinct families: legume lectins, monocot mannose-binding lectins, jacalins, chitin-binding lectins, type 2 RIPs (ribosome-inactivating proteins), amaranthins and cucurbitaceae phloem lectins (Peumans & Van Damme, 1999).

The jacalin family consists of lectins from Moraceae (jacalin, *Maclura pomifera* agglutinin, *Artocarpus hirsuta* lectin, artocarpin), Convulvulaceae (calsepa, conarva), Asteraceae (heltuba), Gramineae (barley and wheat lectins) and Musaceae (banana lectin) (Bourne *et al.*, 1999). They occur in the seeds and vegetative tissues of plants. Irrespective of their low sequence similarity and varying carbohydrate specificity, they share similar structural characteristics. Jacalin, the lectin from *A. integrifolia*, was the first of this family to be sequenced, crystallized and to have its three-dimensional structure determined in complex with methyl-α-D-galactopyranoside (Sankaranarayanan *et al.*, 1996). Subsequently, many further examples were crystallized and three-dimensional structures of at least four of them have been determined. They include *M. pomifera* agglutinin (MPA) in complex with T-antigen disaccharide (Lee *et al.*, 1998), *Helianthus tuberosus* lectin (heltuba) in the native form and complexed with Man-α1-3Man as well as Man-α1-2Man (Bourne *et al.*, 1999), and artocarpin from jack fruit complexed with methyl-α-D-mannose (Pratap *et al.*, 2002). A predicted structure of KM+ lectin is also available (Rosa *et al.*, 1999).

The galactose-specific *A. hirsuta* lectin has been isolated, purified and its sugar binding characterized (Gurjar *et al.*, 1998). The crystallization and preliminary X-ray studies of this lectin have been reported previously (Rao *et al.*, 1999). The detailed binding affinity towards galactose and derivatives has been studied using fluorescence spectroscopy (Gaikwad *et al.*, 1998). The role of this lectin in plant defence has been demonstrated by its insecticidal activity towards the larvae of the red flour beetle (*Tribolium castaneum*; Gurjar *et al.*, 2000). The purified protein has been subjected to both chemical and thermal denaturation studies (Gaikwad *et al.*, 2002). Here, we report the detailed three-dimensional structure of this lectin and its interaction with the sugar methyl-α-D-galactose.

2. Experimental

2.1. Data collection

Details of the purification, crystallization, X-ray diffraction data collection and data processing of the lectin have been reported previously (Gurjar *et al.*, 1998; Rao *et al.*, 1999). However, for crystal form I a new data set collected to 2.5 Å was used. This new data set was collected using an R-AXIS IV^{++} image-plate detector mounted on a Rigaku rotating-anode generator operating at 100 kV and 50 mA. The data were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

2.2. Molecular replacement and refinement

The structures were solved using the molecular-replacement (MR) method implemented in the *AMoRe* program (Navaza & Saludjian, 1997). A jacalin dimer (chains *EF* and *GH*; PDB code 1jac; Sankaranarayanan *et al.*, 1996) was used as the model. The program *REFMAC* (Murshudov *et al.*, 1997) was used for refinement and *QUANTA* (Molecular Simulations Inc.) was used for the display and model fitting of the structures. Programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) were used for crystallographic and other calculations. The *NACCESS* program (Hubbard & Thornton, 1993) was used for calculation of atomic accessible surface, defined by rolling a probe of the given size around a van der Waals surface.

Data in the resolution range 15.0–2.5 Å were used in the calculations for orthorhombic form I and data in the resolution range 20.0–3.0 Å were used for orthorhombic form II. During refinement, 5% of the data was kept aside for calculation of the free R factor. The initial phases obtained from molecular replacement were improved by subsequent rigid-body refinement. By retaining the jacalin sequence for the model, better R and correlation factors could be achieved than by use of a polyalanine sequence.

Rigid-body refinement was followed by several cycles of positional refinement using data in the resolution range 15–2.5 Å for form I and 20–3.0 Å for form II, resulting in the improvement in both the R factor and $R_{\rm free}$. The individual B factors were only refined in the final cycles. After each run of the refinement program *REFMAC*, electron-density maps were calculated and the model was fitted onto a $2F_{\rm o} - F_{\rm c}$ map using the AUTOFIT module of *QUANTA*. The correct sequence and the side-chain conformations were confirmed using OMIT maps. The non-crystallographic symmetry (NCS) present in the structures was used as a restraint in the initial stages of refinement. Similarly, an overall temperature factor was refined in the initial cycles. The bulk-solvent correction was used in *REFMAC*. The X-SOLVATE module of *QUANTA* was used interactively to add solvent molecules. The acceptance of solvent molecules was decided carefully by considering the interactions and short contacts. In regions where the map was not well defined or the residue did not fit the observed density, fresh maps were calculated by omitting the residues and correct residues were chosen based on the calculated difference density. After several cycles of refinement interspersed with model fitting using *QUANTA*, the values of the correlation coefficient CC and the crystallographic R factor improved considerably in both the crystal forms, with no further changes in the final cycles. In the final map no electron density was present for the side-chain atoms of residues Lys21 (α-chain) and Lys18 (β-chain) in all subunits, whilst density was poorly defined for Lys2, Gln42 (both in the α-chain) and Lys6 (β-chain). The program *PROCHECK* (Laskowski *et al.*, 1993) was used to assess the geometry of the refined model. The overall scores and those pertaining to individual parameters were well within accepted limits for both structures. The quality and reliability of the final model satisfies the conditions discussed in validation tests (Kleywegt & Jones, 2002). The data-collection parameters and refinement statistics are given in Table 1.

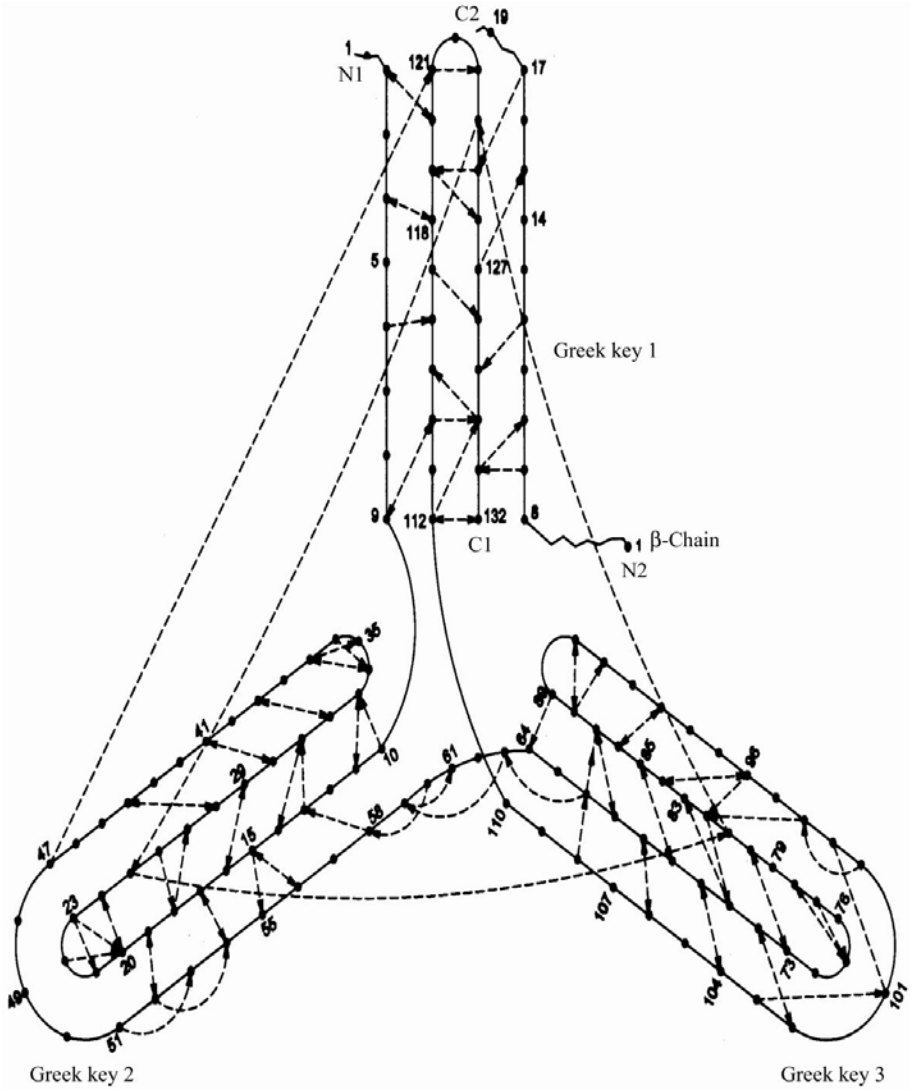

Figure 1
Schematic diagram showing the hydrogen bonds between main-chain atoms of three β-sheets and the organization of the three Greek-key motifs in the subunit of *A. hirsuta* lectin. N1 and N2 and C1 and C2 are the N- and C-termini of the α-chain and β-chain, respectively. The numbering in the diagram corresponds to the residue numbering in the crystal structure.

3. Results and discussion

3.1. Secondary structure and main-chain fold

Two orthorhombic crystal forms that had similar crystal morphologies but distinguishable unit-cell parameters were grown under similar conditions in the presence of methyl-α-D-galactose. A hexagonal form could also be grown under identical conditions. A second hexagonal form that grew in the absence of sugar but with 10% glycerol present in the well solution was of inferior diffraction quality (Rao *et al.*, 1999). The structures were solved in two orthorhombic and one hexagonal form using the molecular-replacement method with jacalin as a search model.

Table 2
Root-mean-square differences (Å) between positions of all atoms when pairs of subunits are exposed.

Form I, orthorhombic form I of *A. hirsuta* lectin. Form II, orthorhombic form II of *A. hirsuta* lectin. I, II, III and IV are the subunits in the tetramer of *A. hirsuta* lectin. Jacalin, PDB code 1jac.

Subunit	II	III	IV	Jacalin
Form I				0.811
I	0.587	0.537	0.560	—
II	—	0.656	0.580	—
III	—	—	0.559	—
Form II				0.742
I	0.283	0.279	0.313	—
II	—	0.328	0.322	—
III	—	—	0.247	—

The structures refined in the two orthorhombic forms are reported here.

The secondary structure of *A. hirsuta* lectin consists of three sets of four-stranded antiparallel β-sheets (Figs. 1 and 2). Almost 70% of the protein residues and a majority of the main-chain hydrogen bonds are in these β-sheets (Fig. 1). The overall fold of *A. hirsuta* lectin is the 'β-prism I fold', which was first identified in the lectins in the structure of jacalin from *A. integrifolia* and is hence termed the 'jacalin fold' (Sankaranarayanan *et al.*, 1996). This fold has three β-sheets arranged to form the faces of a prism. There are three Greek-key motifs in the fold, as in jacalin; the first (GK1) contains 39 residues (Gly1–Phe9 and Leu112–Leu133 of the α-chain and Val10–Ala17 of the β-chain), the second (GK2) contains 52 residues (Thr10–Pro61 of the α-chain) and the third (GK3) contains 46 residues (Glu63–Ile108 of the α-chain) (Fig. 1). Subunit structures viewed from the side and the top and the position of the pseudo-threefold axis are shown in Fig. 2. As in jacalin, no sequence similarity exists between the three β-sheet subdomains. Here, the pseudo-threefold axis is parallel to the prism axis, whereas it is perpendicular to the β-sheets in bulb lectins and at 56° to the barrel axis in type 2 RIPs and amaranthins (Wright, 1997).

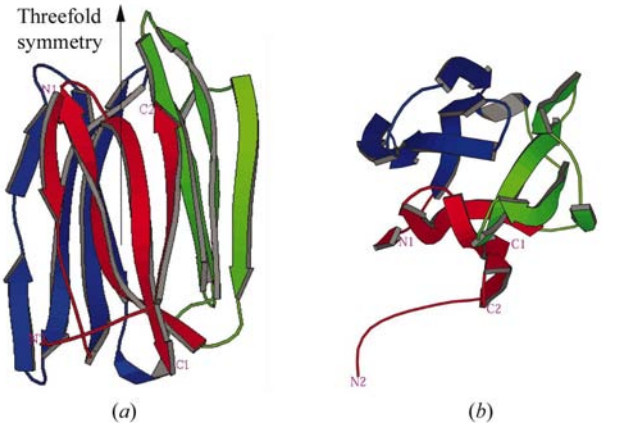

Figure 2
Side and top views of *A. hirsuta* lectin subunit. (*a*) The direction of the pseudo-threefold symmetry axis is parallel to the plane of the paper. (*b*) View down the pseudo-threefold axis.

The amino-acid sequence of *A. hirsuta* lectin was deduced from the electron-density map of the refined structure. A multiple sequence alignment of jacalin-family lectins (Fig. 3) was also used in the selection of correct residues for the observed densities. The residues that differ in the final structure are K45S, T49S, M66T, L124M in the α-chain and S14P in the β-chain. The proline residues 61 and 91 of the α-chain and 14 of the β-chain are in the *cis* conformation. If the lengths of the lectin genes of *A. hirsuta* and *A. integrifolia* are assumed to be the same, then the electron density corresponding to the three N-terminal and two C-terminal residues of the β-chain is absent from the map. No electron density beyond the 18th residue of the β-chain was present in form I, while only Ala could be fitted in the poorly defined electron density at the position of the 19th residue in form II. All 133 residues of the α-chain were traced in the electron-density map. Although *A. hirsuta* lectin is identified as a glycoprotein containing 6.5% neutral sugar (Gurjar *et al.*, 1998), no density for any covalently linked carbohydrates was detected. The closeness of *A. hirsuta* lectin to jacalin is demonstrated by the very low root-mean-square (r.m.s.) deviations between their three-dimensional structures (Table 2).

Of the two orthorhombic forms of *A. hirsuta* lectin, the unit-cell parameters of form I match the unit-cell parameters (a = 92.5, b = 98.7, c = 164.5 Å) reported for an orthorhombic form of jacalin with the same space group. Unfortunately, the three-dimensional structure of this form is not available for further comparison. The unit-cell parameters and packing of recently reported orthorhombic forms of jacalin (Jeyaprakash *et al.*, 2003), however, differ from the structures reported here. The asymmetric unit is a tetramer in both crystal forms of *A. hirsuta* lectin. The packing of the lectin tetramer differs between the two crystals.

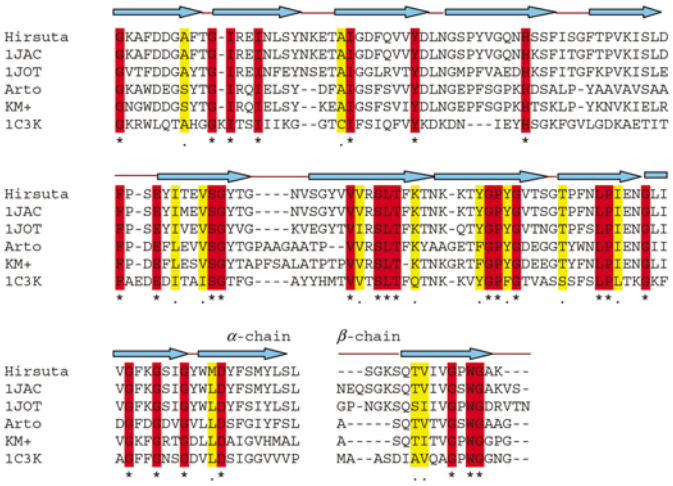

Figure 3
Multiple sequence alignment of *A. hirsuta* lectin with other jacalin-related lectins using *CLUSTALW* (Higgins & Gibson, 1995). The sequences are Hirsuta, *A. hirsuta* lectin; 1JAC, jacalin; 1JOT, MPA; Arto, artocarpin; KM+, KM+ lectin; 1C3K, heltuba lectin. The symbols * and . represent conserved residues and substitution by a similar type of amino-acid residue, respectively. The residues in the β-sheet structure are indicated at the top of the alignment.

3.2. Quaternary association

The gel-filtration experiments indicated that the *A. hirsuta* lectin, like other members of the jacalin family, is a tetramer in solution (Gurjar *et al.*, 1998). The difference between the asymmetric unit of *A. hirsuta* lectin and that of jacalin is that the functional tetramer is the asymmetric unit in the former but not in the latter (Fig. 4). In jacalin, the functional tetramer is generated from two symmetry-related dimers. It may be noted in the case of *A. hirsuta* lectin that using two half-tetramers leads to the generation of an association similar to that found in the asymmetric unit of jacalin. The *A. hirsuta* lectin tetramer had overall dimensions of 20 × 60 × 80 Å and the subunit arrangement showed approximate 222 symmetry. The four carbohydrate-binding sites in the tetramer contributed by each monomer occupy approximately tetrahedral positions with respect to the common centre of the tetramer assembly. The approximate distances of the sugars from the central point are 36 Å and angles made by lines joining the central point to the midpoint of each sugar have values of 81, 116 and 136° instead of the ideal tetrahedral angle of 109.5°. It is interesting to note that the tetrahedral positioning of the sugar-binding sites keeps them equidistant from each other. This disposition of binding sites helps the lectin in its biological role as an agglutinant by facilitating multiple binding of cell-surface receptors by the tetramer (Weis *et al.*, 1991). To assess the stability of this association, the surface area of the subunits that was buried in the tetramer was computed (Hubbard & Thornton, 1993). Each subunit lost an area of around 1760 Å^2, of which a contribution of 1050 Å^2 was from non-polar residues. The total buried surface area of a tetramer is 7065 Å^2, of which 60% is contributed by hydrophobic residues. The buried surface at the interface of subunits *AB* (*A* and *B* chains, subunit I) and *CD* (subunit II) [same as *EF* (subunit III) and *GH* (subunit IV)] is greater than that buried at the interface of *AB* (subunit *I*) and *EF* (subunit III) [same as *CD* (subunit II) and *GH* (subunit IV)], implying a greater stability of dimers of subunits I and II or III and IV compared with the whole tetramer. In this context, it may be of interest to note that the magnitudes of the pseudo-tetrahedral angles, 81° between subunits I and II (same as between III and IV), 116° between I and III (or II and IV) and 136° between I and IV (or II and III), correlate with the degree of association between the individual subunits.

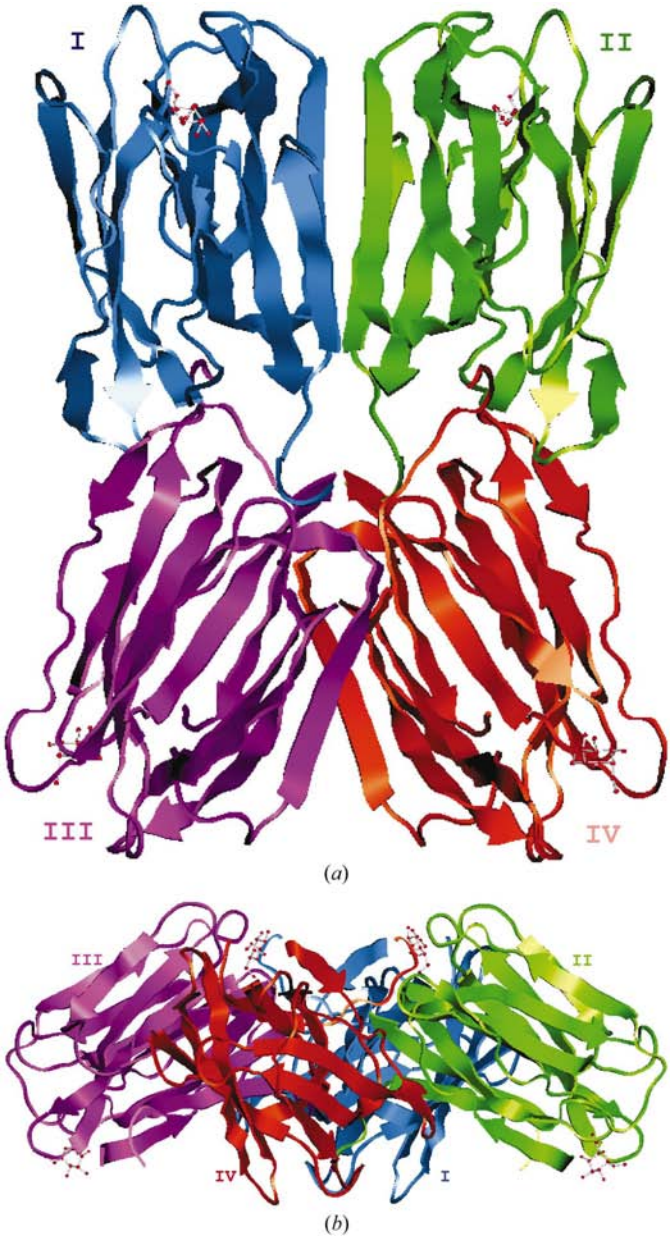

Figure 4
(*a*) The quaternary structure of *A. hirsuta* lectin, showing tetramer association. (*b*) The view of the tetramer that shows the tetrahedral positions of the sugar-binding sites with respect to the centre of the tetramer.

The short β-chain, an integral part of the tertiary structure, plays a key role in subunit association and stabilization of the tetramer. The core of the interaction between subunits I and II (III and IV) is accounted for by the formation of an antiparallel β-sheet involving residues 107–110 of the α-chain and residues 13–10 of the β-chain. A salt bridge between the side chains of Glu109 and Lys117 in the two subunits reinforces this. Towards the edge of the tetramer, in the interface of subunits I and II (III and IV), the main-chain N and C═O atoms of residue Asn105 from the two adjacent subunits interact *via* a water molecule. Their carbonyl groups interact with the side chain of Trp15 (β-chain). There is a network of water molecules hydrogen bonding between these asparagines and threonines 72 and 102 and Pro105. In this interface, towards the centre of the tetramer, the N atom of residue 10 (β-chain) and the C═O atoms of residue 131 in subunits I and II mutually interact through a network of three water molecules. Not all the water molecules in this network are traceable in orthorhombic form II, presumably owing to its low resolution. The association of subunits I and III (same as II and IV) is mainly stabilized by the interpenetrating N-termini of the β-chains. The main-chain N atoms of residues 5, 7 and 8 in the β-chain interact with the C═O groups of residues 61, 10 and 133, respectively. The side chain of Asn35 interacts with the N and C═O atoms of residue 8 in the α-chain and through

three water molecules (absent in orthorhombic form II) with the N atom of residue 9 in the β-chain.

3.3. Sequence comparison of jacalin-family lectins

The amino-acid sequence of *A. hirsuta* lectin, determined solely from the electron-density map, was aligned against the sequences of related lectins such as jacalin, MPA, heltuba, artocarpin and KM+ (Fig. 3). The proportion of glycine in these proteins is slightly greater than is commonly observed and its distribution is random throughout the sequence. The sequence similarity of these lectins varied from 12 to 96%, with the highest similarity between *A. hirsuta* lectin and jacalin, and the lowest between artocarpin and heltuba. The sequence similarity (12–68%) between lectins of the mannose-specific group such as heltuba, artocarpin and KM+ is low compared with the sequence similarity (75–96%) between lectins of the galactose-specific group, such as *A. hirsuta* lectin, jacalin and MPA. A total of 28 residues are invariant in the sequences of these lectins (which may be a requisite for the conservation of the 'β-prism fold'). Similarly, the high content of glycine and hydrophobic residues may help in the packing of the three Greek-key motifs. The aromatic residues Tyr78, Tyr122 and Trp123 are preserved in the galactose-binding site (Fig. 5), but no corresponding aromatic residues are found in the mannose-binding site. The sequence alignment confirms the non-conservation of these aromatic residues. The sequence and structure similarity hints at the apparent evolution of jacalin-family lectins from a common ancestor.

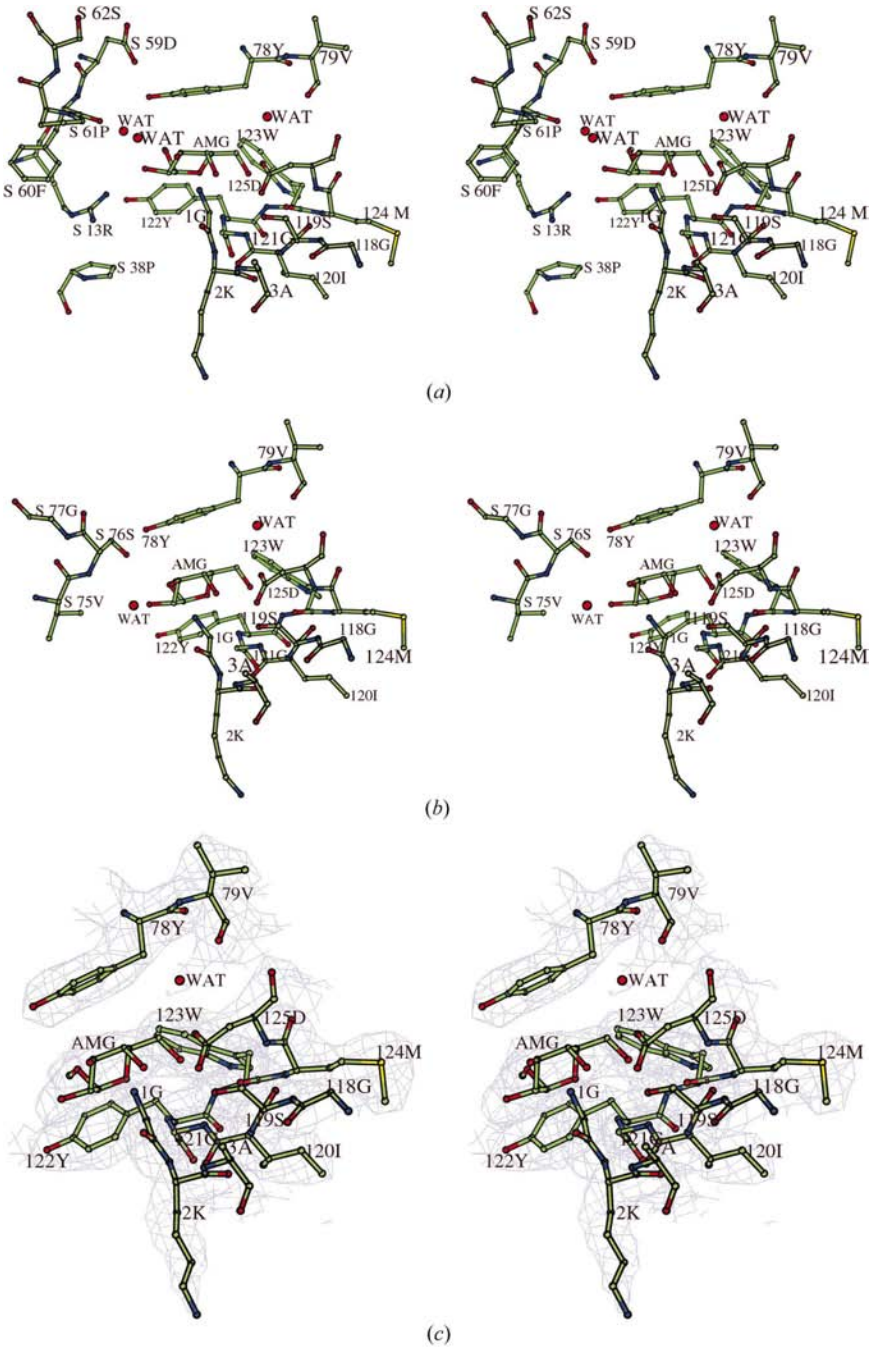

Figure 5
A stereoview of the carbohydrate-binding site of *A. hirsuta* lectin is shown along with the bound methyl-α-galactose (Gal), the amino acids and the solvent molecules around the recognition site. The three different microscopic environments found in the crystals of orthorhombic form I are shown separately. (*a*) The binding site as found in subunits III and IV of the tetramer. (*b*) The binding site in subunit I. (*c*) The sugar binding in subunit II, which is the same as the sugar binding in all subunits of orthorhombic form II. The ($2F_o - F_c$) map is contoured at 1σ. This figure was prepared using *BOBSCRIPT* (Esnouf, 1999).

3.4. Carbohydrate binding and interactions

One of the main goals of the structural studies is the elucidation of a basis for carbohydrate specificity at the atomic level. The difference Fourier map ($F_o - F_c$) clearly showed a methyl-α-D-galactose molecule bound to every subunit of the lectin tetramer. The carbohydrate-binding site is formed by the loops that connect the inner strands of Greek keys 1 and 3. A network of hydrogen bonds, mainly involving the main-chain atoms of the α-chain (Table 3) and van der Waals interactions with a cluster of aromatic amino acids, stabilize the sugar binding. Residues that take part directly in the binding of the sugar molecule are Tyr78 and Val80 of Greek key GK3, and residues Gly1, Gly121, Tyr122, Trp123 and Asp125 of GK1. The atoms of these residues, which belong to the α-chain, are located within 4.0 Å of sugar atoms (Fig. 5). These residues are conserved and also participate in sugar binding in both jacalin and MPA structures. Although the galactose molecule may be considered to be polar, the C atoms at the five epimeric

Table 3
The polar interactions involving sugar and protein atoms at the carbohydrate-binding sites of subunit I of the two crystal forms.

Distances are in Å. Sym, symmetry related.

Sugar atom	Protein atom	Residue No.	Form I	Form II
O1	OH	Tyr78	3.5	3.3
O2 (subunit I)†	OG (sym II)	Ser76	3.3	—
O2 (subunit III)†	NH1 (sym II)	Arg13	2.8	—
O2 (subunit IV)†	NH1 (sym I)	Arg13	2.9	—
O3	N	Gly1	2.9	2.9
O4	OD1	Asp125	2.9	2.6
	OD2	Asp125	3.3	3.5
	N	Gly1	3.2	3.3
O5	N	Tyr 22	2.9	2.9
O6	OD1	Asp125	2.8	3.0
	N	Trp123	3.0	3.1
	O	Trp123	3.1	3.0
	N	Tyr122	3.1	3.0
	N	Gly121	3.9	3.8

† Found only in form I.

positions of the ring and at the adjacent exocyclic position form a contiguous hydrophobic region. This region is found packed against aromatic residues. The shielding of this hydrophobic surface from solvent contact by the sugar also adds to the stability of binding. The aromatic side chains of Tyr78, Tyr122 and Trp123 are within van der Waals distance of the C1, C3, C4 and C5 atoms of the sugar. This is in conformity with the established fact that the galactose binding always involves a stacking interaction with aromatic residues against the B face of the sugar (Pratap *et al.*, 2002). This stacking interaction is mediated by an extended patch of positively charged H atoms on the B face of the sugar and the π-electron cloud of Tyr78 that stacks against the B face. Six of the nine hydrogen bonds formed between the sugar and the lectin involve main-chain atoms. Atoms O3 and O4 of the sugar form hydrogen bonds with the N-terminal N atom of the α-chain, while O4 additionally interacts with the side-chain carboxyl O atoms of Asp125; O5 and O6 interact with the N atom of Tyr122, with O6 also interacting with the carboxyl group of Asp125 and the N and O atoms of Trp123 (Table 3). The methyl group attached to O1 is sandwiched between the phenyl rings of tyrosine residues 78 and 122. The side-chain carboxyl group of Asp125 is held in position by hydrogen bonds to the hydroxyl group of Ser119 and a water molecule that in turn is hydrogen bonded to carbonyl O of Thr79 (Fig. 5). In orthorhombic form II the sugar atom O2 is involved in no other interactions with the lectin molecule. In form I, the sugar atom O2 hydrogen bonds with symmetry-related lectin molecules [with Ser76 OG from the symmetry-related subunit II in the case of subunit I (Fig. 5*b*) and with Arg13 of the symmetry-related subunits II and I in the case of subunits III and IV (Fig. 5*a*)]. The higher affinity of *A. hirsuta* lectin towards methyl-α-galactose compared with α-galactose (Gaikwad *et al.*, 1998) could be because of the stacking of the methyl group against the phenyl rings of Tyr78 and Tyr122 (Fig. 5).

If mannose with O4 in the equatorial configuration is bound in the place of galactose, its interaction with Asp125 will be weakened and that with Gly1 will be lost. Thus, the creation of a free N-terminus at Gly1 (α-chain) by post-translational modification is postulated to be the reason for the galactose-specificity of some jacalin lectins (Sankaranarayanan *et al.*, 1996). Also, modelling studies have shown that the carbohydrate-binding site of jacalin can accommodate but not favourably bind a mannose molecule (Pratap *et al.*, 2002). As predicted above, in a recently reported mannose-bound jacalin structure the interaction of sugar atom O4 with Gly1 was totally absent and that with Asp125 was partially lost (Bourne *et al.*, 2002).

The comparison of structures showed that *A. hirsuta* lectin and jacalin grossly resemble each other in terms of the structure and topology of the sugar-binding site (Table 2). The binding studies, however, showed differences in the binding of disaccharides and oligosaccharides (Gurjar *et al.*, 1998). Similarly, unlike jacalin, the *A. hirsuta* lectin was shown to not be specific for T-antigen. The molecular packing differs in the crystal structures of the two lectins, resulting in the functional tetramer becoming the asymmetric unit in *A. hirsuta* lectin, whereas the same tetramer is generated from two symmetry-related dimers in the jacalin crystal (Fig. 4). One main concern during the structural investigation was to probe into the reason for the differences in sugar specificity. Surprisingly, we could detect little difference between the structures of the two lectins. Except for the packing differences, especially in the crystals of orthorhombic form I in which the residues Arg13 or Ser76 from symmetry-related neighbours also participated in sugar binding (Figs. 5*a* and 5*b*), no other significant differences were detected. Although in terms of additional interactions this provided only one extra hydrogen bond between the sugar and the lectin, the symmetry-related lectin molecule might shield the secondary sugar-binding sites of the lectin and thus prevent disaccharides and oligosaccharides from binding. This association, if stable in solution, could effectively block disaccharide or oligosaccharide binding at these sites. To assess the role of steric effects, we have modelled the molecules of T-antigen (Galβ1,3GalNAcα) and mellibiose (Galα1-6Glc) in the binding site of *A. hirsuta* lectin based on jacalin complexes (PDB codes 1m26 and 1guy). The secondary site A of the lectin is partially blocked by the symmetry-related molecule in crystal form I which can affect the binding of sugars such as mellibiose, whereas the disordered extended C-terminus of the β-chain is capable of obstructing the binding at secondary site B, which can affect the binding of T-antigen. It could not be ascertained whether the electron density for the end residues of the β-chain were not observed because of disorder or whether the residues had been removed as a result of post-translational processing. We are further investigating this at the sequence level. No mutation capable of directly influencing the sugar binding could be detected in the electron-density map. Thus, in the absence of any structural differences between *A. hirsuta* lectin and jacalin, any observed aberration between their sugar specificities can only be assumed to arise from the different molecular association of these lectins and from the influence of the C-terminal end of the β-chain originating from post-

Table 4
Direct contacts (less than 4 Å) between neighbouring or symmetry-related molecules involved in crystal packing.

The corresponding subunits of the tetramers involved in interactions are shown in brackets. Amg, methyl-α-galactose.

Residue of reference functional tetramer	Atom	Residue of symmetry-related functional tetramer	Atom	Contact distance (Å)
A. hirsuta lectin, orthorhombic form I				
Packing along the crystallographic *a* direction				
Phe47 (I)	O	Asn20 (II)	ND2	3.05
Ser76 (I)	O	Tyr122(II)	OH	2.78
Tyr78 (I)	OH	Ser76 (II)	OG	2.80
Tyr122 (I)	OH	Ser49 (II)	O	2.87
Amg200 (I)	O2	Ser76 (II)	OG	3.26
Amg200 (I)	O3	Ser76 (II)	OG	3.54
Packing along the crystallographic *b* direction (same for II and III)				
Arg13 (I)	NH1	Amg200 (IV)	O2	2.93
Asp59 (I)	OD1	Tyr78 (IV)	OH	2.99
A. hirsuta lectin, orthorhombic form II				
Packing along the crystallographic *a* direction				
Asn43 (II)	OD1	Ser76 (IV)	OG	3.66
		Ser76 (IV)	O	3.25
Ser45 (II)	OG	Ser76 (IV)	OG	3.22
Ser49 (II)	N	Glu22 (IV)	OE1	3.44
Packing along the crystallographic *c* direction (same for II and III)				
Glu22 (I)	OE1	Ser100 (IV)	OG	3.59
Glu22 (I)	OE2	Ser100 (IV)	OG	3.52
Ser76 (I)	N	Tyr71 (IV)	OH	2.90
Jacalin (PDB code 1jac)				
Packing along the crystallographic *b* direction (same for II and III)				
Glu22 (I)	OE1	Ser100 (IV)	OG	2.29
Ser76 (I)	N	Tyr71 (IV)	OH	2.75
Packing along the crystallographic *c* direction				
Gly101 (II)	O	Asn74 (IV)	OD1	2.83
Thr102 (II)	N	Asn74 (IV)	OD1	3.02
Amg200 (II)	O1	Ser100 (III)	OG	3.69
Amg200 (II)	O2	Ser100 (III)	OG	3.59
		Thr99 (III)	O	3.87
Amg200 (II)	O3	Thr99 (III)	O	2.94
		Thr99 (III)	OG1	3.30

translational processing as evidenced from simple modelling studies.

3.5. The high solvent content and molecular packing

The crystals of *A. hirsuta* lectin contain an unusually high solvent content of greater than 70%, which is rarely observed in other crystals and is outside the range compiled by Matthews (1968). There are only a few other reports of protein crystals with such high solvent contents (Kim *et al.*, 1999; Izard & Ellis, 2000). As is frequently the case, the majority of water molecules were disordered and only a few could be located in an $(F_o - F_c) > 5\sigma$ map. This resulted in the input of only structurally significant or invariant water molecules in the present structures. Water molecules are considered to be invariant if they interact with common protein atoms and their distances remain less than 1.8 Å in superposed subunits (Sadasivan *et al.*, 1998).

The water molecules may be classified as those trapped inside the subunits, those present at the interface of the subunits and those on the surface of the tetramer. Only two water molecules were found to be structurally important and to be conserved in all subunits. One of them is the water molecule found close to the sugar-binding site that interacts with the side chain of Asp125 and C=O of Val79 and places the only sugar-binding side chain (Asp125) in a favourable position. The other one is the buried water molecule within the subunits which interacts with the side chain of Asp6, the N atom of Val30 and the C=O of Gln42, all of which are from the α-chain. On the face opposite to the sugar-binding site of the β-prism fold, a network of five water molecules is observed that are firmly bound to the main-chain atoms of residues 11–12, 63–65 and 113 of the α-chain. This is also the opening to the hydrophobic core of the lectin molecule. It is debatable whether this pocket corresponds to a lost sugar-binding site in this family of lectins.

To compare the packing of the molecules of *A. hirsuta* lectin in crystals of orthorhombic form I and form II with the crystals of the jacalin structure referred to here, the contacts between residues in symmetry-related molecules in each case are listed in Table 4. It can be seen from the contacts that the only comparable arrangement is between the arrangement along the *c*-axis direction in form II of *A. hirsuta* lectin and the arrangement along the *b*-axis direction in jacalin (PDB code 1jac). In both the cases the unit-cell parameter in the direction of propagation of this arrangement is close to 130 Å. In the case of *A. hirsuta* lectin crystallization, experiments more often yielded the crystals of orthorhombic form I than form II and crystallization only occurred when methyl-α-galactose was present in the protein solution. In both crystal structures of *A. hirsuta* lectin the inter-tetramer interactions are very few, resulting in loose crystal packing, high solvent content and poor diffraction-quality crystals.

3.6. Comparison with homologous proteins from the DALI database

Comparison of the three-dimensional structure of *A. hirsuta* lectin with protein structures from the DALI database (Holm & Sander, 1995) has revealed homology with 12 of them. Four of these proteins (jacalin, MPA, heltuba and artocarpin) have already been discussed. Two insecticidal toxins from *Bacillus thuringiensis* (Li *et al.*, 1991; Morse *et al.*, 2001) and the vitelline membrane outer layer protein I (VMO-I; Shimuzu *et al.*, 1994) from hen egg belonging to the β-prism family showed good overlap with the α-chain of *A. hirsuta* lectin. The β-prism I fold was first identified in the crystal structure of domain II of δ-endotoxin (CryIIIA) from *B. thuringiensis* (Li *et al.*, 1991). The overlap with the remaining five proteins from the DALI database is confined to a small domain region of around 40 residues. It is interesting to note that this 40-residue basic unit constructs the three subdomains in all these proteins. It is possible that nature designed this domain, the 40-residue carbohydrate-binding domain discussed in §1, for diverse functions. Additionally, the presence of domain II of *B. thuringiensis* δ-endotoxin in *A. hirsuta* lectin may be the reason for its insecticidal activity (Gurjar *et al.*, 2000)

The overlap of *A. hirsuta* lectin with other proteins is poorer than its overlap with the members of jacalin lectin family, as

inferred from the r.m.s. deviation in C^{α} positions. Thus, despite the marked differences in their carbohydrate specificity, the Moraceae plant lectins may be considered to be a structurally conserved family.

4. Conclusions

The crystal structure of the *A. hirsuta* lectin shows an interesting quaternary arrangement of monomer subunits in which the sugar-binding sites of the four subunits constituting the tetramer are placed in an approximately tetrahedral arrangement. The pseudo-tetrahedral angle between I and II (the same as between III and IV) is the lowest and that between I and IV (II and III) is the largest. This pseudo-tetrahedral arrangement may have relevance for the biological activity of agglutination in lectin. The distortion of the tetrahedral angles changed an otherwise 23 tetrahedral point-group symmetry of subunit association to an approximate 222 orthorhombic point-group symmetry. In one of the crystal forms, the crystal packing mediated by sugar molecules provides three different microscopic sugar-binding environments. This type of association, extended to solution, can have a bearing on the difference in the binding of disaccharides and oligosaccharides by *A. hirsuta* lectin compared with jacalin. Simple modelling studies showed that molecular association owing to the crystal packing found in orthorhombic crystal form I can affect binding at the secondary binding site A, whereas the extent of the residues present in the C-terminus of the β-chain can influence binding at the secondary binding site B. These steric effects may explain the restricted binding of disaccharides and oligosacharides to *A. hirsuta* lectin. The pseudo-tetrahedral arrangements of sugar-binding sites and the loose packing of the tetramers reflect the biological role of this molecule as an agglutinin. The β-prism I fold that the *A hirsuta* lectin shares with domain II of δ-endotoxin from *B. thuringiensis* may confer insecticidal activity on the lectin.

The authors thank Department of Biotechnology, New Delhi, India for financial support. KNR and UVK thank CSIR, New Delhi, India for research fellowships.

References

Bouckaert, J., Hamelryck, T., Wyns, L. & Loris, R. (1999). *Curr. Opin. Struct. Biol.* **9**, 572–577.
Bourne, Y., Astoul, C. H., Zamboni, V., Peumans, W. J., Menu-Bouaouiche, L., Van Damme, E. J. M., Barre, A. & Rouge, P. (2002). *Biochem. J.* **364**, 173–180.
Bourne, Y., Zamboni, V., Barre, A., Peumans, W. J., Van Damme, E. J. M. & Rouge, P. (1999). *Structure*, **7**, 1473–1482.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Drickamer, K. (1999). *Curr. Opin. Struct. Biol.* **9**, 585–590.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Gaikwad, S. M., Gurjar, M. M. & Khan, M. I. (1998). *Biochem. Mol. Biol. Int.* **46**, 1–9.
Gaikwad, S. M., Gurjar, M. M. & Khan, M. I. (2002). *Eur. J. Biochem.* **269**, 1–5.
Gurjar, M. M., Gaikwad, S. M., Solakhe, S. G., Mukherjee, S. & Khan, M. I. (2000). *Inv. Repro. Dev.* **38**, 95–98.
Gurjar, M. M., Khan, M. I. & Gaikwad, S. M. (1998). *Biochim. Biophys. Acta*, **1381**, 256–264.
Higgins, J. D. & Gibson, T. J. (1995). *Nucleic Acids Res.* **22**, 4673–4680.
Holm, L. & Sander, C. (1995). *Trends Biol. Sci.* **20**, 478–480.
Hubbard, S. J. & Thornton, J. M. (1993). *NACCESS.* Department of Biochemistry and Molecular Biology, University College of London, England.
Izard, T. & Ellis, J. (2000). *EMBO J.* **19**, 2690–2700.
Jeyaprakash, A. A., Katiyar, S., Swaminathan, C. P., Sekar, K., Surolia, A. & Vijayan, M. (2003). *J. Mol. Biol.* **332**, 217–228.
Kim, K., Yokota, H. & Kim, S.-H. (1999). *Nature (London)*, **400**, 787–792.
Kleywegt, G. J. & Jones, A. T. (2002). *Structure*, **10**, 465–472.
Laskowski, R. A., MacArthur, M. W, Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, X., Thompson, A., Zhang, Z., Ton-that, H., Biesterfeldt, J., Ogata, C., Xu, L., Johnston, R. A. Z. & Young, N. M. (1998). *J. Biol. Chem.* **273**, 6312–6318.
Li, J., Carrol, J. & Ellar, D. J. (1991). *Nature (London)*, **353**, 815–821.
Lis, H. & Sharon, N. (1998). *Chem. Rev.* **98**, 637–674.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Morse, R. J., Yamamoto, T. & Stroud, R. M. (2001). *Structure*, **9**, 409–417.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **276**, 581–594.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Peumans, W. J. & Van Damme, E. J. M. (1999). *Seed Proteins*, edited by P. R. Shewry & R. Casey, pp. 657–683. Dordrecht: Kluwer.
Pratap, J. V., Jeyaprakash, A. A., Rani, P. G., Sekar, K., Surolia, A. & Vijayan, M. (2002). *J. Mol. Biol.* **317**, 237–247.
Rao, K. N., Gurjar, M. M., Gaikwad, S. M., Khan, M. I. & Suresh, C. G. (1999). *Acta Cryst.* D**55**, 1204–1205.
Rini, J. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 578–584.
Rosa, J. C., De Oliverira, P. S. L., Garratt, R., Beltramini, L., Resing, K., Roque-Barreira, M. C. & Greene, L. J. (1999). *Protein Sci.* **8**, 13–24.
Sadasivan, C., Nagendra, H. G. & Vijayan, M. (1998). *Acta Cryst.* D**54**, 1343–1352.
Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. & Vijayan, M. (1996). *Nature Struct. Biol.* **3**, 596–603.
Shimuzu, T., Vassylyev, D. G., Kido, S., Doi, Y. & Morikawa, K. (1994). *EMBO J.* **13**, 1003–1010.
Vijayan, M. & Chandra, N. (1999). *Curr. Opin. Struct. Biol.* **9**, 707–714.
Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. (1991). *Science*, **254**, 1608–1615.
Wright, C. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 631–636.

Two orthorhombic crystal structures of a galactose-specific lectin from *Artocarpus hirsuta* in complex with methyl-α-D-galactose. Addendum

K. N. Rao, C. G. Suresh,* U. V. Katre, S. M. Gaikwad and M. I. Khan

Division of Biochemical Sciences, National Chemical Laboratory, Pune-8, India.
Correspondence e-mail: suresh@ems.ncl.res.in

In the paper by Rao *et al.* [(2004), *Acta Cryst.* D**60**, 1404–1412] the Protein Data Bank reference codes were inadvertently omitted from the paper. The coordinates and structure factors for the structures reported in this paper have been depositied with the PDB and the reference codes are as follows: galactose-specific lectin at 2.5 Å resolution, 1toq, r1toqsf, and at 3.0 Å resolution, 1tp8, r1tp8sf.

References

Rao, K. N., Suresh, C. G., Katre, U. V., Gaikwad, S. M. & Khan, M. I. (2004). *Acta Cryst.* D**60**, 1404–1412.

  DOI: 10.1107/S0907444904019134