Structures and binding specificity of galactose- and mannose-binding lectins from champedak: differences from jackfruit lectins

Mads Gabrielsen,^{a*‡}
Puteri Shafinaz Abdul-Rahman,^b
Shatrah Othman,^b Onn H.
Hashim^b and Richard J. Cogdell^a

^aInstitute of Molecular, Cell and Systems
Biology, University of Glasgow, 120 University
Avenue, Glasgow G12 8TA, Scotland, and

^bDepartment of Molecular Medicine, Faculty of
Medicine, University of Malaya, 50603 Kuala
Lumpur, Malaysia

‡ Current address: Beatson Institute for Cancer
Research, Garscube Estate, Switchback Road,
Glasgow G61 1BD, Scotland.

Correspondence e-mail: mgabr@chem.gla.ac.uk

Received 21 February 2014

Accepted 19 April 2014

PDB references: champedak galactose-binding
lectin, 4ak4; complex with galactose, 4akb;
complex with galactose- β 1–3-galactose-NAc,
4akc; champedak mannose-binding lectin, 4akd

© 2014 International Union of Crystallography
All rights reserved

Galactose-binding and mannose-binding lectins from the champedak fruit, which is native to South-east Asia, exhibit useful potential clinical applications. The specificity of the two lectins for their respective ligands allows the detection of potential cancer biomarkers and monitoring of the glycosylated state of proteins in human serum and/or urine. To fully understand and expand the use of these natural proteins, their complete sequences and crystal structures are presented here, together with details of sugar binding.

1. Introduction

Champedak (*Artocarpus integer*) is a fruit native to South-east Asia and Malaysia. The fruit is rich in lectins that bind sugars such as galactose [champedak galactose-binding (CGB) lectin] and mannose [champedak mannose-binding (CMB) lectin], where the CGB lectin in particular is found at high levels.

The basic structure and properties of the CGB lectin are closely related to those of jacalin. When first discovered, the CGB lectin was shown to bind to the complement C1 inhibitor but not to IgA2, IgD, IgG and IgM from human serum. The CGB lectin was later shown to consume complement and to cause a decrease in the complement-mediated haemolytic activity of sensitized sheep erythrocytes. On further investigation, the CGB lectin was also found to be capable of binding to α 1-antichymotrypsin, α 2-HS glycoprotein, haemopexin, inter- α -trypsin inhibitor heavy chain H4, kininogen and leucine-rich glycoprotein of the human serum. Unlike jacalin, which could be isolated as distinctive batches as well as from distinctive geographical origins, the CGB lectins isolated from seeds of six different clones of champedak were shown to be consistent and uniform in terms of structure and interaction with different isotypes of IgA (Hashim *et al.*, 1991, 1993).

The CGB lectin is transcribed as a propeptide and is post-translationally processed into two chains, α and β , where the α -chain has a molecular weight of 13 000 Da (133 amino acids) and the β -chain has a molecular weight of 2100 Da (consisting of 21 amino acids; Abdul Rahman *et al.*, 2002). Although work by Abdul Rahman *et al.* (2002) showed that the first 47 residues of the α -chain and the complete β -chain of CGB are highly similar to the sequence of jacalin, the entire sequence has not yet been determined.

The CMB lectin is a single polypeptide chain, although its full sequence has not yet been determined. Characterization of the CMB lectin has shown that it is capable of binding strongly to IgE and IgM but only interacts weakly with IgA2 (Lim *et al.*, 1997). The CMB lectin has been used to demonstrate the reduced expression of serum α -2 macroglobulin and complement factor B in patients with nasopharyngeal carcinoma, and may consequently serve as a complementary biomarker for this cancer (Seriramalu *et al.*, 2010). Similarly, the CGB lectin has proven to be a useful tool to detect a number of aberrantly expressed serum and urinary O-glycosylated proteins that are associated with different cancers (Abdul-Rahman *et al.*, 2007; Mohamed *et al.*, 2008). Recently, this lectin has also been used in the development of an assay to detect mucin-type O-glycans (mucins and mucin-type glycoproteins), which have been implicated in many important biological functions and pathological conditions, including

Table 1
Data-processing statistics for CGB lectin soaked with Gal and GalNAc.
Values in parentheses are for the highest resolution bin.

	Gal	GalNAc
Space group	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å)	$a = 76.29, b = 121.03,$ $c = 77.64$	$a = 76.29, b = 121.67,$ $c = 77.56$
Resolution (Å)	65.37–1.95 (2.06–1.95)	64.63–2.30 (2.42–2.30)
Observed/unique reflections	539180/52244	436057/32825
R_{meas} † (%)	11.3 (79.5)	20 (83.3)
$R_{\text{p.i.m.}}$ ‡ (%)	3.7 (30.9)	5.7 (23.2)
$\langle I/\sigma(I) \rangle$	17.7 (2.1)	17.1 (3.2)
Completeness (%)	98.5 (96)	100.0 (100.0)
Multiplicity	10.3 (7.4)	13.2 (13.7)

† R_{meas} is defined as $\sum_{hkl} \{ [N(hkl)/[N(hkl) - 1]]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_i I_i(hkl) \}$.
‡ $R_{\text{p.i.m.}}$ is defined as $\sum_{hkl} \{ [1/[N(hkl) - 1]]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_i I_i(hkl) \}$.

malignancy (Lee *et al.*, 2013). Aberrant glycosylation and over-expression of cell-surface glycoconjugates are among the phenotypic changes that are widely recognized in cancer (Ono & Hakomori, 2004). The tumour cells of cancer patients have been shown to actively secrete or shed a repertoire of mucin-type *O*-glycans into the circulation (Wahrenbrock & Varki, 2006; Storr *et al.*, 2008). Hence, the quantitation of the circulating tumour-associated *O*-glycans and their specific detection will be extremely beneficial for the early diagnosis and treatment of cancer.

In view of the many potential clinical applications of the champedak lectins, full characterization of their sequences and structures is of great interest. In the present study, we describe the X-ray crystal structures of CMB and CGB lectins, as well as present the nucleic acid sequences of the genes encoding these two lectins. We also briefly discuss the differences in sugar specificity between the two lectins.

2. Materials and methods

2.1. Generation of cDNA and sequencing of champedak lectins

Total RNA was extracted from the seeds of a mature champedak fruit using an RNeasy Mini Kit (Qiagen GmbH, Hilden, Germany) in accordance with the manufacturer's protocol. The oligonucleotides used for the PCR amplification of the CGB lectin were designed and synthesized based on the known DNA sequence for jacalin (Yang & Czaplá, 1993), whilst the published sequence of artocarpin (da Silva *et al.*, 2005) was used for the CMB lectin. The sequences of the primers for the respective lectins were as follows: CGB-F, 5'-TCT TCA ATA GTT TAA TGC-3'; CGB-R, 5'-TTA CAT AAT TGG CGA TTT-3'; CMB-F, 5'-ATG GCG AGC CAG ACG ATA AC-3'; CMB-R, 5'-AAA GTG CCA TGT GAA CGC CA-3'. PCR amplification was performed using SuperScript One-Step RT-PCR with the Platinum *Taq* System (Invitrogen, Carlsbad, California, USA) according to the manufacturer's instructions. 200 ng of champedak lectin total RNA was reverse-transcribed and PCR-amplified in a total volume of 50 µl

Figure 1
Top, amino-acid sequence alignment of champedak galactose-binding lectin (CGB) deduced as part of this study aligned with its close homologue jacalin from jackfruit. The cleaved fragment of the propeptide is highlighted in grey. Bottom, amino-acid sequence alignment of champedak mannose-binding lectin (CMB), identified as part of this study, aligned with its close homologue artocarpin from jackfruit. The residues that are non-identical have been highlighted with a black background.

Table 2

Refinement statistics.

	CGB			CMB
	High resolution	With Gal	With GalNAc	
PDB code	4ak4	4akb	4akc	4akd
Space group	$P2_1$	$P2_12_12$	$P2_12_12$	$P2_12_12_1$
Unit-cell parameters ($\text{\AA}$, $^\circ$)	$a = 76.17$, $b = 121.73$, $c = 77.74$, $\beta = 90.61$	$a = 76.29$, $b = 121.03$, $c = 77.64$	$a = 76.29$, $b = 121.67$, $c = 77.56$	$a = 76.89$, $b = 86.22$, $c = 95.37$
Resolution ($\text{\AA}$)	76.25–1.65	65.37–1.95	64.63–2.30	47.69–2.10
No. of protein residues	1191	595	593	603
No. of protein atoms	9339	4609	4594	4537
No. of water molecules	972	267	51	277
Ligands	PEG 600	Gal, PEG 600	GalNAc	Cd^{2+}
R_{work} (%)	16.48	17.13	19.82	19.34
R_{free} (%)	18.98	19.24	24.59	23.57
R.m.s.d., bond lengths ($\text{\AA}$)	0.01	0.01	0.01	0.01
R.m.s.d., angles ($^\circ$)	1.01	1.06	1.15	1.19
Wilson B ($\text{\AA}^2$)	20.0§	25.60	44.70	26.30¶
Average isotropic thermal parameters ($\text{\AA}^2$)				
Main chain	22.43	32.63	41.25	46.38 (36.15)††
Side chains	30.40	39.64	46.44	52.73 (42.28)††
Water molecules	37.38	42.48	40.18	43.57
Ligands	39.04	44.29, 44.78‡‡	66.96	94.14, 70.30‡‡
Ramachandran allowed regions (%)	100	100	100	98.63
MolProbity clash score	3.59 (98th percentile)	4.44 (98th percentile)	7.12 (98th percentile)	7.59 (94th percentile)

† R_{work} is defined as $\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$. ‡ R_{free} is calculated as R_{work} , using 5% of the data excluded from refinement. § Published elsewhere (Gabrielsen *et al.*, 2009). ¶ Published elsewhere (Gabrielsen *et al.*, 2010). †† Average B factors for chains A, B and D. ‡‡ Values represent the averaged B factors of the ligands.

reaction mixture containing a final concentration of $1 \times$ Master Mix (containing 1.2 mM MgSO_4 , 200 μM dNTPs), 2 units of Superscript III/Taq DNA polymerase mixture and 2.0 μM gene-specific primers for the respective lectins.

RT-PCR amplification was performed in a programmable Biometra T-Personal Thermocycler (Biometra, Göttingen, Germany). Reverse transcription (RT step) of RNA to cDNA was carried out at 55°C for 30 min and was immediately followed by PCR amplification. The PCR amplification was performed under the following conditions: initial denaturation at 94°C for 2 min followed by 30 cycles of 94°C for 15 s, 55°C for 30 s and 68°C for 45 s, and then final extension at 68°C for 5 min. The reaction products were resolved by agarose gel electrophoresis and purified using a QIAquick Gel Extraction Kit (Qiagen GmbH, Hilden, Germany). The DNA sequences of the CGB and CMB lectins were determined by automated sequencing and translated to amino acids using the ExPaSy Translate Tool. Both the DNA and the amino-acid sequences were compared with the published lectin sequences of jacalin and artocarpin (GenBank accession Nos. L03796 and AY957581, respectively) using the EMBOSS pairwise alignment algorithm software.

2.2. Crystallization, structure solution and refinement

The proteins were isolated and crystallized using previously published protocols (Hashim *et al.*, 1991, 1993; Gabrielsen *et al.*, 2009, 2010). Data collection and processing have been discussed elsewhere (Gabrielsen *et al.*, 2009, 2010). The resolution was based on the statistics following scaling and in particular based on the poor completeness and merging statistics at higher resolution.

The CGB lectin structure was solved by molecular replacement using a polyalanine model based on jacalin (PDB entry 1ku8; Bourne *et al.*, 2002), as the amino-acid sequence of CGB lectin was unknown at this stage, containing a single α -chain using Phaser (McCoy *et al.*, 2007) and the $P2_1$ data set. As the amino-acid sequence became available, this was used to fit the correct residues into the density. Subsequent solutions were determined using the high-resolution structure of CGB lectin as a search model. A total of eight $\alpha\beta$

assemblies were found in the $P2_1$ data set and four were found in the $P2_12_12$ data sets.

Sugars were soaked into the CGB crystals by creating an artificial mother liquor (1:1 ratio of reservoir solution with 20 mM Tris added) with 20 mM Gal (galactose), 5 mM GalNAc (galactose- β 1–3-galactose-Nac) or 20 mM mannose and placing the crystals in these solutions for over 72 h before cooling them in the cryostream. Data from these sugar-soaked crystals were collected on beamline I04 at the Diamond Light Source. All data sets were in space group $P2_12_12$. All relevant data-collection and processing statistics are presented in Table 1.

As the sequence of CMB was undetermined at this point, the structure of CMB lectin was determined by positioning a polyalanine model based on a single subunit of artocarpin (PDB entry 1j4u; Pratap *et al.*, 2002) into the data set using Phaser. The structure contained four chains. As the amino-acid sequence became available, this was used to fit the correct residues into the density.

The structures were refined using REFMAC5 (Murshudov *et al.*, 2011) and BUSTER (Bricogne *et al.*, 2011) and visually inspected using Coot (Emsley & Cowtan, 2004) with inclusion of solvent molecules, sugar moieties and crystallization-related ligands, where required. Side chains were added at this stage as the sequences became available. The models were geometrically validated by MolProbity (Davis *et al.*, 2007; Chen *et al.*, 2010). The relevant statistics for the refined structures are shown in Table 2. All surface-area calculations and identification of hydrogen bonds at the interfaces were calculated by the Protein Interfaces, Surfaces and Assemblies (PDBEPIA) server (Krissinel & Henrick, 2007). Figures were created using PyMOL (<http://www.pymol.org>), topology diagrams were made with TOPDRAW (Bond, 2003) and alignments were prepared with ALINE (Bond & Schüttelkopf, 2009). Ligand-binding figures were created using PDBSum (Laskowski, 2009).

2.3. Accession numbers

The nucleotide sequences for the CGB and CMB lectins reported here have been deposited in the EMBL Nucleotide Sequence Database under accession Nos. FR728240 and FR728241, respectively. To reconfirm some of the ambiguous bases found in the CGB

Figure 2

Cartoon representation of (a) the $\alpha\beta$ subunit of CGB lectin and (b) CMB lectin. The protein structures are coloured to highlight the three sheets making up the prisms [marine for sheet (i), green/cyan for sheet (ii) and warm pink for sheet (iii)]. The cleaved β -chain is highlighted in purple in sheet (iii) for CGB, whereas in CMB this β -strand is part of the single oligopeptide chain. An asterisk (*) highlights the sugar-binding sites of the two proteins. (c) Shows a topology representation of the β -prism that is formed by both proteins. Mature CGB has been post-translationally processed, leaving $\beta 1$ as a separate chain. The processed loop is coloured in grey. The numbering assumes the β -chain to be the first strand of the protein, to make comparison between the two lectins easier.

nucleotide sequence, we performed another round of sequencing and the new sequence was deposited under accession No. HF546978. All protein structures and structure factors have been deposited in the Protein Data Bank (Velankar *et al.*, 2012) as entries 4ak4, 4akb, 4akc and 4akd.

3. Results and discussion

3.1. Sequencing of the lectin genes

The genes were sequenced from extracted RNA converted to cDNA. The nucleotide sequence for the CGB lectin is 95.9% similar to those of the gene encoding jacalin. This in turn translates into 97% identity for their amino-acid sequences. The CMB lectin has 96% nucleotide and 97% amino-acid sequence identities to artocarpin. The amino-acid sequences of CGB and CMB lectins aligned with those of their respective homologues are presented in Fig. 1. As there were some ambiguities in the primary nucleotide-sequencing results, these were repeated and both results were deposited with the EMBL Database.

3.2. CGB lectin structure

The structure of CGB lectin was initially determined at a resolution of 1.65 Å, with crystals exhibiting space group $P2_1$. The asymmetric unit of CGB lectin consists of eight subunits, six partial PEG

600 moieties and 972 water molecules. In the sugar-binding sites, there are tubes of electron density which have been assigned to partially ordered PEG 600 molecules or degraded products thereof (Supplementary Fig. 1[†]). When the individual chains within the asymmetric unit are superposed, the average root-mean-square deviation (r.m.s.d.) is 0.23 Å and the structures overlay well, with no significant differences. As a result, the following description of the monomeric unit will focus on a single $\alpha\beta$ -chain pair.

The monomer unit comprises a heavy α -chain consisting of 133 amino acids and a light β -chain consisting of 21 amino acids. The observed electron density covers all of the α -chains, whereas the β -chains are somewhat less well defined, with only residues 3–18 accounted for in the most ordered chains and residues 4–18 in the poorest ordered chain (chain *D*). The heavy chain consists of 11 β -strands (numbered $\beta 2$ – $\beta 12$ to make comparison with CMB easier).

The CGB lectin monomer makes up a β -prism, in which three antiparallel β -sheets, each consisting of four β -strands, share an interface (Fig. 2a). The β -chain makes up the fourth strand in the third β -sheet and will be referred to as $\beta 1$ throughout the structural description, to avoid confusion (Fig. 2c). The β -chain is bound to the α -chain with approximately 50% of its surface buried in this interface. Strands 5, 4, 3, 6 form sheet (i), strands 9, 8, 7, 10 form sheet (ii) and

[†] Supporting information has been deposited in the IUCr electronic archive (Reference: TT5050).

strands 2, 11, 12, 1 forms sheet (iii). The β -prism is kept in place by hydrogen bonds between the outer strands ($\beta 2$ and $\beta 5$, $\beta 6$ and $\beta 9$, $\beta 10$ and $\beta 1$; see Fig. 2c).

As this crystal was the only one exhibiting space group $P2_1$ and all others, including those reported here, exhibited $P2_12_12$, it was initially assumed that the monoclinic space group was due to accidental annealing during crystal transfer (Gabrielsen *et al.*, 2009). Superposing the eight subunits of the monoclinic space group onto the content of the orthorhombic cell, the extraneous monoclinic subunits are positioned where the symmetry-related copies of the $P2_12_12$ structure are located. This suggests that annealing is the likely cause of the difference in space group.

3.3. CMB structure

The structure of the CMB lectin was determined using crystals of space group $P2_12_12_1$ to a resolution of 2.0 Å (Gabrielsen *et al.*, 2010). As with CGB, the structure of the CMB monomer is a β -prism made up of a single polypeptide chain (Figs. 2b and 2c). There are four chains making up the asymmetric unit and 277 water molecules. Electron density was observed in the interfaces between adjacent asymmetric subunits, without being located on a symmetry axis. Based on the anomalous difference maps calculated and the presence of cadmium chloride in the crystallization conditions, these large clouds of unaccounted for density were assigned to be Cd^{2+} ions. The four chains superpose well, with an r.m.s.d. below 0.3 Å for chains A, B and D, and 0.37 Å for any superposition with chain C. The difference of r.m.s.d. for chain C is due to the poorer quality of the electron-density map for this region of the asymmetric unit, compared with the rest, which is also reflected by the much higher average *B* factor for this chain (77.0 Å² for the main chain of C *versus* 46.4 Å² for the average of all four chains). The loop between $\beta 8$ and

$\beta 9$ is disordered in all chains, although it is possible to model into the electron density in chains A, B and D. Lys35 is in the disallowed region of the Ramachandran plot for all chains, although it is clearly defined in the electron density for two.

3.4. Quaternary structures

Both lectins are homotetramers made up of four subunits (named subunits I, II, III and IV throughout the discussion), each subunit consisting of an α -chain and a β -chain or a single chain for the CGB and CMB lectin, respectively, similar to jacalin and artocarpin (Bourne *et al.*, 2002; Pratap *et al.*, 2002). The four subunits are related by 222 symmetry (Figs. 3 and 4).

Figure 3

(a) Cartoon representation of the CGB lectin tetramer, with the main parts of the four subunits coloured green, dark purple, marine and warm pink, and the separate β -chains in lighter shades. (b) Close-up of the interface between subunits I (green/cyan) and II (dark purple). (c) Close-up of the interface between subunits I and III (marine). The residues are named by subunit (in roman numerals) and by α -chain or β -chain (no prime or prime, respectively).

This configuration is kept in place by two different interfaces between the subunits. The first interface, between I and II (Fig. 3*b*; repeated between III and IV), is the largest, involving 4.3% of the accessible surface area of each of the α -chains. Hydrogen bonds are formed symmetrically between I-Glu109 on one subunit and II-Lys117 and II-Ser128 on the opposing subunit. This interface is further strengthened by the interaction between the β -chains themselves (involving three residues forming hydrophobic contacts) and the β -chain from one $\alpha\beta$ unit and the α -chain from the opposite one (involving 20% of the light-chain accessible surface area and 6% of the heavy-chain surface), with hydrogen bonds formed between I-Ile'11, I-Gly'13, I-Trp'15, I-Gln'8, I-Thr'9 on the β -chain and their respective partners II-Asn110, II-Pro107, II-Ser105 and II-Asn110 on the neighbouring α -chain (roman numerals identify which subunit the residues belong to and a prime highlights β -chain residues). The second interface, between subunits I and III (repeated between II and IV), is slightly smaller, involving 3.7% of the surface area of the two chains (Fig. 3*c*). Ten residues are involved in this interface, mostly forming a hydrophobic area, with hydrogen bonds being formed symmetrically between I-Val8 and III-Asn35. The β -chain of one subunit also forms three hydrogen bonds with the α -chain of the opposing subunit, between I-Gly'5, I-Ser'7 and I-Gln'8 and III-Pro61, III-Phe9/III-Thr10 and III-Leu133. There are no direct contacts between the two β -chains. Subunits I and IV (or II and III) have no significant interface between them. These interactions leave an appearance of a closely knitted structure formed by the intertwining of the different chains.

CMB lectin has a similar, albeit simpler, oligomeric structure (Fig. 4*a*). The same four subunits are positioned similarly, with some differences in the interfaces based on the sequence differences between the two homologues (Figs. 4*b* and 4*c*). The interface between subunits I and II is made by 15% of the solvent-accessible area of each subunit. The interface is stabilized by symmetrical salt bridges

between Lys134 and Glu126 on opposing chains, as well as a number of symmetrical hydrogen bonds between Thr7, Gly9 and Trp11 on one chain and Glu126/Asn127, Pro124 and Asn122 on the opposing chain, respectively.

The interface between subunits I and III involves 9% of the solvent-accessible area of each subunit. There are a number of symmetrical hydrogen bonds keeping the two chains together, namely Ser3, Gln4 and Ser23 on one chain and their bonding partners Thr25, Leu15, Asn48 on the opposing chain, respectively.

In contrast to the oligomeric structure of CGB, there is a shared interface between subunits I and IV in CMB. The interface makes up 4.8% of the solvent-accessible area of each chain and is stabilized by salt bridges between I-Lys60 and IV-Glu69 and between I-Arg136 and IV-Glu74, in addition to hydrogen bonding between I-Ser14 and IV-Lys10, I-Asn17 and IV-Glu74, and I-Arg136 and IV-Glu74. Owing to the compact nature of the tetramer, 30% of the tetrameric

Figure 4

(*a*) Cartoon representation of the CMB lectin tetramer, with the four subunits coloured green, dark purple, marine and warm pink. (*b*) Close-up of the interface between subunit I (green/cyan) and subunit II (dark purple). For clarity, only one of the two symmetrical parts of the interface is shown. (*c*) Close-up of the interface between subunits I and III (marine). The residues are named by subunit (in roman numerals).

accessible surface area is buried in the interfaces of both the CGB and CMB lectins.

3.5. Sugar binding by CGB lectin

Work by Jeyaprakash *et al.* (2003) has identified three components of the sugar-binding site of jacalin: a primary binding site and two secondary sites, named *A* and *B*. As there is strong sequence and structural homology between the CGB lectin and jacalin, we here base our following discussion using their terminology, where the primary binding site is responsible for binding the galactose part of the carbohydrate targets of CGB. The secondary site *A* can bind any α -linked sugar moiety but cannot tolerate any β -substitutions. In cases with a β -substituted disaccharide, the nonreducing end is located at secondary site *B* (Jeyaprakash *et al.*, 2003). The overall binding site is made up by loops between $\beta 5$ and $\beta 6$, $\beta 7$ and $\beta 8$, and $\beta 11$ and $\beta 12$.

As part of the structural analysis of the CGB lectin, galactose (Gal) or galactose- $\beta 1$ -3-galactose-Nac (GalNac) were soaked into crystals, data were collected and the structures of CGB in complex with the sugars were solved (CGB-Gal and CGB-GalNac, respectively). Attempts to introduce other sugars such as mannose were unsuccessful. The two structures superpose well onto the native structure,

with r.m.s.d. values below 0.2 Å. In both CGB-Gal and CGB-GalNac, the majority of the binding sites are occupied by a sugar moiety (three of four sites in CGB-Gal and all four sites in CGB-GalNac).

The differences in the conformation of the primary site are small when comparing the sugarless binding site with the occupied sites. The only noticeable change is the side chain of residue Phe122, which has rotated approximately 45° around the χ_2 angle. This shift in the side chain is due to the hydrophobic interactions formed between the side chain and the bound sugar molecule.

The binding of galactose occurs *via* a number of hydrogen bonds between the O atoms on the sugar ring and with side-chain and main-chain N and O atoms on the α chain (O3 and Gly1 N, O4 and Gly1 N and Asp125 OD1, O6 and Trp123 O, Trp123 N and Tyr122 N, O5 and Tyr122 N). There is also a stacking of Tyr78 against the *B* face of galactose and hydrophobic interactions from the aromatic group of Tyr122. These interactions are completely conserved in jacalin and the binding is identical to that described elsewhere (Jeyaprakash *et al.*, 2003; Fig. 5*a*).

The disaccharide GalNac binds in the same region, with the β -substituted sugar GalNac taking up the position of galactose in the monosaccharide soak. Similar hydrogen bonds are formed (O3 and Gly1 N, O4 and Gly1 N and Asp125 OD1, O6 and Asp125 OD1).

Figure 5 Sugar-binding sites of the lectins. (a) CGB-Gal structure with galactose observed in the binding site. The upper part shows the output from PDBsum (Laskowski, 2009), with hydrogen bonds shown as dashed lines and hydrophobic interactions shown as red arches. The lower part shows the same residues in the X-ray structure. (b) CGB-GalNac structure, with Gal observed in the binding site. (c) The CMB structure with superposed Me- α -Man from artocarpin. Superposition is based on protein backbone. For the sake of clarity artocarpin is not shown.

There are a number of hydrophobic interactions contributed by Tyr78, Gly121 and Tyr122 (Fig. 5*b*). The galactose points into the secondary site *B*, but does not make any specific interactions, as described for jacalin by Jeyaprakash *et al.* (2003).

Isothermal calorimetry (ITC) experiments showed that CGB binds galactose with a K_b of 0.34×10^{-3} M (data not shown), which is comparable with previous binding studies performed on jacalin, which has a K_b of 0.8×10^{-3} M (Arockia Jeyaprakash *et al.*, 2005). In our present experiments we could not show binding of mannose to CGB, and we were unsuccessful in obtaining a structure of CGB with mannose in the sugar-binding site. This is in contrast to previous results for jacalin, where weak binding of mannose was seen, and a crystal structure of jacalin with bound methylated mannose (Me- α -Man) in the sugar-binding site has been solved (Arockia Jeyaprakash *et al.*, 2005). Although there are no differences in structure or sequence that can account for this observed difference, the results are consistent with the earlier published work performed on CGB lectin and jacalin. Despite having closely related structures, these lectins seem to show different interactions with these sugars (Hashim *et al.*, 1991, 1993; Pineau *et al.*, 1991). It is likely that the change in specificity is caused by subtle differences in the environment near the sugar-binding site, including solvent molecules. Recent experiments conducted using mucins that were chemically treated to expose their glycan structures also showed a dissimilar pattern of binding when allowed to interact with these two lectins (Lee *et al.*, 2013).

3.6. Sugar binding by CMB lectin

The sugar-binding properties of CMB are opposite to those of CGB. CMB binds mannose with a K_b of 0.18×10^{-3} M, but shows no sign of binding galactose (data not shown). As of yet, we have not been able to obtain a crystal structure of CMB with mannose in its active site. However, as there is a high sequence similarity between CMB and artocarpin, with all the sugar-binding residues (Gly15, Asp138, Leu139 and Asp141) conserved, it is probable that the binding in these two cases is similar. Based on this hypothesis a model has been made by positioning mannose bound in a structure of artocarpin (PDB entry 1j4u; Pratap *et al.*, 2002) into the structure of the CMB binding pocket (Fig. 5*c*). The binding site is formed by the loops between $\beta 5$ and $\beta 6$, $\beta 7$ and $\beta 8$, and $\beta 11$ and $\beta 12$, as in the CGB lectin. The modelled mannose forms hydrogen bonds between O6 and Asp139, Leu140 and Asp142, O4 and Asp142, and O3 and Gly16. There are also hydrophobic interactions between the sugar molecule and residues Gly15 and Gly138. These residues are conserved in artocarpin and bind the sugar in a similar fashion.

4. Conclusions

We have determined the nucleotide sequence encoding the CGB and CMB lectins from champedak and determined their crystal structures to resolutions of 1.6 and 2.10 Å, respectively. Whilst similar in structure, the binding specificities of these two proteins are opposite, with the CGB lectin binding galactose but showing no sign of binding mannose, and *vice versa* for the CMB lectin. These natural products are becoming increasingly important in medical applications, where they can be used as diagnostic tools for a number of disorders which affect the glycosylation of proteins or their levels in serum or urine samples from patients.

This work was carried out with the support of the Diamond Light Source (Proposal Nos. MX1229 and MX6683). We acknowledge Dr Alan Riboldi-Tunnicliffe and Dr Alexander Schüttelkopf for helpful

discussions, Ms Margaret Nutley and Professor Alan Cooper for performing the ITC experiments, and Dr Kate Townson and Dr Hugh J. Willison for the gift of the disaccharide. We also thank Ms Wan Izlina Wan Ibrahim and Ms Jaime Jacqueline Jayapalan of University Malaya Centre for Proteomics Research for performing the MS analysis to determine the purity of both lectins. The work was partially funded by a grant from the University of Malaya (MOHE-HIR H-20001-E000009), Kuala Lumpur, Malaysia (OHH).

References

- Abdul Rahman, M., Anuar Karsani, S., Othman, I., Shafinaz Abdul Rahman, P. & Haji Hashim, O. (2002). *Biochem. Biophys. Res. Commun.* **295**, 1007–1013.
- Abdul-Rahman, P. S., Lim, B.-K. & Hashim, O. H. (2007). *Electrophoresis*, **28**, 1989–1996.
- Arockia Jeyaprakash, A., Jayashree, G., Mahanta, S. K., Swaminathan, C. P., Sekar, K., Surolia, A. & Vijayan, M. (2005). *J. Mol. Biol.* **347**, 181–188.
- Bond, C. S. (2003). *Bioinformatics*, **19**, 311–312.
- Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* **D65**, 510–512.
- Bourne, Y., Astoul, C. H., Zamboni, V., Peumans, W. J., Menu-Bouaouiche, L., Van Damme, E. J., Barre, A. & Rougé, P. (2002). *Biochem. J.* **364**, 173–180.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonnrhein, C. & Womack, T. O. (2011). *BUSTER v2.8.0*. Cambridge: Global Phasing Ltd.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. III, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids Res.* **35**, W375–W383.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Gabrielsen, M., Abdul-Rahman, P. S., Isaacs, N. W., Hashim, O. H. & Cogdell, R. J. (2010). *Acta Cryst.* **F66**, 592–594.
- Gabrielsen, M., Riboldi-Tunnicliffe, A., Abdul-Rahman, P. S., Mohamed, E., Ibrahim, W. I. W., Hashim, O. H., Isaacs, N. W. & Cogdell, R. J. (2009). *Acta Cryst.* **F65**, 895–897.
- Hashim, O. H., Gendeh, G. S. & Jaafar, M. I. (1993). *Biochem. Mol. Biol. Int.* **29**, 69–76.
- Hashim, O. H., Ng, C. L., Gendeh, S. & Nik Jaafar, M. I. (1991). *Mol. Immunol.* **28**, 393–398.
- Jeyaprakash, A. A., Katiyar, S., Swaminathan, C. P., Sekar, K., Surolia, A. & Vijayan, M. (2003). *J. Mol. Biol.* **332**, 217–228.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Laskowski, R. A. (2009). *Nucleic Acids Res.* **37**, D355–D359.
- Lee, C. S., Muthusamy, A., Abdul-Rahman, P. S., Bhavanandan, V. P. & Hashim, O. H. (2013). *Analyst*, **138**, 3522–3529.
- Lim, S. B., Chua, C. T. & Hashim, O. H. (1997). *J. Immunol. Methods*, **209**, 177–186.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Mohamed, E., Abdul-Rahman, P. S., Doustjalali, S. R., Chen, Y., Lim, B.-K., Omar, S. Z., Bustam, A. Z., Singh, V. A., Mohd-Taib, N. A., Yip, C.-H. & Hashim, O. H. (2008). *Electrophoresis*, **29**, 2645–2650.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Ono, M. & Hakomori, S. (2004). *Glycoconj. J.* **20**, 71–78.
- Pineau, N., Aucouturier, P., Preud'homme, J. L., Hagiwara, K. & Kobayashi, K. (1991). *Mol. Immunol.* **28**, 185–187.
- Pratap, J. V., Jeyaprakash, A. A., Rani, P. G., Sekar, K., Surolia, A. & Vijayan, M. (2002). *J. Mol. Biol.* **317**, 237–247.
- Seriramalu, R., Pang, W. W., Jayapalan, J. J., Mohamed, E., Abdul-Rahman, P. S., Bustam, A. Z., Khoo, A. S.-B. & Hashim, O. H. (2010). *Electrophoresis*, **31**, 2388–2395.
- daSilva, L. L. P., de Molfetta-Machado, J. B., Panunto-Castelo, A., Denecke, J., Goldman, G. H., Roque-Barreira, M. C. & Goldman, M. H. S. (2005). *Biochim. Biophys. Acta*, **1726**, 251–260.
- Storr, S. J., Royle, L., Chapman, C. J., Hamid, U. M., Robertson, J. F., Murray, A., Dwek, R. A. & Rudd, P. M. (2008). *Glycobiology*, **18**, 456–462.
- Velankar, S. *et al.* (2012). *Nucleic Acids Res.* **40**, D445–D452.
- Wahrenbrock, M. G. & Varki, A. (2006). *Cancer Res.* **66**, 2433–2441.
- Yang, H. & Czaplá, T. H. (1993). *J. Biol. Chem.* **268**, 5905–5910.