

Ashok K. Patel,^{a,b} ‡ Vijay K. Singh,^{b,‡} Ulrich Bergmann,^{a,c} Medicherla V. Jagannadham^b and Petri Kursula^{a,d,*}

^aDepartment of Biochemistry, University of Oulu, Oulu, Finland, ^bMolecular Biology Unit, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221 005, India, ^cBioCenter Oulu Proteomics Core Facility, University of Oulu, Finland, and ^dCentre for Structural Systems Biology, Helmholtz Centre for Infection Research (CSSB-HZI), DESY, Hamburg, Germany

‡ These authors made equal contributions.

§ Present address: Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA.

Correspondence e-mail:
petri.kursula@gmail.com

Received 23 February 2011
Accepted 17 March 2011

Purification, crystallization and preliminary X-ray crystallographic analysis of MIL, a glycosylated jacalin-related lectin from mulberry (*Morus indica*) latex

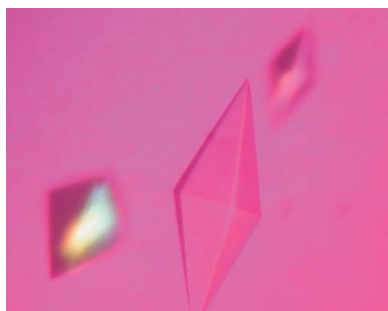
A quantitatively major protein has been purified from the latex of *Morus indica*. The purified previously uncharacterized protein, *M. indica* lectin (MIL), was further shown to be a glycosylated tetramer and belongs to the family of jacalin-related lectins. Crystallization of MIL was also accomplished and the tetragonal crystals diffracted synchrotron X-rays to a resolution of 2.8 Å.

1. Introduction

Lectins are a diverse group of proteins that are able to specifically recognize cell-surface carbohydrates, thus playing central roles in biological processes and being of considerable interest with regard to practical applications. One of the lectin subfamilies is comprised of jacalin-related lectins (JRLs), which fold into a type I β -prism containing three Greek-key motifs (Sankaranarayanan *et al.*, 1996). JRLs exhibit low sequence similarity and differ with respect to their carbohydrate-binding specificity. Members of the JRL family are thought to be involved, for example, in plant defence mechanisms (Chisholm *et al.*, 2000; Williams *et al.*, 2002; Wang & Ma, 2005; Ma *et al.*, 2010). There are presently >300 JRLs in the Pfam database, the vast majority of which are from plants. Some members of the JRL family have a modular structure comprising multiple copies of JRL domains in tandem or fused to other domains (Raval *et al.*, 2004). A number of JRLs have been characterized with respect to sugar specificity and structure, including three-dimensional structure (Sankaranarayanan *et al.*, 1996; Bouckaert *et al.*, 1999; Bourne *et al.*, 2002; Houles Astoul *et al.*, 2002; Pratap *et al.*, 2002; Jeyaprakash *et al.*, 2003; Rougé *et al.*, 2003; Rabijns *et al.*, 2005). Recently, the first crystal structure of a mammalian JRL was also reported (Kanagawa *et al.*, 2011).

Based on sugar specificity, JRLs can be subdivided into two groups: mannose-specific lectins (mJRLs) and galactose-specific lectins (gJRLs). A fundamental difference between the mJRLs and gJRLs concerns their biosynthesis and processing. The gJRLs (*e.g.* jacalin) undergo post-translational modification of a precursor polypeptide chain, generating a long α -chain and a short β -chain (Yang & Czaplá, 1993). In contrast, mJRLs have a single polypeptide chain and are synthesized on free ribosomes and localized to the cytosol (Peumans *et al.*, 2000). These differences in post-translational processing directly affect the size of the sugar-binding cavity (Houles Astoul *et al.*, 2002).

The mulberry tree (genus *Morus*) is probably best known for its use in raising silkworms (*Bombyx mori*) in sericulture, where it serves as the sole food for the worms; in this symbiotic setup the mulberry defence mechanisms do not attack silkworms, while they are detrimental to most other insects. The mulberry tree is also valuable for its edible fruits, which contain, for example, high concentrations of antioxidants. Several species of mulberry exist, growing in areas around the globe, and there is a large genetic variation between different species (Awasthi *et al.*, 2004). The latex of *M. indica* contains only a few proteins at high concentration, of which the protease indicain has previously been described (Singh *et al.*, 2008). Here, we have isolated and crystallized a quantitatively major protein directly from its natural source, *M. indica* latex, and identified it as a member of the jacalin-related lectin family.



2. Experimental

2.1. Materials

Hi-Load Superdex S-200 was purchased from Amersham Pharmacia. Ether-Toyopearl 650S (Tosoh) and Coomassie Brilliant Blue R-250 were obtained from Sigma. Ampholine carrier ampholytes were obtained from LKB. All other chemicals were of the highest purity commercially available.

2.2. Purification

The lectin MIL was purified from the latex of *M. indica* using a protocol modified from that used for the purification of indicain (Singh *et al.*, 2008). Fresh latex secretions induced by superficial incisions made on young stems and apical buds of *M. indica* plants were collected in 10 mM Tris–HCl pH 8.0 and stored at 253 K for 24 h. The latex was thawed at room temperature and centrifuged at 24 000g for 50 min to remove all insoluble cell debris and gum. The crude supernatant was treated with ammonium sulfate at 85% saturation at 277 K and the resultant precipitate was collected by centrifugation at 24 000g for 30 min. The pellet was redissolved in 25 mM MES pH 6.5 containing 1.5 M ammonium sulfate and subjected to hydrophobic interaction chromatography on an Ether-Toyopearl column (7 × 2 cm) pre-equilibrated with the same buffer. The column was washed with the same buffer and the bound proteins were eluted with a decreasing linear gradient of 1.5–0 M ammonium sulfate at a flow rate of 3 ml min⁻¹. All fractions were assayed for protein content, extent of homogeneity and haemagglutination activity.

The active fractions were desalted, concentrated using a 10 kDa cutoff Vivaspin (Vivascience) concentrator and subjected to size-exclusion chromatography on a Superdex S-200 column (Pharmacia, 1.2 × 120 cm) with a buffer consisting of 25 mM MES pH 6.5 and 0.5 M NaCl at a flow rate of 0.25 ml min⁻¹. The active and homogenous fractions were pooled, concentrated by ultrafiltration, dialyzed and stored in 50 mM MES pH 6.0 at 277 K. During the purification and analysis, SDS–PAGE gels were generally run under mildly denaturing conditions (no reducing agents and no heating of the sample); to obtain MIL monomers, regular denaturing conditions were used (10 mM DTT in the sample buffer, heating of the sample to 368 K for 5 min prior to the run).

2.3. Haemagglutination assay

Fresh whole rabbit blood (5.0 ml) was washed three times and suspended in 15 ml phosphate-buffered saline pH 7.4 (PBS). Trypsin (50 µl, 10 mg ml⁻¹ solution in PBS) was added to the washed erythrocytes and the sample was incubated at 310 K for 20 min. The trypsinized erythrocytes were washed three times with PBS and then suspended in PBS to give a 2% suspension. Haemagglutination assays were carried out in small glass tubes or in 96-well plates in a final volume of 50 µl consisting of 40 µl of a 1%(v/v) suspension of red blood cells and 10 µl lectin fraction. The haemagglutination titre, defined as the reciprocal of the highest dilution exhibiting visible haemagglutination, was considered as one haemagglutination unit. Specific activity is defined in this case as the number of haemagglutination units per milligram of protein.

2.4. Carbohydrate-content determination and deglycosylation

The carbohydrate content of MIL was determined using the phenol–sulfuric acid method (Hounsell *et al.*, 1997); the carbohydrate content of the purified lectin was obtained from a standard plot generated using galactose standards. Deglycosylation of the enzyme

was performed by a chemical method using trifluoromethanesulfonic acid (Edge, 2003). Electrophoresis was carried out with glycosylated and deglycosylated protein samples using the method described above. The gel was stained with Schiff's reagent, which is specific for glycoproteins.

2.5. Isoelectric focusing

The isoelectric point (pI) of purified MIL was determined by isoelectric focusing on polyacrylamide disc gels (0.6 × 10.0 cm). Electrophoretic runs were carried out with ampholine carrier ampholytes (pH range 4–6) at 300 V for 2 h using 0.1 M NaOH as the catholyte and 0.1 M orthophosphoric acid as the anolyte. The protein was visualized by Coomassie G-250 staining.

2.6. Mass spectrometry

The molecular weight of the purified enzyme was determined by MALDI–TOF as previously described for indicain (Singh *et al.*, 2008). Tryptic peptide analysis and partial sequencing of the fragmented peptides were used to obtain the first sequence information on MIL. An SDS–PAGE band representing MIL was excised from a Coomassie-stained gel and destained by washing three times for 5 min with 50 mM ammonium bicarbonate in 50% acetonitrile. The sample was then alkylated by incubation with 20 mM DTT followed by 45 mM iodoacetamide, washed and dehydrated by incubation for several minutes in acetonitrile and subsequent drying in a SpeedVac for 5 min. The dried gel piece was incubated with trypsin [20 ng µl⁻¹ trypsin (Roche, recombinant proteomics grade) dissolved in 50 mM ammonium bicarbonate containing 10%(v/v) acetonitrile] in a PCR instrument using a program that kept the sample at 275 K for the first hour and subsequently at 308 K. After overnight incubation, tryptic peptides were extracted by incubation with 20 µl 1% trifluoroacetic acid (TFA) in 10% acetonitrile for 5 min in an ultrasonic bath. Extraction was repeated with 20 µl 50% acetonitrile/0.1% TFA and with 20 µl 70% acetonitrile/0.1% TFA. The combined extracts were diluted with 0.1% TFA to 200 µl and passed over a VivaPure C-18 microcartridge (Sartorius) following the instructions of the manufacturer. Peptides were eluted into 5 µl 0.2% formic acid/50% acetonitrile and subjected to offline nano-electrospray mass spectrometry on a Q-Tof 2 (Waters). The emitters were medium-length filament borosilicate capillaries from Proxeon. A survey spectrum was obtained using 30 V cone voltage and 800 V capillary potential. Several double- or triple-charged ions were selected for MSMS interrogation with collision energies between 30 and 45 V. MSMS spectra were interpreted manually to read sequence tags, which were then used in a *BLAST* search.

2.7. Crystallization of MIL

Prior to crystallization screening, the purity and conformational homogeneity of the protein were confirmed by SDS–PAGE and dynamic light scattering (DynaPro 801, Wyatt Technology). Crystallization was performed with 15 mg ml⁻¹ pure MIL in 0.1 M Tris–HCl pH 8.0, 5.0 mM DTT, 10 mM CaCl₂ and 20 mM NaCl. The original crystals grew in a condition from Hampton Research Index screen consisting of 0.2 M Li₂SO₄ and 25% PEG 3350 in 0.1 M Bis-Tris propane pH 6.0. To obtain diffracting crystals of sufficient size, further crystallization experiments were performed using the hanging-drop vapour-diffusion technique at 291 K. Single crystals were grown in 4–5 d from drops that were obtained by mixing 2 µl protein solution and 2 µl precipitant solution consisting of 0.2 M Li₂SO₄ and 23% PEG 3350 in 0.1 M Bis-Tris propane pH 5.5. The

Table 1
Purification of the lectin MIL from the latex of *M. indica*.

Step	Total protein (mg)	Total activity (units)	Specific activity (units mg ⁻¹)	Purification (fold)	Yield (%)
Crude extract	64.5	1200	18.6	1.0	100
Ammonium sulfate precipitation (85%)	47.0	1050	22.3	1.2	88
HIC ETP column unbound	19.5	505	25.9	1.4	42
Gel filtration (Superdex S-200)	8.2	310	37.8	2.0	26

tetragonal bipyramid-shaped crystals had typical dimensions of 300 × 200 × 200 μm.

2.8. Data collection and processing

Prior to data collection, a single crystal was cryocooled by plunging it into liquid nitrogen. Diffraction data were collected on the BL14-1 synchrotron beamline at BESSY (Berlin) at a wavelength of 0.918 Å in a cryostream of gaseous nitrogen at 100 K. Data were collected at a crystal-to-detector distance of 316 mm and 360 images with an oscillation range of 0.5° were recorded using an exposure time of 3 s per frame. The diffraction data were processed using *XDS* (Kabsch, 2010) and *XDSi* (Kursula, 2004). Programs from the *PHENIX* (Adams *et al.*, 2010) and *CCP4* (Winn *et al.*, 2011) packages were used for further data analysis.

3. Results and discussion

3.1. Purification and biochemical characterization of *M. indica* lectin

The previously uncharacterized MIL lectin was purified from the latex of *M. indica*. The starting material contained only a few proteins with high abundance, as seen on SDS-PAGE; the most abundant protein had an apparent molecular weight of 65 kDa (Fig. 1*a*). The purification procedure involved ammonium sulfate precipitation, hydrophobic interaction chromatography and size-exclusion chromatography. The purification was followed by a haemagglutination

assay. In hydrophobic interaction chromatography, MIL appeared in the unbound fraction; note that the corresponding bound fraction was used previously to purify indicain (Singh *et al.*, 2008). There were three peaks in size-exclusion chromatography, which was the last purification step (Fig. 1*b*); the second peak showed a high purity on SDS-PAGE and haemagglutination activity (not shown) and fractions from this peak were thus pooled and used for further experiments. The purification is summarized in Table 1.

The purified lectin showed an apparent molecular weight of ~70 kDa on SDS-PAGE under mildly denaturing conditions (no heating or DTT) and a single band of ~18 kDa under standard denaturing conditions (heating with 10 mM DTT), indicating that the protein is likely to be a homotetramer consisting of four identical subunits of 18 kDa (Fig. 1*c*). Mass spectrometry revealed the presence of a mixture of polypeptides with molecular masses of 17.350 kDa (monomer), 34.709 kDa (dimer) and 69.489 kDa (tetramer) (not shown), further proving the presence of MIL as a tetramer.

The minimal concentration required for visible haemagglutination of purified MIL was 3.0 ng ml⁻¹ with rabbit red blood cells and preliminary results from inhibition assays suggest that MIL has a preference for galactose over maltose (unpublished data). The chemical carbohydrate analysis indicated that MIL has ~6–7% covalently bound carbohydrate. To further confirm the glycosylation, an SDS-PAGE gel was stained with Schiff's reagent (Fig. 1*c*) and the staining also identified MIL as a glycoprotein. After deglycosylation, the enzyme aggregated and lost its haemagglutination activity (not shown), indicating a possible role of the glycans in structural integrity. The purified protein appeared as single band on isoelectric focusing with an isoelectric point of ~4.6, suggesting that the surface of MIL is highly populated by acidic amino-acid residues (Fig. 1*c*).

3.2. Tryptic peptide mapping and partial sequence from fragmentation

No sequence exists for MIL in the sequence databases. Thus, in order to confirm the identity of the protein and to obtain preliminary sequence information, CID tandem mass spectrometry was used.

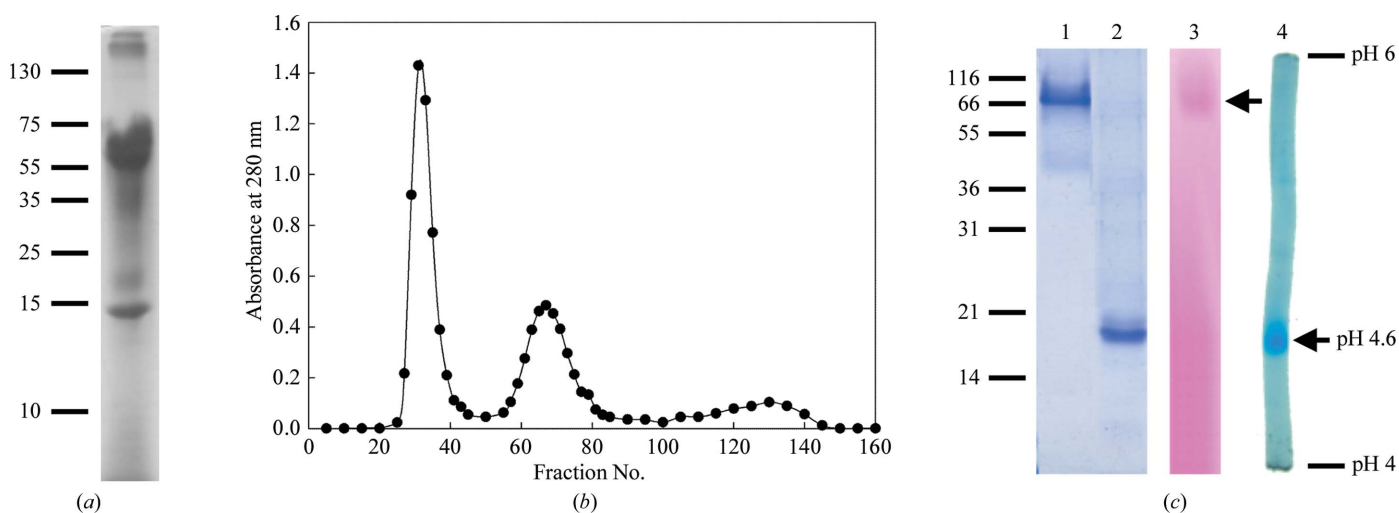


Figure 1
Purification of MIL. (a) SDS-PAGE analysis of the crude material, *i.e.* the clarified latex sample. Positions of molecular-weight markers are shown on the left (in kDa). (b) Size-exclusion chromatography elution profile of MIL at the last step of the purification on a Superdex S-200 column. The fractions from the second peak (65–75) contained pure MIL. (c) Electrophoretic analyses of the purified MIL lectin from *M. indica*. Lane 1, MIL under mildly denaturing conditions; lane 2, MIL under fully denaturing conditions; lane 3, as in lane 1 but stained with Schiff's reagent; lane 4, isoelectric focusing of MIL. The positions of molecular-weight markers are indicated on the left (in kDa; for lanes 1–3); the pH values for isoelectric focusing are indicated on the right. MIL in lanes 3 and 4 is indicated by arrows.

Table 2
MIL diffraction data-processing statistics.

Values in parentheses are for the highest resolution shell

Space group	$P4_12_1/P4_32_12$
Unit-cell parameters	
$a = b$ (Å)	118.9
c (Å)	128.9
$\alpha = \beta = \gamma$ (°)	90
Resolution range (Å)	20.0–2.80 (2.87–2.80)
Total No. of observations	128573
No. of unique reflections	23073
Multiplicity	5.6 (5.6)
Completeness (%)	98.7 (97.4)
R_{merge}^\dagger	0.149 (0.924)
$\langle I/\sigma(I) \rangle$	8.8 (1.8)
Wilson B factor (Å ²)	50.1

$$^\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

Rather reliable sequence tags were obtained for two tryptic peptides (Fig. 2). The first of these peptides, which was fully sequenced, is nearly identical to a segment of jacalin-family lectins from different plants, including a galactose-binding lectin from *M. nigra* (MornigaG), while the second peptide is more remotely homologous to the very C-terminus of MornigaG, suggesting that this peptide represents the MIL C-terminus. The peptides are less similar to peptides from MornigaM, a mannose-specific lectin from *M. nigra*. While the results confirm the presence of a jacalin-type lectin in the sample, they also indicate that large differences in the sequence may exist between such lectins from *M. nigra* and *M. indica*, two related plant species. The crystal structure of MornigaM has been determined, also with a bound mannose ligand (Rabijns *et al.*, 2005).

3.3. Crystallization

Prior to the crystallization experiments, the purity and homogeneity of MIL were analysed by SDS–PAGE with Coomassie R-250 and silver staining and by dynamic light scattering (data not shown). The protein was found to be highly pure and homogeneous and to be suitable for crystallization. During crystallization screening, original hits were obtained using sparse-matrix screens. During optimization, large bipyramidal crystals were obtained that routinely diffracted X-rays to approximately 3.2 Å resolution using synchrotron radiation (Fig. 3a).

3.4. Data collection and processing

One of the approximately 50 crystals tested diffracted synchrotron X-rays to a resolution of 2.8 Å on beamline BL14-1 at BESSY, Berlin (Fig. 3b). This crystal was used for data collection and the data-processing statistics are presented in Table 2. Based on the symmetry of the diffraction data, the crystal belonged to the primitive

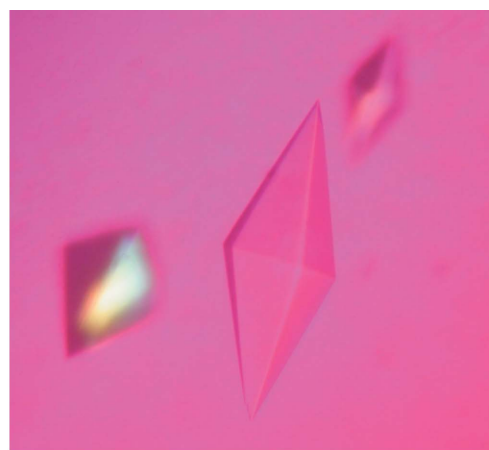
MIL	K-TYGPYGH	TSGR																
R																		
MornigaG	E-TYGPYGV	TSGT																
MornigaM	K-TYGPYGK	EEGT																
MIL	I	G	G	V	F	L	E	N	L	G	F	H	L	A	L			
I																		
MornigaG	K	G	S	V	G	Y	W	L	D	Y	I	G	F	H	L	S	L	-COOH
MornigaM	K	G	R	S	G	F	V	V	D	A	I	G	V	H	L	S	L	-COOH

Figure 2
Alignment of the two determined peptide sequences and the corresponding regions of the galactose-specific and mannose-specific jacalin-related lectins from *M. nigra* (MornigaG and MornigaM, respectively). Identical residues are shaded grey.

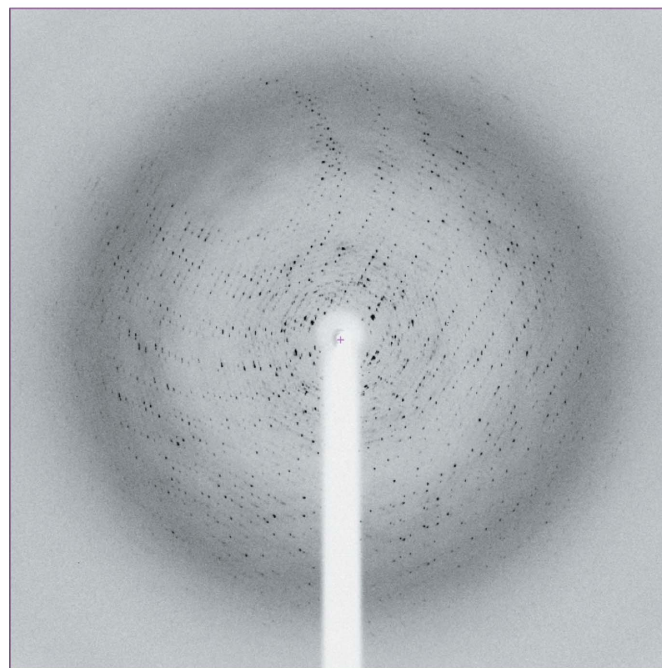
tetragonal point group $P422$ (unit-cell parameters $a = 118.9$, $b = 118.9$, $c = 128.9$ Å, $\alpha = \beta = \gamma = 90^\circ$). The systematic absences indicate that the space group is $P4_12_12$ or $P4_32_12$.

Assuming one tetrameric assembly per asymmetric unit, the Matthews coefficient ($V_M = 3.28$ Å³ Da⁻¹) falls within the generally observed range (Matthews, 1968). The solvent content is estimated to be 63% in this case. It is possible that up to six MIL monomers may occupy the asymmetric unit, but a higher than average solvent content seems to be likely from the relatively high Wilson B factor and the weak diffraction obtained from the crystals despite extensive optimization of the crystallization conditions.

In addition to our molecular-replacement efforts using several plant lectins of the JRL family as models, we are attempting to soak crystals with heavy atoms and to determine the structure of MIL using experimental phasing approaches. Simultaneously, attempts are being made to obtain more complete coverage of the amino-acid



(a)



(b)

Figure 3
Crystals of MIL and preliminary diffraction data. (a) Tetragonal crystals grown using hanging-drop vapour diffusion with 0.2 M Li₂SO₄ and 23% PEG 3350 as precipitant in 0.1 M Bis-Tris propane pH 5.5. (b) Diffraction pattern, with spots visible to a resolution of 2.8 Å.

sequence of MIL, a quantitatively major protein from the economically important plant *M. indica*. Based on the unique features of the identified peptide sequences, the three-dimensional structure of MIL is likely to provide novel information on the fine structural details of the JRL family.

Funding from the Council of Scientific and Industrial Research (India), the University Grants Commission (India) and the Academy of Finland is gratefully acknowledged. AKP and VKS received a travel grant from the Boehringer Ingelheim Fonds (Germany) to carry out crystallization experiments and data collection at the RUG, Groningen, The Netherlands and the University of Oulu, Oulu, Finland. We also acknowledge Dr A. J. Moir (University of Sheffield, England) for the MALDI-TOF analysis. Data collection at BESSY was supported by the European Community Seventh Framework Programme (FP7/2007–2013) under grant agreement 226716.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Awasthi, A. K., Nagaraja, G. M., Naik, G. V., Kanginakudru, S., Thangavelu, K. & Nagaraju, J. (2004). *BMC Genet.* **5**, 1.
- 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., Barre, A. & Rougé, P. (2002). *Biochem. J.* **364**, 173–180.
- Chisholm, S. T., Mahajan, S. K., Whitham, S. A., Yamamoto, M. L. & Carrington, J. C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 489–494.
- Edge, A. S. (2003). *Biochem. J.* **376**, 339–350.
- Houlès Astoul, C., Peumans, W. J., van Damme, E. J., Barre, A., Bourne, Y. & Rougé, P. (2002). *Biochem. J.* **367**, 817–824.
- Hounsell, E. F., Davies, M. J. & Smith, K. D. (1997). *The Protein Protocols Handbook*, edited by J. M. Walker, pp. 633–634. Totowa: Humana Press.
- Jeyaprakash, A. A., Katiyar, S., Swaminathan, C. P., Sekar, K., Surolia, A. & Vijayan, M. (2003). *J. Mol. Biol.* **332**, 217–228.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kanagawa, M., Satoh, T., Ikeda, A., Nakano, Y., Yagi, H., Kato, K., Kojima-Aikawa, K. & Yamaguchi, Y. (2011). *Biochem. Biophys. Res. Commun.* **404**, 201–205.
- Kursula, P. (2004). *J. Appl. Cryst.* **37**, 347–348.
- Ma, Q.-H., Tian, B. & Li, Y.-L. (2010). *Biochimie*, **92**, 187–193.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Peumans, W. J., Hause, B. & Van Damme, E. J. (2000). *FEBS Lett.* **477**, 186–192.
- Pratap, J. V., Jeyaprakash, A. A., Rani, P. G., Sekar, K., Surolia, A. & Vijayan, M. (2002). *J. Mol. Biol.* **317**, 237–247.
- Rabijns, A., Barre, A., Van Damme, E. J., Peumans, W. J., De Ranter, C. J. & Rougé, P. (2005). *FEBS J.* **272**, 3725–3732.
- Raval, S., Gowda, S. B., Singh, D. D. & Chandra, N. R. (2004). *Glycobiology*, **14**, 1247–1263.
- Rougé, P., Peumans, W. J., Barre, A. & Van Damme, E. J. (2003). *Biochem. Biophys. Res. Commun.* **304**, 91–97.
- Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. & Vijayan, M. (1996). *Nature Struct. Biol.* **3**, 596–603.
- Singh, V. K., Patel, A. K., Moir, A. J. & Jagannadham, M. V. (2008). *Phytochemistry*, **69**, 2110–2119.
- Wang, X.-M. & Ma, Q.-H. (2005). *Plant Physiol. Biochem.* **43**, 185–192.
- Williams, C. E., Collier, C. C., Nemacheck, J. A., Liang, C. & Cambron, S. E. (2002). *J. Chem. Ecol.* **28**, 1411–1428.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Yang, H. & Czaplá, T. H. (1993). *J. Biol. Chem.* **268**, 5905–5910.