Research Article | Open Access

Crystal Structure of *Colocasia esculenta* Tuber Agglutinin at 1.74 Å Resolution and Its Quaternary Interactions

Rajagopal Chattopadhyaya*, Himadri Biswas and Apurba Sarkar

Department of Biochemistry, Bose Institute, Calcutta, India

Abstract

Many members of the monocot mannose-binding lectin family, characterized by specificity towards mannose have been characterized and cloned. A majority of these lectins molecules contain 1-4 polypeptides of about 110 residues each. From the previously solved crystal structures of a few such lectins, mostly from non-edible plants, these lectins are thought to possess a common β-prism II fold structure. The major tuber storage protein of *Colocasia esculenta* is a monocot mannose-binding, widely used, dietary lectin. This tuber agglutinin contains two polypeptides of 12.0 and 12.4 kDa by matrix assisted laser desorption ionisation time-of-flight mass spectrometry. By gel filtration at pH 7.2, the purified lectin has a $\alpha_2\beta_2$ form of apparent molecular mass of 48.2 kDa in solution but at pH 3, it has the heterodimeric $\alpha\beta$ form. Lectin crystals were obtained by hanging-drop, vapor-diffusion method at room temperature and high-resolution X-ray diffraction data were collected using a home X-ray source. Among previously solved crystal structures of this family are garlic, Solomon's seal, snowdrop, daffodil and Spanish blue-bell lectins, but the protein sequence of the *Colocasia esculenta* tuber agglutinin was found to be closest to that of the *Remusatia vivipara* lectin having no simple mannose-binding property. Using the previously solved 2.4Å crystal structure of the *Remusatia vivipara* lectin, that of *Colocasia esculenta* has been solved by molecular replacement and subsequent crystallographic refinement and root mean square deviations between various lectins are tabulated and rationalized. The asymmetric unit in our lectin crystal structure contains four β-prism II domains or two $\alpha\beta$ heterodimers, each forming a $\alpha_2\beta_2$ heterotetramer with a symmetry related unit. The tetrameric interface obtained from our crystal structure is used to explain the conversion to dimers in acidic pH. Five ordered magnesium ions were located in the asymmetric unit and the presence of magnesium verified by atomic absorption spectroscopy.

Abbreviations: CEA: *Colocasia esculenta* Agglutinin; CVL: *Crocus vernus* Lectin; F.O.M.: Figure of Merit; MALDI-TOF: Matrix Assisted Laser Desorption Ionisation Time-of-Flight Mass Spectrometry; PCL: *Polygonatum cyrtonema* Agglutinin; RVL: *Remusatia vivipara* Lectin; Scafet: *Scilla campanulata* Heterodimeric Agglutinin

Introduction

The recognition of carbohydrate moieties by lectins has important applications in a number of biological processes such as cell-cell interaction, signal transduction, cell growth and differentiation [1]. The functionality of lectin molecules depend on the specific carbohydrate recognition domain, a part of the three dimensional structure of the protein, normally held together by non-covalent interactions and disulfide linkages.

Among the twelve new families of plant lectins [2] or the seven families by the older classification [3], the monocot mannose-binding lectin family comprises lectins with an exclusive specificity towards mannose, also called GNA-related lectins. Numerous members of this family of lectins have been characterized and cloned from Alliaceae, Amaryllidaceae, Araceae, Bromeliaceae, Liliaceae and Orchidaceae species, as summarized by Barre et al. [4]. The majority of all these lectins consist of one, two or four identical polypeptide(s) of about 110 amino-acid residues. Each subunit possesses a novel pseudo-3-fold symmetry having three 4-stranded anti parallel β-sheets oriented as 3 sides of a trigonal prism forming a 12-stranded β-barrel, referred to as the β-prism II fold, first reported in the homotetrameric *Gallanthus nivalis* crystal structure in complex with methyl-α-D-mannoside (PDB ID 1MSA) [5]. The core of this β-barrel is lined with conserved hydrophobic side chains, which stabilize the fold. Later crystal structures of complexes of the heterodimeric *Allium sativum* lectin (1BWU) [6], homotetrameric *Narcissus pseudonarcissus* lectin (1NPL) [7], homodimeric *Scilla campanulata* lectin (1B2P) [8] and homodimeric *Polygonatum cyrtonema* (3A0C) [9] all show the β-prism II fold and are monocot mannose binding lectins. Except for the garlic (*A. sativum*) and Solomon's seal (*P. cyrtonema*) lectins which are from edible/medicinal plants, other mannose binding lectins mentioned above are from non-edible, gardening plants like snowdrop (*G. nivalis*), daffodil (*N. pseudonarcissus*) and Spanish blue-bells (*S. campanulata*), the last two being highly poisonous. In view of possible intestinal or immunological toxicity/distress, the study of lectins from edible plants assumes greater importance than those from poisonous ones.

Colocasia esculenta of the Araceae family is a tuberous monocotyledonous Asian plant growing in tropical and subtropical climates; it is widely used for human consumption as a supplementary food source [10]. Colocasia extracts (from taro, corm) possess important pharmacological properties including anti-inflammatory, anti-cancer, anti-fungal, anti-viral [11], while the lectin has insecticidal activities [12]. Another group reported several isoforms of the very similar lectin tarin and its covalent modification [13], but in our study of the storage protein, covalently bound sugar was never found [12]. We find the intact protein is a $\alpha_2\beta_2$ heterotetramer of 49 kDa composed of two different polypeptides, with small subunits of 12.0 kDa and large subunits of 12.4 kDa though slightly different masses were reported earlier [14].

*Corresponding author: Rajagopal Chattopadhyaya, Department of Biochemistry, Bose Institute, Calcutta-700054, India, Tel: 91-33-2569-3239; Fax: 91-33-2355-3886; E-mail: raja@jcbose.ac.in

Our 1.74Å crystal structure of a β-prism II (BP2) fold of the Colocasia lectin (PDB ID 5D5G) describes the dimeric and tetrameric interactions.

Materials and Methods

Materials

Syringe-driven filters, 0.22 µm pore size, were purchased from Merck Millipore (Mumbai, India). Boxes for setting hanging drop crystals were bought from Nunc. (Roskilde, Denmark) and cover slips from Blue Star (Mumbai, India). Most of the buffering agents (Hepes, Na-cacodylate, MOPS, etc.), precipitants (PEGs) and Sigmacote were procured from the Sigma Chemical Company, Missouri, USA. All other chemicals, obtained from Merck (Mumbai, India), were of molecular biology or analytical grade.

Protein purification

CEA was purified to homogeneity following a modification of the known protocol [14]. The tubers were homogenized in 0.2 M NaCl containing 1 gl^{-1} ascorbic acid (5 ml per gram of fresh weight) at pH 7.0 using a Waring blender. The homogenates were filtered through cheesecloth and centrifuged (12,000 rpm for 10 min). After it was brought to 20 mM in CaCl$_2$, the pH was adjusted to 9.0 (with 1 N NaOH), kept overnight in the cold, and re-centrifuged at 12,000 rpm for 10 min. Then it was adjusted to pH 4.0 (with 1 N HCl) and re-centrifuged at 12,000 rpm for 10 min. Subsequently, the clear supernatant was adjusted to pH 7.5 (with 1 N NaOH) and solid ammonium sulphate was added to reach a final concentration of 1.5 M. After standing overnight in the cold room, the precipitate was removed by centrifugation at 12,000 rpm for 30 min. The final supernatant was decanted, filtered through filter paper (Whatman 3 MM). A column of mannose-Sepharose 4B was equilibrated with 1.5 M ammonium sulphate (in 50 mM sodium acetate, pH 6.5). After passing the extract through the column, it was next washed with 1.5 M ammonium sulphate (in 50 mM sodium acetate, pH 6.5) until the A280 decreased below 0.01. Then, the lectin was eluted by a gradient of ammonium sulphate decreasing from 1.5 M to 0 M. The elution profile was monitored online in a Waters 2489 detector at 280 nm. The concentrations of the protein solutions, stated in our text, were determined from their heterodimeric extinction coefficient, ε_{280}, of 46,660 M^{-1} cm^{-1}. Hence all concentrations mentioned here mean the heterodimeric concentration.

MALDI-TOF analysis

Matrix assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) analysis of the crystal was performed on an Ultraflex TOF/TOF (Bruker Daltonics, Germany) mass spectrometer equipped with a UV nitrogen laser of 337 nm. A protein solution (1 µL) was mixed with 1 µL matrix solution (saturated solution of sinapinic acid in acetonitrile/0.1% aqueous trifluoroacetic acid) and 1 µL of this mixture was deposited on the probe plate. The spectra were recorded in the reflectron positive ion mode after the evaporation of the solvent and were acquired and analyzed by Bruker Daltonics Flex control software.

pH treatment of CEA

To study the effect of pH on the secondary and tertiary structure of CEA, the following buffers were used, all having 150 mM NaCl: 10 mM Glycine-HCl (pH 1-3), 10 mM Na-acetate (pH 4-5), 10 mM Na-phosphate (pH 6-7.2), 10 mM Tris-Cl (pH 8-9) and 10 mM Glycine-NaOH (pH 10-12). Protein was incubated for 12 h at room temperature at different pH values before recording various size exclusion chromatography as described below.

Size-exclusion chromatography

The size-exclusion chromatography experiments were performed on a Superdex-75 10/300 GL column attached to a Waters HPLC system. An aliquot of 200 µl of protein samples (0.5 mg/ml) prepared by incubation with varying pH was injected on to the column. The column was pre-equilibrated with the appropriate buffer of varying pH. The flow rate was adjusted to 0.5 ml/min and elutant was detected on-line by Waters 2489 UV visible detector at 280 nm. To determine the size, the column was calibrated with the following proteins in PBS, pH 7.2: *Helix promita* lectin (79 kDa), *Galanthus nivalis* lectin (52 kDa), *Narcissus pseudonarcissus* lectin (26 kDa), *Pseudomonas aeruginosa* lectin (13.7 kDa).

Crystallization and X-ray diffraction data collection

CEA crystals were obtained by hanging-drop, vapor-diffusion method at room temperature, in about 3 to 4 weeks using 0.1 M Hepes pH 6.0 and 1.8 M ammonium sulfate as precipitant. Data collections at cryogenic temperature (−170ºC) were carried out at the home X-ray source. For data collection, the crystals were soaked in the cryo-protectant solution (respective buffer containing additional 30% ethylene glycol). The collected data were indexed, processed and scaled using Crystal Clear™ (Version 2.0) and the implemented program d*TREK®.

Molecular replacement, refinement and interface analysis

The structure solution and analyses were carried out using various modules of CCP4i; the structures were solved using molecular replacement [15] using the crystal structure of the two domain RVL (PDB ID: 3R0E) [16] having 94.6% sequence identity to CEA. The model was monitored and modified using the molecular graphics program Coot within the CCP4i package, substituting the side chains for the residues differing between CEA and RVL. The refinement was carried out mostly using Refmac5 [17] also within the CCP4i package, switching to PHENIX 1.8.1 [18] at the very end.

The physicochemical features of the dimeric and tetrameric interfaces were calculated using the PISA web server available within the PDB [19], including listing of salt bridges, hydrogen bonds, interfacing residues, etc. The program Coot [15] was used for monitoring the progress of the crystallographic refinement as well as for the display of models superposed with electron density maps, as shown here. Refer to Supplementary Table 1 for crystallography data collection and refinement statistics.
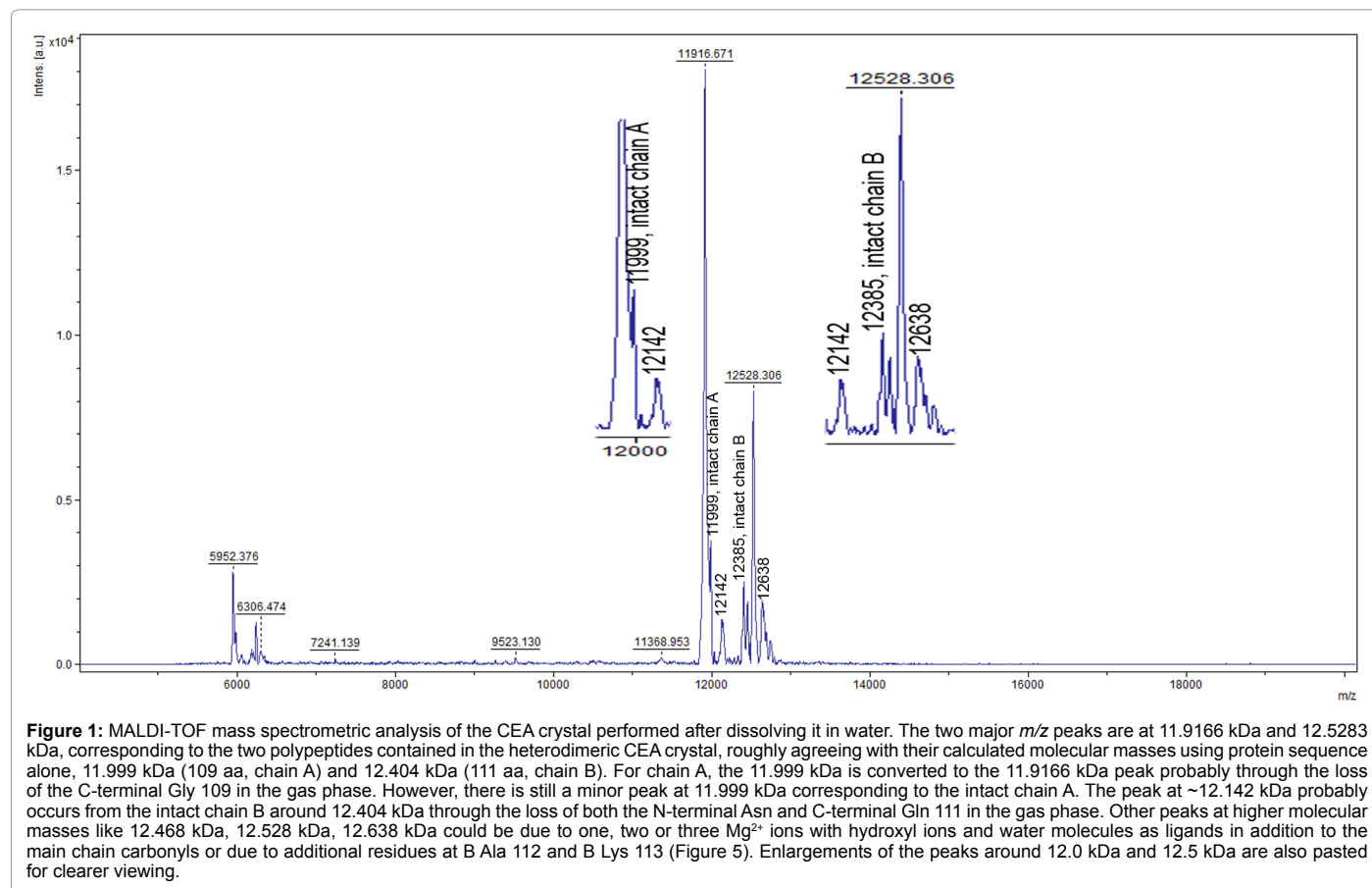
Magnesium detection by atomic absorption

After finding the ordered magnesium ions in our electron density maps, the detection of magnesium (absorption at 285.4 nm) and manganese (absorption at 279.7 nm) from some CEA samples dissolved in 0.01 N hydrochloric acid were attempted by atomic absorption spectroscopy in a Thermo Scientific iCE3000 series.

Results

Mass spectrometric analysis

The mass spectrum (Figure 1) of the solubilised CEA crystal ascertained that two chains of masses 11.9166 kDa and 12.5283 kDa were crystallized and these correspond roughly with the masses calculated from the sequences of the two protein chains in CEA, 11.999 kDa (109 aa) and 12.404 kDa (111 aa), respectively, differences attributed to either loss of terminal residues in the gas phase or addition of hydrated magnesium ions particularly for chain B.

Figure 1: MALDI-TOF mass spectrometric analysis of the CEA crystal performed after dissolving it in water. The two major *m/z* peaks are at 11.9166 kDa and 12.5283 kDa, corresponding to the two polypeptides contained in the heterodimeric CEA crystal, roughly agreeing with their calculated molecular masses using protein sequence alone, 11.999 kDa (109 aa, chain A) and 12.404 kDa (111 aa, chain B). For chain A, the 11.999 kDa is converted to the 11.9166 kDa peak probably through the loss of the C-terminal Gly 109 in the gas phase. However, there is still a minor peak at 11.999 kDa corresponding to the intact chain A. The peak at ~12.142 kDa probably occurs from the intact chain B around 12.404 kDa through the loss of both the N-terminal Asn and C-terminal Gln 111 in the gas phase. Other peaks at higher molecular masses like 12.468 kDa, 12.528 kDa, 12.638 kDa could be due to one, two or three Mg^{2+} ions with hydroxyl ions and water molecules as ligands in addition to the main chain carbonyls or due to additional residues at B Ala 112 and B Lys 113 (Figure 5). Enlargements of the peaks around 12.0 kDa and 12.5 kDa are also pasted for clearer viewing.

Gel filtration chromatography

Gel permeation chromatography of the Colocasia lectin shows a single peak corresponding to a calculated molecular mass of 48.2 kDa at pH 7.2. There is an appreciable change in the position of the CEA peak at pH 3 relative to higher pH measurements and elution volume increases progressively from 10.17 ml to 11.65 ml in the range of pH 7.2 to 3 (Figure 2). However, at pH 5, no change in elution profile is observed with respect to pH 7.2. At pH 3.5, the presence of two peaks eluting, ~10.17 ml and ~11.65 ml are noticed. After lowering pH to 3, only the peak eluting ~11.65 ml is observed. This implies, upon lowering pH from 7.2 to 3, CEA converts from heterotetramer to heterodimer (Figure 2).

Previous high resolution lectin crystal structures

The family of monocot mannose-binding lectins has been described [2-4]. Among the few other high resolution (2 Å or better) protein structures of the same family preceding our study are the homodimeric *Polygonatum cyrtonema* agglutinin (PCL, 3A0C 28,647 reflns, 2.0 Å) [9] and homodimeric *Scilla campanulata* lectin (ScaMan, 1B2P, 33,837 reflns, 1.7 Å) [8]. Another high resolution structure was the heterodimeric *Crocus vernus* Dutch bulb lectin from the chitinase family (CVL, 3MEZ, 30,977 reflns, 1.94 Å); for this a preliminary paper was published [20].

Among previously solved crystal structures, *Remusatia vivipara* lectin (RVL, PDB ID: 3R0E, 25,784 reflns, 2.40 Å) [16] has the highest sequence identity of 93.8% with the *Colocasia esculenta* agglutinin. Other lectins like Scafet (32% or lower), PCL (45% or lower) and CVL

(45% and 50%) have lower sequence identity with CEA. This justifies our use of RVL as the search model for molecular replacement, though RVL is supposedly not a simple sugar-binding lectin.

Structure solution by molecular replacement

The crystal was primitive orthorhombic, with unit cell parameters a=122.01 Å, b=47.20 Å, c=82.25 Å and α=β=γ=90.00°. A listing of the axial reflections post-processing showed it belonged to $P222_1$, by noting the systematic absences. The website http://www.ruppweb.org predicts a 93.2% probability of having two dimers in the asymmetric unit, V_m=2.52 Å³/Da by submitting the cell parameters and an effective resolution of 1.6 Å, by comparing with protein structures of similar resolution. Most of the molecular replacement jobs were carried out with chains A and B only from 3R0E as the input model. A total of ~60 separate molecular replacement jobs were carried out within the program MOLREP, CCP4 package [15] by varying resolution ranges of data used for the rotation function and translation function, and different input models selected from entries 3R0E, 1KJ1 and 1MSA either using a heterodimer or a tetramer (dimer of a heterodimer) after deletion of the water molecules/ligands, but were judged to be ultimately unsuccessful as the final R-value using 32,328 reflections in the 8.0-2.0 Å range varied in the narrow range of 0.5363 to 0.5417 for five different rotation and translation function solutions found using MOLREP, with a common position of the first dimer and different positions of the second in the unit cell, after restrained refinement using REFMAC5 [15].

At this point, the input model was refined after making the necessary amino acid substitutions for changing the protein sequence
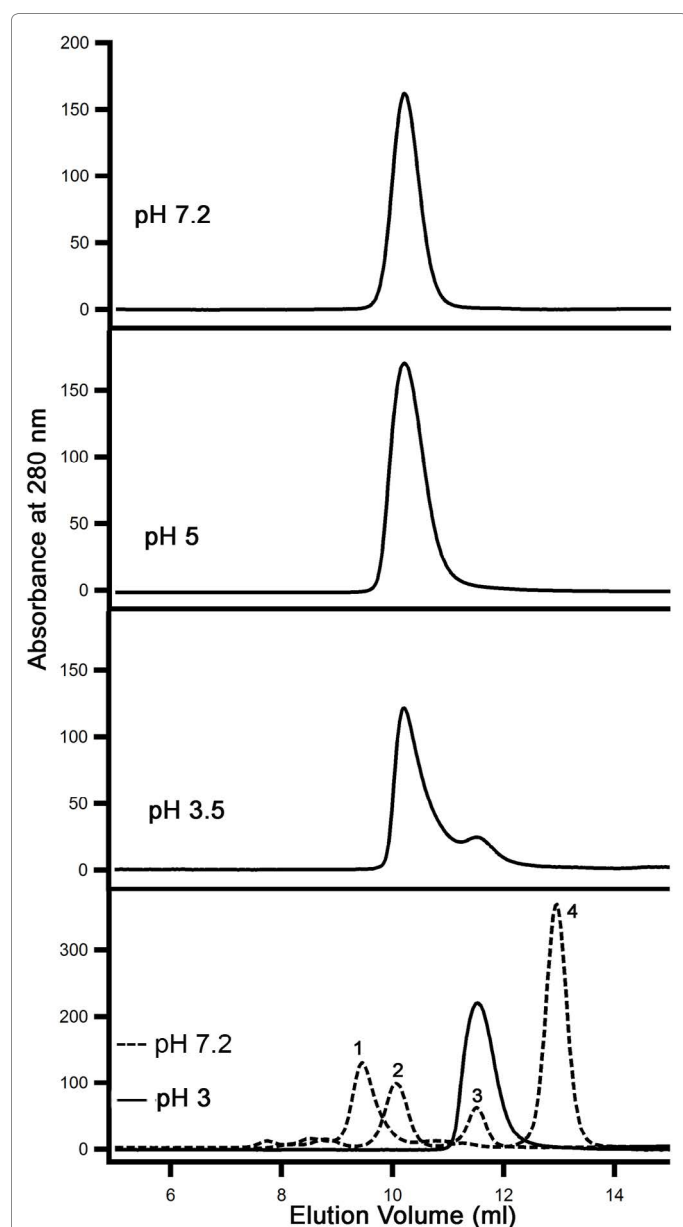
Figure 2: Size-exclusion chromatography of CEA at neutral and acidic pH. Gel permeation of the *Colocasia* lectin showed a single peak at elution volume of 10.17 ml at both pH 7.2 and pH 5. However, at pH 3.5, a second minor peak corresponding to a smaller molecular mass appears for CEA. The bottom panel superposes results from CEA (solid curve, pH 3) and various standard lectins of similar shape and known oligomerization status at pH 7.2 (dotted curves): *H. promita* lectin hexamer, 79 kDa (peak 1), *G. nivalis* lectin homotetramer, 52 kDa (peak 2), *N. pseudonarcissus* lectin homodimer, 26 kDa (peak 3), *P. aeruginosa* lectin monomer, 13.7 kDa (peak 4). Thus the higher molecular mass CEA peak is close to the GNA homotetramer (peak 2) and the lower molecular mass CEA peak is close to the *N. pseudonarcissus* lectin homodimer (peak 3). These are interpreted to be the $\alpha_2\beta_2$ heterotetramer and αβ heterodimer respectively.

from that in 3R0E to that of *Colocasia esculenta* agglutinin. Molecular replacement was run afresh using the Phaser MR package [15]. This run was successful and rigid body refinement within REFMAC5 reduced the R-factor to 0.4196 using 12.0-3.0 Å data and a F.O.M. of 0.5519; using 16.0-2.5 Å data the R-factor became 0.4381 and F.O.M. of 0.4380. This new solution seemed even more probable/trusted as both the dimers formed tetramers through their B chains (as in [16]) in the arrangement resulting from these rotations and translations within the

P222$_1$ unit cell (Figure 3). Only *one* of the two dimers had shown this property in the previous solutions found by MOLREP.

By inspecting the 2Fo–Fc and Fo–Fc maps at the end of the rigid body refinement, the N-termini of the A and C chains were repositioned into density, as they were sitting in negative density in the difference map for the trial model (Figure 4); the C-termini of the B and D chains were adjusted manually using real space refinement in Coot [15]. Several side chains were also manually positioned into densities visible after the rigid refinement stage. Near the beginning of the restrained refinement, using 9,776 reflections in the 16.0-3.0 Å range, the R-value reduced to 0.2946, F.O.M 0.6113 using positional refinement with a constant overall B-factor of 20 Å^2 within REFMAC5 and near the end, using 49,194 reflections in the 21.0-1.74 Å range R-value became 0.3441 and R-free remained at 0.4272. The Fo, Fc correlation coefficient is calculated to be 0.79 for the final model. PDB calculated the Wilson B of our dataset to be 12.40 Å^2 using data till 1.74 Å but that calculated in program Xtriage by the PDB yielded a Wilson B of 6.5 Å^2 using all data up to 1.536 Å.

### Deviation from the *Remusatia vivipara* lectin structure

Indeed, the structure of CEA is also therefore expected to be close to that of RVL, obtained from an ornamental monocotyledonous flowering plant (Araceae family) tuber. However, the RVL crystallized in a tetragonal space group P4$_1$ and its 2.4 Å structure included 3510 protein atoms and 129 water molecules [16].

The comparison of 5D5G with the RVL structure which was used as the input model in molecular replacement deserves special mention (Figure 4). On superposing residues A:1-109 and B:1-110 of CEA with those of 3R0E, including a total of 1,752 backbone atoms, the RMSD between the structures is 0.8945 Å. However, if we omit a few residues from the termini of chain A, including only A:3-106 and B:1-110 instead, the RMSD between the two structures in the 1,712 backbone atoms is only 0.5872 Å (Table 1 and Figure 4). This was actually apparent at the start of crystallographic refinement, as maps had been calculated after rigid body refinement and the termini of the A and C chains were sitting in negative densities in the Fo–Fc map.

Colocasia esculenta crystal structure

The average temperature factor of an atom including waters and ligands was 19.0 Å^2 for our submitted structure, as noted in the PDB file for 5D5G, but the backbone atoms like Cα have much lower average B values (Figures 5 and 6). These are much lower than those obtained for 3R0E [16], as shown.

Our crystal structure of mannose-free CEA also contains two independent heterodimers in the asymmetric unit, each forming a $\alpha_2\beta_2$ tetramer by associating with its symmetry mate within the crystal. Among the 3,577 non-hydrogen atoms in the asymmetric unit, were the protein atoms, one Hepes molecule, one sulphate ion, 5 magnesium ions and 134 waters located in the electron density maps (Figure 6). Crystallization conditions, details of X-ray data and its collection, refinement, characteristics of final model are described in some detail in the header of our PDB entry 5D5G. The structure factor files are also publicly available from the PDB. Whereas synchrotron X-ray data collected at the ESRF beamline gave the 2.4 Å dataset consisting of 25,784 reflections for RVL, our dataset was collected to 1.536 Å in a home X-ray source, though our data was used till 1.74 Å as completeness declined afterwards [21,22]. Though the R$_{merge}$ in intensities for our 1.536 Å dataset was on the high side (0.38) we decided to keep the high resolution reflections to 1.74 Å (R$_{merge}$=0.35) as this was to be a structure solution by molecular replacement, neither by multiple isomorphous
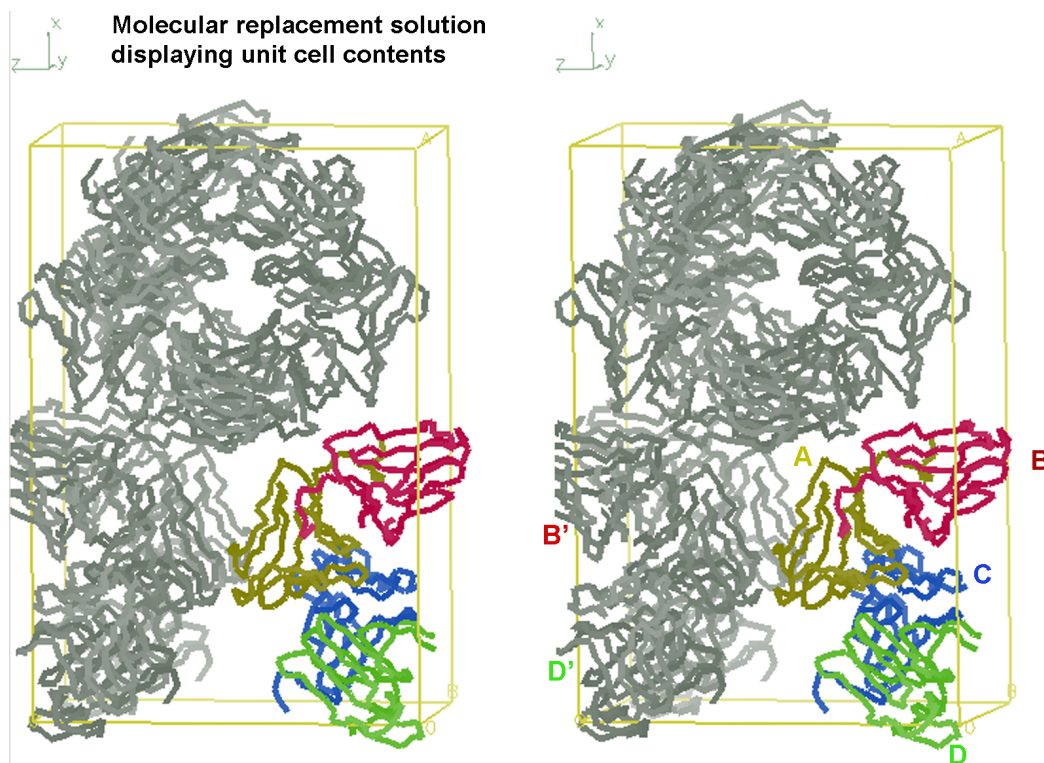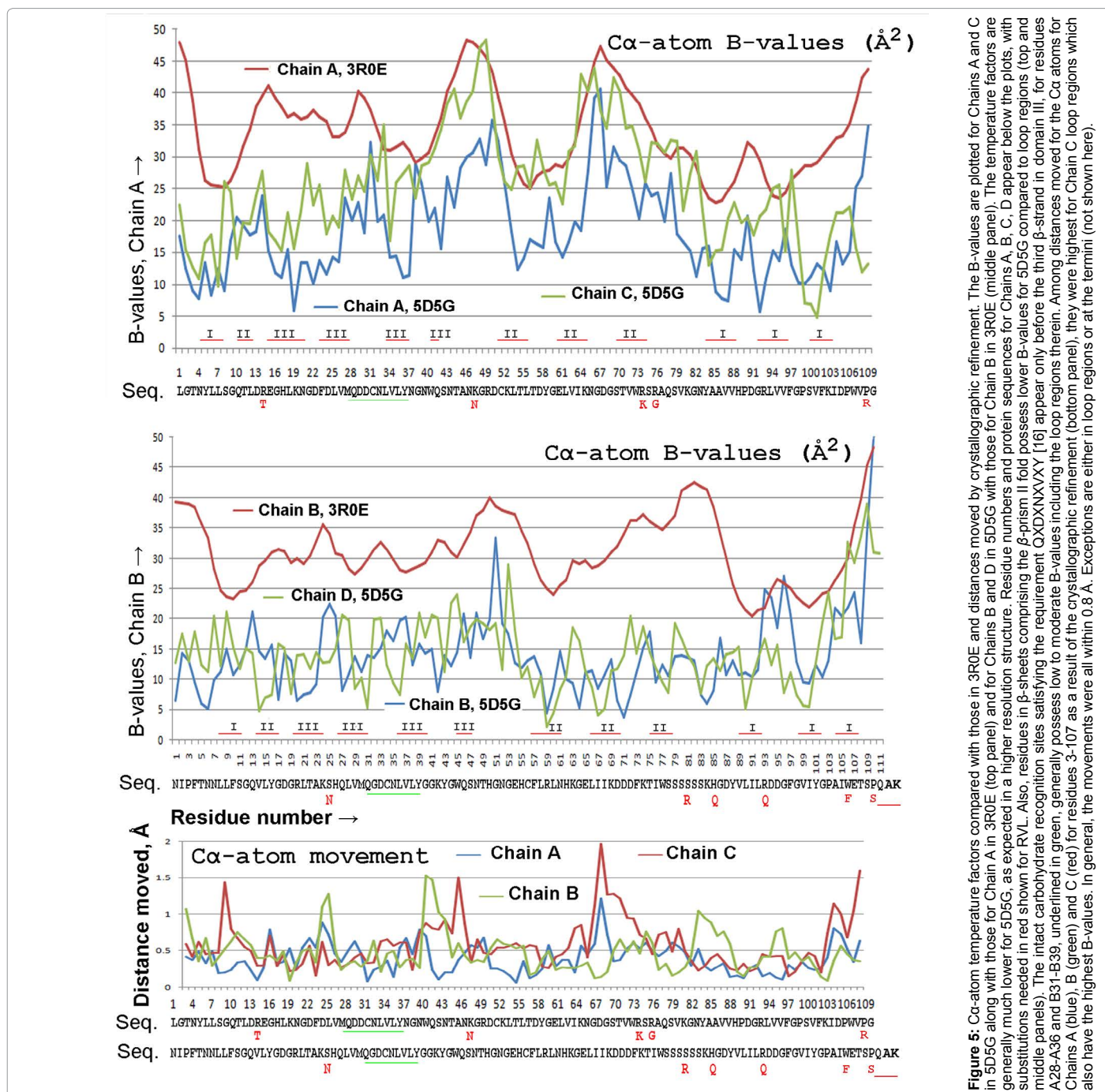
Figure 3: Packing diagram for correct molecular replacement solution obtained by Phaser MR Package [15], as Cα models in stereo. Chains A, B in yellow and red comprise one heterodimer and chains C, D in blue and green comprise the other and together they make the asymmetric unit in this P222$_1$ crystal; orientation of the three crystallographic axes indicated. There is an exact 2-fold about vertical axis **a**, relating the two parts of a tetramer separately for each heterodimer shown in color. Other symmetry related molecules filling up the unit cell are shown in grey. A copy of Chain B' related by a unit translation along **–c** pairs with Chain B and a copy of Chain D' related by the same unit translation pairs with Chain D, forming the tetramers as shown in Figure 9.
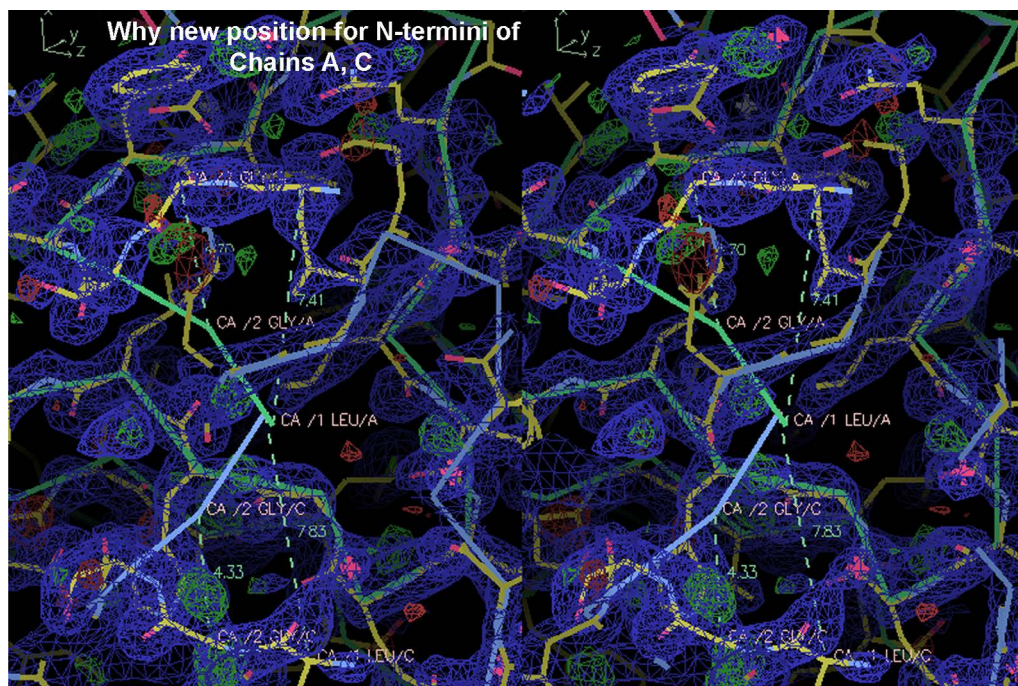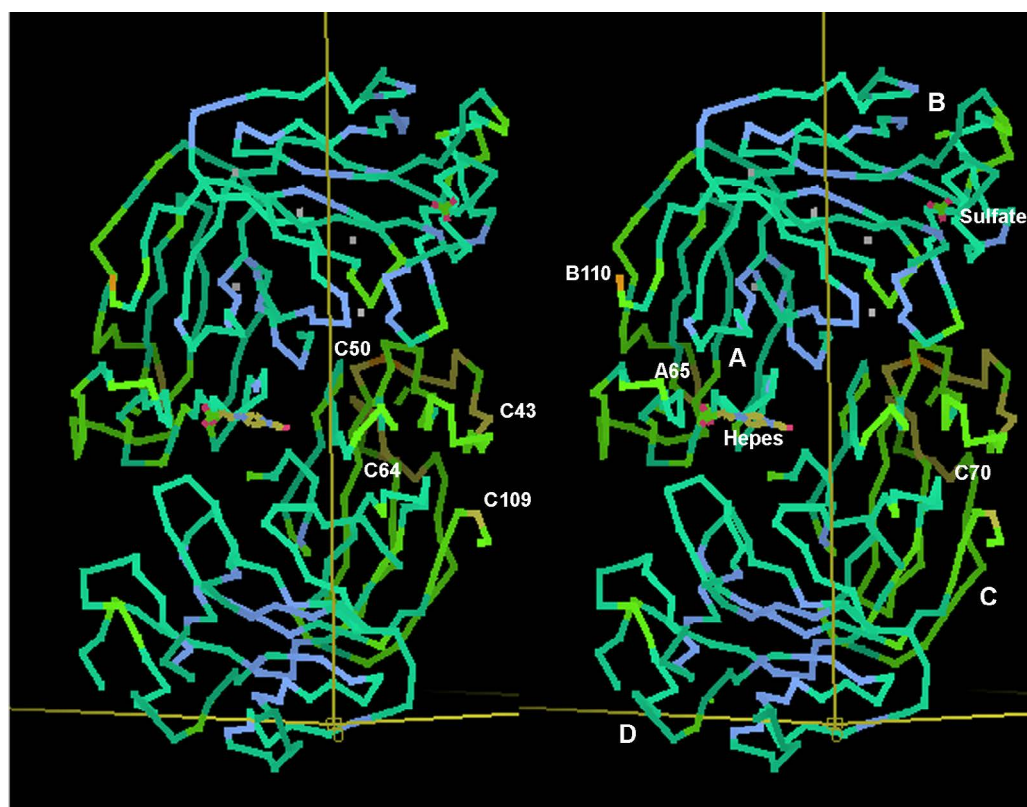

Figure 4: New positions for residues 1-3 in Chains A and C relative to those in 3R0E, shown in stereo with 2Fo–Fc map (blue). The largest differences between the CEA structure and that of RVL occur at the N-termini of the A- and C-chains. The molecular replacement solution shows Cα models of Chain A (green) and Chain C (light blue), placed in a clashing position. In fact, these N-terminal residues were also sitting in negative density after rigid body refinement. All non-hydrogen atoms in the refined model are superimposed, with carbon atoms in yellow, nitrogen in blue and oxygen in red; 2Fo–Fc map is displayed at 1.5σ level, and Fo–Fc map at 3σ. Distances in Å that these Cα atoms had to be moved to fit the X-ray data are displayed for both Chains A and C, between their original and final positions.

Regions from 5D5G used	Other structure, regions used respectively, reference	Atoms compared	RMSD
3MEZ or CVL [20]			
A3-A36, B45-B110	A2-A35, B48-B113 (remaining regions do not superpose well)	800	0.9310Å
1DLP or Scafet [27]			
A3-A8, A10-A37, A40-A46, A49-A102; B7-B11, B13-B40, B44-B108	A1-A5, A15-A42, A48-A54, A56-A109; A125-A129, A139-A166, A171-A235	1,544	1.1140Å
3A0C or PCL [9]			
A3-A21, A22-A36, A40-A105	A1-A19, A21-A35, A42-A107	800	1.0574Å
B8-B17, B19-B24, B26-B39, B51-B108	B3-B12, B14-B19, B22-B35, B48-B105	704	0.9567Å
When both the regions from A, B subunits are used to compare		1,504	1.6134Å
3R0E or RVL [16]			
A:1-109, B:1-110	A:1-109, B:1-110	1,752	0.8945Å
A:3-106, B:1-110	A:3-106, B:1-110	1,712	0.5872Å

Table 1: Comparison of A, B dimer in 5D5G with previous heterodimeric β-prism II lectin crystal structures, using backbone atoms only.

Figure 5: Cα-atom temperature factors compared with those in 3R0E and distances moved by crystallographic refinement. The B-values are plotted for Chains A and C in 5D5G along with those for Chain A in 3R0E (top panel) and for Chains B and D in 5D5G with those for Chain B in 3R0E (middle panel). The temperature factors are generally much lower for 5D5G, as expected in a higher resolution structure. Residue numbers and protein sequences for Chains A, B, C, D appear below the plots, with substitutions needed in red shown for RVL. Also, residues in β-sheets comprising the β-prism II fold possess lower B-values for 5D5G compared to loop regions (top and middle panels). The intact carbohydrate recognition sites satisfying the requirement QXDXNXVXY [16] appear only before the third β-strand in domain III, for residues A28-A36 and B31-B39, underlined in green, generally possess low to moderate B-values including the loop regions therein. Among distances moved for the Cα atoms for Chains A (blue), B (green) and C (red) for residues 3-107 as a result of the crystallographic refinement (bottom panel), they were highest for Chain C loop regions which also have the highest B-values. In general, the movements were all within 0.8 Å. Exceptions are either in loop regions or at the termini (not shown here).

Figure 6: Model in 5D5G color coded according to temperature factors. The asymmetric unit is again shown as a Cα model in stereo, color coded here according to the temperature factor at each residue, color-coded as follows: blue, less than 10 Å²; turquoise, 10 to 20 Å²; green, 20 to 30 Å²; yellow, 30 to 40 Å²; orange, 40 to 50 Å². The four domains A, B, C, D are labeled, as are the Hepes molecule, the sulfate ion and five Mg²⁺ as white dots. On average, chains B and D have the lowest, C has the highest and A intermediate in temperature factors. Also, regions comprising the dimeric or the tetrameric interfaces have the lowest temperature factors in all four chains. The tetrameric interfaces are all generated around the vertical **a** axis (along x) as it is a crystallographic 2-fold, hence domains B, D are closest to the viewer and A, C away. The highest B-values (>36 Å²) occur in labeled residues A66, A67, B110, region C43-C50, region C64-C70 and C109, seen in yellow/orange color, accompanied by weaker electron densities especially for their side chains; these are mostly regions which do not press against a neighboring molecule but are free to move in the disordered solvent in the crystal.

replacement nor anomalous methods. Since the input model in 3R0E was quite reliable for our molecular replacement solution, using data till 1.74 Å is justified in order to maintain a high data-to-parameter ratio of 49,194/(4 × 3,577) or roughly 3.5/1. We believe in the advice of keeping as much X-ray data as possible as long as the shells are high in completeness, though much of the data is weak, as this is beneficial regarding the accuracy of the structure [22]. As a result, our R value for the 49,194 reflections is 0.344, as opposed to 0.209 for RVL [16]. The refined structure in 5D5G has no bond length, bond angle, chirality or planarity outliers and has 0.9% Ramachandran outliers.

The mannose-free CEA structure solved using 49,194 reflections in the 20.99-1.74 Å resolution range contains four β-prism II domains, each with a β-barrel made by three subdomains of amphipathic anti-parallel β-sheets arranged around a pseudo 3-fold axis like three faces of an equilateral prism (Figures 3, 4 and 6). Electron density for residue Gln 111 was observed only in chain D, but it was not observed for chain B having the same sequence. Side chain atoms for which density could not be observed at all in the electron density map are given in our PDB file for 5D5G as "missing" atoms, particularly from A Arg 49, B Ala 98, C Leu 1, C Arg 49 and D Ala 98. It was particularly observed that some aliphatic side chains at the core of these β-prisms had less than unit occupancy, probably due to motion allowed inside the cores.

Cys 31, Cys 51 in the A, C subunits and Cys 34, Cys 56 in the B, D subunits form disulphide bridges. Cis peptide bonds exist at Gly 97-Pro 98 in the A, C subunits and at Gly 102-Pro 103 in the B, D subunits near the C-termini as found with RVL [16].

As previously observed in RVL and other β-prism II domains, the two domains (A, B or C, D) have their C-terminal β-strands hydrogen bonding with a β-sheet in their partner domain, called C-terminal exchange [23].

Presence of magnesium ions in CEA

Colocasia esculenta tubers were homogenized in 0.2 M NaCl, 20 mM CaCl₂, pH 9 adjusted with NaOH and finally adjusted to 1.5 M ammonium sulfate pH 6.5 with 50 mM sodium acetate used for the mannose-Sepharose B affinity column. Prior to crystallization, the CEA solution was made 20 mM in Tris pH 8, 150 mM NaCl. Thus the buffers used for purification or for crystal growth all lacked magnesium. Magnesium was detected in CEA by atomic absorption spectroscopy, confirming our finding of ordered magnesium in the electron density maps (Figure 7). Though our sample was tested for manganese also, the result was negative for that element.

It was found that for the five Mg²⁺ ions located in electron density map, both residues that contributed a carbonyl as a ligand to such an ion had to possess B-values in the 5-10 Å² range. Otherwise, the magnesium ions could not be observed. We looked near the same residues in subunits C (Leu 6) and D (Ala 85, Asn 6, Thr 22, Val 89, Ile 91, Phe 97, Val 99, Gly 102) but as those had B-values exceeding 10 Å²,
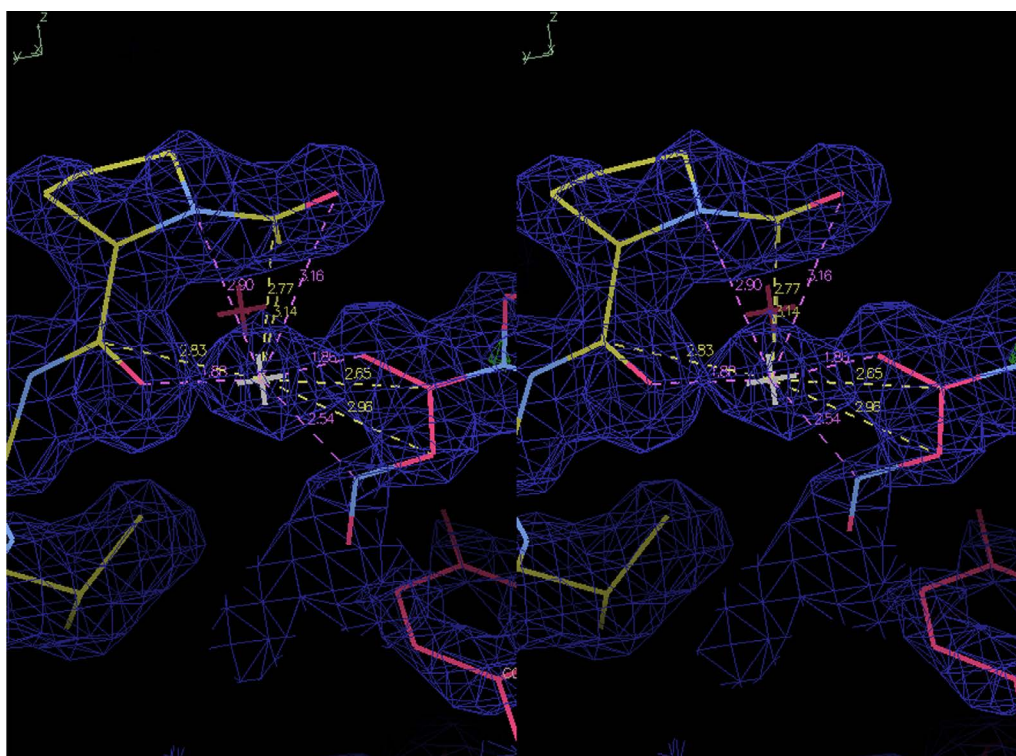
Figure 7: One of the five Mg^{2+} ions to be located at the final stages of refinement, bound between the carbonyls of A Pro 98 and B Gly 102, at distances of 1.88Å and 1.86Å from those oxygen atoms. These had between 5.14 σ to 5.86 σ levels in the Fo–Fc map. While this magnesium sits at the border of Chains A (carbons in yellow) and B, three others are embedded between pairs of carbonyl oxygens within Chain B and one between those in Chain A. Though Mg^{2+} and Na^{+} are isoelectronic with divalent oxygen, these difference densities cannot be explained by water molecules due to their proximity with the carbonyl oxygen atoms. Na^{+} has an average ligand distance of 2.46 (14) Å with its oxygen ligands as observed from an analysis of the protein structures in the PDB [28], while Mg^{2+} exhibits smaller metal-ligand distances. The presence of bound magnesium was also confirmed by atomic absorption spectroscopy.

the densities corresponding to equivalent magnesium ions could not be observed. However, a difference density indicated a Mg^{2+} ion with C Pro 98 (as with A Pro 98), but its other carbonyl partner was from D Ile 100 (instead of B Gly 102) as here, D Ile 100 has the lower B-value compared to D Gly 102.

Ordered sulfate in the crystal

A strong tetrahedral density sitting adjacent to B His 55 and B' His 62 whose side chains stack with each other, became apparent at the intermediate stages of our crystallographic refinement. Since there was no phosphate in the concentrated CEA stock solution (containing 20 mM Tris-HCl) used for crystallization and the protein solution eventually had at least ~0.5 M ammonium sulfate in the drop from which crystals had resulted, we interpreted this as a sulfate anion and this was incorporated in our refinement thenceforth. Figure 6 shows the position of this ordered sulfate in the unit cell and Figure 8 shows how it explains the electron density map and justifies its location by hydrogen bonding to the two histidine side chains near the tetrameric interface and a main chain amide in B Gly 17. Similarly, we located a molecule of Hepes, used as a buffer during crystal growth (Figure 6).

Tetramer and dimer formation

The presence of two distinct N-termini is a consequence of the original polypeptide chain folding into two β-prism II domains with two cleavages near the middle. The distance between the C-terminus at A Gly 109 and the N-terminus at B Asn 1 negates a possibility of the 7-residue linker remaining in a disordered form. The two domains or subunits A and B have 44% sequence identity between them. Using

PYMOL [24], Figure 9 shows a view and a close up highlighting specific interactions stabilizing the tetramer formation; Figure 10 shows the various interactions stabilizing the dimer between subunits A and B.

Analyses of αβ and ββ' interfaces in 5D5G

Using the PISA web server [19], both interfaces scored the maximum possible complex formation significance score of 1.00. For the interface between chain B and its symmetry mate B', 22 hydrogen bonds and 12 salt bridges were found (Figure 9); for that between chains A and B, 21 hydrogen bonds and 1 salt bridge were found. Similarly, for the interface between chains C and D, 29 hydrogen bonds and 1 salt bridge were found; for that between D and D', 20 hydrogen bonds and 10 salt bridges were found. In lieu of the numerous salt bridges seen for the B-B' or D-D' interfaces, the A-B and C-D interfaces having 1644.9 Å^2 and 1813.5 Å^2 interface areas respectively, are stabilized by interactions between numerous hydrophobic residues (Figure 10), characteristic of the β-prism II lectin structures [16].

Structural comparisons within the four chains in the asymmetric unit

As each of the four subunits in the asymmetric unit exists in a β-prism II structure, it is useful to record the similarity and dissimilarity among all possible pairs between the subunits A, B, C, D as shown in Table 2. These root mean square deviations were generally calculated for the pairs for the backbone atoms only, but when the sequence is identical, as for (A, C) and (B, D) pairs, it may also include the side chains. Thus the structural similarity is the highest between the B, D pair followed by that for the A, C pair, regardless of the residue ranges
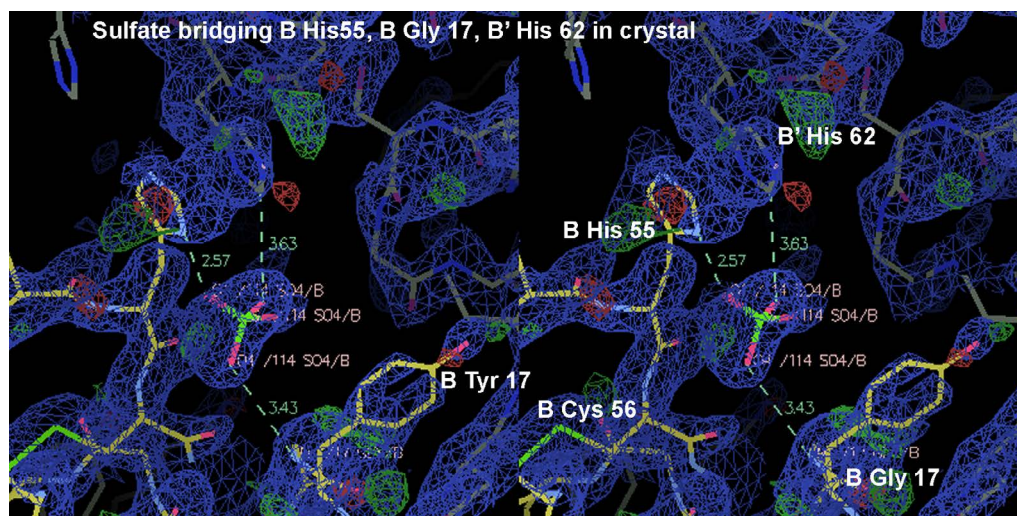
Figure 8: Sulfate bridging parts of Chain B and also Chain B'. The ordered sulfate ion seen in the 2Fo–Fc density contoured at 1.6 σ level, forming hydrogen bonds with nitrogen atoms in two histidine side chains belonging to two different subunits related by crystal symmetry and with the main chain amide of B Gly 17.

Pair used	Regions superposed in InsightII	No. atoms compared	RMSD
A, B	A2-A36 with B5-B39, A41-A102 with B46-B107	776	0.923 Å
A, C	A1-A109 with C1-C109 (backbone only)	872	0.993 Å
A, C	A3-A106 with C3-C106 (backbone only)	832	0.706 Å
A, C	A2-A109 with C2-C109 (with side chains)	1,664	1.287 Å
A, D	A2-A36 with D5-D39, A41-A102 with D46-D107	776	0.932 Å
B, C	B5-B39 with C2-C36, B46-B107 with C41-C102	776	0.924 Å
B, D	B1-B110 with D1-D110 (backbone only)	880	0.609 Å
B, D	B1-B110 with D1-D110 (with side chains)	1,734	1.002 Å
B, D	B1-B108 with D1-D108 (backbone only)	864	0.529 Å
B, D	B1-B108 with D1-D108 (with side chains)	1,708	0.953 Å
C, D	C2-C36 with D5-D39, C41-C102 with D46-D107	776	0.893 Å
(A, B), (C, D)	A3-A106 with C3-C106, B1-B108 with D1-D108	1,696	0.664 Å
(A, B), (C, D)	A3-A106 with C3-C106, B1-B108 with D1-D108 (with side chains)	3,326	1.061 Å

Table 2: RMSD between all possible pairs of CEA subunits in 5D5G.

compared, or whether side chains were included in the comparison (Table 2). The other pairs have dissimilar sequences, hence the greater values of RMSD are expected. Some regions were kept out of the superposition calculations as they show greater deviations and their inclusion will needlessly worsen the deviation observed.

### Pairwise comparisons with other lectin crystal structures

In the backbone comparisons with other β-prism II heterodimeric lectins, the RMSD is lowest for 3MEZ (*Crocus vernus* lectin or CVL) as there are only a few regions of overlap (Table 1); the rest of this pair do not superpose well, though CVL has 45% and 50% sequence identity in the two chains with CEA. Among the remaining two, overall RMSD is lower for the 3.3 Å structure 1DLP (heterodimeric *Scilla campanulata* agglutinin or Scafet) compared to that with 3A0C (*Polygonatum cyrtonema* agglutinin or PCL) (Table 1). Individual domains in 3A0C superpose better than the dimer as a whole, as PCL has higher sequence identity with CEA compared to Scafet, but the dimer does not, due to a rotation observed between the A and B subunits in 3A0C relative to 5D5G. All of these are however, less similar with 5D5G compared to the RVL structure.

The RMSD values obtained in all these cases are close to 1 Å,

around the same values obtained in Table 2 for chains having different sequences (like A, B or C, D or A, D or B, C), since the two dissimilar chains within CEA share 44% sequence identity between them, about the same level as with these other pairs of lectins.

Discussion

Gel filtration of the lectin suggested its presence in the $\alpha_2\beta_2$ form at neutral pH, eluting at the same place where the homotetrameric *G. nivalis* standard elutes (peak 2), in agreement with gel filtration data (Figure 2). Though our *Colocasia esculenta* crystal structure in 5D5G was the first such released by the PDB in September, 2015, two later studies on the similar lectins of other variants [26,27] did not mention it, their PDB entries released June and September, 2016, respectively. One of these discusses a CEA crystal structure solved at 2.1 Å refined with 22,646 reflections only to the level of the heterodimer [25], completely ignoring the fact that in their P3₁21 crystal, an almost identical heterotetramer is observed between αβ dimers related by symmetry, just as in our study. Despite data collection at an Argonne Synchrotron, the temperature factor average is 56.5 Å² for that structure. For our X-ray data, the Wilson B factor was calculated to be 12.4 Å² to 1.74 Å by our refinement or 6.5 Å² using data to 1.536 Å by the PDB. This lectin has 7 amino acid substitutions and a 2 residue deletion relative to ours
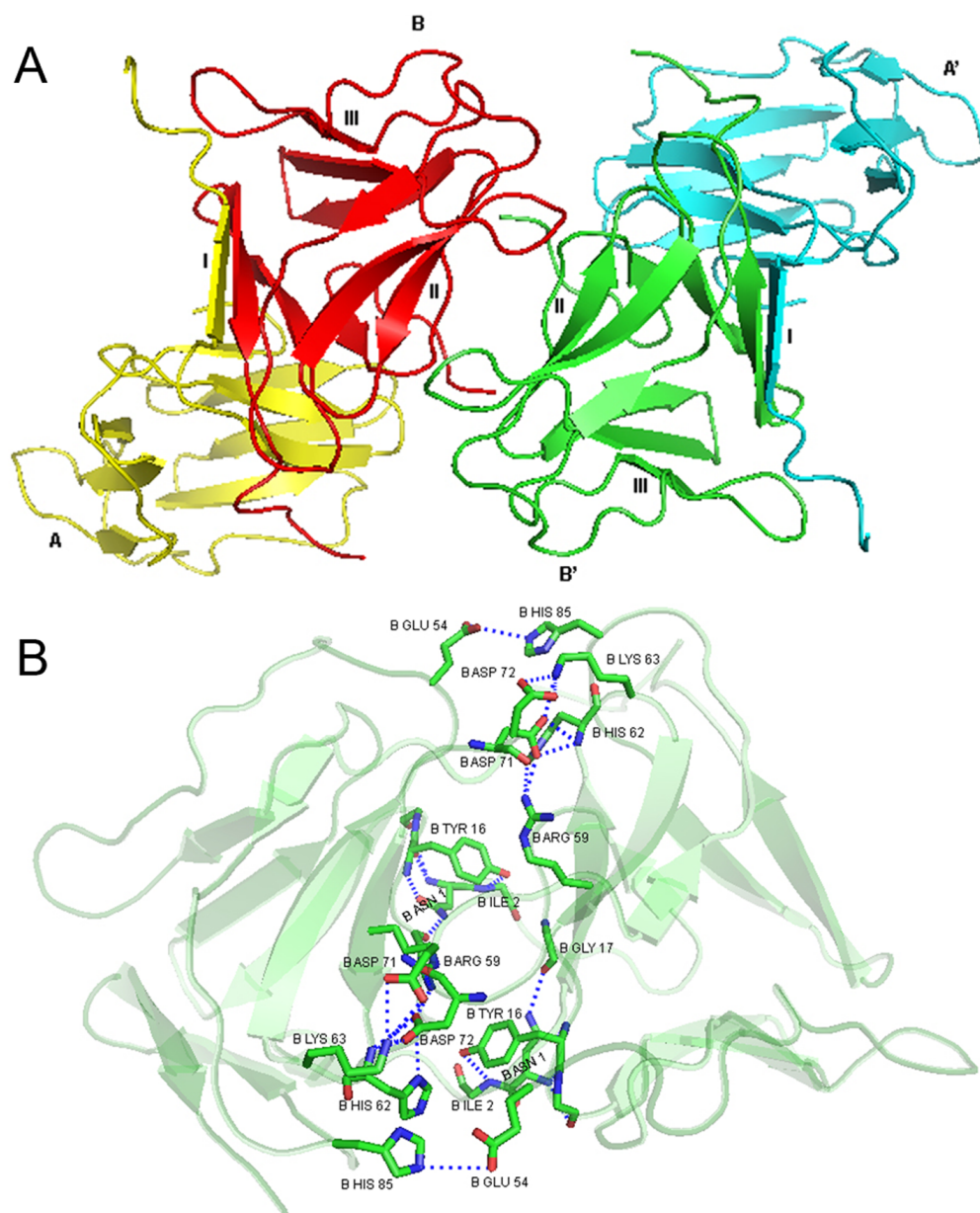
Figure 9: Tetramer formation within the CEA crystal between two adjacent copies of the AB dimer using subunit B. Subunit A is in yellow, subunit B is in red. In the second copy of the AB dimer, subunit A' is in light blue and subunit B' is in light green. (A) In the top half, we are looking down the crystallographic **a** axis, a perfect 2-fold, passing through the centre in this diagram, between the red and green molecules. (B) In the bottom half, displaying both hydrogen bonds/salt bridges and hydrophobic interactions important in the dimer-dimer interface using two copies of subunit B, looking down the **a** axis. Here the two copies of Sheet II are facing each other across a crystallographic 2 fold. The crystallographically independent molecules C and D within the asymmetric unit also form an identical tetramer with a symmetry-related C' and D', though not shown here. Figure generated using PYMOL [24].

(positions 13, 75, 82, 87 in Chain A; positions 25, 85, 98 in Chain B, lack of residues B112-113) [25].

Yet another recent study describes the CEA lectin crystallized in P2₁ space group diffracting to 1.72 Å and its trimannoside-bound complex in P1 space group diffracting to 1.91 Å [26]. In this study, the X-ray data was also collected at an Argonne Synchrotron, and has better data statistics, in the R_{sym} of 6.7% or 8.5% [26]. For the native crystal, the cell parameters are also quite close to those obtained for our crystal in P222₁ space group. According to Figure 1D in that paper [26], we have solved the structure of a slightly modified CESP here (6 substitutions relative to our sequence), whereas they have TarinA and TarinB. There are a total of 11 amino acid substitutions and a deletion between our protein sequence and that of the tarin (positions 13, 14, 15, 47, 77, 87 in Chain A; positions 6, 25, 80, 85, 110 in Chain B, lack of residues B111-113). So RVL is equally close to our CEA lectin with 11 substitutions and the same deletion (Figure 5). However, tarin also exhibits ββ' interactions [26], like RVL [16] and our CEA lectin.

Our crystal structure shows there are several salt bridges between side-chains in the ββ' interface (Figure 9). Upon reaching pH 3, several of these, e.g. B Glu 54, B Asp 71, B Asp 72 are protonated from both the participating subunits, hence at least 6 of the salt bridges are broken, thereby explaining the observed transition from heterotetramer to
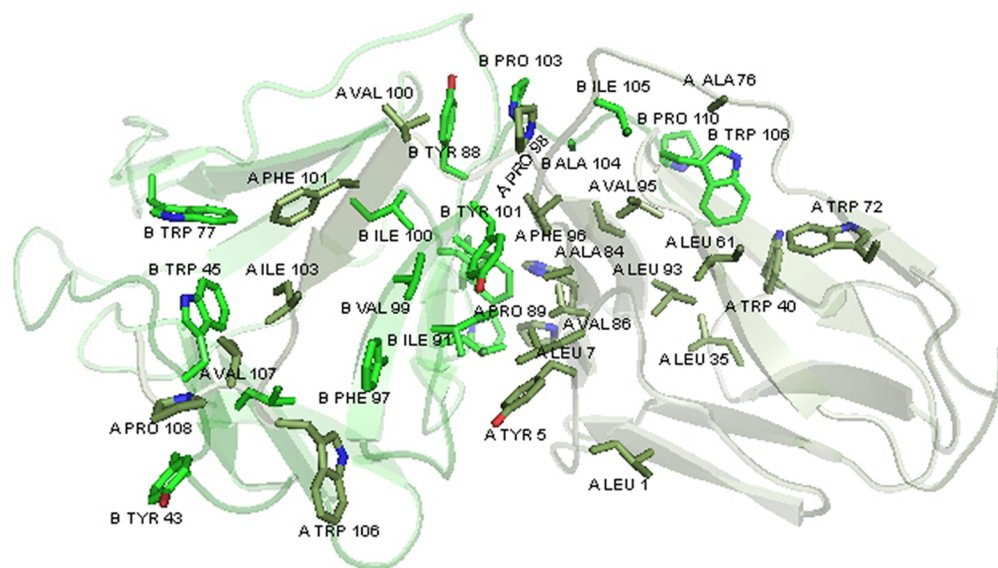
Figure 10: Dimer formation between subunits A and B seen in the CEA crystal is shown above, with the top half emphasizing the hydrophobic residues important in that interface, displayed as stick models. Subunit A is mainly to the right, in dark green or grey color; subunit B is mainly to the left in green. The sheet I of subunit A faces and interacts with the sheet I of subunit B, and this is the stabilizing interaction between the two subunits. As in other lectins in this category, the C-terminal tail of each subunit (top right side for subunit B, lower left for subunit A) crosses over to the other side and hydrogen bonds with the other subunit, thereby completing sheet I and spreading the volume over which the two subunits interact [23]. Oxygen atoms are in red and nitrogen atoms in blue displayed for some residues in this interface. Numerous hydrophobic side chains are involved in this interaction, most of them marked here. There is a similar interface between independent subunits C and D.

heterodimer. The pH dependence on the stability of CEA as seen in our gel filtration chromatography (Figure 2) suggested that the electrostatic interactions in the salt bridges make a significant contribution to the conformational stability of the heterotetramer.

The classical carbohydrate recognition site satisfying the sequence requirement QXDXNXVXY [16] appears only once in each of Chains A and B, both located on domain III (Figure 5, green underline). These are at 28-QDDCNLVLY-36 in Chain A and at 31-QGDCNLVLY-39 in Chain B, just as with RVL [16]. In both Chains A and B, domain I is involved in the formation of the dimeric interface and domain II from two B chains are involved in the formation of the tetrameric interface (Figures 9 and 10). So domain III containing the carbohydrate recognition site is not directly involved in these interfaces.

Legume lectins are known to contain Mn^{2+} and Ca^{2+}, but there is probably no previous report of Mg^{2+} in plant lectins. However, we did not test for lectin activity as a function of increasing amounts of EDTA concentration. The previous study on tarin showed its activity does not depend on metal ions [13]. However the existence of several isoforms with a heterogeneous population of different covalently attached glycans [13], usually does not yield such highly resolution X-ray diffraction as reported by almost the same authors [26]. Their purification scheme [26] shows they used the natural tarin for crystallography.

Since the heterodimeric *Scilla campanulata* agglutinin or Scafet diffracted only to 3.3 Å [27], the CEA becomes the first heterodimeric structure to be studied at such a high resolution, the comparable resolution being for *Polygonatum cyrtonema* agglutinin or PCL [9] and the tarin [26]. Our biophysical study correlating with the crystal structure described here is expected to be published shortly.

Acknowledgement

Funding

References

1. Lis H, Sharon N (1991) Lectin-carbohydrate interactions. Curr Opin Struct Biol 1: 741-749.

2. Van Damme EJM, Lanoo N, Peumans WJ (2008) Plant lectins. Adv Botanical Res 48: 107-209.

3. Biswas H, Chattopadhyaya R (2014) Thermal, chemical and pH induced unfolding of turmeric root lectin: Modes of denaturation. PLoS ONE 9: e103579.

4. Barre A, Van Damme EJM, Peumans WJ, Rougé P (1996) Structure-function relationship of monocot mannose-binding lectins. Plant Physiol 112: 1531-1540.

5. Hester G, Kaku H, Goldstein IJ, Wright CS (1995) Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. Nature Struct Mol Biol 2: 472-479.

6. Chandra NR, Ramachandraiah G, Bacchawat K, Dam TK, Surolia A (1999) Crystal structure of a dimeric mannose-specific agglutinin from garlic: Quaternary association and carbohydrate specificity. J Mol Biol 285: 1157-1168.

7. Sauerborn MK, Wright LM, Reynolds CD, Grossmann JG, Rizkallah PJ (1999) Insights into carbohydrate recognition by *Narcissus pseudonarcissus* lectin: The crystal structure at 2 Å resolution in complex with alpha1-3 mannobiose. J Mol Biol 290: 185-199.

8. Wood SD, Wright LM, Reynolds CD, Rizakallah PJ, Allen AK, et al. (1999) Structure of the native (unligated) mannose-specific bulb lectin from *Scilla campanulata* (bluebell) at 1.7 Å resolution. Acta Crystallogr D Biol Crystallogr 55: 1262-1272.

9. Ding J, Bao J, Zhu D, Zhang Y, Wang DC (2010) Crystal structures of a novel anti-HIV mannose-binding lectin from *Polygonatum cyrtonema* Hua with unique ligand-binding property and super-structure. J Struct Biol 171: 309-317.

10. de Castro LA, Carneiro M, Neshich Dde C, de Paiva GR (1992) Spatial and temporal gene expression patterns occur during corm development. Plant Cell 4: 1549-1559.

11. Prajapati R, Kalariya M, Umbarkar R, Parmar S, Sheth N (2011) Colocasia esculenta: A potent indigenous plant. Int J Nutr Pharmacol Neurol Dis 1: 90-96.

12. Das A, Roy A, Hess D, Das S (2013) Characterization of a highly potent insecticidal lectin from *Colocasia esculenta* tuber and cloning of its coding sequence. Am J Plant Sci 4: 408-416.

13. Pereira PR, Winter HC, Vericimo MA, Meagher JL, Stuckey JA (2015) Structural analysis and binding properties of isoforms of tarin, the GNA-related lectin from *Colocasia esculenta*. Biochim Biophys. Acta 1854: 20-30.

14. Van Damme EJM, Goossens K, Smeets K, Leuven FV, Verheart P, et al. (1995) The major tuber storage protein of Araceae species is a lectin. Plant Physiol 107: 1147-1158.

15. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50: 760-763.

16. Shetty KN, Bhat GG, Inamdar SR, Swamy BM, Suguna K (2012) Crystal structure of a β-prism II lectin from *Remusatia vivipara*. Glycobiology 22: 56-69.

17. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240-255.

18. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213-221.

19. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774-797.

20. Akrem A, Meyer A, Perbandt M, Voelter W, Buck F (2011) Isolation, purification, crystallization and preliminary crystallographic studies of a chitinase from *Crocus vernus*. Acta Crystallogr F67: 340-343.

21. Wlodawer A, Minor W, Dauter Z, Jaskolski M (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J 275: 1-21.

22. Wlodawer A, Minor W, Dauter Z, Jskolski M (2013) Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. FEBS J. 280: 5705-5736.

23. Liu W, Yang N, Ding J, Huang R, Hu Z (2005) Structural mechanism governing the quaternary organization of monocot mannose-binding lectin revealed by the novel monomeric structure of an orchid lectin. J Biol Chem 280: 14865-14876.

24. DeLano WL, Bromberg S (2004) The PyMOL molecular graphics system. DeLano Scientific SSC, San Carlos, California, USA.

25. Vajravijayan S, Pletnev S, Pletnev VZ, Nandhagopal N, Gunasekaran K (2016) Structural analysis of β-prism lectin from *Colocasia esculenta* (L.) S chott. Int J Biol Macromol 91: 518-523.

26. Pereira PR, Meagher JL, Winter HC, Goldstein IJ, et al. (2017) High-resolution crystal structures of *Colocasia esculenta* tarin lectin. Glycobiology 27: 50-56.

27. Wright LM, Reynolds CD, Rizakallah PJ, Allen AK, Van Damme EJ, et al. (2000) Structural characterization of the native fetuin-binding protein of *Scilla campanulata* Agglutinin: A novel two-domain lectin. FEBS Lett 468: 19-22.

28. Zheng H, Chruszcz M, Lasota P, Lebioda L, Minor W (2008) Data mining of metal ion environments present in protein structures. J Inorg Biochem 102: 1765-1776.