

Mannose in Complex with *Colocasia esculenta* Tuber Agglutinin

Rajagopal Chattopadhyaya*, Avisek Mondal

Department of Biochemistry, Bose Institute, Calcutta 700054, India

*Corresponding author: Rajagopal Chattopadhyaya, Department of Biochemistry, Bose Institute, Calcutta, 700054, India, Tel: 91-3323214568; Fax: 91-3323553886; E-mail: rchatto2001@yahoo.co.in

Received date: May 10, 2018; Accepted date: June 07, 2018; Published date: June 14, 2018

Copyright: © 2018 Chattopadhyaya R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The major tuber storage protein of *Colocasia esculenta* is a monocot mannose-binding, widely used, dietary lectin, for which a crystal structure was previously shown to consist of four β -prism II domains or two $\alpha\beta$ heterodimers, each forming a $\alpha_2\beta_2$ heterotetramer with a symmetry related unit. In the present study, the same lectin crystallized in the presence of a 50 molar excess of free mannose, persists in forming such $\alpha_2\beta_2$ heterotetramers. Lectin crystals apparently complexed to mannose 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. Using the crystal structure of the *Remusatia vivipara* lectin, the structure of the complex has been solved by molecular replacement and subsequent crystallographic refinement. Five different mannose binding sites per heterodimer with partial occupancy have been found, among which only one appears to bind close to a classical carbohydrate recognition site characterized earlier for these lectins. Mannose binding at these five sites is described.

Keywords: Monocot mannose-binding lectins; β -prism II fold; Edible *Colocasia esculenta* tuber agglutinin; Crystal structure with mannose

Abbreviations: CEA: *Colocasia esculenta* agglutinin; FOM: Figure of Merit; RVL: *Remusatia vivipara* lectin

Introduction

The recognition of carbohydrate moieties by lectins, which are proteins of non-immune origin found in most living cells, has important applications in a number of biological processes such as cell-cell interaction, signal transduction, cell growth and differentiation [1].

Among the twelve families of plant lectins [2], the monocot mannose-binding lectin family or GNA-related lectins possess an exclusive specificity towards mannose. 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. [3]. By analyzing the binding characteristics and features in the protein molecules capable of binding mannose non-covalently, both in animals and bulb lectins in plants, the latter family was found to contain a consensus sequence motif QXDXNXVXY responsible for such recognition [4]. As mannose is an abundant cell surface monosaccharide involved in a variety of cellular interactions, it is of some interest to understand the features responsible for such recognition [3,4].

The β -prism II fold was first reported in the homotetrameric *Galanthus nivalis* (snowdrop) 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* (garlic) lectin with α -D-mannose (1BWU) [6], homotetrameric *Narcissus pseudonarcissus* (daffodil) lectin with α 1-3 Mannobiose (1NPL) [7], unligated homodimeric *Scilla campanulata* (Spanish blue-bells) lectin (1B2P) [8] and homodimeric *Polygonatum*

cyrtonema Hua unligated or with methyl- α -D-mannoside or α 1-3 Mannobiose (3A0C, 3A0D, 3A0E) [9] all show the β -prism II fold for each of their subunits, of about 110 residues. A mannopentaose complex in two binding modes was also reported quite early with the snowdrop or *Galanthus nivalis* lectin [10]. Thus for the monocot mannose-binding lectin family, a majority of the lectins could be crystallized with modified mannose or mannose dimer or mannopentaose [5,7,9,10], while mannose was bound only to *Allium sativum* lectin [6].

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 [11]. Its corm extracts possess important pharmacological properties including anti-inflammatory, anti-cancer, anti-fungal, anti-viral [12], while the lectin [13] has insecticidal activities [14]. Another group reported several isoforms of the very similar lectin tarin and its covalent modification [15], and more recently high resolution crystal structures of the tarin lectin alone and in complex with a trimannoside, Man α (1,3)Man α (1,6)Man [16]. The first *Colocasia esculenta* tuber agglutinin crystal apo structure released by the PDB was also described in detail in a recent publication [17] and related to its observed solution properties using gel filtration, dynamic light scattering, circular dichroism, fluorescence emission as a function of pH, temperature and denaturant concentration [18].

It was shown that the classical carbohydrate recognition site [4] satisfying the sequence QXDXNXVXY occurs only once in each of chains A and B of the *Colocasia esculenta* lectin [17], while it occurs thrice in each chain of the lectins in snowdrop, garlic and daffodil [5-7]. The classical site occurs twice in each chain of *Scilla campanulata* [8] and *Polygonatum cyrtonema* [9] lectins, as one of the three sites is modified. Our crystal structure of the mannose-free *Colocasia esculenta* lectin showed that each subunit possesses a 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 a β -prism II (BP2) fold [17], common for all these GNA-

related lectins. The *Colocasia esculenta* lectin is very close in sequence to the *Remusatia vivipara* lectin of the *Araceae* family, both having the classical site only in sheet III [17,19]. For the *Remusatia vivipara* lectin it was speculated, 'It is possible that the binding of mannose moiety is weak because of steric hindrance but binding to higher oligosaccharides may be facilitated by further interactions with other amino acids of the protein(s)' [19].

Mannose commonly exists in the major pyranose (six-membered ring) form and the minor furanose (five-membered ring) form. Each ring closure can have either an α or β configuration at the anomeric position, C1. The molecule rapidly undergoes isomerization among these four forms, the furanose forms estimated to be less than 1% in population, while the pyranose forms are estimated as $\sim 66\%$ in α and $\sim 33\%$ in β configuration (Wikipedia). To fix mannose in the α pyranose form, researchers have employed methyl- α -D-mannoside instead of mannose, as the mannoside lacks the β anomer [5,9]. Though the electron densities remain less definitive for mannose used in our study, it is found from our 1.85 Å crystal structure of the complex that there appears to be at least five different sites per heterodimer of the *Colocasia esculenta* lectin, with occupancies in the 0.6 to 0.9 range.

Materials and Method

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. Crystallization reagents, called Crystal Screen™, in particular HR2-110, were obtained from Hampton Research (Aliso Viejo, California, USA).

Protein purification

CEA was purified to homogeneity following a modification of the known protocol [13]. The tubers were homogenized in 0.2 M NaCl containing 1 g L^{-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 mins.). After it was brought to 20 mM in CaCl_2 , the pH was adjusted to 9.0 (with 1N NaOH), kept overnight in the cold, and re-centrifuged at 12,000 rpm for 10 mins. Then it was adjusted to pH 4.0 (with 1 N HCl) and re-centrifuged at 12,000 rpm for 10 mins. 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 mins. 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 $\text{M}^{-1} \text{cm}^{-1}$. Hence all concentrations mentioned here mean the heterodimeric concentration. Finally the pooled fractions were collected together and concentrated by ultrafiltration using Vivaspin Turbo 15 units through a 3,000 molecular weight cutoff membrane (Sartorius Stedim Biotech, Goettingen, Germany).

Crystallization and X-ray diffraction data collection

All 50 conditions within the HR2-110 reagent formulation of Hampton Research were set up with 2 μL of protein stock containing 8.10 mg/mL of purified CEA in 20 mM Tris pH 8.5 and a 50 molar excess of D-mannose, mixed with 1 μL of the crystallizing agent, as hanging drops. The best crystals were obtained by hanging-drop, vapor-diffusion method at room temperature, in about 3 to 4 weeks using 0.1 M sodium citrate tribasic dihydrate pH 5.6 and 1 M ammonium phosphate monobasic as the crystallizing agent, the growth carried out at 298 K. Data collections at cryogenic temperature (-162°C) were carried out at the home X-ray source. For data collection, the crystals were soaked in the cryo-protectant solution (70% of 0.1M sodium citrate tribasic dihydrate pH 5.6 and 1 M ammonium phosphate monobasic mixed with 30% ethylene glycol, v/v). The collected data were indexed, processed and scaled using Crystal Clear™ (Version 2.0) and the implemented program d*TREK®.

Molecular replacement, refinement and locating the bound mannose molecules

During processing of the diffraction data with the program d*TREK the preliminary analysis of the frames had suggested either trigonal or orthorhombic or monoclinic as the various possibilities, the unit cell being twice as large for orthorhombic/monoclinic systems compared to that in trigonal. Regarding symmetries observed during processing, Laue group 3m signified both $P3_121$ (152) and $P3_221$ (154), whereas Laue group 3 signified both $P3_1$ (144) and $P3_2$ (145) as possible space groups. Due to the uncertainties about the correct space group within the trigonal system, the raw diffraction data was processed in monoclinic space group C2 having the lowest symmetry; then molecular replacement was easily solved with three heterodimers in the asymmetric unit (or 12 heterodimers in the C2 unit cell). The R-factor obtained in C2 was 0.32 using 8.0-3.0 Å data after carrying out rigid refinement. In this solution, the heterodimers were seen to form right handed helices around an axis parallel to c.

As the reflections along 00l showed extinction unless $l=3n$, the actual space group had to be among the trigonal ones; orthorhombic/monoclinic do not possess this index requirement. Hence, trigonal space groups 145 and 154 associated with left handed helices could be ruled out, as the correct molecular replacement solution showed right handed helices. Among the two other trigonal space groups 152 and 144, the latter was rejected as the molecular replacement solution obtained for C2 seemed to mimic space group 152 with one heterodimer in the asymmetric unit and six such within the unit cell.

The structure solution and analyses were carried out using various modules of CCP4i; the structures were solved using molecular replacement [20] using the crystal structure of the two domain RVL (PDB ID: 3R0E) [19] having 93.8% sequence identity to CEA, using only chains A and B. In the correct space group, the molecular replacement was straightforward with only one heterodimer in the asymmetric unit as explained above, the symmetry related units forming a right handed helix about c. It was found to mimic the same

solution found in space group C2 as described above. After applying the correct rotation and translation, the model was modified using the molecular graphics program Coot within the CCP4i package, substituting the side chains for the residues differing between CEA and RVL. After rigid refinement using 8,441 reflections in the 16.0-3.0 Å range and 2% additional 'free' reflections, the R-value was 0.3514 and R_{free} was 0.3172. Restrained refinement was next, carried out mostly using Refmac5 [21] also within the CCP4i package, switching to PHENIX 1.8.1 [22] at the very end. At the end of several cycles of Refmac5 the R-value was 0.3630 for 34,494 reflections in the 16.0-1.85 Å range, F.O.M. 0.560, correlation coefficient of F_o , F_c being 0.886, the mean B of the structure being 32.8 Å². At this point the Fo-Fc map was examined in detail and side chain occupancies of both chains A and B manually adjusted for about 20% of the residues. This lowered the R-value to 0.3540 and F.O.M. increased to 0.5865 for the same number of reflections.

The program Coot [20] was used for monitoring the progress of the crystallographic refinement as well as for the display of models superposed with electron density maps. Examining both 2Fo-Fc and Fo-Fc maps, several blobs of unexplained density were identified at 6.13σ, 4.69σ, 4.59σ and 4.55σ in the latter, too large in size to be explained as water molecules or water clusters. Of these, the 6.13σ blob, not as large as the others in volume, was tentatively identified as a phosphate group with partial occupancy. Mannose molecules were fitted in the other large blobs of density before the introduction of any water molecules in stages, with three, four and five mannoses per asymmetric unit. The 33 water molecules were introduced at the end satisfying hydrogen bonding criteria.

Results

Structure solution by molecular replacement

The crystal was trigonal, in space group P3121 with unit cell parameters $a=b=75.896$ Å, $c=124.065$ Å and $\alpha=\beta=90^\circ$, $\gamma=120^\circ$. Other crystallographic details and refinement statistics are presented in Table 1. It is seen from this table that although data were collected to 1.7 Å resolution, data till 1.85 Å were actually used in our refinement, as completeness declines afterwards [23,24]. Though the R_{merge} in intensities for our 1.7 Å dataset was on the high side (0.313), we decided to use all reflections till 1.85 Å as this turned out to be a structure solution by molecular replacement instead of multiple isomorphous replacement or anomalous methods. As with the 1.74 Å mannose-free CEA crystal structure [17], the input model in 3R0E was quite reliable for our molecular replacement, and use of 35,187 unique reflections allows us to maintain a high data-to-parameter ratio of 35,187/(4×1,824) or roughly 4.82/1. The crystal has a solvent content of 70.89% or 4.23 Å³/Dalton as the Matthews parameter. It has been suggested that weak data should be kept for accuracy of the structure [24]. As a result, our R value for the 35,187 reflections is 0.3162, as opposed to 0.209 for RVL [19] for which Synchrotron X-ray data at the ESRF beamline was used, having a much lower value of R_{merge} . The refined structure in 5D9Z has no bond length, bond angle, chirality or planarity outliers and has 2.3% Ramachandran outliers using our home X-ray source data (Table 1), notably at A3 Thr, A46 Asn, A68 Gly, B24 Lys and B111 Gln.

Deviation from the *Remusatia vivipara* lectin structure

Indeed, the structure of CEA in 5D9Z is also therefore expected to be close to that of RVL (the input model), obtained from an

ornamental monocotyledonous flowering plant (*Araceae* family) tuber. Maps had been calculated after rigid body refinement and the terminal residues of the A chain were sitting in negative densities in the Fo-Fc map, as found also for the mannose-free CEA [17]. During refinement, the positions of A1 Leu and A2 Gly had to be moved 6.18 Å and 3.77 Å away from their starting positions, otherwise the maximum movement in Ca atoms was always <1.0 Å, and often within 0.5 Å.

	Space Group	P3 ₁ 21
	Mosaicity	2.21°
	Unit Cell a, b, c	75.896 Å, 75.896 Å, 124.065 Å
	α, β, γ (°)	90.00, 90.00, 120.00
	Wavelength	1.5419 Å
	Resolution range (Å)	29.04-1.70 (1.76-1.70) ^a
Data collection	Total reflections	3,60,689
	Unique reflections	39,501
	Average redundancy	9.13 (3.62)
	Completeness (%) ^d	85.6 (24.7)
	R_{merge} ^b	0.313
	Output $\langle I/\sigma \rangle^c$	4 (1.3)
	Reduced Chi Squared	1.17
	Resolution range	19.80-1.85 Å
	R-factor	0.3162
	R-free	0.3699
	Protein atoms	3,577
	Water molecules	33
	Unique reflns used	35,187
Refinement	Bonds=0.011 Å	
	Angles=1.538°	
	R.m.s.d.	Chirality=0.086 Å ³
		Planarity=0.005 Å
		Dihedrals=15.366°
	Ramach. Outliers	2.30%
	All atom clash score	15

Table 1: Crystallography data collection and refinement statistics.

Colocasia esculenta crystal structure in complex with mannose

Our crystal structure of mannose-bound CEA contains one heterodimer in the asymmetric unit, forming a $\alpha_2\beta_2$ tetramer by associating with its symmetry mate within the crystal (Figure 1). Among the 1,824 non-hydrogen atoms in the asymmetric unit, were 1,716 belonging to the lectin, 60 atoms in mannose, 4 in one phosphate ion and 33 in waters located in the electron density maps.

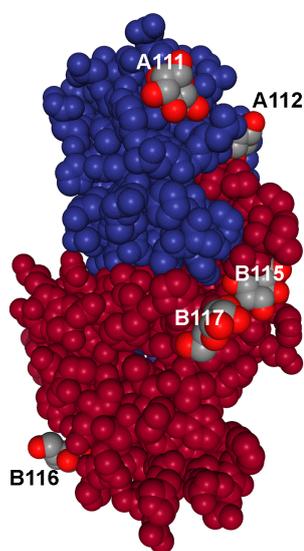


Figure 1: Packing diagram for correct molecular replacement solution obtained, as Ca models in stereo. Chains A, B in yellow and red comprise one heterodimer with fractional coordinates (x,y,z) is the asymmetric unit in this P₃₁2₁ crystal, displaying four out of the six possible asymmetric units in one unit cell; orientation of the three crystallographic axes indicated. Three other symmetry related molecules filling up the unit cell are shown in grey, with fractional coordinates (y,x,-z), (-y,x,-z+1/3) and (-x,y,-z+1/3) as marked. Tetramer formation through Chain B is seen for the two grey heterodimers at the left with fractional coordinates (y,x,-z) and (-y,x,-z+1/3). The first heterodimer in yellow and red similarly forms a tetramer with another symmetry-related asymmetric unit not displayed here, with fractional coordinates (x-y,-y,-1/3-z).

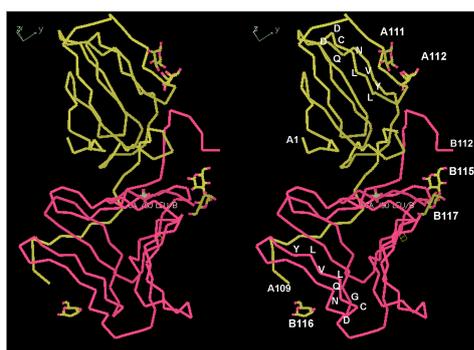


Figure 2: Temperature factors of Ca atoms in CEA lectins compared in chains A (upper) and B (lower). The curves for the mannose-bound CEA in 5D9Z are given in red in both plots, and in blue for the first heterodimer (chains A,B) and green for the second heterodimer (chains C,D) in mannose-free 5D5G. The sequences of chains A and B run along the x-axis, including green underlines at their classical carbohydrate recognition sites.

The average temperature factor of an atom including waters and ligands was 36.0 Å² for our submitted structure, but a majority of

backbone Ca atoms have lower average B values, particularly in chain B (Figure 2). However, average B values were in general higher in the complex compared to those in the mannose-free lectin (Figure 2). For our X-ray data, the Wilson B factor was calculated to be 20.3 Å² to 1.85 Å by our refinement or 15.3 Å² using data to 1.7 Å by the PDB.

The mannose-bound CEA structure at 1.85 Å resolution range contains two β-prism II domains in the asymmetric unit (Figure 1), though the mannose-free CEA crystal has four such [17]. Each has 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 [16,17,19]. In mannose-bound CEA, Cys 31, Cys 51 in the A subunits and Cys 34, Cys 56 in the B subunits form disulphide bridges. Cis peptide bonds exist at Gly 97-Pro 98 in the A subunits and at Gly 102-Pro 103 in the B subunits near the C-termini as found with RVL [19], the mannose-free CEA [17] and tarin in both forms [16]. As previously observed in RVL and other β-prism II domains, the two domains (A,B) have their C-terminal β-strands hydrogen bonding with a β-sheet in their partner domain, called C-terminal exchange [25].

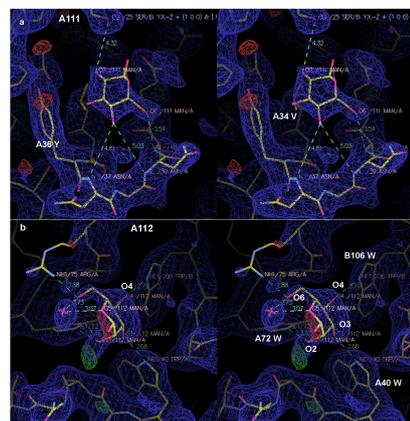


Figure 3: Mannose molecules superimposed on electron density maps. (A-E) Mannose molecules A111, A112, B115, B116 and B117 are respectively displayed superimposed on 2Fo - Fc (blue, 1σ level) and Fo-Fc (green, 3σ level) maps. This Fo-Fc map was calculated right before the introduction of the mannose and water molecules in our study. Mannose oxygen atoms are labeled. Models with green bonds are after final stage of Refmac5 while those with yellow bonds are after the Phenix refinement, differing a bit from each other in atom positions. Due to isomerization in mannose, the uncertainty in the O1 position generally leads to lack of electron density in that portion. However, O3 and O4 atoms seem more inside electron density as they are more ordered. Such big blobs of density which are not entirely within hydrogen bonding distance from protein atoms are not representing water molecule networks. The 5 mannoses have been renamed A201, A202, B201, B202 and B203 in the PDB file 5D9Z respectively.

Tetramer and dimer formation

Figures in reference [17] shows a view and a close up highlighting specific interactions stabilizing the tetramer formation in mannose-free CEA; it also shows the various interactions stabilizing the dimer between subunits A and B in mannose-free CEA. Tetramer and dimer formation for the mannose-bound CEA are very similar, except that

two of the mannose molecules seem to bind at the tetrameric interface formed by two B chains.

Analyses of $\alpha\beta$ and $\beta\beta'$ interfaces in 5D9Z

Using the PISA web server [26], for the interface between chain B and its symmetry mate B', 18 hydrogen bonds (22 for mannose-free lectin [17]) and 12 salt bridges were found; for that between chains A and B, 21 hydrogen bonds and 1 salt bridge were found. Other than the number of hydrogen bonds for chain B and its symmetry mate B' as noted above, the number of other bonds in the interfaces is the same in the mannose-free and mannose-bound CEA. In lieu of the numerous salt bridges seen for the B-B' interfaces, the A-B interfaces in the complex with 1728.9 Å² interface area, are stabilized by interactions between numerous hydrophobic residues, characteristic of the β -prism II lectin structures [16].

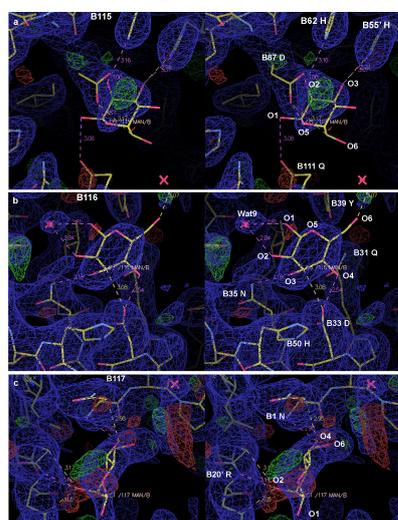


Figure 4: Two mannose molecules bound to chain A with electron density maps. Mannoses A111 and A112 are seen in parts A (top) and B (bottom) with surrounding atoms in the CEA structure, all superimposed on the final 2Fo-Fc map displayed at 1 σ level. (A) A111 is seen to bind at one edge of the classical site which ends at A36 Tyr. Atom O2 forms a contact with a serine O γ atom, O4 forms two long contacts with main chain amides of A37 Asn, A39 Asn and O6 forms a hydrogen bond with the main chain amide of A40 Trp. (B) A112, through its atoms O2, O3, O4 makes hydrogen bonds with N ϵ 1 atoms in three tryptophans A72, A40, B106 respectively, as also found earlier in the carbohydrate recognition domain IV in the daffodil lectin crystal [7], while O6 makes a hydrogen bond to the arginine side chain in A75 and a water molecule. The O1 side of mannose sticks out of the electron density blob due to disorder stemming from isomerization.

Five mannose binding sites per heterodimer in the crystal

In Figure 3, close-up views of the five mannose molecules found as superposed on 2Fo-Fc and Fo-Fc maps, are presented. Mannose oxygen atoms O1 through O6 are labeled. Models with green bonds represent the final stage of Refmac5 while those with yellow bonds represent the models after the Phenix refinement. Due to isomerization in mannose, the uncertainty in the O1 position generally

leads to lack of electron density in that portion. Added to this is partial occupancy of all the mannose molecules. Their occupancies have been estimated to vary from 0.6 to 0.9 by trial and error, trying to make the B-values equal to those of nearby protein side chains. However, O3 and O4 atoms seem more inside electron density as they are more ordered.

In Figures 4 and 5, the immediate neighbourhood provided by CEA for the five bound mannose molecules are shown in stereo. Both of the two molecules binding to chain A represent binding at non-classical sites, though A111 is at the edge of a classical site (Figure 4A). The other mannose binds to the triangular face of chain A (Figure 4B) as found earlier for the daffodil lectin [7]. Only one of the remaining three mannoses binding to chain B, B116, seems to be near a classical site (Figure 5B). The two other mannoses bound to chain B seem to be at the tetrameric interface and should stabilize the tetramer structure (Figure 5A, 5C).

In Figures 6 and 7, overall views of the heterodimer in CEA with five mannose molecules bound are shown. In Figure 6, the model is shown as an alpha carbon trace in stereo for the lectin, and stick models for the five mannose molecules. It is in fact a good summary of our study. In Figure 7, a space filling model of the heterodimer is seen in nearly the same orientation.

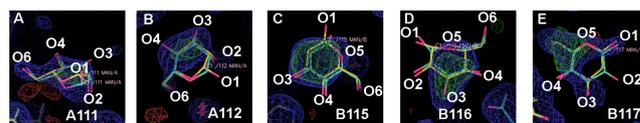


Figure 5: Three mannose molecules bound to chain B with electron density maps. Mannoses B115, B116 and B117 are seen in parts A (top), B (middle) and C (bottom) with surrounding atoms in the CEA structure, all superimposed on the final 2Fo-Fc map displayed at 1 σ level. (A) B115 is seen to hydrogen bond via O2 and O3, to B62 His and B55' His (symmetry related molecule) respectively, thus aiding tetramer formation, virtually like the role played by a sulfate bridging chain B with chain B' in the mannose-free CEA crystal [17]. Atoms O1 and O6 are outside the density blob, though O1 may be within hydrogen bonding distance of the B111 Gln side chain. O5 is seen hydrogen bonding to B87 Asp side chain. (B) B116 is bound to the most 'classical' site in this study, hydrogen bonding via O3 and O4 to B34 Asn and B33 Asp side chains respectively, while O1, O2 are hydrogen bonded to an ordered water molecule. (C) B117 binds near the N-terminus of chain B, through O4 to B1 Asn main chain; through O2, it also presumably binds to a B20' Arg main chain (symmetry related molecule, part of tetramer), though this portion lacks density. Surprisingly, O1 sits within electron density in this mannose.

Discussion

Various features of CEA like mass spectrometric analysis, presence of magnesium ions, ordered sulfate in the crystal, gel filtration of the lectin suggesting its presence in the $\alpha_2\beta_2$ form at neutral pH, and gradual conversion to the $\alpha\beta$ form at acidic pH, were rationalized by the mannose-free crystal structure [17] and its denaturation [18]. Pairwise comparisons of CEA with other heterodimeric β -prism II lectin crystal structures as measured by RMSD values were also presented [17]. The RMSD between chains A, B together in 5D9Z with the same chains in 5D5G was found to be 0.504 Å; that between chains

A, B in 5D9Z with chains C, D in 5D5G was found to be 0.513 Å, using all backbone atoms for the calculations performed in PYMOL [27]. Hence there does not appear to be significant changes in the lectin structure upon mannose binding.

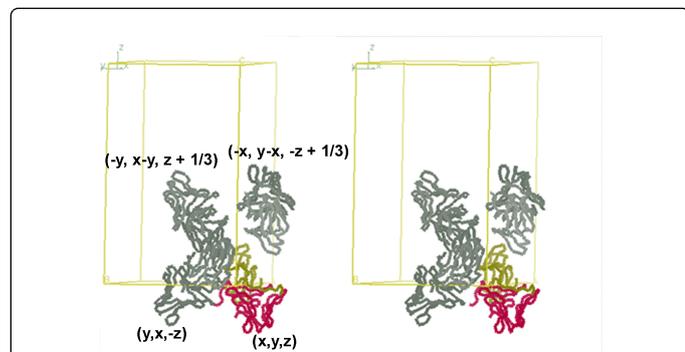


Figure 6: Alpha carbon trace in stereo for the $\alpha\beta$ heterodimer in our crystal. Chain A is shown in yellow and chain B in red, with stick models for the five mannose molecules found. One letter amino acid codes for residues in the classical carbohydrate binding sites, 28-QDDCNLVLY-36 in chain A and 31-QGDCNLVLY-36 in chain B are superposed on the C_{α} traces respectively. Here it is readily seen that mannoses A111, A112 bind to chain A and B115, B116 and B117 bind to chain B. A111 is sitting on one edge of the classical site while B116 is more centrally located in its site. The remaining 3 mannoses are not near the classical sites, with A112 sitting on the 'triangular face' of chain A, B115 bridging the two histidines at B62 and B55' help tetramer formation, while B117 is also bridges chains B and B' thus helping tetramer formation, though not apparent from this figure as chain B' is not shown.

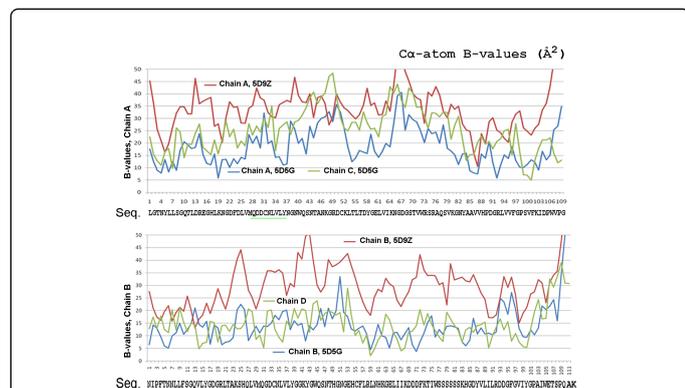


Figure 7: A space filling model for the $\alpha\beta$ heterodimer in our crystal. Using nonhydrogen atoms only, this is roughly in the same view as in Figure 6, but chain A is in dark blue and chain B is in dark red. The five mannose molecules are labeled, their carbon atoms in grey and oxygen atoms in red.

The quality of the electron density maps for the bound mannose molecules are admittedly not ideal due to the existence of isomers of mannose and due to partial occupancy. We used mannose, not methyl- α -D-mannoside as done in some earlier studies [5,9]; use of the methyl derivative avoids isomerization. It should be noted that no electron density maps were shown for the bound mannose monomers for snowdrop [5], garlic [6] or *Polygonatum cyrtoneuma* Hua [9], but such

maps were included for α 1-3 mannobiose [7,9] and for mannopentaose [10] binding. Indeed, by glycan array analysis, it was reported that the highly similar RVL preferred to bind complex derivatives, N-linked glycans and fucosylated mannose [19]. However, based on our crystal structure, CEA does seem to offer about 5 sites for mannose binding per heterodimer (Figure 6). B115 and B117 appear to stabilize the tetramer by bridging the two symmetry related B-chains.

A question may be raised as to whether the densities attributed here to the five mannose molecules could be explained by water molecules instead. Lee and Kim have analyzed the interstitial water structures in 1500 protein structures solved at greater than 1.5 Å resolution deposited in the PDB [28]. However, water polygons studied by them appear to have different characteristics [28] regarding their electron density distribution, quite unlike the large blobs noticed in our study explained as mannose molecules (Figures 3-5).

Conclusion

A recent study describes the CEA lectin tarin crystallized in P21 space group diffracting to 1.72 Å and its trimannoside-bound complex in P1 space group diffracting to 1.91 Å [16]. In this study, the X-ray data was collected at an Argonne Synchrotron, and has better data statistics [16], compared to our data collected at a home X-ray source. 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. The trimannoside binding in the tarin A and tarin B chains was also structurally compared with those in GNA and Concanavalin A, concluding the position of the bound 1,6-terminal mannose between tarin and GNA differ by 4.5 Å [16]. The carbohydrate binding surface in tarin was found to be larger and flatter, and due to the proximity in sequence and structure with CEA, the same holds for the latter.

The existence of several nonclassical binding sites for CEA as found in our study contributes in understanding the carbohydrate-binding modes of this lectin and its specificity.

Acknowledgement

We thank Mr. Pallab Chakraborty for maintaining our Institute X-ray diffraction data collection facility and for helping with crystallographic data collection, Mr. Himadri Biswas of our department for help with CEA purification.

This study was funded by the Department of Science and Technology, Government of India through Bose Institute.

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. Barre A, Van Damme EJM, Peumans WJ, Rougé P (1996) Structure-Function Relationship of Monocot Mannose-Binding Lectins. Plant Physiol 112: 1531-1540.
4. Ramachandriah G, Chandra NR (2000) Sequence and Structural Determinants of Mannose Recognition. Proteins Struct Funct Genet 39: 358-364.

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, et al. (1999) Crystal structure of a dimeric mannose-specific agglutinin from garlic: Quaternary association and carbohydrate specificity. J Mol Biol 285: 1157-1168.
7. 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.
8. Ding J, Bao J, Zhu D, Zhang Y, Wang DC (2010) Crystal structures of a novel anti-HIV mannose-binding lectin from *Polygonatum cyrtoneuma* Hua with unique ligand-binding property and super-structure. J Struct Biol 171: 309-317.
9. Wright CS, Hester G (1996) The 2.0 Å structure of cross-linked complex between snowdrop lectin and branched mannopentaose: evidence for two unique binding modes. Structure 4: 1339-1352.
10. de Castro LA, Carneiro M, Neshich DdeC, 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. 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.
13. 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.
14. Pereira PR, Winter HC, Vericimo MA, Meagher JL, Stuckey JA, et al. (2015) Structural analysis and binding properties of isoforms of tarin, the GNA-related lectin from *Colocasia esculenta*. Biochim Biophys Acta 1854: 20-30.
15. Pereira PR, Meagher JL, Winter HC, Goldstein IJ, Paschoalin VM, et al. (2017) High resolution crystal structures of *Colocasia esculenta* tarin lectin. Glycobiology 27: 50-56.
16. Chattopadhyaya R, Biswas H, Sarkar A (2017) Crystal Structure of *Colocasia esculenta* Tuber Agglutinin at 1.74 Å Resolution and Its Quaternary Interactions. J Glycobiol 6: 126.
17. Biswas H, Chattopadhyaya R (2017) Thermal and chemical denaturation of *Colocasia esculenta* tuber agglutinin from $\alpha 2\beta 2$ to unfolded state. J Biomol Str Dyn 36: 2179-2193.
18. Shetty NK, Bhat GG, Inamdar RS, Swamy MB, Suguna K (2012) Crystal structure of a β -prism II lectin from *Remusatia vivipara*. Glycobiology 22: 56-69.
19. Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: programs for protein crystallography, Acta Crystallogr D Biol Crystallogr 50: 760-763.
20. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240-255.
21. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, et al. (2010) PHENIX: A comprehensive python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213-221.
22. Wlodawer A, Minor W, Dauter Z, Jskolski 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.
23. 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.
24. Liu W, Yang N, Ding J, Huang R, Hu Z, et al. (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.
25. Krissinel, E. & Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774-797.
26. DeLano WL, Bromberg S (2004) The PyMOL Molecular Graphics System. DeLano Scientific SSC, San Carlos, California, USA.
27. Lee J, Kim S-H (2009) Water polygons in high-resolution protein crystal structures. Protein Sc 18: 1370-1376.