J. Biochem. 2012;152(1):87–98 doi:10.1093/jb/mvs047

JB THE JOURNAL OF BIOCHEMISTRY

Crystal structure of the lectin of *Camptosema pedicellatum*: implications of a conservative substitution at the hydrophobic subsite

Received December 15, 2011; accepted March 15, 2012; published online May 2, 2012

Claudener Souza Teixeira[1], Helton Colares da Silva[1], Tales Rocha de Moura[1], Francisco N. Pereira-Júnior[1], Kyria Santiago do Nascimento[1], Celso Shiniti Nagano[2], Alexandre Holanda Sampaio[1,2], Plinio Delatorre[3], Bruno Anderson Matias Rocha[1] and Benildo Sousa Cavada[1,*]

[1]BioMol-Lab, Laboratório de Moléculas Biologicamente Ativas, Departamento de Bioquímica e Biologia Molecular; [2]Departamento de Engenharia de Pesca, Universidade Federal do Ceará, Fortaleza, Ceará 60440-970, Brazil; and [3]Departamento de Biologia Molecular, Universidade Federal da Paraíba, João Pessoa, Paraíba 58059-900, Brazil

*Benildo Sousa Cavada, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Ceará, Mister Hull Avenue s/n. Campus do Pici, Bloco 907; P.O. Box 6043, Fortaleza, Ceará 60440-970, Brazil. Tel: 55 085 4008 9818, email: bscavada@ufc.br

Lectins have been used as models for studies of the molecular basis of protein–carbohydrate interaction and specificity by deciphering codes present in the glycan structures. The purpose of the present study was to purify and solve the complete primary and crystal structure of the lectin of *Camptosema pedicellatum* (CPL) complexed with 5-bromo-4-chloro-3-indolyl-α-D-mannose (X-Man) using tandem mass spectrometry. CPL was purified by single-step affinity chromatography. Mass spectrometry findings revealed that purified CPL features a combination of chains weighing 25,298 ± 2 (α-chain), 12,835 ± 2 (β-chain) and 12,481 ± 2 Da (γ-chain). The solved crystal structure of CPL features a conservative mutation in the hydrophobic subsite, a constituent of the carbohydrate recognition domain (CRD), indicating the relevance of hydrophobic interactions in the establishment of interactions with carbohydrates. The substitution and the analysis of the interactions with X-Man also revealed that the hydrophobic effect caused by a minor change in the hydrophobic subsite interferes in the formation of H-bonds due to the reorientation of the indolyl group in the CRD.

Keywords: Camptosema pedicellatum/hydrophobic subsite/lectin.

Abbreviations: CFL, *Cratylia floribunda* lectin; ConA, concanavalin A; ConBr, *Canavalia brasiliensis* lectin; ConM, *Canavalia maritima* lectin; CPL, *Camptosema pedicellatum* lectin; CRD, carbohydrate recognition domain; DGL, *Dioclea grandiflora* lectin; DGui, *Dioclea guianensis* lectin; DVL, *Dioclea violacea* lectin; DwL, *Dioclea wilsonii* lectin; HU, hemagglutinating units; NO, nitric oxide; X-Man, 5-bromo-4-chloro-3-indolyl-α-D-mannose.

The ability of plant extracts to agglutinate cells has been reported since the early 19th century. This activity has subsequently been shown to be the most common characteristic of lectins. Lectins were long regarded as mere curiosities of nature, but in the 1960s, when they were discovered to trigger an array of biological events, lectins became the object of intensive study (*1*). Today they are known to play a major role in cell communication, host defence, fertilization, development, parasite infection and tumour metastasis (*2*).

Lectins are a structurally heterogeneous group of proteins or glycoproteins with at least one non-catalytic domain binding reversibly to a specific mono- or oligosaccharide (*3*). Lectins are ubiquitous in nature and are found in all types of living organisms. They have been used as models for the study of the molecular basis of protein–carbohydrate interaction and specificity by deciphering codes present in glycan structures (*4*).

Many different and evolutionarily unrelated lectin families have been identified. Among these, lectins purified from species of the family Leguminosae represent the largest and most thoroughly studied family (*5*). Despite their relatively conserved primary structures, Leguminosae lectins exhibit considerable diversity in glycan-binding specificities. Because they are often expressed in high yields in legume seeds, they can easily be purified in amounts suitable for experimental approaches that require large amounts of protein such as microcalorimetry and X-ray crystallography. Legume lectins have traditionally represented a paradigm for the study of protein–carbohydrate interactions (*4*). Thus, concanavalin A (ConA), extracted from the seeds of *Canavalia ensiformis* (family Leguminosae, tribe Phaseoleae, subtribe Diocleinae), is the best characterized plant lectin so far.

Legume lectins of the Diocleinae subtribe, often referred to as ConA-like lectins, present glucose/mannose-monosaccharide binding specificities, and chemical and physicochemical similarities reveal considerable homology in amino acid sequence and 3D

structure (*6*, *7*). Nevertheless, Diocleinae lectins may differ substantially in biological activities, as shown in studies on human lymphocyte proliferation and interferon γ production (*8*), induction of paw oedema and vasodilator effects in rats (*9*), nitric oxide production (*10*, *11*), histamine release from rat peritoneal mast cells (*12*), *in vivo* lymphocyte activation and apoptosis (*13*), antidepressant-like effect in mice (*14*) and differentiation of human colon cancer cells (*15*), among others.

Differences in biological activity are associated with factors such as pH-dependent oligomerization, differences in the position of amino acids in the carbohydrate-binding site (*6*, *7*, 16) and minor changes in residues in key positions of the quaternary structure (*17*, 18).

Structural studies can help clarify the molecular mechanisms involved in the biological activities of different lectins, identify important aspects of protein–carbohydrate interaction and show how lectins decode the large amounts of information stored in glycan structures.

Thus, the purpose of the present study was to purify and solve the complete primary and crystal structure of the lectin of *Camptosema pedicellatum* (CPL) complexed with 5-bromo-4-chloro-3-indolyl-α-D-mannose (X-Man) using tandem mass spectrometry.

Materials and Methods

Plant material

Seeds of CPL were collected from plants grown in Northeastern Brazil (Crato, Ceará). The botanical identification was carried out at the Department of Biology, Federal University of Ceará (UFC).

Lectin purification

Mature seeds of *C. pedicellatum* were ground into a fine powder using a coffee mill. The flour was stirred with 0.15 M NaCl (1:10, m/v) for 3 h at room temperature. The mixture was centrifuged at $10{,}000 \times g$ for 20 min at 4°C. The clear supernatant (crude extract) was applied to a Sephadex G-50 column (a cross-linked dextran) (15 cm× 2 cm) equilibrated with 150 mM NaCl pH 7.0 containing 5 mM $CaCl_2$ and 5 mM $MnCl_2$. After washing of the unbound material in the equilibrium solution, the lectin was eluted from the gel with 100 mM glucose, pooled, exhaustively dialysed against distilled water and freeze-dried. Absorbance at 280 nm was used to estimate the protein concentration in all chromatographic fractions. The purification process was monitored by SDS–PAGE as described elsewhere (19) and the purified lectin was used for mass spectrometry analysis and crystallization trials.

Hemagglutination activity and inhibition assays

Hemagglutination assays were carried out as described elsewhere (20) using serial dilutions with rabbit erythrocytes, either native or treated with proteolytic enzymes (trypsin or papain). Results were expressed in hemagglutinating units (HU), with one HU being defined as the smallest amount (mg) of protein per millilitre capable of inducing visible agglutination. The lectin carbohydrate-binding specificity was defined as the smallest sugar concentration capable of fully inhibiting agglutination. Two-fold serial dilutions (initial concentration: 100 mM) of D-glucose, D-galactose, D-mannose, D-arabinose, D-fructose, D-fucose, D-xylose, *N*-acetyl-D-glucosamine, α-methyl-D-galactopyranoside, α-methyl-D-mannopyranoside, lactose and carrageenan were prepared in 150 mM NaCl. Lectin (4 HU) was added to each dilution.

MW determination by mass spectrometry

The molecular mass of CPL was determined by electrospray ionization mass spectrometry (ESI–MS) using a hybrid mass spectrometer (the Synapt HDMS system, Waters Corp., Milford, USA) operating in positive ion mode at 10,000 resolution. Protein solution (10 pmol/μl) was infused into the system using the built-in syringe drive at a flow rate of 10 μl/min. The capillary voltage and the cone voltage were set at 3 kV and 40 V, respectively. The source temperature was maintained at 100°C and nitrogen was used as a drying gas (flow rate: 150 l/h). Data acquisition was done with the software Mass Lynx 4.0 and the multiply charged spectra were deconvoluted using maximum entropy techniques (21).

Protein digestion and tandem mass spectrometry analysis

The protein digestion and tandem mass spectrometry analysis were carried out as described by Shevchenko *et al.* (22). Thus, protein was submitted to SDS–PAGE and the bands were excised and bleached in a solution of 50 mM ammonium bicarbonate in 50% acetonitrile. The bands were then dehydrated in 100% acetonitrile and dried in a speedvac (LabConco). The gel was rehydrated with a solution of 50 mM ammonium bicarbonate containing trypsin (Promega) or chymotrypsin (Sigma) (1:50 w/w; enzyme:substrate) at 37°C overnight. Subsequently, the peptides were extracted in a solution of 50% acetonitrile with 5% formic acid and concentrated in a speedvac. In addition, a lectin aliquot was dissolved in 2.0 M urea in 50 mM sodium phosphate (pH 6.8) and digested for 24 h at 37°C with endoproteinase Glu-C (Roche) obtained from *Staphylococcus aureus* (1:100 w/w; enzyme:substrate). The peptides were separated with a C18 chromatography column (75 μm × 100 mm) using a nanoAcquity^{TM} system and eluted with acetonitrile gradient (10–85%) containing 0.1% formic acid. The liquid chromatograph was connected to a nanoelectrospray mass spectrometer source (Synapt HDMS system, Waters Corp., Milford, USA). The mass spectrometer was operated in positive mode, using a source temperature of 80°C and a capillary voltage of 3.5 kV. The instrument was calibrated with double-protonated [Glu1]-fibrinopeptide B ions $(M + 2H^+)$ $2^+ = 785.84$. The LC-MS/MS experiment was done with the DDA (data-dependent acquisition) function selecting for experiments with double or triple-charged precursor ions, which were fragmented by collision-induced dissociation (CID) with the ramp collision energy level adjusted according to the charge state of the precursor

ion. The data were managed and analysed with the software Proteinlynx (Waters), using 'peptide fragmentation pattern' as search parameter. Some peptide sequences were obtained by *de novo* manual sequencing followed by manual interpretation of CID spectra.

Primary structure analysis

The primary sequence alignments and secondary structure predictions were made with the software ESPript 2.2 (23). The theoretical pI was calculated with ProtParam (24).

Crystallization and data collection

The freeze-dried purified lectin was resuspended in Milli-Q^{TM} water containing 3 mM X-Man at a final concentration of 10 mg ml^{-1} and incubated at 310 K for 1 h prior to the crystallization experiments. The crystals grew on 24-well Linbro^{TM} plates at room temperature (293 K) by the vapour-diffusion method (25) in hanging drops using Crystal Screens I and II (Hampton Research, Riverside, CA, USA). Each well contained 300 µl reservoir solution and the drops were composed of equal amounts (2 µl) of protein solution and reservoir solution.

X-ray data were collected from a single crystal cooled to 100 K. Crystals were previously soaked in a cryoprotectant solution of water (70%) and PEG 400 (30% v/v) to avoid ice formation and submitted to X-ray diffraction at a wavelength of 1.42 Å using a synchrotron-radiation source (MX1 station, Laboratório Nacional de Luz Síncrotron, Campinas, Brazil). A complete data set was obtained using a MarCCD 165 mm (MAR Research) in 180 frames with an oscillation range of 1°. The data set was indexed and integrated using MOSFLM (26). Intensities were reduced through SCALA (27).

Molecular replacement and refinement

The crystal structure of CPL was determined by molecular replacement using the program MolRep (28). The atomic coordinates used as model were those of recombinant *Dioclea grandiflora* (lectin of DGL) complexed with X-Man (PDB code 2JEC) (*7*), with a final correlation coefficient of 62.4% and an R_{factor} of 42.1% after calculation of rotation and translation functions.

The initial structure was submitted to rigid body and restrained refinement with the program REFMAC5 (*29*). R_{factor} and R_{free} converged to 35.96 and 36.02%, respectively. Subsequently, the structure was modelled using WINCOOT. The loop region (117–123) was adjusted to satisfy the electron density map and 77 water molecules were added to the structure. The second restrained refinement yielded an R_{factor} of 19.0% and an R_{free} of 23.6%. The stereochemical quality of the model was inspected by Ramachandran plot using PROCHECK program (29), Van der Walls contacts, polar contacts and hydrophobic interaction were analysed with the CCP4 software CONTACT, adopting the cut-off distances 3.5, 3.5 and 5.0 Å, respectively. All figures and superpositions were performed with the program PyMOL (30). The atomic coordinates for the structure were deposited in PDB with access code 3U4X.

Results and Discussion

CPL was purified in a single step using a Sephadex G-50 affinity chromatography column, in which the lectin was quantitatively retained in the cross-linked dextran gel column and desorbed with D-glucose, providing strong evidence of carbohydrate-binding properties (Fig. 1a). This procedure has been widely used for the purification of Diocleinae lectins (16, *31–33*). Hemagglutinating activity towards native and enzyme-treated rabbit erythrocytes was fully inhibited by D-mannose (12.5 mM), D-glucose and α-methyl-D-mannopyranoside (3.16 mM), but not by as much as 100 mM D-galactose, D-arabinose, D-fructose, D-fucose, D-xylose, *N*-acetyl-D-glucosamine, α-methyl-D-galactopyranoside, lactose or carrageenan (Table I). Similar carbohydrate-binding specificities have been reported for other Diocleinae lectins (16, *34*). Furthermore, CPL displayed greater affinity for α-methyl-D-mannopyranoside than for D-mannose, indicating the presence of a more strongly hydrophobic substituent at C-1, conceivably allowing for additional interactions with hydrophobic regions of the lectin-binding site (*34*).

The electrophoretic profile obtained with SDS–PAGE of the affinity-purified CPL, both in the presence and absence of β-mercaptoethanol, revealed a major band of 30 kDa (α-chain) and two minor components (β-chain and γ-chain) of 16 and 13 kDa, respectively (Fig. 1b), suggesting CPL is a typical ConA-like lectin subject to the posttranslational process of circular permutation described by Carrington *et al.* (*35*). Thus, CPL is expressed as a pre-pro-protein ($_{Nterm}$signal peptide + γ-chain + linker peptide + β-chain + $_{Cterm}$signal peptide) cleaved into a γ product and a β product. The active protein is a final fused product (α-chain) with the two fragments in inverse order and no signal or linker peptides (*35–37*). The β- and γ-fragments observed on SDS–PAGE (Fig. 1b) are unlinked products of this process while the α-chain is the mature protein. Electrospray ionization mass spectrometry confirmed purified CPL consists of a combination of chains weighing 25,298 ± 2 (α-chain), 12,835 ± 2 (β-chain) and 12,481 ± 2 Da (γ-chain) (Fig. 2).

The complete CPL protein sequence was obtained from overlaps of 27 digested peptides sequenced by tandem mass spectrometry, resulting in 236 amino acid residues (β-chain: 1–118. γ-chain: 119–236) (Fig. 3). The isotope-averaged molecular masses calculated for the full-length α-chain (25,297 Da) and its derived fragments β and γ (12,835 and 12,480 Da, respectively) are in agreement with the experimentally determined mass. Table II shows all sequenced peptides and their respective molecular mass. The theoretical pI based on the final sequence was 5.4. The protein sequencing data reported in this article are deposited in the UniProt Knowledgebase under the accession number P86894.

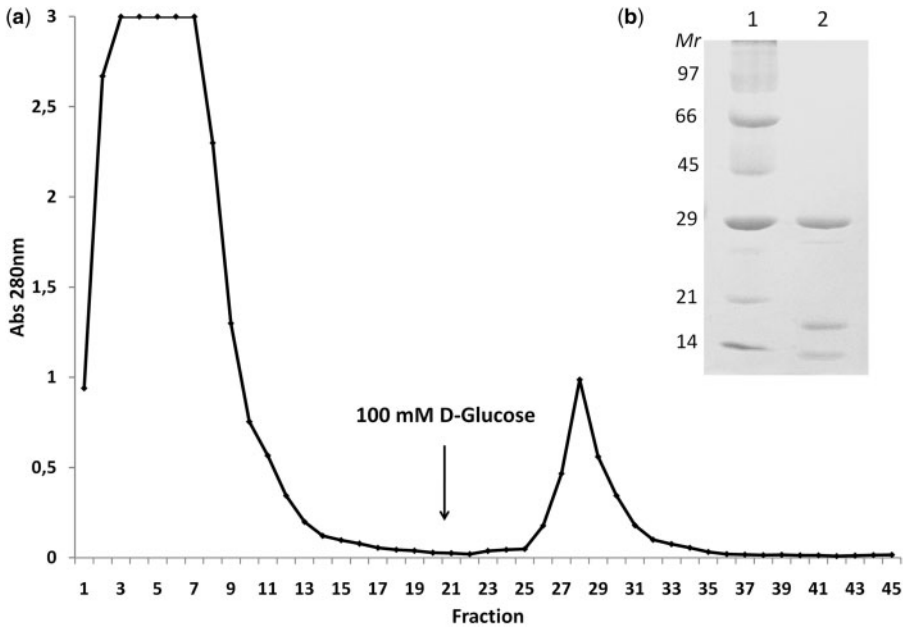

Fig. 1 Purification of CPL. (**a**) Elution profile of Sephadex G-50 affinity chromatography. Approximately 10 ml crude extract was applied to the Sephadex G-50 column (15 cm× 2 cm) equilibrated with 150 mM NaCl containing 5 mM $CaCl_2$ and 5 mM $MnCl_2$. The lectin was eluted with 100 M D-glucose included in the above buffer at a flow rate of 1 ml/min. Fractions (2.0 ml) were collected and monitored for protein content at 280 nm. (**b**) SDS–PAGE. Lane 1: Molecular mass markers (phosphorylase b, 97 kDa; bovine serum albumin, 66 kDa; ovalbumin, 45 kDa; carbonic anhydrase, 29 kDa; trypsin inhibitor, 20.1 kDa and α-lactalbumin, 14.4 kDa); Lane 2: CPL.

Table I. Inhibitory effect of monosaccharides and disaccharides on haemagglutinating activity of CPL.

Sugar	MIC (mM)
D-Glucose	50
D-Galactose	NI
D-Mannose	12.5
D-Arabinose	NI
D-Fructose	NI
D-Xilose	NI
D-Fucose	NI
N-Acetyl-D-glucosamine	NI
α-methyl-D-galactopyranoside	NI
α-methyl-D-mannopyranoside	3.16
Lactose	NI
Carrageenan	NI

MIC, minimum inhibitory concentration; NI, sugar not inhibitory until a concentration of 100 mM.

The primary structure of CPL is quite homologous to that of other Diocleinae lectins, such as *D. guianensis* (Dgui) (86%) (SwissProt accession code: P81637), *D. grandiflora* (DGL) (85%) (SwissProt accession code: A9J251), *D. violacea* (DVL) (81%) (PDB accession code: 2GDF), *D. wilsonii* (DwL) (85%) (SwissProt accession code: P86624), *C. ensiformis* (ConA) (81%) (SwissProt accession code: P02866), *C. brasiliensis* (ConBr) (81%) (SwissProt accession code: P55915) and *Cratylia floribunda* (CFL) (97%) (SwissProt accession code: P81517). CFL differs from CPL by only five residues: 99 (V/L), 136 (S/N), 201 (P/T), 205 (P/I) and 211 (F/W) (Fig. 4).

Some Diocleinae lectins exhibit pH-dependent dimer–tetramer equilibrium. The physiological relevance of this phenomenon remains poorly understood. However, pH-dependent dimer/tetramer equilibrium has been reported to exert a considerable influence on oligosaccharide recognition and biological activity (*7*), suggesting residues 123 (A/E), 131 (N/H) and 132(Q/K) play an important role in dimer/tetramer equilibrium properties. Similarly to Dgui (17) and CFL (*38*), which exhibit pH-dependent dimer–tetramer transition, CPL contains the residues Ala123, Asn131 and Gln132.

One of the most important primary structural differences between CPL and other Diocleinae lectins is the substitution of Val99 for Leu99. This conservative amino acid substitution was confirmed by MS/MS analyses of the peptides T5 (Fig. 5) and E2. The product-ion spectra of the doubly charged ion at $m/z = 541.30$ and of the singly charged ion at $m/z = 1651.87$ (E2) produced sequence-specific y-ion series, from which the primary structures were interpreted as VGLSASTGVYK and WVRVGLSASTGVYKE, respectively. The tryptic peptide which comprises amino acids from V91 to K101 of other Diocleinae lectins sequenced by MS/MS (*39–41*) presented an m/z ratio of 555.45 (M+H+ 1108.90), corresponding to the peptide with Leu99. Legume lectins are known to possess three types of hydrophobic subsites based on different ligand affinities. The fact that one of these subsites is adjacent to the conserved monosaccharide binding site explains why hydrophobic glyco/mannosides and other hydrophobic derivative monosaccharides bind more strongly (10–50 times) than their non-hydrophobic analogues (*42*). Kanellopoulos *et al.* (*43*) determined the structures of two ConA complexes, ConA/4′-nitrophenylalpha-D-mannopyranoside

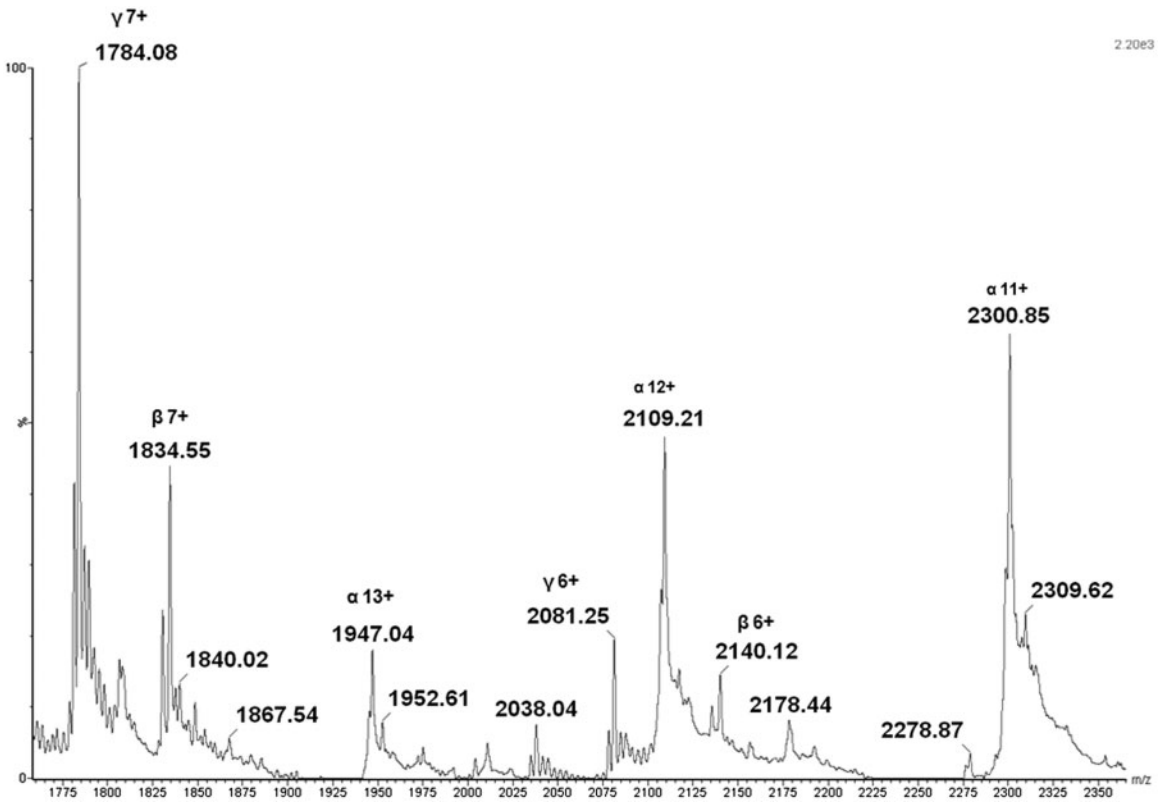

Fig. 2 Mass spectrometry analysis of CPL. Multicharged mass spectra of CPL acquired with ESI-MS in a hybrid quadrupole/ion mobility separator/orthogonal acceleration-time of flight mass spectrometer. Ions with different charge state were identified and numbered.

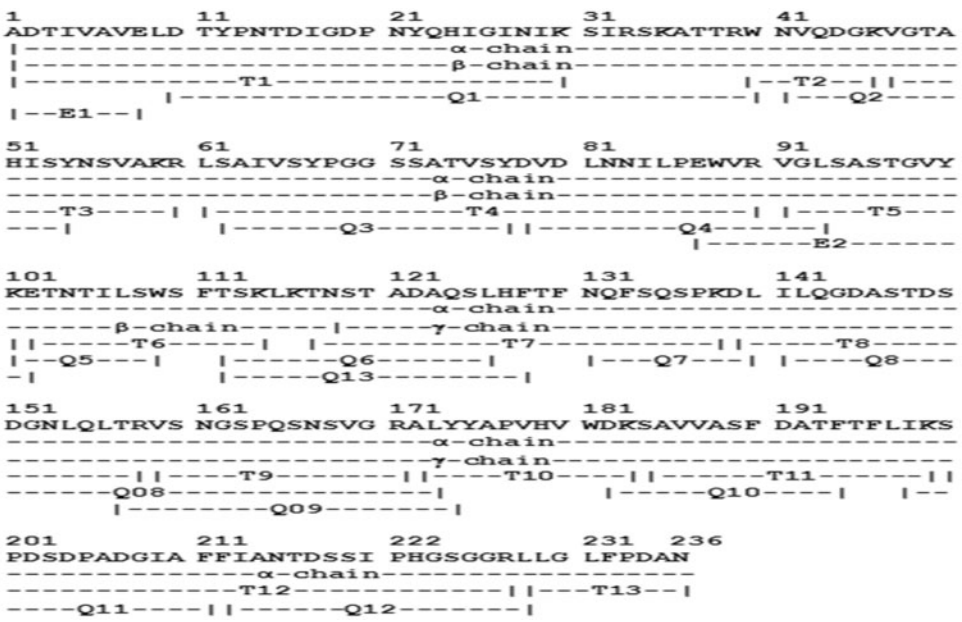

Fig. 3 Amino acid sequencing of CPL. Amino acid sequence of CPL assembled from sequences of overlapping degradation products generated by cleavages with trypsin (T), chymotrypsin (Q) and endoproteinase Glu-C (E). The amino acid sequences of the α-chain (1–236), β-chain (1–118) and γ-chain (119–236) are highlighted.

and ConA/4′-nitrophenyl-alpha-D-glucopyranoside, and found that the hydrophobic moiety of these molecules interacts with a hydrophobic subsite formed by Tyr12, Leu99 and Tyr100. These amino acid residues and their interactions are conserved in all Diocleinae lectins (*44–46*) but in CPL, the only lectin in the group with Val instead of Leu in position 99. This difference translates directly into specific fine carbohydrate-binding properties and biological activities.

Table II. Sequenced peptides of *C. pedicellatum* lectin and their respective molecular masses.

Peptide	Experimental mass (Da)	Sequence
T1	3,298.6201	ADTIVAVELDTYPNTDIGDPNYQHIGINIK
T2	845.3484	WNVQDGK
T3	1,345.6206	VGTAHISYNSVAK
T4	3,221.5845	LSAIVSYPGGSSATVSYDVDLNNILPEWVR
T5	1,080.5006	VGLSASTGVYK
T6	1,512.7410	ETNTILSWSFTSK
T7	2,454.8164	TNSTADAQSLHFTFNQFSQSPK
T8	2,131.0090	DLILQGDASTDSDGNLQLTR
T9	1,287.5410	VSNGSPQSNSVGR
T10	1,460.7125	ALYYAPVHVWDK
T11	1,715.9370	SAVVASFDATFTFLIK
T12	2,786.9766	SPDSDPADGIAFFIANTDSSIPHGSGGR
T13	958.5063	LLGFPDAN
Q1	3,573.7866	DTYPNTDIGDPNYQHIGINIKSIRSKATTRW
Q2	1,487.6444	NVQDGKVGTAHISY
Q3	1,544.8213	SAIVSYPGGSSATVSY
Q4	1,850.9443	DVDLNNILPEWVGL
Q5	1,090.3043	KETNTILSW
Q6	1,847.8566	TSKLKTNSTADAQSLHF
Q7	1,162.5244	NQFSQSPKDL
Q8	3,356.6765	ILQGDASTDSDGNLQLTRVSNGSPQSNSVGRAL
Q9	1,891.8994	TKVSNSGSPQGNSVGRAL
Q10	1,356.6044	DKSAVVASFDATF
Q11	1,431.6642	IKSPDSPADGIAF
Q12	1,840.7765	FIANTDSSIPHGSGGRL
E1	816.9266	ADTIVAVE
E2	1,651.8729	WVRVGLSASTGVYKE

CPL yielded small irregular crystals in 100 mM HEPES (pH 7.5) containing 2.0 M ammonium sulphate and 2% PEG 400 after 1 week (Fig. 6a) (Hampton Research Crystal Screen I, condition No. 39), while plate crystals were obtained in 200 mM ammonium sulphate, 100 mM sodium cacodylate pH 6.5 and 30% PEG 8000 (Fig. 6b) (Hampton Research Crystal Screen I, condition No. 15). These crystals were not suitable for X-ray diffraction experiments. However, several optimization steps were performed combining these two conditions, changing the pH and precipitant concentration. CPL crystals grew in 100 mM HEPES (pH 8.5) containing 2.0 M ammonium sulphate and 2% PEG 400 to maximum dimensions of 0.1 mm× 0.2 mm× 0.15 mm (Fig. 6c).

CPL crystals diffracted to a maximum resolution of 2.15 Å using a Synchrotron radiation source (LNLS, Campinas, Brazil). The complete data set of 180 frames was indexed, integrated and scaled in the 2.15 Å resolution range. The crystals belonged to the orthorhombic space group I222 with cell parameters $a = 65.9$ Å, $b = 66.7$ Å, $c = 107.7$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$ and $\gamma = 90^\circ$. The Matthews coefficient 2.37 $Å^3\,Da^{-1}$ (*47*) was calculated based on a molecular weight of 25,298 ± 2 Da and implies that the crystal contains 48.18% solvent (indicating the presence of a monomer in the asymmetric unit) and that the biological assembly is a 222 tetramer. The data collection statistics are shown in Table III.

The refined monomer structure of CPL complexed with X-Man consists of 236 amino acids folded as a β-sandwich, as observed in several other legume lectins (Fig. 7a). The CPL tetramer consists of two canonical dimers linked by salt bridges between β-strands (Fig. 7b). CPL features a metal-binding site with the conserved residues Asn14 and Tyr12 (interacting with calcium), Glu8 and His24 (interacting with manganese) and Asp10 and Asp19 (interacting with both). In addition, the peptide bond of Ala206 and Asn207 in the *cis* configuration is isomerized due to the presence of divalent metals changing the side chain orientation of Asn14 and Asp207—a common observation in legume lectins. The substitution of a single residue can significantly affect biological activity, as observed with site-directed mutagenesis applied to other lectins (*48*).

The carbohydrate-binding site is occupied by X-Man (Fig. 8). The complex involves 10 intermolecular H bonds formed between O3, O4, O5 and O6 of the X-Man complex, amino acid side chains and 8 van der Waals interactions (Table IV). Legume lectin structures feature a conserved hydrophobic subsite composed of Tyr12, Leu99 and Tyr100 (*44*). However, the primary structure of CPL has a conserved substitution at this site (Val99 for Leu99) (Fig. 8). This change in the hydrophobic subsite reduces the hydrophobic interactions between X-Man and the lectin. Due to the absence of a methyl group, Val99 is ~1.0 Å closer than Leu99 to the indolyl group of X-Man. As a result, the hydrophobic interactions in this subsite strengthen the carbohydrate association in the carbohydrate binding site (*42*).

The hydrophobic amino acid in position 99 not only enhances lectin–carbohydrate interaction, it also affects the orientation of X-Man in the carbohydrate recognition domain (CRD). By placing the ligand at a distance of 3.98 Å, Val99 causes a change in the

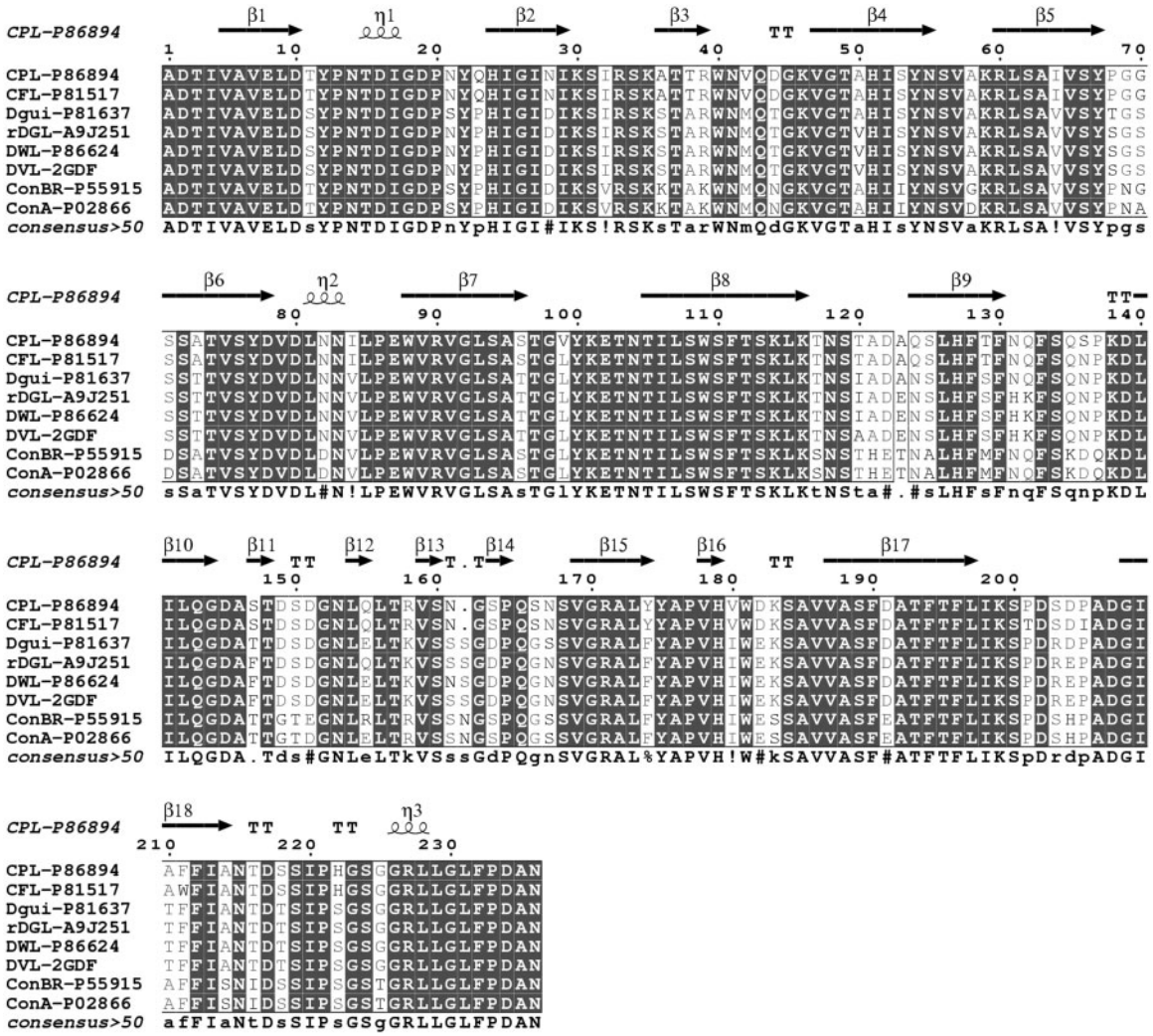

Fig. 4 Multiple sequence alignment of ConA-like lectins. Alignment of CPL, CFL, Dgui, DGL, DVL, DwL, ConBr and ConA reveals a highly conserved sequence in the secondary structure.

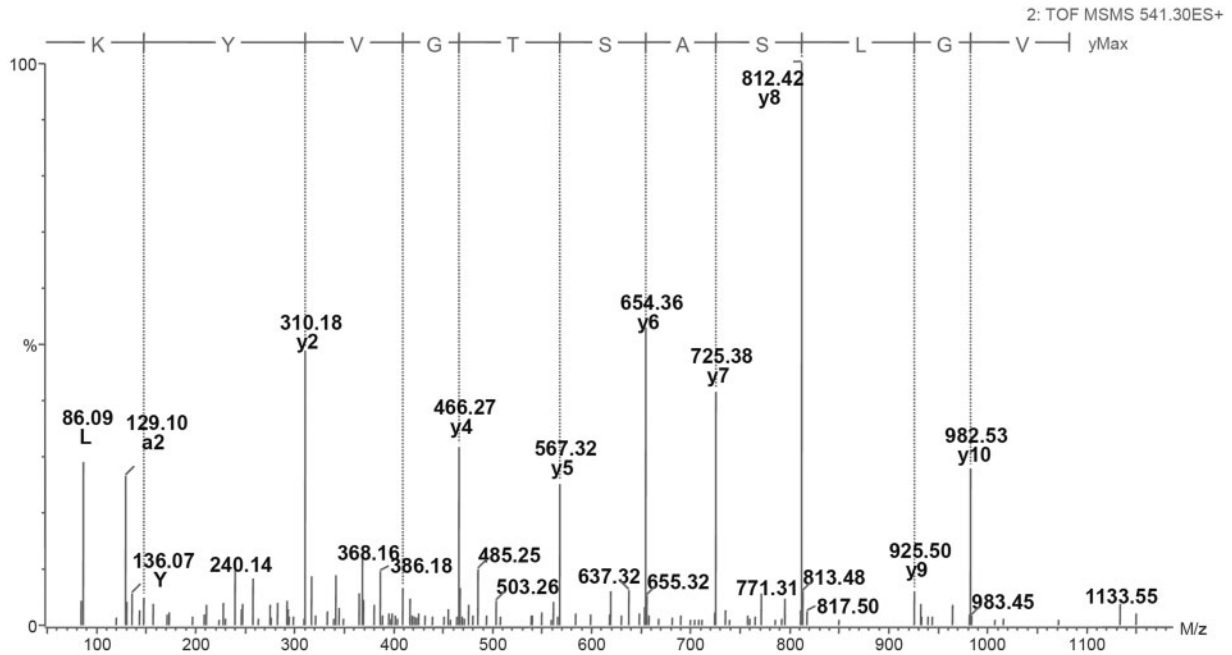

Fig. 5 Tandem mass spectrometry of T5 peptide of CPL. Collision-induced fragmentation of the doubly charged ion at $m/z = 541.3$ corresponding to the peptide T5 of CPL (see Fig. 2 and Table II). Sequence-specific y ions used for structure determination are indicated.

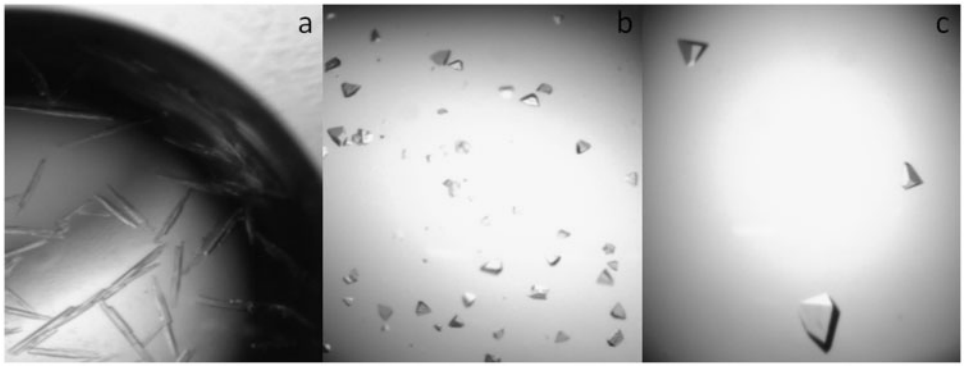

Fig. 6 Crystals of CPL. (**a**) Plate crystals obtained with Crystal Screen I condition No. 15, (**b**) crystals obtained with Crystal Screen I condition No. 15 and (**c**) single crystal diffracted at 2.15 Å obtained with the optimization of condition No. 39.

Table III. Statistics of data collection, refinement and structure quality.

Parameter	Value
Data collection	
Beamline wavelength	1.42 Å
Space group	I222
Unit cell parameters (Å)	
a	65.9
b	66.7
c	107.7
Total reflections	85,239
Number of unique reflections	12,740
Molecules per asymmetric unit	1
R_{merge}[a] (%)	14.7 (25.0)[d]
Completeness (%)	96.4 (96.9)[d]
Multiplicity	6.7 (6.2)[d]
I/σ	3.4 (2.8)[d]
Molecular replacement	
Correlation coefficient	62.4
R_{factor}[b] (%)	42.1
Refinement	
Resolution range (Å)	21.5–2.1 (2.2–2.1)[d]
R_{factor}[b](%)	19.0
R_{free}[c] (%)	23.6
Number of residues in asymmetric unit	236
Number of water molecules	77
RMS deviations from ideal values	
Bond lengths (Å)	0.019
Bond angles (degrees)	1.924
Temperature factor	
Average B value for whole protein chain (Å)	17.84
Ramachandran plot	
Residues in most favoured regions (%)	95.7
Residues in additional allowed regions (%)	4.3
Residues in generously allowed regions (%)	0

[a] $R_{merge} = \frac{\sum_{hkl}\sum_i \left|\bar{I}(hkl) - \langle I(hkl)i\rangle\right|}{\sum_{hkl}\sum_i \langle I(hkl)i\rangle}$ where $I(hkl)_i$ is the intensity of i-th measurement of reflection h and $I(hkl)$ is the mean value of the $I(hkl)_i$ for all I measurements.

[b] $R_{factor} = \frac{\sum_h ||F_{obs}| - |F_{calc}||}{\sum_h |F_{obs}|}$

[c] Calculated with 5% of the reflections omitted from refinement.

[d] Values in parentheses represent the high resolution shell.

position of the indolyl group in a way unfavourable to the establishment of a bond between the OH molecule in Tyr12 and the N1 component of X-Man (Fig. 9a), thereby making H-bonds with Tyr12 impossible (Fig. 9b). In other Diocleinae lectins complexed with X-Man, the distance to the hydroxyl group in Tyr12 is reduced and interaction is facilitated due to the presence of Leu99 in the hydrophobic subsite (Table V).

The CRD superposition analysis of DVL (PDB code: 3AX4) and CPL (PDB code: 3U4X) (Fig. 9b) revealed differences in the position of the indolyl group and the hydroxyls in Tyr12. Gadelha *et al.* demonstrated the importance of the hydroxyl-Tyr12

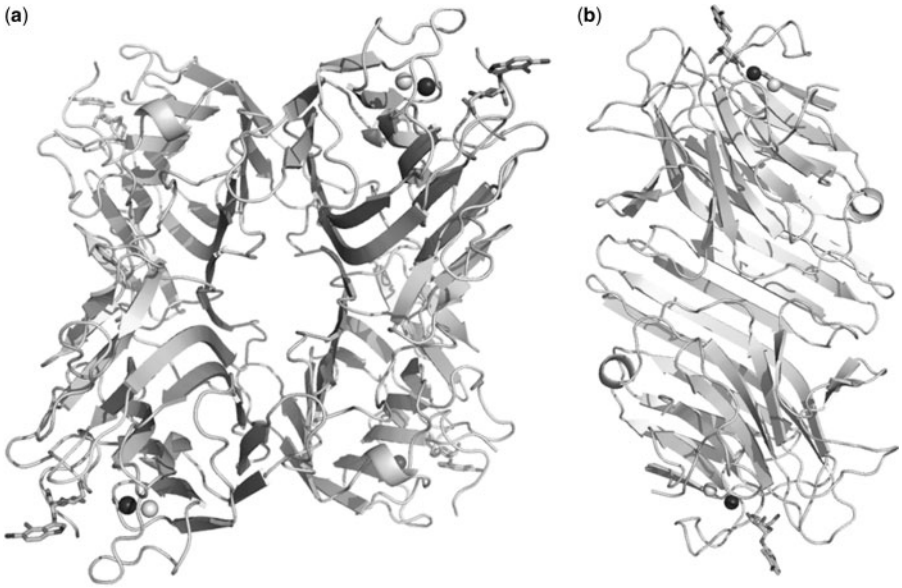

Fig. 7 The canonical lectin dimer presented in the CPL tetramer. (**a**) Overall structure of CPL complexed with X-Man. (**b**) The two subunits of the canonical dimmer. Ca^{2+} (black) and Mn^{2+} (white) are shown as spheres.

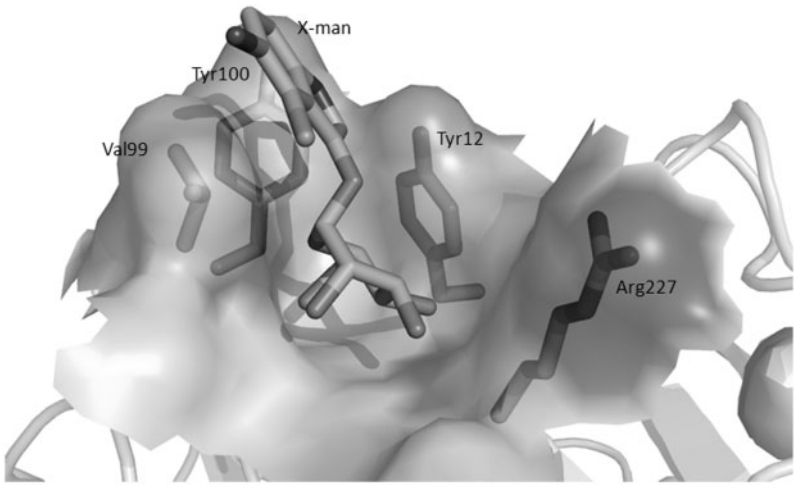

Fig. 8 Representation of the surface area of the carbohydrate recognition domain of CPL complexed with X-Man. Amino acids that compose the hydrophobic subsite are labelled: Tyr12, Val99 and Tyr100.

Table IV. Van der Waals interaction and polar contacts between the lectin from *C. pedicellatum* seeds and X-Man.

Amino acids	X-Man	Distances (Å)
Van der Waals interactions		
Tyr12 OH	C11	3.38
Asp207 OD1	C4	3.39
Asp207 CG	O4	3.48
Arg227 CB	O4	3.40
Val99 CB	O5	3.44
Gly98 CA	O6	3.21
Gly98 C	O6	3.43
Ala206 CB	O6	3.27
Polar contacts		
Arg227 N	O3	2.93
Asn14 ND2	O4	3.01
Asp207 OD1	O4	2.60
Arg227 N	O4	3.19
Val99 N	O5	3.06
Gly98 N	O6	3.33
Val99 N	O6	3.01
Tyr100 O	O6	3.36
Tyr 100 N	O6	3.01
Asp207 OD2	O6	2.97

to the configuration of CRD in the lectin of *C. maritima* (ConM), associated with the substitution of Pro for Ser in position 202, bringing Tyr12 closer to the disaccharides forming H-bonds. This was later adduced to explain why ConM can interact with endothelial cell glycans and stimulate NO production by the activation of endothelial nitric oxide synthase (*11*). Surprisingly, the difference in position between ConM and ConA was enough to increase NO production at least 3-fold, despite 96% homology (*49*).

The position of the indolyl group in CPL and other ConA-like lectins differs by 0.9 Å, enough to prevent the formation of H-bonds with Tyr12. In fact, the hydrophobic interactions of Val/Leu significantly increase the association constant of the ligand in the CRD. The residue can model the ligand inside the site by hydrophobic interaction, possibly modifying the affinity of Val99-lectins and Leu99-lectins for carbohydrate complexes. This is a small but seemingly

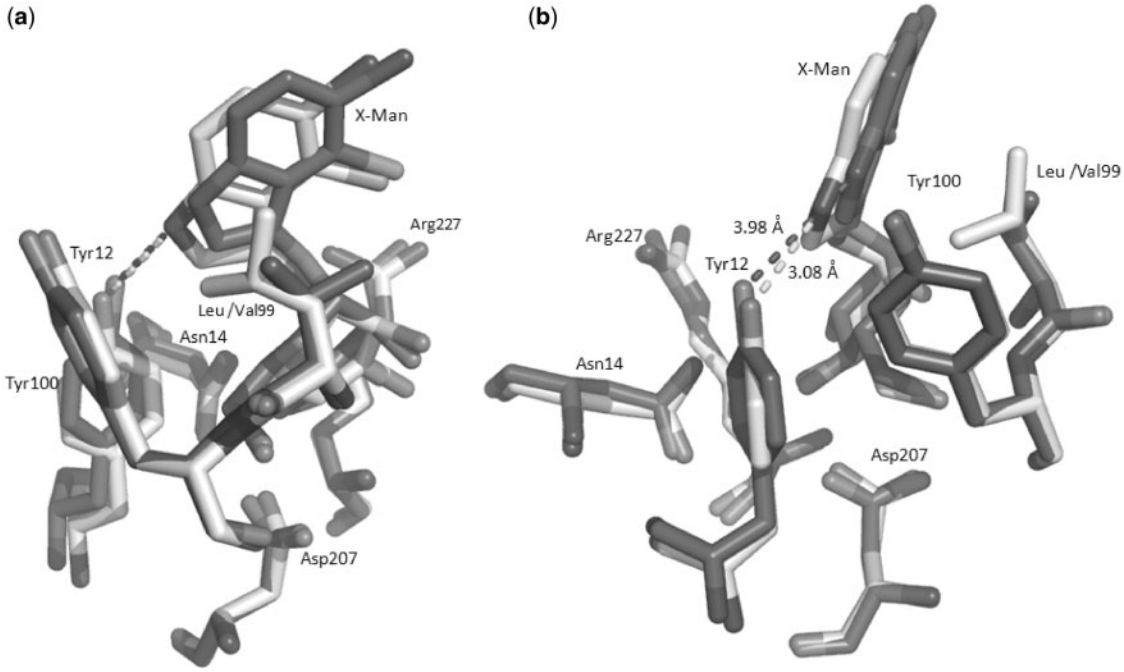

Fig. 9 Structural comparison between CPL (black) and DVL (PDB code: 3AX4) (grey). (**a**) difference in the position of X-Man in CPL and DVL due to the substitution of Leu for Val in position 99. (**b**) Hydrogen bond formation between the hydroxyl group of Tyr12 and the N1 component of X-Man oriented by Leu99 in DVL (grey dashes). The presence of Val99 does not favour H-bond formation due to the reorientation of X-Man in the CRD. The ligand is positioned at a distance of 3.98 Å from the hydroxyl group of Tyr12 (black dashes).

Table V. Distances between OH Tyr12 and N1 of the X-Man in carbohydrate recognition domain.

Diocleinae lectin	Amino acid 99	Distance (Å)	PDB code
Camptosema pedicellatum	Valine	3.98	3U4X
Dioclea violacea	Leucine	3.08	3AX4
Dioclea wilsonii	Leucine	3.20	3SH3
Dioclea guianensis	Leucine	3.19	2JDZ

important variation within these otherwise quite homologous molecules.

The diversity of biological activities induced by Diocleinae lectins have been shown to be related to several factors, alone or in combination, including biological assembly and dimer-tetramer equilibrium (*38*), changes in CRD configuration induced by punctual amino acid modifications (*49*), CRD volume (*46*) and, as shown in the present study, substitutions in hydrophobic subsite amino acid residues.

The crystal structure of CPL features a conservative mutation in the hydrophobic subsite, a constituent of the CRD. This modification shows the relevance of hydrophobic interactions in the establishment of interactions between carbohydrates and hydrophobic substituents and how these interactions determine the ligand orientation at the carbohydrate binding site, enhancing the lectin-carbohydrate interaction specificity. The substitution and the analysis of the interactions with X-Man revealed that the hydrophobic effect caused by the presence in the hydrophobic subsite of Val99 instead of Leu99 made H-bond formation difficult due to the reorientation of the indolyl group in the CRD. These structural aspects are responsible for part of the observed variability in biological activity in otherwise relatively homologous proteins.

Acknowledgements

We also thank the Laboratório Nacional de Luz Síncrotron (LNLS), Campinas—Brazil. K.S.N., C.S.N., A.H.S., P.D., B.A.M.R. and B.S.C. are senior investigators of CNPq. We thank Jesper Sampaio for English language editing of the manuscript.

Funding

Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP, partial), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, partial) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, partial).

Conflict of interest

None declared.

References

1. Sharon, N. and Lis, H. (2004) History of lectins: from hemaglutinins to biological recognition molecules. *Glycobiology* **4**, 53–62
2. Gabius, H.J. and Gabius, S. (1997) *Glycoscience. Status and Perspectives.*, Chapman & Hall, Weinheim, Germany
3. Peumans, W.J. and Van Damme, E.J.M. (1995) Lectins as plant defense proteins. *Plant Physiol.* **109**, 347–352
4. Loris, R., Van Walle, I., De Greve, H., Beeckmans, S., Deboeck, F., Wyns, L., and Bouckaert, J. (2004) Structural basis of oligomannose recognition by the *Pterocarpus angolensis* seed lectin. *J. Mol. Biol.* **335**, 1227–1240

5. Van Damme, E.J.M., Peumans, W.J., Barre, A., and Rougé, P. (1998) Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* **17**, 575–692
6. Calvete, J.J., Thole, H.H., Raida, M., Urbanke, C., Romero, A., Grangeiro, T.B., Ramos, M.V., Rocha, I.M.A., Guimarães, F.N., and Cavada, B.S. (1999) Molecular characterization and crystallization of Diocleinae lectins. *Biochim. Biophys. Acta* **1430**, 367–375
7. Nagano, C.S., Calvete, J.J., Barettino, D., Pérez, A., Cavada, B.S., and Sanz, L. (2008) Insights in to the structural basis of the pH-dependent dimer–tetramer Equilibrium through crystallographic analysis of recombinant Diocleinae lectins. *Biochem. J.* **409**, 417–428
8. Barral-Netto, M., Santos, S.B., Barral, A., Moreira, L.I.M., Santos, C.F., Moreira, R.A., Oliveira, J.T.A., and Cavada, B.S. (1992) Human lymphocyte stimulation by legume lectins from the Diocleae tribe. *Immunol. Invest.* **21**, 297–303
9. Assreuy, A.M., Fontenele, S.R., Freitas, P.A., Fernandes, D.C., Rodrigues, N.V., Bezerra, E.H., Moura, T.R., Nascimento, K.S., and Cavada, B.S. (2009) Vasodilator effects of Diocleinae lectins from the *Canavalia* genus. *N-S Arch. Pharmacol.* **380**, 509–521
10. Andrade, J.L., Arruda, S., Barbosa, T., Paim, L., Ramos, M.V., Cavada, B.S., and Barral-Neto, M. (1999) Lectin-induced nitric oxide production. *Cell. Immunol.* **194**, 98–102
11. Gadelha, C.A.A., Moreno, F.B.M.B., Santi-Gadelha, T., Cajazeiras, J.B., Rocha, B.A.M., Assreuy, A.M.S., Mota, M.R.L., Pinto, N.V., Meireles, A.V.P., Borges, J.C., Freitas, B.T., Canduri, F., Souza, E.P., Delatorre, P., Criddle, D.N., Azevedo, W.F. Jr, and Cavada, B.S. (2005) Native crystal structure of a nitric oxide releasing lectin from the seeds of *Canavalia maritima*. *J. Struct. Biol.* **152**, 185–194
12. Gomes, J.C., Rossi, R.R., Cavada, B.S., Moreira, R.A., and Oliveira, J.T.A. (1994) Histamine release induced by glucose (mannose)-specific lectins isolated from Brazilian beans. Comparison with concanavalin A. *Agent Actions* **41**, 132–135
13. Barbosa, T., Arruda, S., Cavada, B.S., Grangeiro, T.B., Freitas, L.A.R., and Barral-Netto, M. (2001) In vivo lymphocyte activation and apoptosis by lectins of the diocleinae subtribe. *Mem. Inst. Oswaldo Cruz.* **96**, 673–678
14. Barauna, S.C., Kaster, M.P., Nascimento, K.S., Rossi, F.M., Teixeira, E.H., Cavada, B.S., Rodrigues, A.L.S., and Leal, R.B. (2006) Antidepressant-like effect of lectin from *Canavalia brasiliensis* (ConBr) administered centrally in mice. *Pharmacol. Biochem. Behav.* **85**, 160–169
15. Pinto, V.P.T., Teixeira, E.H., Teixeira, A.H., Carneiro, V.A., Cristino-Filho, G., Dus, D., Debray, H., Sampaio, A.H., and Cavada, B.S. (2010) Lectins isolated from Brazilian beans as markers of membrane glycoconjugates of human colon cancer cells. *J. Cancer Res. Exp. Oncol.* **2**, 54–59
16. Cavada, B.S., Barbosa, T., Arruda, S., Grangeiro, T.B., and Barral Netto, M. (2001) Revisiting proteus: do minor changes in lectin structure matter in biological activity? Lessons from and potential biotechnological uses of the Diocleinae subtribe lectins. *Curr. Protein Pept. Sci.* **2**, 123–135
17. Wah, D.A., Romero, A., Gallego del Sol, F., Cavada, B.S., Ramos, M.V., Grangeiro, T.B., Sampaio, A.H., and Calvete, J.J. (2001) Crystal structure of native and Cd/Cd-substituted *Dioclea guianensis* seed lectin. A novel manganese-binding site and structural basis of dimer-tetramer association. *J. Mol. Biol.* **310**, 885–894
18. Brinda, K.V., Mitra, N., Surolia, A., and Vishveshwara, S. (2004) Determinants of quaternary association in legume lectins. *Protein Sci.* **13**, 1735–1749
19. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
20. Moreira, R.A. and Perrone, J.C. (1977) Purification and partial characterization of a lectin from *Phaseolus vulgaris*. *Plant Physiol.* **59**, 783–787
21. Ferrige, A.G., Seddon, M.J., Green, B.N., Jarvis, S.A., and Skilling, J. (1992) Disentangling electrospray spectra with maximum entropy. *Rapid Commun. Mass Spectrom.* **6**, 707–711
22. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006) In gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860
23. Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structure of proteins. *Nucleic Acids Res.* **31**, 3320–3323
24. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005) Protein identication and analysis tools on the ExPASy Server in *The Proteomics Protocols Handbook* (Walker, J.M., ed.), pp. 571–607, Humana Press, Totowa, NJ
25. Jancarik, J. and Kim, S.H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409–441
26. Leslie, A.G.W. (1997) *MOSFLM User'sGuide: MOSFLM Version 550*. MRC Laboratoryof Molecular Biology, Cambridge, England
27. Evans, P.R. (1997) Scala. joint CCP4 and ESF-EACBM. *Newsletter* **33**, 22–24
28. Vargin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025
29. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760–763
30. Delano, W.L. (2002) *The Pymol Molecular Graphics System*. DeLano Scientific, San Carlos, CA
31. Cavada, B.S., Grangeiro, T.B., Ramos, M.V., Cordeiro, E.F., Oliveira, J.T.A., and Moreira, R.A. (1996) Isolation and partial characterization of a lectin from *Dioclea rostrata* Benth seeds. *R. Bras. Fisiol. Veg.* **8**, 31–36
32. Oliveira, J.T.A., Cavada, B.S., and Moreira, R.A. (1991) Isolation and partial characterization of a lectin from *Cratylia floribunda* seeds. *Revist. Brasil. Bot.* **14**, 61–66
33. Ceccatto, V.M., Cavada, B.S., Nunes, E.P., Nogueira, N.A.P., Grangeiro, M.B., Moreno, F.B.M.B., Teixeira, E.H., Sampaio, A.H., Alves, M.A.O., Ramos, M.V., Calvete, J.J., and Grangeiro, T.B. (2002) Purification and partial characterization of a lectin from Canavalia grandiflora Benth. seeds. *Protein Peptide Lett.* **9**, 67–73
34. Ramos, M.V., Moreira, R.A., Cavada, B.S., Oliveira, J.T.A., and Rouge, P. (1996) Interaction of lectins from the sub tribe Diocleinae with specific ligands. *R. Bras. Fisiol. Veg.* **8**, 193–199

35. Carrington, D.M., Auffret, A., and Hanke, D.E. (1985) Polypeptide ligation occurs during post-translational modification of Concanavalin A. *Nature* **313**, 64–67
36. Cunningham, B.A., Hemperly, J.J., Hopp, T.P., and Edelman, M. (1979) Favin versus concanavalin A: circularly permuted amino acid sequences. *Proc. Natl Acad. Sci. USA* **76**, 3218–3222
37. Chrispeels, M.J., Hartl, P.M., Sturm, A., and Faye, L. (1986) Characterization of the endoplasmic reticulum-associated precursor of concanavalin A. *J. Biol. Chem.* **261**, 10021–10024
38. Gallego del Sol, F., Cavada, B.S., and Calvete, J.J. (2007) Crystal structures of *Cratylia floribunda* seed lectin at acidic and basic pHs. Insights into the structural basis of the pH-dependent dimer–tetramer transition. *J. Struct. Biol.* **158**, 1–9
39. Delatorre, P., Rocha, B.A.M., Simões, R.C., Pereira-Júnior, F.N., Silva, H.C., Bezerra, E.H.S., Bezerra, M.J.B., Marinho, E.S., Gadelha, C.A.A., Santi-Gadelha, T., Farias, D.L., Assreuy, A.M.S., Marques-Domingos, G.F.O., Nagano, C.S., and Cavada, B.S. (2011) Mass spectrometry and X-ray diffraction analysis of two crystal types of *Dioclea virgata* lectin: an antinociceptive protein candidate to structure/function analysis. *Appl. Biochem. Biotechnol.* **164**, 741–754
40. Rangel, T.B.A., Assreuy, A.M.S., Pires, A.F., Carvalho, A.U., Benevides, R.G., Simões, R.C., Silva, H.C., Bezerra, M.J.B., Nascimento, A.S.F., Nascimento, K.S., Nagano, C.S., Sampaio, A.H., Delatorre, P., Rocha, B.A.M., Fernandes, P.M.B., and Cavada, B.S. (2011) Crystallization and characterization of an inflammatory lectin purified from the seeds of *Dioclea wilsonii*. *Molecules* **16**, 5087–5103
41. Correia, J.L.A., Nascimento, A.S.F., Cajazeiras, J.B., Gondim, A.C.S., Pereira, R.I., Sousa, B.L., Silva, A.L.C., Garcia, W., Teixeira, E.H., Nascimento, K.S., Rocha, B.A.M., Nagano, C.S., Sampaio, A.H., and Cavada, B.S. (2011) Molecular characterization and tandem mass spectrometry of the lectin extracted from the seeds of *Dioclea sclerocarpa* Ducke. *Molecules* **16**, 9077–9089
42. Sharon, N. and Lis, H. (1990) Legume lectins- A large family of homologous proteins. *FASEB J.* **4**, 3198–3208
43. Kanellopoulos, P.N., Pavlou, K., Perrakis, A., Agianian, B., Vorgias, C.E., Mavrommatis, C., Soufi, M., Tucker, P.A., and Hamodrakas, S.J. (1996) The crystal structure of the complexes of concanavalin A with 40-nitrophenyl-alpha-D-mannopyranoside and 40-nitrophenylalpha-D-glucopyranoside. *J. Struct. Biol.* **116**, 345–355
44. Bezerra, G.A., Oliveira, T.M., Moreno, F.B.M.B., Souza, E.P., Rocha, B.A.M., Benevides, R.G., Delatorre, P., Azevedo, W.F. Jr, and Cavada, B.S. (2007) Structural analysis of canavalia maritima and canavalia gladiata lectins complexed with different dimannosides: new insights into the understanding of the structure-biological activity relationship in legume lectins. *J. Struct. Biol.* **160**, 168–176
45. Oliveira, T.M., Delatorre, P., Rocha, B.A.M., Souza, E.P., Nascimento, K.S., Bezerra, G.A., Moura, T.R., Benevides, R.G., Bezerra, E.H.S., Moreno, F.B.M.B., Freire, V.N., Azevedo, W.F. Jr, and Cavada, B.S. (2008) Crystal structure of dioclea rostrata lectin: insights into understanding the pH-dependent dimer-tetramer equilibrium and the structural basis for carbohydrate recognition in Diocleinae lectins. *J. Struct. Biol.* **164**, 177–182
46. Bezerra, E.H.S., Rocha, B.A.M., Nagano, C.S., Bezerra, G.A.A., Moura, T.R., Bezerra, M.J.B., Benevides, R.G., Sampaio, A.H., Delatorre, P., and Cavada, B.S. (2011) Structural analysis of Conbr reveals molecular correlation between the carbohydrate recognition domain and endothelial NO synthase activation. *Biochem. Biophys. Res. Commun.* **408**, 566–570
47. Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497
48. Hirabayashi, J. and Kasai, K. (1991) Effect of amino acid substitution by sited-directed mutagenesis on the carbohydrate recognition andstability of human 14-kDa beta-galactoside-binding lectin. *J. Biol. Chem.* **266**, 23648–23653
49. Delatorre, P., Rocha, B.A.M., Gadelha, C.A.A., Santi-Gadelha, T., Cajazeiras, J.B., Souza, E.P., Nascimento, K.S., Freire, V.N., Sampaio, A.H., Azevedo, W.F. Jr, and Cavada, B.S. (2006) Crystal structure of a lectinfrom *Canavalia maritima* (ConM) in complex with threalose and maltose reveals relevant mutation in ConA-like lectins. *J. Struct. Biol.* **154**, 280–286