Biochemical Systematics and Ecology 60 (2015) 46–55

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

journal homepage: www.elsevier.com/locate/biochemsyseco

Lectin genes and their mature proteins: Still an exciting matter, as revealed by biochemistry and bioinformatics analyses of newly reported proteins

Andreia Varmes Fernandes [a], Márcio Viana Ramos [b], José Hélio Costa [b], Ilka Maria Vasconcelos [b], Renato de Azevedo Moreira [c], Frederico Bruno Mendes Batista Moreno [c], Maria Eliza Caldas dos Santos [a], José Francisco de Carvalho Gonçalves [a, *]

[a] Laboratory of Plant Physiology and Biochemistry, National Institute for Research in the Amazon (MCTI-INPA), Manaus, Amazonas, Brazil
[b] Federal University of Ceará (UFC), Biochemistry and Molecular Biology Department, Fortaleza, Ceará, Brazil
[c] Center for Experimental Biology (NUBEX), University of Fortaleza (UNIFOR), Health Sciences Center, Fortaleza, Ceará, Brazil

ARTICLE INFO

Article history:
Received 1 October 2014
Accepted 15 February 2015
Available online 28 March 2015

Keywords:
Hemagglutinating activity
New lectins
N-terminal protein sequencing
Seeds of Amazonian legumes

ABSTRACT

Two new lectins were purified through affinity chromatography after crude extract preparation under high ionic strength. The hemagglutinating activity of these lectins from the seeds of the legumes *Dioclea bicolor* (DBL) and *Deguelia scandens* (DSL) was inhibited by galactose and glucose, respectively, and the molecular masses were estimated at 24 and 22 kDa (via SDS-PAGE), respectively. The alignment of internal peptides of DBL (MS/MS) with known protein sequences revealed similarity to other legume lectins. The N-terminal amino acid sequence of DSL also aligned with legume lectins. Cross-similarities among the two studied lectins were observed only after sequence permutation. More than a dozen lectins have been reported for the genus Dioclea but none that recognize galactose. DSL is the first lectin reported for the Deguelia genus in the tribe Millettieae. With the aid of bioinformatics tools and searches for genome/transcriptome information about closely related sequences, new lectin members of Millettieae were also identified. Electrophoresis profiling and amino acid sequence analysis suggested that DBL-Gal and DSL do not undergo post-transcriptional ConA-like circular permutation. Molecular modeling of the deduced amino acid sequences of the Millettieae lectins suggested that the overall folding of the monomeric structures of legume lectins is conserved. This and other recent studies highlight native plants of the Amazon as renewed sources of lectins.

1. Introduction

Lectins are primarily defined as carbohydrate-binding proteins (Wittmann and Pieters, 2013) that are thought to be ubiquitous in nature and display a broad range of cellular and physiological functions in different organisms (Ingale and Hivrale, 2013). Plant lectins, notably those belonging to legume seeds, have become suitable models for molecular and

* Corresponding author.
E-mail address: jfc@inpa.gov.br (J.F. de Carvalho Gonçalves).

http://dx.doi.org/10.1016/j.bse.2015.02.002
0305-1978/

structural studies involving protein-carbohydrate interactions at the atomic level (de Bentzmann et al., 2014). Indeed, their ability to discriminate specific structural features of carbohydrate-containing molecules highlights them as interesting tools for a wide range of applications, including diagnostics and the treatment of metabolic diseases (Kontro et al., 2014). Accordingly, screening for new lectins is a still attractive approach (Nascimento et al., 2012). However, screenings are still based on the evaluation of the erythrocyte agglutinating activity of protein extracts, and the success of this strategy has become marginal. With new DNA sequences of plant species now available, the use of bioinformatics allows the prospecting of lectin genes *in silico*, substantially improving bio-prospecting. In this study, we report two new purified legume lectins and two new lectin genes (cDNAs). Based on the amino acid sequence information of the new proteins, interesting insight on their phylogenetic clustering and overall folding are reported.

2. Material and methods

2.1. Chemicals

Chemicals for electrophoresis, α-lactose-agarose, Sephadex G-50, and carbohydrates were purchased from Sigma Chemical (São Paulo, Brazil). Chemicals for amino acid sequencing were obtained from Waco (Tokyo, Japan). Molecular weight markers were obtained from PROMEGA (São Paulo, Brazil). All other chemicals were of analytical grade.

2.2. Plant material

Seeds of *Dioclea bicolor* Benth. (Fabaceae Lindl.)were obtained of plants located at the following coordinates (04° 22′ 59″ S and 70° 01′ 52″ W), locally named Benjamim Constant, Amazon, Brazil. Seeds of *Deguelia scandens* Aubl. (Fabaceae Lindl.) were harvested of plants found on Judge Island in the downstream region of the Black River, São Gabriel da Cachoeira, Amazon, Brazil, at the following coordinates (00° 08′ 49.0″ S and 67° 04′ 29.0″). The collections were performed in August 2009.

2.3. Protein extraction

Fine seed flour of *D. bicolor* was suspended in 250 mM glycine-HCl (pH 2.6)-buffered saline (150 mM) and maintained at 8 °C. The resulting mixture (1:10; v/w) was shaken for 2 min and centrifuged for 10 min at 10,000 × g and 4 °C. The soluble phase was stored at 8 °C, and the precipitated material was subjected to a new cycle of extraction. The new soluble phase was combined with the first and subjected to dialysis in distilled water for 48 h at 8 °C, with the water being renewed eight times.

Fine seed flour of *D. scandens* was mixed with 50 mM Tris–HCl (pH 7.5)-buffered saline (150 mM) and shaken for 2 h at 25 °C. The resulting extract (1:10; v/w) was centrifuged and dialyzed as stated for the first sample described. The materials were freeze-dried stored at 25 °C and used in all analyses.

2.4. Lectin activity

Lectin activity in both samples was first assessed by hemagglutinating assays and further confirmed by using glucose and galactose as lectin inhibitors. The overall procedure described by Moreira and Perrone (1977) was followed. The freeze-dried materials were dissolved in 100 mM Tris–HCl (pH 7.6)-buffered saline (150 mM) added to 5 mM $CaCl_2/MnCl_2$. Rabbit erythrocytes (2%) in 150 mM saline were used for agglutination. The inhibition of agglutination was investigated by first incubating the samples [extracts (10 mg/ml) or purified lectins (1 mg/ml)] with different concentration of the following carbohydrates prepared in 150 mM saline at 100 mM, prior to adding the red blood cells: N-acetyl-D-galactosamine; α-lactose; 2-nitrophenyl-ß-D-galactopyranoside; D-galactose; ß-methyl-D-galactopyranoside; α-methyl-D-galactopyranoside; D-mannose; N-acetyl-D-glucosamine; D-glucose; α-methyl-D-glucopyranoside; D-fructose; L-arabinose; sucrose and maltose.

2.5. Affinity chromatography

The choice of the matrices for lectin purification was according to the results of the hapten inhibition of the extracts by the monosaccharides glucose and galactose. Accordingly, the extract of *D. bicolor* was subjected to fractionation through an α-lactose-agarose matrix. The extract of *D. scandens* was applied to a Sephadex G-50 resin. Both gels were first washed with Tris–HCl buffer, and then the samples, prepared in the same buffer, were applied to the columns and first eluted with Tris–HCl buffer until the unbound proteins were recovered. The lectins, which were expected to be retained in the columns, were recovered by exchanging the Tris–HCl buffer for 100 mM glycine (pH 2.6)-buffered saline (150 mM) in the α-lactose-agarose column and 100 mM glucose in the Sephadex G-50 resin. The elution of the proteins was monitored at 280 nm. All runs were performed at 25 °C. The lectin-containing fractions were dialyzed as described above and freeze-dried.

2.6. Electrophoresis

The crude extracts (10 mg/ml) and the purified lectins (1 mg/ml) were electrophoresed through a 12.5% polyacrylamide gel under denaturing conditions. The gels were stained with Coomassie brilliant blue R-250 to visualize the proteins. The experimental protocol was performed according to Laemmli (1976).

2.7. Amino acid sequence analysis

The N-terminal sequencing of *Deguelia scandens* lectin was performed according to the Edman degradation protocol using an automated peptide sequencer (Shimadzu PPSQ-23).

DBL (1 mg/ml) was treated with dithiothreitol (10 mM) and alkylated (10 mM iodoacetamide). The treated lectin was enzymatically digested (trypsin, 1:100 enzyme/protein ratio), and the resulting peptides were *de novo* sequenced using a Synapt HDMS mass spectrometer (Waters, Manchester, UK). The experimental conditions were those proposed by Fernandes et al. (2012). The set of peptides sequencing was used to perform a BLAST search against the nr or refseq-protein databases (NCBI) for protein identification.

2.8. Bioinformatics

Data from transcript libraries [Sequence Read Archive (SRA)] available at GenBank (http://www.ncbi.nlm.nih.gov) were used to retrieve homologous sequences of Millettieae lectins by BLAST searches using the DSL data as a reference. The SRA sequences were assembled into contigs (full-length cDNAs) using the CAP3 Sequence Assembly Program (Huang and Madan, 1999). The cDNAs were translated into amino acid sequences using the translate tool at the ExPASy web server (http://web.expasy.org/translatehttp://www.infobiogen.fr/services/analyseq/cgi-bin/traduc_in.pl).

2.9. Sequence alignment and phylogenetic analysis

A multiple sequence alignment was carried out using Clustal X (Larkin et al., 2007). The phylogenetic tree of lectins was constructed using the neighbor-joining method (Saitou and Nei, 1987) in the MEGA5 program (Tamura et al., 2011). The alignment was performed using the "blosum" substitution matrix, and the phylogenetic tree was constructed using the following parameters: bootstrap method with 1000 replications, "number of differences" as the substitution model and "complete deletion" for gaps/missing data treatment. All proteins used for phylogenetic tree are available in the supplementary data (Supplementary material txt file).

2.10. 3D structure modeling of lectins from Millettia pinnata (L.) Panigrahi (Fabaceae Lindl.)

The construction of the 3D structures of lectins from *M. pinnata* was carried out by employing interactive implementation of the threading assembly refinement (I-TASSER) (Zhang, 2008; Roy et al., 2010, 2012).

3. Results

Images of the seeds and certain characteristics of them are given as supplementary material (Supplementary Material Table S1). The crude extracts agglutinated rabbit erythrocytes at the minimum concentration of 0.15 mg/ml (DBL) and 0.31 mg/ml (DSL). Galactose (10 mM) but not glucose (100 mM) eliminated the lectin activity of the *Dioclea* extract. Accordingly, the lectin was retained on α-lactose agarose resin and was purified to homogeneity after release under acid conditions (Fig. 1A). In contrast, glucose (10 mM) but not galactose (100 mM) inhibited the agglutination of the *Deguelia* extract. Thus, to purify the lectin, the crude extract was chromatographed using a Sephadex G-50 column. As observed in Fig. 1B, a protein fraction was retained on the column and recovered after adding 100 mM glucose. The proteins isolated by affinity chromatography were named DBL-Gal (*Dioclea bicolor* lectin – galactose specific) and DSL (*Deguelia scandens* lectin). As attempts to determine the N-terminal amino acid sequence of DBL-Gal by Edman degradation failed, it was assumed that the N-terminal amino acid residue was blocked. The N-terminal amino acid sequence of DSL suggested sequence identity with other legume lectins (Table 1), and the best identity scores were found with other lectins from the same taxonomic group (Millettieae). To search for similarities to DBL-Gal, the protein was digested with trypsin, and the resulting fragments were subjected to MS/MS analyses. Four independent short peptides sequences were obtained (Supplementary Material Table S2). However, using this set of sequences to perform a BLAST search yielded no protein identification. Randomly assembling the individual sequences to obtain their correct position in the protein sequence allowed us to generate three longer peptides, which were used to perform a new BLAST search. The longer peptides adequately aligned with sequences of other legume lectins (Table 2) using local (BLAST) and global (ClustalW2) alignments. The molecular masses of DBL-Gal and DSL estimated by electrophoresis were 24 and 22 kDa, respectively (Fig. 2). These values were consistent with data reported for many legume lectins (Peumans et al., 2001). Glucose, N-acetyl-D-glucosamine, α-methyl-D-glucopyranoside, D-fructose, L-arabinose, sucrose and maltose inhibited DSL at concentrations lower than 25 mM but had no effect on the activity of DBL-Gal. In contrast, glycosides of the galactose series, galactose,

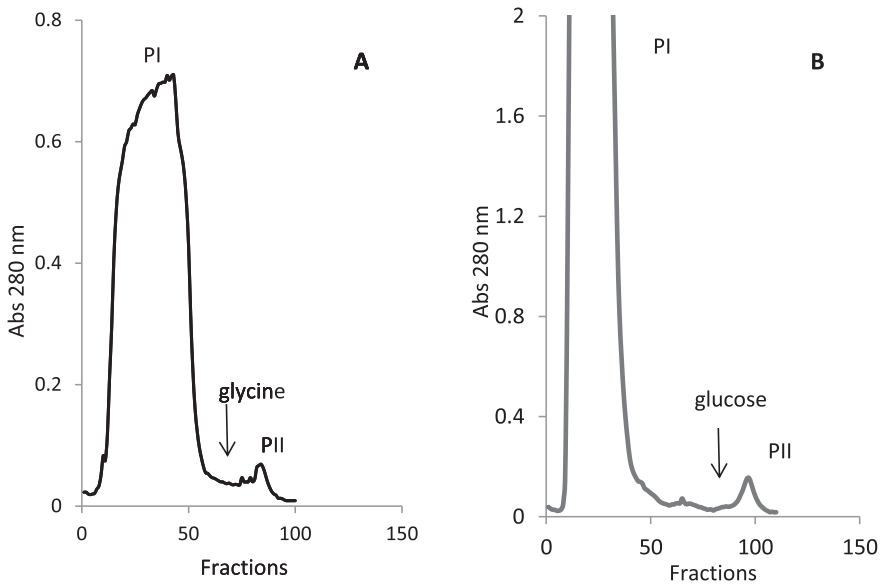

Fig. 1. Affinity Chromatography. A - *Dioclea bicolor* extract on α-lactose – agarose column (20 ml). B - *Deguelia scandens* extract on Sephadex G 50 (10 ml). The columns and the samples were prepared in 50 mM Tris–HCl pH 7.6 with 150 mM NaCl, 5 mM $MnCl_2$ and $CaCl_2$. For the analyses, the samples [20 ml, 2.5 mg/ml] and [10 ml, 10 mg/ml] were applied to the columns, respectively. Unbound proteins (PI) were washed form the columns with the starting buffer Tris. Proteins interacting to the columns (PII) were eluted with 100 mM glycine-HCl pH 2.6 (A) and 100 mM glucose (B). Fractions (2 ml) were collected at a flow rate of 0.2 ml/min and monitored by spectroscopy at 280 nm. Fractionated proteins were further analyzed by electrophoresis.

N-acetyl-D-galactosamine, α-lactose, 2-nitrophenyl-ß-D-galactopyranoside, ß-methyl-D-galactopyranoside and α-methyl-D-galactopyranoside, were strong inhibitors of DBL-Gal at concentrations lower than 12.5 mM.

To our knowledge, DSL is the first lectin reported for the taxon Millettieae. Therefore, we searched for genomic and transcript data for other members of Millettieae in GenBank (NCBI) using the amino terminal sequence of DSL as a query. Two full-length cDNA sequences from *M. pinnata* were identified as putative lectin cDNAs (Supplementary Material Fig. S1). The deduced amino acid sequences exhibited sequence identity to that of DSL and other legume lectins (Table 1). The three internal peptides of DBL-Gal also aligned with both *Millettia* sequences (Fig. 3). It is worth noting that the two deduced amino acid sequences of *Millettia* lectins shared only 35.82% sequence identity (Fig. 3), and they clustered in distinct clades compared with known lectin sequences (Fig. 4). Curiously, DBL-Gal grouped close to MpL2, whereas DSL was closer to MpL1. Three dimensional models built for MpL1 and MpL2 revealed the conserved dome-shaped structure commonly reported for legume lectins (Fig. 5).

Table 1
Comparative analysis of the N-terminal amino acid sequence of seed lectin from *Deguelia scandens* with other amino acid sequences of plant lectins. The bold and underlined amino acids refer to those which are conserved in lectins of *Deguelia scandens* and in other plants.

Species	Accession number	Alignment of N-terminal amino acid sequence	Identity (%)
Deguelia scandens	**This study**	**ADS**V**S**F**N**F**P**D**N**Y**P**GR**DQ**G**NLIFQGDA**	–
Pterocarpus angolensis	CAD19803.1	Q**DS**L**SF**G**FP**-TF**P**S-**DQ**K**NLIFQGDA**	66.6
Pterocarpus angolensis	gi|46015347	X**DS**L**SF**G**FP**-TF**P**S-**DQ**K**NLIFQGDA**	66.6
Platypodium elegans	gi|371927765	T**DS**L**SF**SFI-**N**FDR **D**ER**NLIFQGDA**	62.5
Cladrastis kentukea	Q39529.1	S**DS**L**SF**T**F**D-**N**FRP-**DQ**RD**LI**L**QGDA**	58.3
Sophora flavescens	ACD13798.1	**ADS**L**SF**T**F**S-DFDP-NGED**L**L**FQGDA**	58.3
Sophora flavescens	AAG00508.1	**ADS**L**SF**T**F**S-DFDP-NGED**L**L**FQGDA**	58.3
Sophora alopecuroides	AAY68291.1	**ADS**L**SF**T**F**S-DFNQ-NEED**L**L**FQGDA**	54.1
Millettia pinnata 1	**This study**	**AN**S**SF**T**F**D-**N**FTP-Q**Q**ED**LI**L**QGDA**	54.1
Sophora japonica	AAB51442.1	S**DS**L**SF**T**F**N-**N**FGP-**DQ**RD**LI**L**QGDA**	54.1
Arachis hypogaea	ABJ15831.1	L**DS**L**SF**SYN-**N**FEQD**D**ER**NLI**L**QGDA**	52.0
Sophora alopecuroides	ACD39390.1	**ADS**L**SF**T**F**S-**N**FDQ-NEED**L**L**FQGDA**	50.0
Cladrastis kentukea	Q39528.1	S**DS**L**SF**T**F**N-**NF**P**P**-NSED**LIFQ**K**DA**	50.0
Pterocarpus rotundifolius	AAT57665.1	S**DS**FP**F**G**FF**-**N**FDQ-**DERNLI**Y**QGDA**	50.0
Millettia pinnata 2	**This study**	T**D**TT**SF**S**F**T-KFHS-N**Q**P**NLI**L**QGDA**	50.0
Sophora japonica	AAB51457.1	S**DS**L**SF**TYE-**N**FQP-NPED**LI**L**QRDA**	41.7

Sequences were aligned using the ClustalW2 to obtain the amino acid identity (%).

Table 2
Comparative analysis of 3 internal peptides of DBL-Gal isolated from *Dioclea bicolor* seeds with corresponding peptides of other leguminous plant lectins.

Species	Accession Number	Alignment of peptides			Identity (%)
		Peptide 1	Peptide 2	Peptide 3	
Dioclea bicolor	**This study**	VNSAEILSFSFPK	HIGIDVNSVK	SILPEWVRVGFSATSGLSR	–
Sophora japonica	P93538.1	VNSAEILSFSFPK	HIGIDVNSVK	SILPEWVRVGFTATTGLTT	90.5
Sophora japonica	P93535.1	VNSAEILSFSFPK	HIGIDVNSIN	SILPEWVRVGFTAATGLTT	83.3
Vatairea macrocarpa	P81371.1	–SEVVSFSFTK	HIGINVNSIE	SALPEWVRVGFSATSGLSR	79.5
Vatairea guianensis	P86893.1	–SEVVSFSFTK	HIGINVNSIE	SALPEWVRVGFSATSGLSR	79.5
Cladrastis kentukea	Q39527.1	VNSEEALSFTFTK	HIGIDVNSIE	SILPEWVRVGFSAATGRSA	73.8
Robinia pseudoacacia	Q42372.1	VNSTGSLSFSFPK	HMGIDVNSIQ	KVLPEWVRVGFTATTGLSE	73.8
***Millettia pinnata* 1**	**This study**	–TDTTSFSFTK	HIGIDVNSIK	SVLPEYVSVGFSATTGASD	69.2
Robinia pseudoacacia	BAA36414.1	VNSTGSLSFSFPK	HLGINVNSIK	DVLPEWVRFGFSATTGIDK	69.1
Robinia pseudoacacia	BAA36415.1	VNSTESVSFSFTK	HIGIDVNSIN	DVLPQFVDVGFSATTGLSE	69.1
Robinia pseudoacacia	Q41159.1	VNSTGSLSFSFPK	HMGINVNSIV	DVLPEWVRFGFSATTGIDK	66.7
Robinia pseudoacacia	Q41161.1	VNSTGSLSFSFPK	HMGINVNSIV	DVLPEWVRFGFSATTGIDT	66.7
Robinia pseudoacacia	Q41162.1	VNSTGSLSFSFPK	HMGINVNSIV	DVLPEWVRFGFSATTGIDT	66.7
Sophora japonica	AGA94529.1	VNSTGSLSFSFPK	HMGINVNSIV	DVLPEWVRFGFSATTGIDT	66.7
Robinia pseudoacacia	gi|15826665	–TGSLSFSFPK	HLGINVNSIE	IVLPEWVRFGFSATTGIDK	64.1
Ulex Europaeus	gi|6573436	-NLSDDLSFNFDK	HIGIDVNSIK	AILPEWVSVGFSGGVGNAA	63.4
Dioclea guianensis	gi|16085810	IADANSLHFSFSQ	HIGIDIKSIR	NVLPEWVRVGLSATTGLYK	57.1
Dioclea grandiflora	CAM91962.1	IADENSLHFSFHK	HIGIDIKSIR	NVLPEWVRVGLSATTGLYK	57.1
***Millettia pinnata* 2**	**This study**	–ANSLSFTFDN	HVGVNVNSIR	EVLPEWVRVGFSGSTGQYT	53.9

Amino acid numbering in the peptides using *Sophora japonica* lectin (P93538.1) as reference is: peptide 1 = 13 to 25; peptide 2 = 155 to 164; peptide 3 = 215 to 233. Peptides were aligned using the ClustalW2 to obtain the amino acid identity (%).

4. Discussion

Fabaceae has been historically screened for lectins since the 1980s. The search for new lectins in this taxon have been motivated by the pioneer study on a lectin (ConA) from the legume seeds of *Canavalia ensiformis*, which included the characterization of its sugar-biding specificity, structural data (amino acid sequence and crystallography) and unusual post-transcriptional events (Carrington et al., 1985). Two legume taxa (Vicieae and Diocleinae) have primarily been studied because a similar and practical protocol for purifying legume lectins became available from studies with ConA. Seeds of the legumes of tribe Vicieae, comprising the genera Pisum, Lathyrus and Vicia, and the many species belonging to the sub-tribe Diocleinae (Phaseoleae), comprising the genera Canavalia, Dioclea and Cratylia became the main sources of newly purified Fabaceae lectins (Calvete et al., 1999; Peumans et al., 2001). In the following years, extensive molecular studies with new purified Fabaceae lectins permitted to identify some of their common properties such as primary sequence homology and a much conserved folding of their monomers. The molecular diversity however seems to be justified by the multiplicity of way legume lectins assemble their quaternary associations, resulting in distinct quaternary folding (Srinivas et al., 2001; Calvete et al., 1999). The new lectins reported here are compared to other lectins from the Fabaceae. To conclude, the new lectin DBL-Gal differs, in its carbohydrate-binding specificity, from other lectins from the same clade.

Dioclea bicolor lectin (DBL-Gal). Taxonomic position: Fabaceae (family); Faboideae (subfamily); Phaseoleae (tribe); Diocleinae (subtribe). Approximately thirteen genera are included in clade Diocleinae. All lectins reported in the genera Canavalia and Cratylia exclusively bind glucose/mannose and their methyl or (α, β) *p*-nitrophenyl derivatives (Ramos et al., 1996). With the exception of *Dioclea lehmanii*, for which two distinct lectins have been reported, all other Dioclea lectins share very similar carbohydrate-binding properties to those of Canavalia and Cratylia (Calvete et al., 1999). The lectins belonging to *D. lehmanii* bind to glucose/mannose (DLL-I) and lactose/sucrose and melezitose (DLL-II) and therefore are distinct proteins (Pérez et al., 1990; Pérez, 1998). Additionally, two different lectins were later purified from the seeds of *Cymbosema roseum* (sub-tribe Diocleinae), with sugar distinct specificity toward mannose (Cavada et al., 2006) or lactose (Rocha et al., 2009). Nonetheless, our attempts to detect glucose/mannose-binding lectins in the seed protein extract of *Dioclea bicolor* failed. Hence, this is the first case in which a unique galactose/lactose-binding lectin is reported in Dioclea seed extracts.

Deguelia scandens lectin (DSL). Taxonomic position: Fabaceae (family); Faboideae (subfamily); Millettieae (tribe). Almost thirty-five genera are included in this later clade. Although legume seeds have been screened in detail for lectins, DSL is the first lectin reported for the genus Deguelia. For that reason, only limited comparisons with other legume lectins are initially possible. However, DSL was classified as a glucose/mannose-binding lectin, and the analysis of its N-terminal amino acid sequence revealed similarities to other legume lectins. Curiously, using the sequence data of DSL to perform an *in silico* bioinformatics analysis led us to the genome/transcriptome data of another genus of Millettieae with lectins not previously reported.

Millettieae lectins (MpL1/MpL2). Taxonomic position: Fabaceae (family): Faboideae (subfamily): Millettieae (tribe). Thirty tribes are included in Faboideae, and the genera Millettia and Deguelia are thus placed in the same clustering. To our knowledge, there is no previous report of lectins in Millettia, and the unique report of lectin screening in Deguelia was based on our previous study (Fernandes et al., 2011). The N-terminal amino acid sequence of DSL shares similarity with the two deduced amino acid sequences of Millettia cDNAs (see Table 1). MpL1 and MpL2 also exhibit similarity with DBL-Gal (see

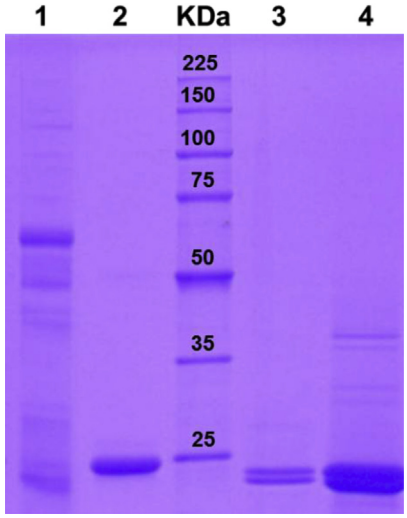

Fig. 2. Denaturing polyacrylamide gel electrophoresis (12.5%) of protein extracts from seeds of *Dioclea bicolor* (1) and *Deguelia scandens* (4). The purified lectins are shown in lane 2 (DBL-Gal) and 3 (DSL). Lectins (10 μg) were loaded on the gel. Molecular Weight markers (KDa) were from PROMEGA (Broad Range Protein Molecular Weight Markers) and the proteins were visualized by staining with Coomassie brilliant blue R-250.

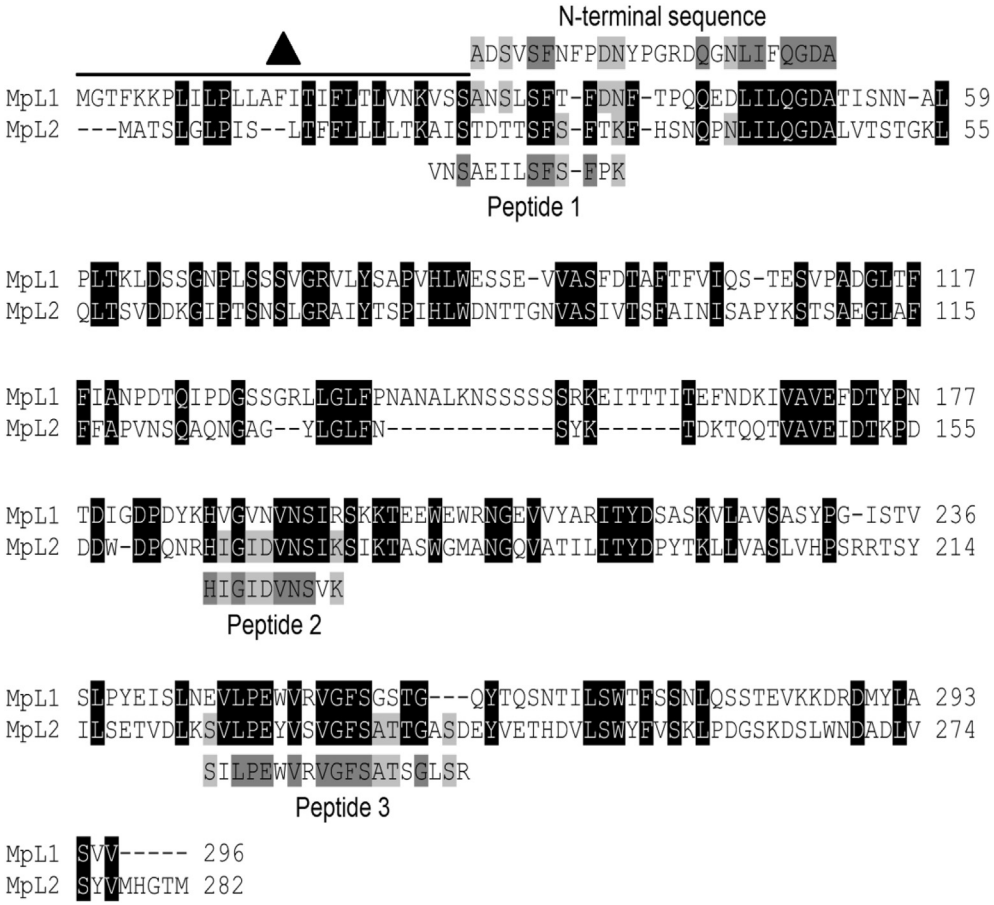

Fig. 3. Alignment of the deduced amino acid sequences of *Millettia pinnata* lectins 1 (MpL1) and 2 (MpL2). The internal peptides from *Dioclea bicolor* lectin as well as the N-terminal amino acid sequence from *Deguelia scandens* lectin are indicated. The alignment was generated by the CLUSTALW2 algorithm. Identical amino acids residues between MpL1 and MpL2 are shown on a black background. Complete conserved amino acids between lectins of *Millettia pinnata* and *Dioclea bicolor* or *Deguelia scandens* are shown in gray while partial conserved amino acids are shown in light gray. The predicted signal peptides of MpL1 and MpL2, deduced by PSORT software (Nakai and Kanehisa, 1992), are indicated by one filled underlined triangle.

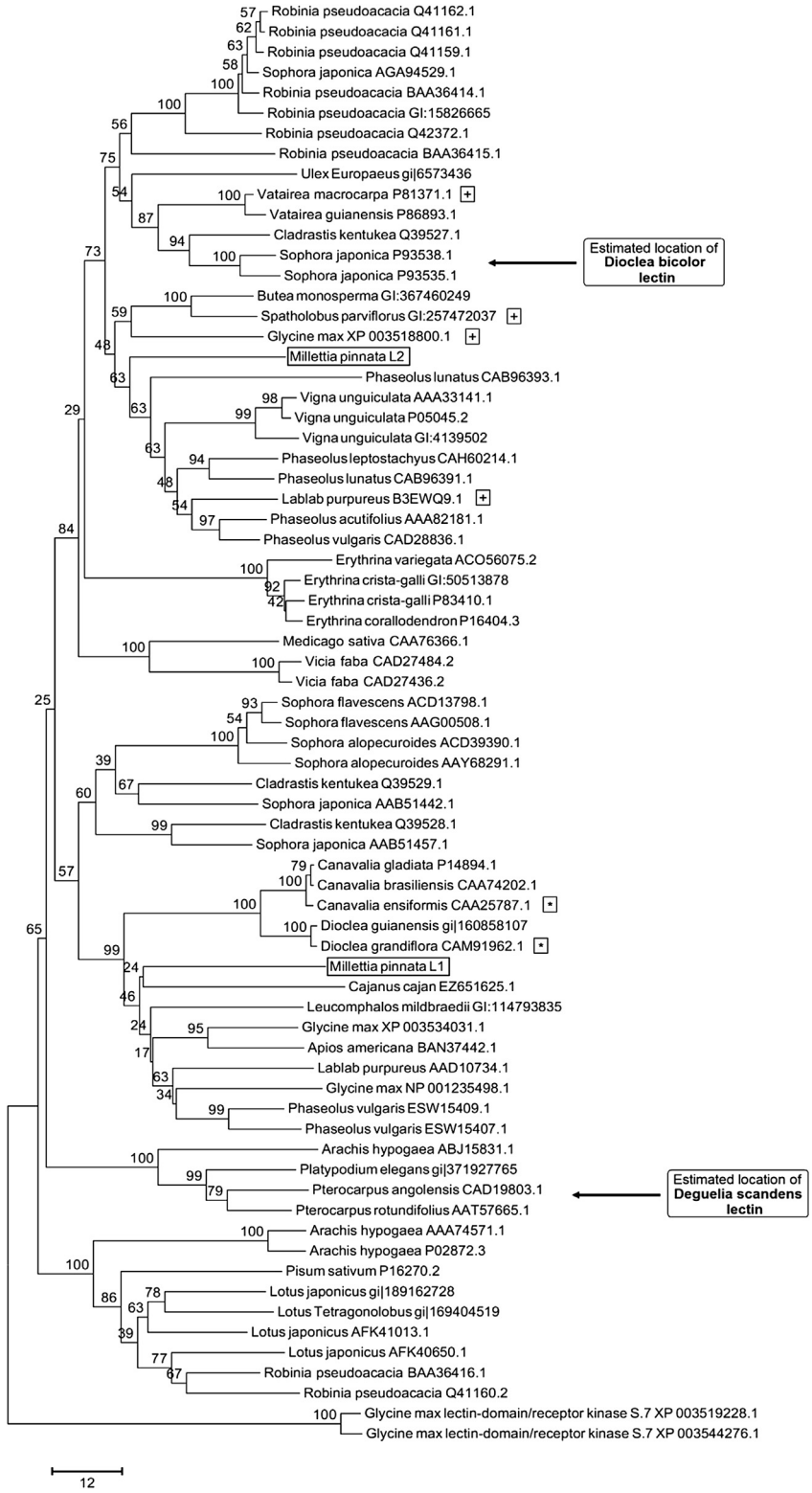

Fig. 4. Phylogenetic tree of representative lectins from species of the taxon Fabales (order). Estimated locations of *Deguelia scandens* and *Dioclea bicolor* lectins are indicated by arrows. The two *Millettia pinnata* lectins are indicated within rectangles. * (gray) indicate the lectin structures used in modeling of *Millettia pinnata* L1. + (gray) indicate the lectin structures used in modeling of *Millettia pinnata* L2. Two lectin-domain/receptor kinase S.7 were used as out-group. Accession numbers of the sequences in GenBank (NCBI) are shown after scientific names of each species.

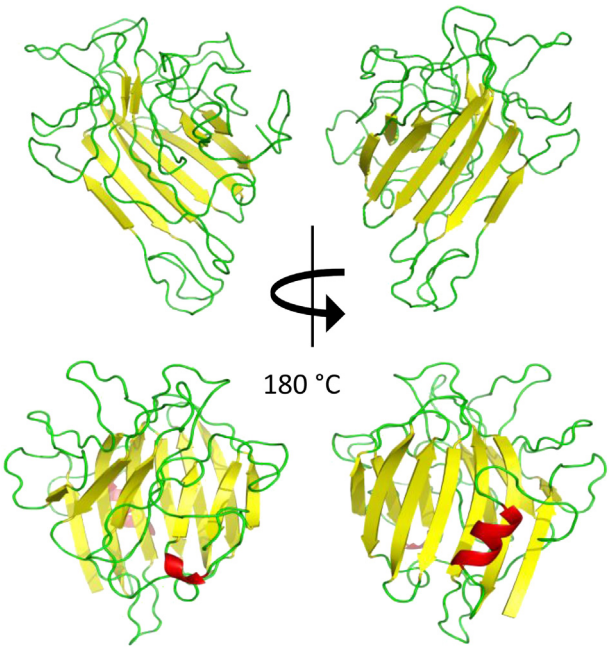

Fig. 5. 3-D overall view of MpL-1 (up) and MpL-2 (down) folding. Front (left) and back (right) views of the models. Images generated and handled by PyMol software. Molecular modeling was performed as described in material and methods. Both models adopted the highly conserved shape found in legume lectins with two sets of β-sheet sandwich segments and almost free of α-helix segments.

Table 2). Therefore, the N-terminal amino acid sequence of DSL and the internal peptide sequences of DBL-Gal aligned with the deduced amino acid sequences of MpL1 and MpL2 (Fig. 3). The identity of MpL1 with MpL2 (Fig. 3) was only partial (35.82%), suggesting two mature lectins with distinct properties. However, we cannot comment on the sugar-binding characteristics of MpL1 and MpL2.

Clustering of new Faboideae lectins. It is noteworthy that the estimated grouping of DBL-Gal in Faboideae was closer to Sophora lectins, belonging to the tribe Sophoreae, instead of grouping closer to the lectins of the subtribe Diocleinae, belonging to the tribe Phaseoleae. As observed in the phylogenetic tree shown in Fig. 4, DBL-Gal was placed far from the lectins belonging to Dioclea and Canavalia. This unusual clustering will be discussed later. The estimated location of DSL was closer to the Pterocarpus genus, belonging to tribe Dalbergieae. Remarkably, MpL1 and MpL2 were segregated into distinct clusters: MpL1 was placed almost close to Dioclea lectins, whereas MpL2 was placed within the vicinity of DBL-Gal. This is in agreement with the poor similarity between MpL1 and MpL2 observed for their deduced amino acid sequences, as shown in Fig. 3, and strongly supports that two distinct mature proteins exist in *M. pinnata*. In addition, both proteins were deduced from transcriptome data of different tissues (MpL1 from roots and MpL2 from leaves), indicating that the corresponding genes are functional and differentially expressed.

Structural homology. To build 3-D models of MpL1 and MpL2, distinct crystallography datasets were used (see Fig. 4). To obtain the models presented in Fig. 5, the crystallographic coordinates of ConA were used as a template to model MpL1, and the crystallographic data of the lectin from *Vatairea macrocarpa* (VML) was the template to MpL2. Both templates were adopted according to the best similarity of each with *Millettia* lectins and the availability of crystallographic coordinates (Supplementary Material Fig. S2). Not surprisingly, the 3-D models of MpL1 and MpL2 closely resembled the overall folding of ConA and VML and displayed the known features observed in almost all legume lectins, highlighting the antiparallel β-sheet segments in the monomer. Remarkably, the 3-D models of MpL1 and MpL2 also resemble each other, in agreement with another study reporting the comparative structural analysis of legume lectins (Sinha et al., 2007). Overall, a decrease in the sequence identity among legume lectins (as observed here for MpL1 and MpL2) does not indicate significant alteration in the overall topology of the legume lectin monomers. Indeed, structural diversity among legume lectins is typically observed in their quaternary associations (Sinha et al., 2007). Thus, the analysis performed here primarily supports that two distinct lectin genes are present in *M. pinnata* and that the two putative proteins share structural similarities to other legume lectins. To a lesser extent, the same conclusion is applicable to DBL-Gal and DSL, as suggested by their peptide amino acid sequences. However, more experimental evidence is needed to strengthen this hypothesis.

Sequence permutation. An aspect of legume lectins that continues to attract attention is their posttranslational processing. Despite the homology found among the nucleotide sequence of legume lectins and their highly conserved three-dimensional structures, a variety of posttranslational and posttranscriptional events are known (Calvete et al., 1998). The most intriguing event however is the circular permutation of peptide segments observed during the biosynthesis of ConA (Carrington et al.,

1985; Cunningham et al., 1979); as a result, the amino acid sequence of the mature protein is poorly correlated with the gene sequence. The proteolytic processing of the lectin VML was reported to resemble that of ConA in many aspects, though peptide permutation was not observed (Calvete et al., 1998). In this regard, little can be discussed regarding the two new legume lectins DBL-Gal and DSL reported here due to our limited information about the entire amino acid sequences. However, the electrophoretic analysis of these proteins revealed that neither DBL-Gal or DSL exhibited additional fragments as a result of natural proteolytic processing, as observed for ConA (and almost all Canavalia and Dioclea lectins) and VML (Wang et al., 1971). This observation suggests that both proteins (DBL-Gal and DSL) are composed of a unique polypeptide chain and most likely do not undergo ConA-like posttranscriptional events (Fig. 2). This hypothesis is better supported in the case of DSL because a unique N-terminal amino acid sequence was obtained by Edman degradation. The best sequence alignment of MpL1 was achieved when circular peptide permutation was used with the ConA sequence (Supplementary Material Fig. S2); MpL2 aligned better (and MpL1) with VML without sequence permutation (Supplementary Material Fig. S2.). However, this information is not solid enough to reveal whether MpL1 (and MpL2) undergoes peptide permutation. In both cases, the homology observed among the sequences with or without simulation of circular permutation supported the 3-D models exhibiting rather similar folding.

Concluding Remarks. New legume lectins are reported. DBL-Gal differs from other lectins studied in the same clade in terms of its carbohydrate-binding specificity. This is the first report of the absence of glucose-binding lectins in the Dioclea genus. The electrophoretic profile and phylogenetic analysis suggest that DBL-Gal does not share similarities with ConA-like lectins in terms of posttranscriptional events, and sequence homology was only observed by applying the circular permutation of the ConA sequence.

DSL is the first lectin reported for the genus Deguelia. An unique N-terminal amino acid sequence was determined for DSL. However, internal micro heterogeneity is likely to occur in DSL full primary sequence. This is likely because the electrophoresis profile of the protein exhibited two protein bands with quit similar predicted molecular masses. Closely related isoforms arising of micro heterogeneity in the amino acid sequence is probably the most reported characteristic of legume lectins (Peumans et al., 2001).

An *in silico* structural analysis of the two lectin genes of *M. pinnata* suggested two distinct lectins are coded. Despite the poor homology revealed by sequence alignment, both predicted sequences were similarly folded at the 3-D level. This study contributes to the understanding of lectins, lectin genes and their use as molecular tools in legume phylogeny, a matter frequently revisited.

Acknowledgments

This study is supported by grants from the following Brazilian Agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq: Universal, project number 480233/2011-0) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). This study is part of the consortium Pro-Amazonia-CAPES (project number 52 – AUXPE n° 3390/2013). J.F.C. Gonçalves, M.V. Ramos and Ilka Maria Vasconcelos are researchers of the Brazilian Council for Research and Development (CNPq). We also thank AJE for revising the English manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bse.2015.02.002.

References

Calvete, J.J., Santos, C.F., Mann, K., Grangeiro, T.B., Nimtz, M., Urbanke, C., Sousa-Cavada, B., 1998. Amino acid sequence, glycan structure, and proteolytic processing of the lectin of *Vatairea macrocarpa* seeds. FEBS Lett. 425, 286–292. http://dx.doi.org/10.1016/S0014-5793(98)00243-9.

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., Cavada, B.S., 1999. Molecular characterization and crystallization of Diocleinae lectins. Biochim. Biophys. Acta 1430, 367–375.

Carrington, D.M., Auffret, A., Hanke, D.E., 1985. Polypeptide ligation occurs during post-translational modification of concanavalin A. Nature 31, 64–67. http://dx.doi.org/10.1038/313064a0.

Cavada, B.S., Marinho, E.S., Souza, E.P., Benevides, R.G., Delatorre, P., Souza, L.A., Nascimento, K.S., Sampaio, A.H., Moreno, F.B., Rustiguel, J.K., Canduri, F., Azevedo Jr., W.F., Debray, H., 2006. Purification, partial characterization and preliminary X-ray diffraction analysis of a mannose-specific lectin from *Cymbosema roseum* seeds. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 62, 235–237. http://dx.doi.org/10.1107/S174430910600371X.

Cunningham, B.A., Hemperly, J.J., Hopp, T.P., Edelman, G.M., 1979. Favin versus concanavalin A: circularly permuted amino acid sequences. Proc. Natl. Acad. Sci. U. S. A. 76, 3218–3222.

de Bentzmann, S., Varrot, A., Imberty, A., 2014. Monitoring lectin interactions with carbohydrates. Methods Mol. Biol 1149, 403–414. http://dx.doi.org/10.1007/978-1-4939-0473-0_32.

Fernandes, A.V., Ramos, M.V., Gonçalves, J.F.C., Maranhão, P.A., Chevreuil, L.R., Souza, L.A.G., 2011. Seeds of *Amazonian Fabaceae* as a source of new lectins. Braz. J. Plant Physiol. 23, 237–244.

Fernandes, A.V., Ramos, M.V., Vasconcelos, I.M., Moreira, A.C., Moreno, F.B., Pereira, J.O., Gonçalves, J.F.C., 2012. Purification and characterization of a lectin of the Swartzieae Legume Taxa. Protein Pept. Lett. 19, 1082–1088. http://dx.doi.org/10.2174/092986612802762679.

Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877. http://dx.doi.org/10.1101/gr.9.9.868.

Ingale, A.G., Hivrale, A.U., 2013. Plant as a plenteous reserve of lectin. Plant Signal Behav. 8, e26595. http://dx.doi.org/10.4161/psb.26595.

Kontro, H., Joenväärä, S., Haglund, C., Renkonen, R., 2014. Comparison of sialylated N-glycopeptide levels in serum of pancreatic cancer patients, acute pancreatitis patients and healthy controls. Proteomics 14, 1713–1723. http://dx.doi.org/10.1002/pmic.201300270.

Laemmli, U.K., 1976. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948. http://dx.doi.org/10.1093/bioinformatics/btm404.

Moreira, R.A., Perrone, J.C., 1977. Purification and partial characterization of a lectin from *Phaseolus vulgaris*. Plant Physiol. 59, 783–787. http://dx.doi.org/10.1104/pp.59.5.783.

Nakai, K., Kanehisa, M., 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14, 897–911. http://dx.doi.org/10.1016/S0888-7543(05)80111-9.

Nascimento, K.S., Cunha, A.I., Nascimento, K.S., Cavada, B.S., Azevedo, A.M., Aires-Barros, M.R., 2012. An overview of lectins purification strategies. J. Mol. Recognit. 25, 527–541.

Pérez, G., 1998. Isolation and characterization of a novel lectin from *Dioclea lehmanni* (Fabaceae) seeds. Int. J. Biochem. Cell. Biol. 30, 843–853. http://dx.doi.org/10.1016/S1357-2725(98)00045-4.

Pérez, G., Hernandez, M., Mora, E., 1990. Isolation and characterization of a lectin from the seeds of *Dioclea lehmanni*. Phytochemistry 29, 1745–1749.

Peumans, W.J., Van Damme, E.J.M., Barre, A., Rouge, P., 2001. Classification of plant lectins in families of structural and evolutionary related proteins. The molecular immunology of complex carbohydrates-2. Adv. Exp. Med. Biol. 491, 27–54. http://dx.doi.org/10.1007/978-1-4615-1267-7_3.

Ramos, M.V., Moreira, R.A., Cavada, B.S., Oliveira, J.T.A., Rougé, P., 1996. Interaction of lectins from the Diocleinae sub-tribe with specific ligands. Rev. Bras. Fisiol. Veg. 8, 193–199.

Rocha, B.A., Moreno, F.B., Delatorre, P., Souza, E.P., Marinho, E.S., Benevides, R.G., Rustiguel, J.K., Souza, L.A., Nagano, C.S., Debray, H., Sampaio, A.H., de Azevedo, W.F.J., Cavada, B.S., 2009. Purification, characterization, and preliminary X-ray diffraction analysis of a lactose-specific lectin from *Cymbosema roseum* seeds. Appl. Biochem. Biotechnol. 152, 383–393. http://dx.doi.org/10.1007/s12010-008-8334-9.

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738. http://dx.doi.org/10.138/nprot.2010.5.

Roy, A., Yang, J., Zhang, Y., 2012. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res. 40, 471–477.

Saitou, N., Nei, M., 1987. The Neighbor-Joining method-a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Sinha, S., Gupta, G., Vijayan, M., Surolia, A., 2007. Subunit assembly of plant lectins. Curr. Opin. Struct. Biol. 17 (5), 498–505. http://dx.doi.org/10.1016/j.sbi.2007.06.007.

Srinivas, V.R., Reddy, G.B., Ahmad, N., Swaminathan, C.P., Mitra, N., Surolia, A., 2001. Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. Biochim. Biophys. Acta 1527, 102–111. http://dx.doi.org/10.1016/s0304-4165(01)00153-2.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, Maximum Parsimony Methods. Mol. Biol. Evol. 28, 2731–2739.

Wang, J.L., Cunningham, B.A., Edelman, G.M., 1971. Unusual fragments in the subunit structure of concanavalin A. Proc. Natl. Acad. Sci. U. S. A. 68, 1130–1134.

Wittmann, V., Pieters, R.J., 2013. Bridging lectin binding sites by multivalent carbohydrates. Chem. Soc. Rev. 42, 4492–4503. http://dx.doi.org/10.1039/c3cs60089k.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinforma. 9, 40. http://dx.doi.org/10.1186/1471-2105-9-40.