In Silico Comparative Analysis of Legume Lectins

Moraes Filho RM[1,2]*, Rossiter JG[1,2], Cavalcanti Junior EA[1,2] and Martins LSS[2,3]

[1]Department of Agronomy of the Federal Rural University of Pernambuco, Recife, PE, Brazil
[2]Department of Agronomy - Plant Breeding, Federal Rural University of Pernambuco, Recife, PE, Brazil
[3]Department of Biology, Federal Rural University of Pernambuco, Recife, PE, Brazil

Abstract

Legume lectins are a large group of proteins with similar structures, considerable amino acid sequence homology, and a variety of carbohydrate-binding specificities. Over 363 structures of plant lectins have been characterized, and they are the most extensively studied proteins of this class. In this study, *in silico* analyses were performed on legume lectin DNA and protein sequences of 35 species, aiming to find their conserved domains and to predict and compare their different tertiary structures, functions and molecular interactions. The protein sequences were aligned with ClustalW algorithm implemented in MEGA 6.06 software and a phylogenetic tree was constructed using the Neighbor-joining method. Functional domain analysis revealed eight functional domains present in the four selected sequences representative of the main phylogenetic groups. The identified domains are related to carbohydrate binding, cytokinin binding and protein serine/threonine kinase activities. A model of tertiary structure of *Vigna unguiculata* was generated by Phyre2 server with the multitemplate feature, and its quality was verified by Molprobity and ProSA-web servers. Results of the molecular docking analysis revealed interaction sites with monosaccharides and a cytokinin from two different binding pockets. These results may provide theoretical informations into the molecular basis of legume lectin functions and structure.

Introduction

Originally defined as proteins of non-immune origin that agglutinate cells and/or precipitate glycoconjugates [1], lectins, or agglutinins, are a structurally heterogeneous group of proteins or glycoproteins, with the ability to bind selectively, free or conjugated saccharides in a specific and reversible manner, by two or more binding sites [2-4].

Most plant lectins exhibit other properties. Some also bind plant growth regulators, the cytokinins, which are small hydrophobic molecules derivate from adenine [5,6]. They are secreted proteins and are found in vacuoles, cell walls, or intercellular spaces, of just about every plant organ [7]. Legume root lectins, are involved in the recognition and binding of *Rhizobium* and *Brudyrhizobium* sp. for symbiotic purposes. Based on the host specificity between legume and rhizobial symbiotic partners, the lectin-recognition hypothesis [8-10], explains the strong correlation between the Rhizobiaceae family bacteria and their legume hosts, and the ability of host produced lectins to bind to *Rhizobium* cells [7].

Based on their overall structure, lectins are classified as "merolectins," "hololectins," or "chimerolectins". While merolectins are built exclusively of a single carbohydrate binding domain, hololectins contain two or more identical or very similar domains, and are capable of agglutinating cells and precipitating glycoconjugates. Chimerolectins possess a carbohydrate-binding domain tandemly arrayed with an unrelated domain that acts independently and has a well-defined biological activity [3]. Lectins are synthesized as inactive precursors and undergo post-translational processing on its N-terminal and C-terminal regions, which determines their quaternary association, physiological targeting and carbohydrate binding activity [11]. Tertiary and quaternary structures of a great number of lectins have been determined by X-ray crystallography [12-15]. Over 363 structures of plant lectins and their complexes with sugars are listed in the 3D Lectine Database [16]. The majority of the entries are from legume lectins, or L-type lectins, with 227 structures. Legume lectins are a large group of proteins with similar structures, but distinct carbohydrate specificities. Primarily

found in the seeds of leguminous plants where they can represent about 10% of the total soluble protein content, they are synthesized during seed development and transported to the vacuole. Legume lectins have also been found in the bark of some leguminous trees, and in very low amounts in other vegetative tissues. In some species lectins have been found to be encoded by separate but very similar genes [17]. The architecture of the legume lectin monomer is usually described as consisting of two β -sheets. Their quaternary interfaces are also formed between β -strands. The monomers have highly similar sequences and share the same tertiary structure, with minor variations in loop lengths or lengths of strands. Most legume lectins are known to exist mainly as homodimers or homotetramers, with the tetramers being dimmers of dimers [12,18,19]. Found in the seeds of *Canavalia ensiformis*, the first lectin to be sequenced and to have the three-dimensional structure determined by X-ray crystallography, was the Concanavalin A (ConA) [20-23]. ConA and other similar lectins from the *Canavalia* and *Dioclea* genus undergo a complex post-translational modification in which the order of the N- and C-domains of the protein is reversed [19,24,25]. Other kind of lectins such as the soybean agglutinin (SBA), *Dolichos biflorus* lectin (DBL), peanut agglutinin (PNA) and *Erythrina corallodendron* lectin (EcorL) shows a simpler process, based on the removal of the N-terminal signal peptide followed by cleavage of the C-terminal peptide [11].

Biological, molecular, biochemical and evolutionary arguments indicate that lectins are of great importance in plant defense. A strong

argument for this role is their capacity of binding glycoconjugates of other organisms. Although many plant lectins are able to bind monosaccharides such as glucose, mannose or galactose, they have a much higher affinity for oligosaccharides, which are not common or totally absent in plants [3]. However, legume lectins are not only of interest because of their carbohydrate binding properties. Their various dimeric and tetrameric quaternary structures, with the capacity to agglutinate cells and to precipitate multivalent carbohydrates can serve as excellent model systems for the investigation of folding and association reactions of oligomeric proteins and to analyze the details of protein–protein interaction [18,26].

In silico analysis can be of great value for predicting structures and functions of proteins, and has been used to characterize many proteins and enzymes from diverse eukaryotic and prokaryotic species [11,27-31].

In this study, *in silico* analyses were performed on legume lectin DNA and protein sequences of 35 species, aiming to find their conserved domains and to predict and compare their tertiary structures, functions and molecular interactions. The comparative analysis will provide valuable theoretical insights for future studies with these proteins.

Materials and Methods

Database search and sequence retrieval

35 DNA and amino acid sequences of the analyzed legume lectins were downloaded from GenBank from the National Center for Biotechnology Information. The accession numbers are listed in the Table 1.

ORF identification and multiple alignments

Open reading frames were detected by the ORF Finder tool. Exon numbers of the sequences were analyzed by using GSDS2 server [32].

Sequence analysis

Physico-chemical parameters of lectin sequences were analyzed by ProtParam [33]. Subcellular localizations were predicted by the CELLO2GO server [34]. Signal peptide cleavage sites were predicted using the TOPCONS server [35]. Functional domains analysis was carried out by by ProDom server [36].

Conserved motif analysis

Conserved motif structure was analyzed by using the MEME SUITE tool [37], with the following parameters: maximum number of motifs to find=7; minimum width of motif=6; and maximum width of motif=50.

Multiple alignment and phylogenetic analysis

Sequence alignment of lectin protein sequences was performed with ClustalW algorithm implemented in Molecular Evolutionary Genetic Analysis (MEGA 6.06) [38], with default parameters and visualized by BioEdit Sequence Alignment Editor. The phylogenetic tree was constructed using the neighbor-joining method for 2000 bootstrap replicates.

Tertiary structure prediction, evaluation, and validation of the model

3-D models of legume lectin protein sequences were predicted using the Phyre2 server [39] in multi-template intensive mode and visualized by UCSF Chimera package [40]. Model quality was evaluated using the Molprobity server [41], by Ramachandran plot analysis. Z-score was calculated using interactive ProSA-web server to recognize

errors in 3-D structures, which indicated model quality and total energy deviation of the structure with respect to energy distribution derived from random conformations [42].

Molecular docking and prediction of binding sites

Prediction of molecular pockets was carried out using DoGSiteScorer server [43]. Ligand binding sites were predicted using the Patch Dock server [44].

Results

Identification and characterization legume lectin proteins

In this study, a total of 35 legume lectin DNA sequences of five tribes from the Papilionoidae subfamily (Phaseoleae, Dalbergieae, Sophoreae, Cicerae, Trifoleae) and one Caesalpinioidae tribe (Cercideae) were retrieved from NCBI in FASTA format (Table 1). The analysis of DNA sequences by GSDS server showed that all sequences are composed of a single exon and have no introns. A search for conserved domains was performed with the DELTA Blast Tool and all translated protein sequences exhibited similarity with the *lectin_legume_LecRK_Arcelin_ConA* family (Cd06899). This alignment model includes the legume lectins (also known as agglutinins), the arcelin (also known as phytohemagglutinin-L) family of lectin-like defense proteins, and the LecRK family of lectin-like receptor kinases, concanavalinA (ConA), and an alpha-amylase inhibitor. Homotetramer interaction sites (HTis), homodimer interaction sites (HDis), N-linked glycosylation sites (Nlgs) and Metal binding sites (MBs) were also present in the sequences.

Identification of conserved motifs

The analyses of conserved motifs of legume lectin proteins were performed by MEME Suite tool. Based on the results, four conserved motifs of Legume lectins were discovered (Table 2; Figure 1). Three of the four motifs occurred in all analyzed sequences. The motif 1 was absent in *B. variegata, B. purpurea, M. truncatula* and *C. arietinum* sequences. All motifs sequences were analyzed by The Delta Blast tool and exhibited similarities with Lectin L-Type Superfamily domains.

Structural and functional predictions

The primary structure of the legume lectin precursor sequences was deduced by ProtParam server. Table 3 shows the Molecular weight, theoretical pI, and grand average of hydropathicity (GRAVY) of the evaluated protein sequences. TOPCONS server was used to predict Post-translational modifications. This tool predicts the presence and location of signal peptide cleavage sites in protein sequences incorporating a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks. TOPCONS showed that 31 of the 35 evaluated sequences have an N-terminal signal peptide (Table 3). The identified signal peptides were [21-34] a long and with the exception of *C. arietinum* and *L. purpureous* lectins, the cleavage site occurred just after a conserved serine.

As shown in Table 2, after the removal of signal peptides, legume lectin sequences varied in size from 242 (*P. tetragonolobus*) to 263 (*B. variegata; D. guianensis*) amino acids with 25.4 kDa (*D. guianensis*) to 27.8 (*C.* arietinum; *M. truncatula*). The isoeletric point (pI) was between 4.60 (*P. glabellus*) and 6.36 (*G. max*). The grand average of hydropathicity was between -0.529 (*A. hypogaea*) and 0.165 (*P. augusti*). These results indicate that the evaluated legume lectins are hydrophilic in nature, with only six proteins being slightly hydrophobic (Table 3). Functional domain analysis was performed by ProDom (Protein Domain Database) server. ProDom is a comprehensive set of protein

Species	GeneBank Acession Number (NCBI)	GeneBank Identification	ORF size (bp)	Subfamily/tribe
Arachis hypogaea	U22469.1	NMLA mannose/glucose-binding lectin precursor	843	Papilionoideae; Dalbergieae
Bauhinia purpurea	D12481.1	Lectin	873	Caesalpinioideae, Cercideae
Bauhinia variegata	EU596376.1	Lectin I (BVL)	876	Caesalpinioideae, Cercideae
Canavalia brasiliensis	Y13904.1	ConBr	873	Papilionoideae, Phaseoleae
Canavalia ensiformis	AF308777.1	ConA	873	Papilionoideae, Phaseoleae
Cicer arietinum	XM 004509655.1	Seed lectin-like	843	Papilionoideae, Cicerae
Cladrastis lutea	U21959.1	LECCLA2	873	Papilionoideae, Sophoreae
Dioclea guianensis	AM701772.1	Lectin percursor	876	Papilionoideae, Phaseoleae
Dolichos biflorus	M23216.1	DBL (Dolichos biflorus lectin)	828	Papilionoideae, Phaseoleae
Erythrina corallodendron	X52782.1	Lectin (EcorL)	846	Papilionoideae, Phaseoleae
Glycine max	NM 001250281.2	Lectin (LOC732576)	849	Papilionoideae, Phaseoleae
Lablab purpureus	DQ534060.1	B11 alpha amylase inhibitor precursor	825	Papilionoideae, Phaseoleae
Medicago truncatula	XM 003612921.1	Legume lectin beta domain protein	861	Papilionoideae; Trifolieae
Phaseolus acutifolius	U10416.1	Phytohemagglutinin	831	Papilionoideae, Phaseoleae
Phaseolus augusti	AJ843877.1	Lectin	837	Papilionoideae, Phaseoleae
Phaseolus coccineus	AJ438774.1	phytohemagglutinin	822	Papilionoideae, Phaseoleae
Phaseolus costaricensis	AJ849453.1	Phytohemagglutinin - L	822	Papilionoideae, Phaseoleae
Phaseolus filiformis	AJ844592.1	Lectin precursor	825	Papilionoideae, Phaseoleae
Phaseolus glabellus	AJ844593.1	Lectin precursor	837	Papilionoideae, Phaseoleae
Phaseolus leptostachyus	AJ844591.1	Lectin precursor	843	Papilionoideae, Phaseoleae
Phaseolus lunatus	AJ271874.1	Lectin (LBL-6)	837	Papilionoideae, Phaseoleae
Phaseolus maculatus	AJ845972.1	Lectin precursor	834	Papilionoideae, Phaseoleae
Phaseolus microcarpus	AJ845193.1	Lectin precursor	837	Papilionoideae, Phaseoleae
Phaseolus oligospermus	AJ845195.1	Lectin precursor (lec 1)	837	Papilionoideae, Phaseoleae
Phaseolus parvulus	AJ845971.1	Lectin precursor	837	Papilionoideae, Phaseoleae
Phaseolus vulgaris	AJ439715.1	Lec4-B17 gene	828	Papilionoideae, Phaseoleae
Platypodium elegans	JN133278.1	Lectin (pel A)	786	Papilionoideae, Dalbergieae
Psophocarpus tetragonolobus	U60765.1	Basic Agglutinin (WBAI)	727	Papilionoideae, Phaseoleae
Pterocarpus angolensis	AJ426054.1	Lectin	783	Papilionoideae; Dalbergieae
Pterocarpus rotundifolius	AY639659.1	Lectin	750	Papilionoideae; Dalbergieae
Robinia pseudoacacia	AB012633.1	Lectin	858	Papilionoideae; Robinieae
Sophora alopecuroides	DQ011517.1	Lectin	843	Papilionoideae, Sophoreae
Sophora flavescens	AF285121.1	Lectin	855	Papilionoideae, Sophoreae
Vigna aconitifolia	JF501650.1	Lectin	843	Papilionoideae, Phaseoleae
Vigna unguiculata	AJ621421.1	Lectin precursor	846	Papilionoideae, Phaseoleae

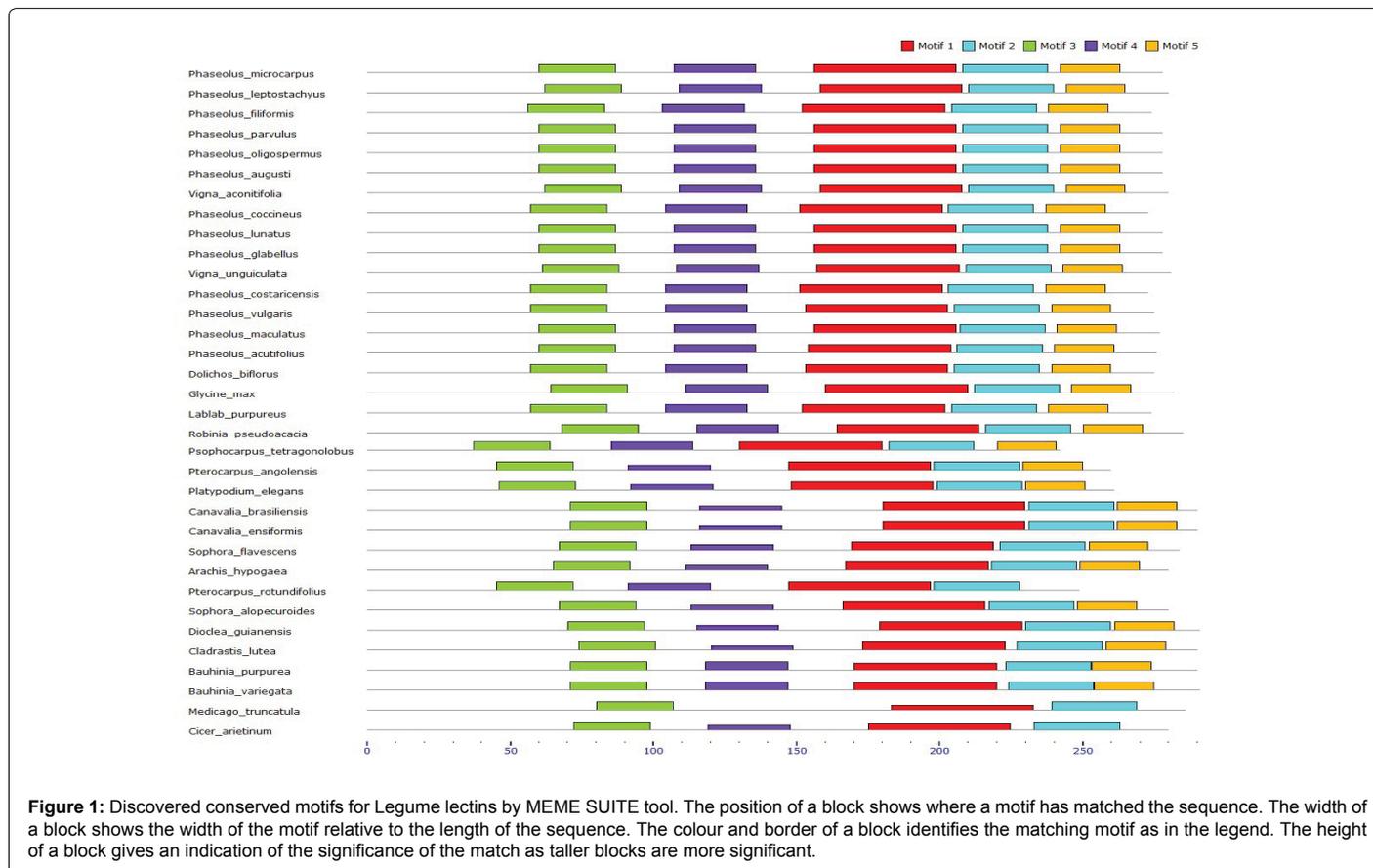**Table 1:** Thirty five evaluated legume lectin sequences and their species of origin.

Motif	E-value	Sites	Width	Best possible match	Conserved domain
1	8.3e-884	31	50	WDPEYRHIGIDVNCIKSIKTARWDMRNGQNAEVLITYDASTKLLVASLVY	Lectin L-Type Superfamily
2	2.5e-545	35	30	RRTSYCVSERVDLKSVLPEWVRVGFSATTG	Lectin L-Type Superfamily
3	4.0e-464	35	27	GNGEPTRASLGRAFYSAPIQIWDKTTG	Lectin L-Type Superfamily
4	2.5e-330	34	29	SADGLAFALVPVGSQPKDNGGCLGLFDNA	Lectin L-Type Superfamily
5	8.5e-262	32	21	YYETHDVLSWSFASKLSDGTT	Lectin L-Type Superfamily

Table 2: Discovered conserved motifs by MEME tool in 35 legume lectin precursor sequences.

Origin Species	Protein Size (aa)	MW (kDa)	pI	GRAVY	SPCS	Localization Prediction	Biological Process
A. hypogaea	280/254	31.0/26.6	5.71/4.71	-0.325/-0.529	A26-A27	Ext	Rs, St, RLs,CPm
B. purpurea	290/262	32.2/26.7	7.81/5.61	-0.088/-0.161	A28-A29	Ext, PM	Rs, Ca, St, RLs, CPm
B. variegata	291/263	32.2/26.7	7.81/5.61	-0.109/-0.185	A28-A29	Ext, PM	Rs, Ca, St, RLs, CPm
C. brasiliensis	290/261	31.4/25.5	5.65/5.44	-0.018/-0.197	A29-A30	Ext, PM	Rs, St,
C. ensiformis	290/261	31.4/25.6	5.47/5.63	-0.021/-0.205	A29-A30	Ext, PM	Rs, St CPm
C. arietinum	280/230	31.1/27.8	5.25/5.00	-0.087/-0.238	A30-A31	Cy, Va	Ca, Rs, St, RLs, CPm
C. lutea	290/255	32.0/26.5	6.42/5.11	-0.307/-0.434	A35-A36	Ext, PM	Ca, St, Rs, CPm
D. guianensis	291/263	31.3/25.4	6.06/5.26	-0.024/-0.181	A28-A29	Ext,	Ca, St, Rs, CPm

D. biflorus	275/253	29.5/25.8	4.81/4.79	0.074/-0.107	A22-A23	Ext, PM, Chl	Rs, St, RLs, CPm
E. corallodendron	281/255	30.7/26.2	4.91/4.70	-0.048/-0.191	A26-A27	Ext	Ca, Rs, St, RLs, CPm
G. max	282/256	30.0/25.6	5.93/6.38	0.089/-0.068	A26-A27	Ext,	Rs, St, RLs, CPm
L. purpureus	274/252	29.9/27.5	5.62/5.47	0.022/-0.117	A22-A23	Ext, PM	Rs, St, RLs, CPm
M. truncatula	286/253	31.5/27.8	5.46/5.01	-0.152/-0.223	A33-A34	Cy	Ca, Rs, St, Ls, CPm
P. acutifolius	276/252	29.7/27.2	4.75/4.67	-0.046/-0.140	A24-A25	Ext	Rs, St, RLs, CPm
P. augusti	278/256	29.6/27.1	5.48/5.13	0.264/0.165	A22-A23	Ext, PM	Rs, St, RLs, CPm
P. coccineus	273/252	29.5/27.3	5.80/5.63	0.017/0.088	A21-A22	Ext, PM	Rs, St, RLs, CPm
P. costaricensis	273/252	29.3/27.0	4.99/4.87	0.088/-0.22	A21-A22	Ext, PM	Rs, St, RLs, CPm
P. filiformis	274/254	29.4/27.3	5.32/5.00	0.020/-0.06	A20-A21	Ext, PM	Rs, St, RLs, CPm
P. glabellus	278/254	29.8/27.3	4.68/4.60	0.120/0.019	A24-A25	Ext, PM	Rs, St, RLs, CPm
P. leptostachyus	280/254	30.3/27.5	5.54/5.32	-0.028/-0.167	A26-A27	Ext, PM	Rs, St, RLs, CPm
P. lunatus	278/256	29.6/27.0	5.69/5.30	0.244/0.136	A22-A23	Ext, PM	Rs, St, RLs, CPm
P. maculatus	277/253	29.8/27.3	5.14/5.03	0.140/0.041	A24-A25	Ext,PM	Rs, St, RLs
P. microcarpus	278/254	29.9/27.3	5.93/5.77	0.005/-0.095	A24-A25	PM	Rs, St, RLs, CPm
P. oligospermus	278/254	29.8/27.2	4.82/4.64	0.124/0.025	A24-A25	Ext, PM, Chl	Rs, St, RLs,C Pm
P. parvulus	278/254	29.9/27.3	5.03/4.92	0.017/-0.094	A24-A25	Ext, PM	Rs, St, RLs, CPm
P. vulgaris	275/273	29.5/27.3	4.83/4.73	0.064/-0.062	A21-A22	PM	Ca, Rs, St, RLs, CPm
P. elegans	261	29.2/26.7	5.73/5.95	-0.347/-0.294	-	Ext, PM	Rs, St, Ca, RLs, CPm
P. tetragonolobus	242	26.6/26.4	5.70/5.70	-0.129/-0.137	-	Ext, Cy, Chl	Rs, St, RLs, CPm
P. angolensis	260	28.4/26.0	5.19/4.97	-0.304/-0.285	-	Ext	Ca, Rs, St, CPm
P. rotundifolius	249	27.7/26.6	4.93/4.82	-0.381/-0.408	-	Ext	Rs, St
R. pseudoacacia	285/254	30.9/25.7	6.83/5.53	0.160/0.096	A31-A32	Ext, PM	Rs,St, RLs, CPm
S. alopecuroides	280/251	31.1/26.4	5.85/5.31	-0.162/-0.334	A29-A30	Ext,	Rs, St, RLs
S. flavescens	284/254	31.3/26.4	5.65/5.12	-0.071/-0.264	A30-A31	Ext	Ca, Rs, St, RLs, CPm
V. aconitifolia	280/256	30.0/27.5	5.35/5.35	0.114/-0.037	A24-A25	PM	Rs, St, RLs, CPm
V. unguiculata	281/255	30.0/27.3	5.62/5.62	0.071/-0.051	A26-A27	PM	Rs, St, RLs, CPm

Table 3: Primary structure analysis, signal peptides, subcellular location and function prediction of legume lectins.

Figure 1: Discovered conserved motifs for Legume lectins by MEME SUITE tool. The position of a block shows where a motif has matched the sequence. The width of a block shows the width of the motif relative to the length of the sequence. The colour and border of a block identifies the matching motif as in the legend. The height of a block gives an indication of the significance of the match as taller blocks are more significant.

Citation: Moraes Filho RM, Rossiter JG, Cavalcanti Junior EA, Martins LSS (2017) *In Silico* Comparative Analysis of Legume Lectins. J Genet Genom 1 103.
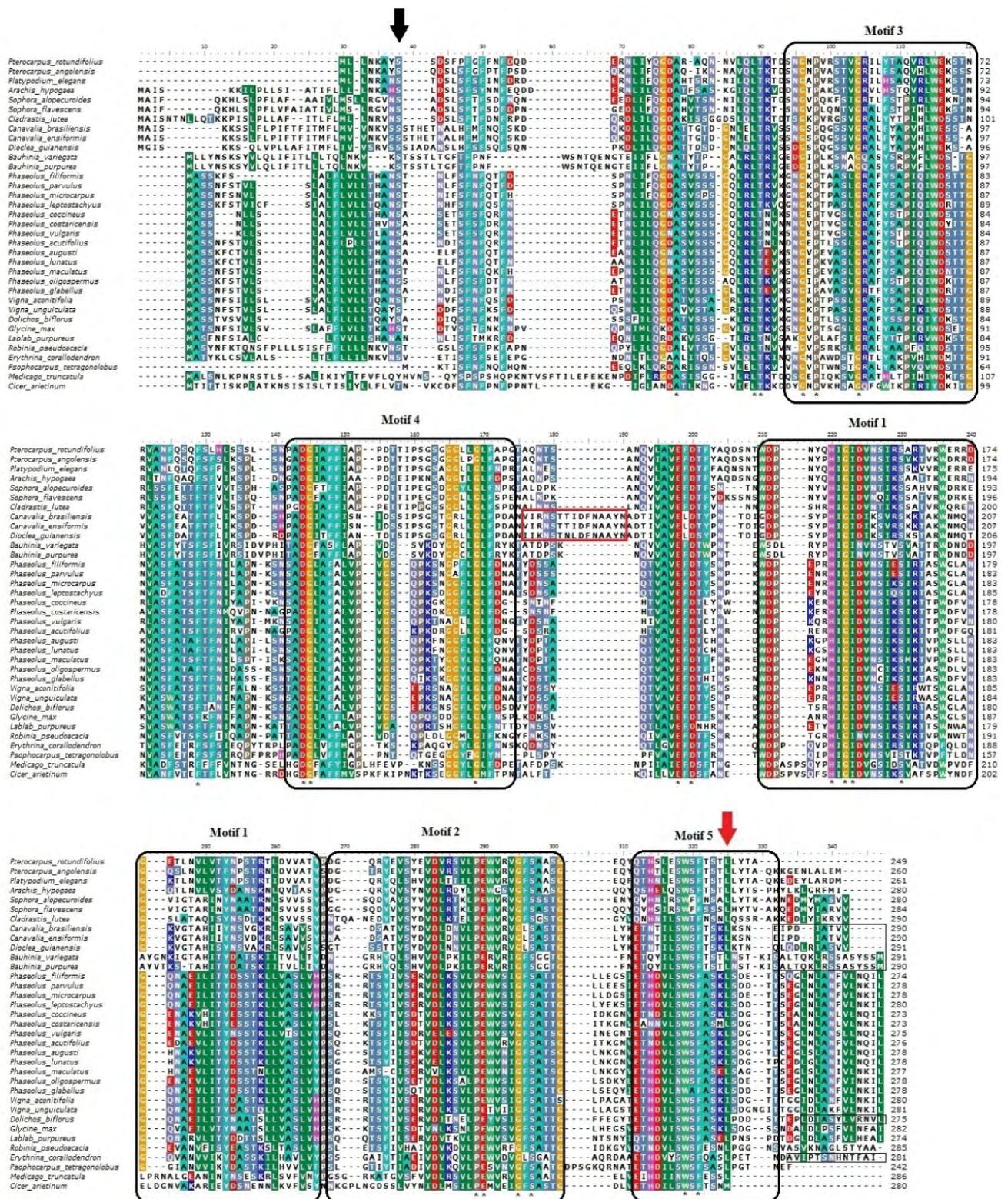
Page 5 of 11

Figure 2: Sequence alignment of legume lectin proteins in 35 plant species. Sequences were aligned by ClustalW, and identical and similar residues are displayed in the same color. Motifs discovered by the MEME SUITE tool are indicated above the sequences. Known C-Terminal cleaved peptides are indicated by black rectangles, and Linker glycosylated peptide from Canavalia and Dioclea species are indicated by a red rectangle. Red arrow indicates the conserved Leucine identified by Moreira et al. as a signal for C-Terminal cleavage sites.

domain families automatically generated from the Universal Protein Resource Knowledge Database (Uniprot) by clustering homologous segments. For this analysis four representative sequences were selected.

Phylogenetic analysis

To examine the phylogenetic relationships of the evaluated legume lectin proteins sequences, a multiple alignment was performed by ClustalW algorithm implemented in MEGA 6.06 software. A phylogenetic tree was constructed using the Neighbor-joining method and the bootstrap test carried out with 2000 replicates. By this analysis the legume lectin protein sequences from the 35 species were divided into four groups, designated from Group 1 to Group 4 (Figure 2).

Tertiary structure prediction, evaluation and validation of the model

For tertiary structure prediction *V. unguiculata* lectin (VuL) sequence was selected. The model was constructed using the Phyre2 server. This server uses the alignment of hidden Markov models via HHsearch [45] to improve the accuracy of alignment and detection rate. It also incorporates the Poing tool [46], which is an *ab initio* folding simulation to model regions of the proteins with no detectable homology to known structures. Poing tool also combines multiple templates to improve model accuracy. Verification of stereochemical quality of the model using Ramachandran plot analysis was performed by the Molprobity server. The generated model had 84.7% of amino

acid residues were in favored regions. ProSA-web (Protein Structure Analysis web) was used to recognition errors in the tertiary structure prediction of PLs. The Z-score was used to measure the energy, as it indicated overall quality of the model. Positive Z-score values show that the structure is not stabilized while zero and negative scores represent one of the ideal structures. Due to the presence of Ramachandran outliers, model refinement was carried out with the KiNG software [47]. After refinement, the model exhibited 100% of its amino acid residues in favored regions. In ProSA server analysis, the Z-score of model was of -7.64 which were highlighted as large black dot in Figure 6b. The Plot of residue scores shows local model quality by plotting energies as a function of amino acid sequence position. Positive values correspond to problematic or erroneous parts of the input structure. As was shown in the graph of Figure 6c, all amino acid residues of VuL are below the zero on x-axis.

Prediction of molecular pockets and molecular docking

DogSiteScorer web server [43] was employed to detect potential binding pockets of the selected protein model. DoGSiteScorer is a grid-based method which uses a Difference of Gaussian filter to detect potential binding pockets - solely based on the 3D structure of the protein. Subsequently, global properties describing the size, shape, and chemical features of the predicted pockets are calculated [43]. The server identified 9 potential molecular pockets in the protein model. The objective of computational docking is to determine how two molecules
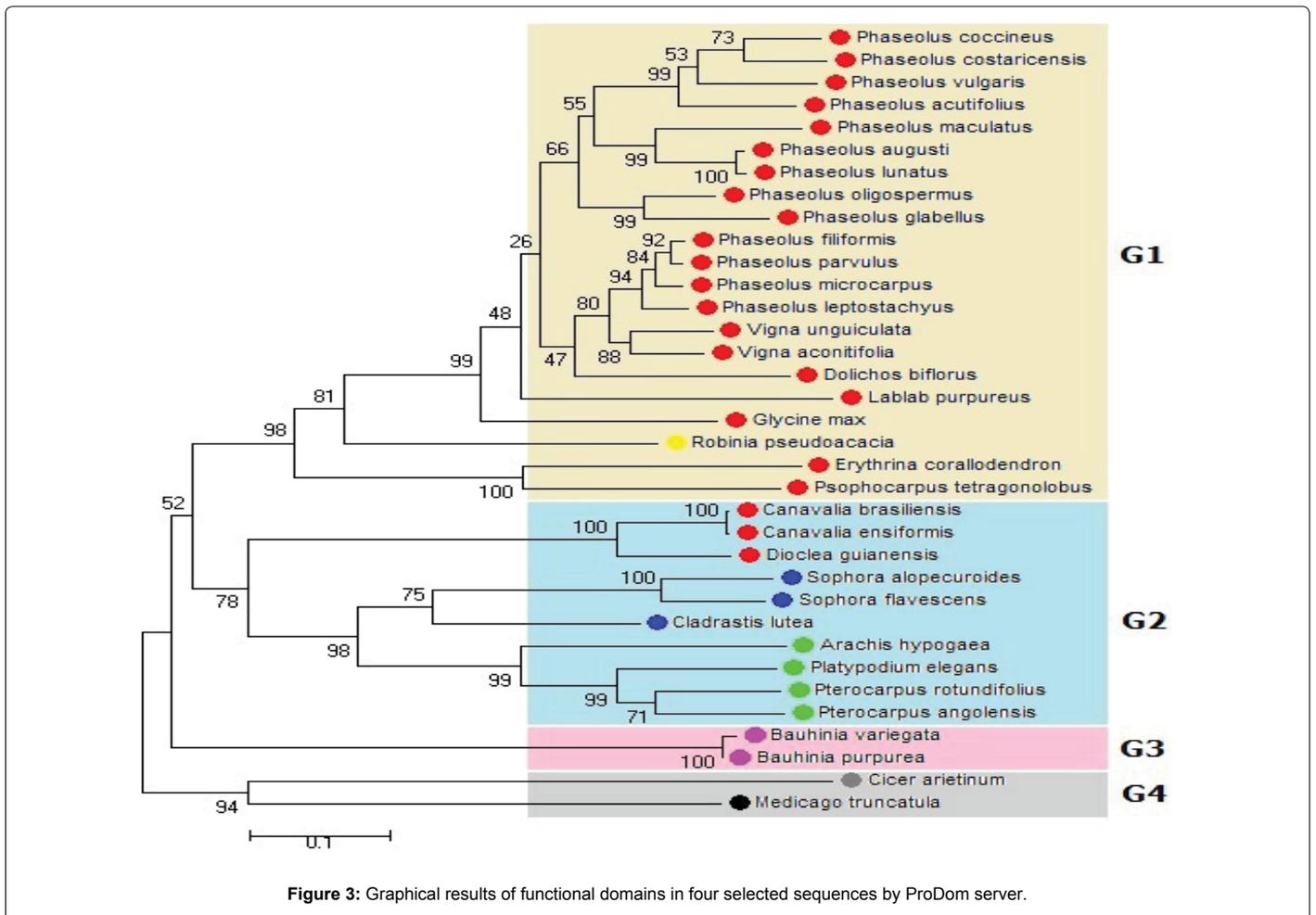

Figure 3: Graphical results of functional domains in four selected sequences by ProDom server.

will interact. Molecular docking is often employed to determine how receptors and ligands interact to form a binding pocket. The selected model was used to predict interaction sites with two monosaccharides (Glucose, Mannose) and a cytokinin (Kinetin) by the Patchdock server. Patchdock algorithms are inspired by image segmentation and object recognition techniques, which are used in computer vision. Given two molecules, their surfaces are divided into patches according to the surface shape. The patches are then filtered so that only patches with hot-spot residues are retained. Once the patches are identified, they are superimposed using a shape-matching algorithm [44]. After the docking procedure, interaction refinement was performed using the Firedock server. FireDock is an efficient method for refinement and re-scoring of rigid-body protein-protein docking solutions. Results from patchdock were refined by Firedock, generating results ranked on the basis of global energy (Table 4). Patchdock results showed that Glucose, Mannose and Kinetin interact with VuL in binding pockets identified by DoGSitescorer server. Glucose and Mannose interact with the protein in the binding pocket 3, formed by amino acids present in domains PD000671, PD506531, PDB0U9L5. Kinetin interaction site is

an independent binding site (Binding pocket 1), formed by PD000671, PD506531, PDC662J4. Figures 3 and 4 shows the solutions with best results of global energy values by the Firedock analysis.

Discussion

First discovered in seeds of leguminous plants, L-type lectins have structural motifs present in a variety of glycoproteins from other eukaryotic organisms [17]. A great number of these proteins have been characterized and employed in biomedical and analytical procedures [11-19,48,49]. Many lectins have been isolated from seeds and other storage tissues in plants where they contribute for a very large proportion of the total tissue protein content. Many of these lectins behave similarly to storage proteins, and some of these are degraded during germination and the development process [50,51]. In some cases, aside from their carbohydrate binding properties, plant lectins exhibit specific interactions with small molecules that are predominantly hydrophobic in nature, such as phythormones, like citokinins and auxins [52]. Lectins also may play a role of great importance in plant defense as they are capable of binding oligosaccharides, which are
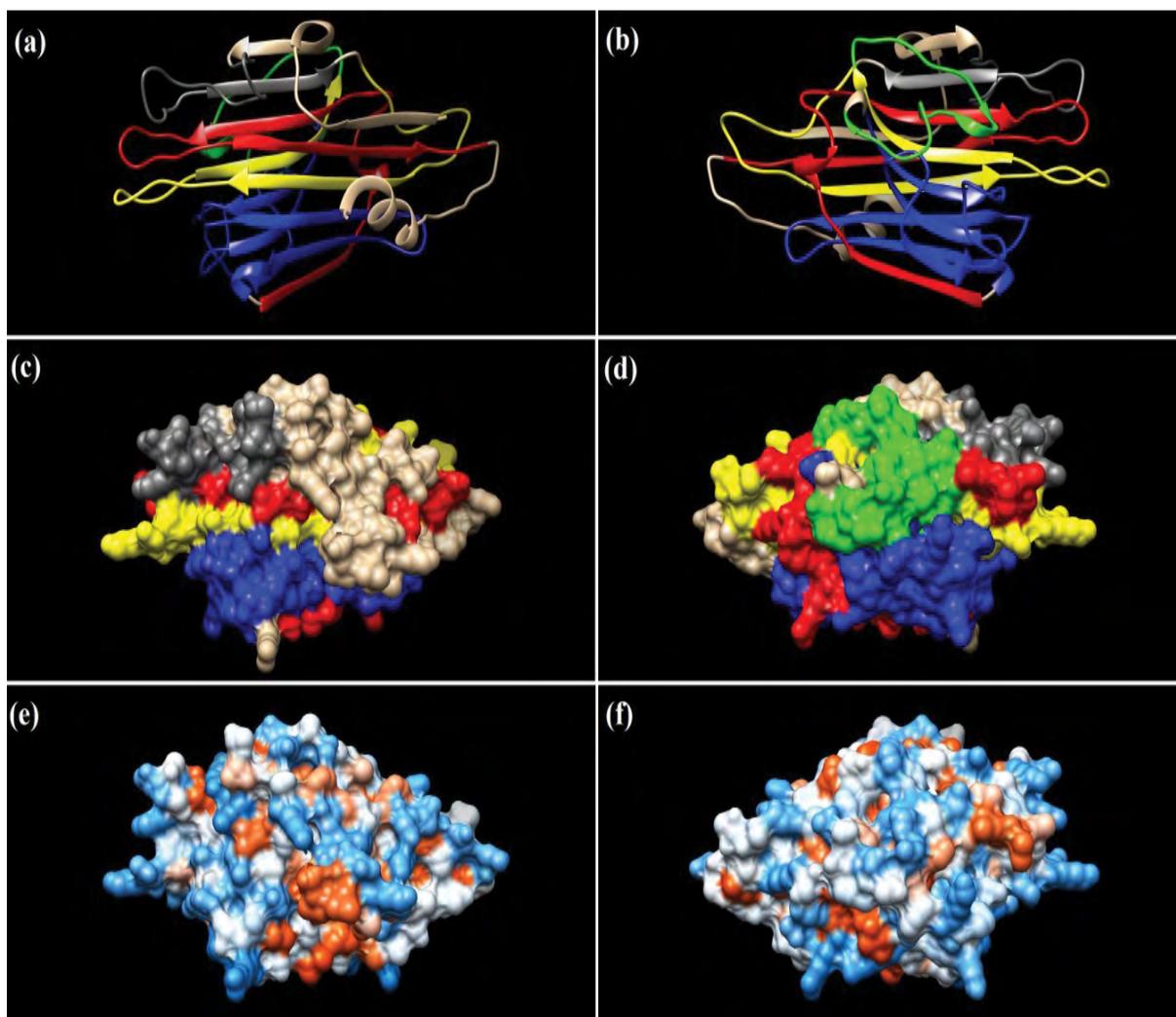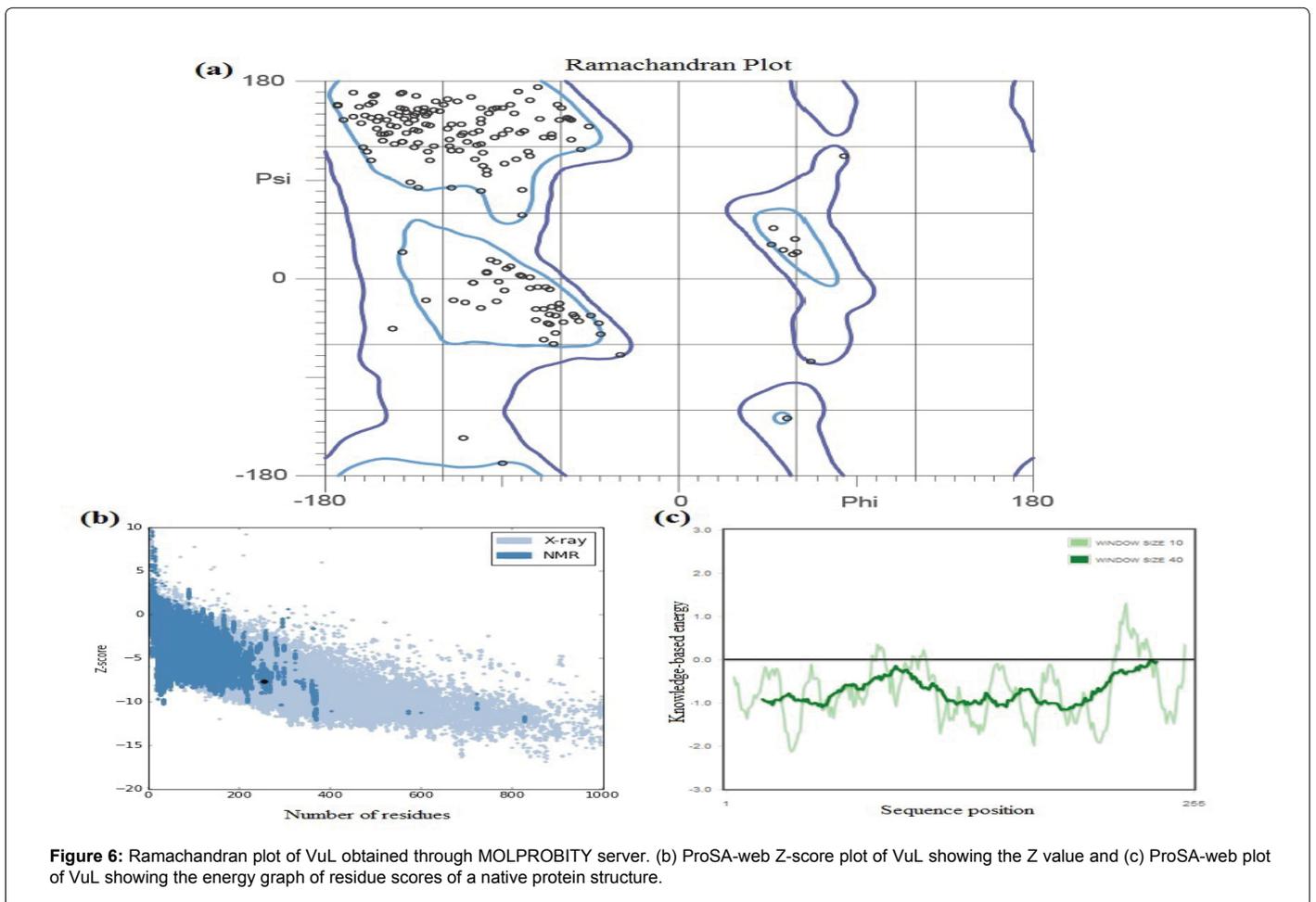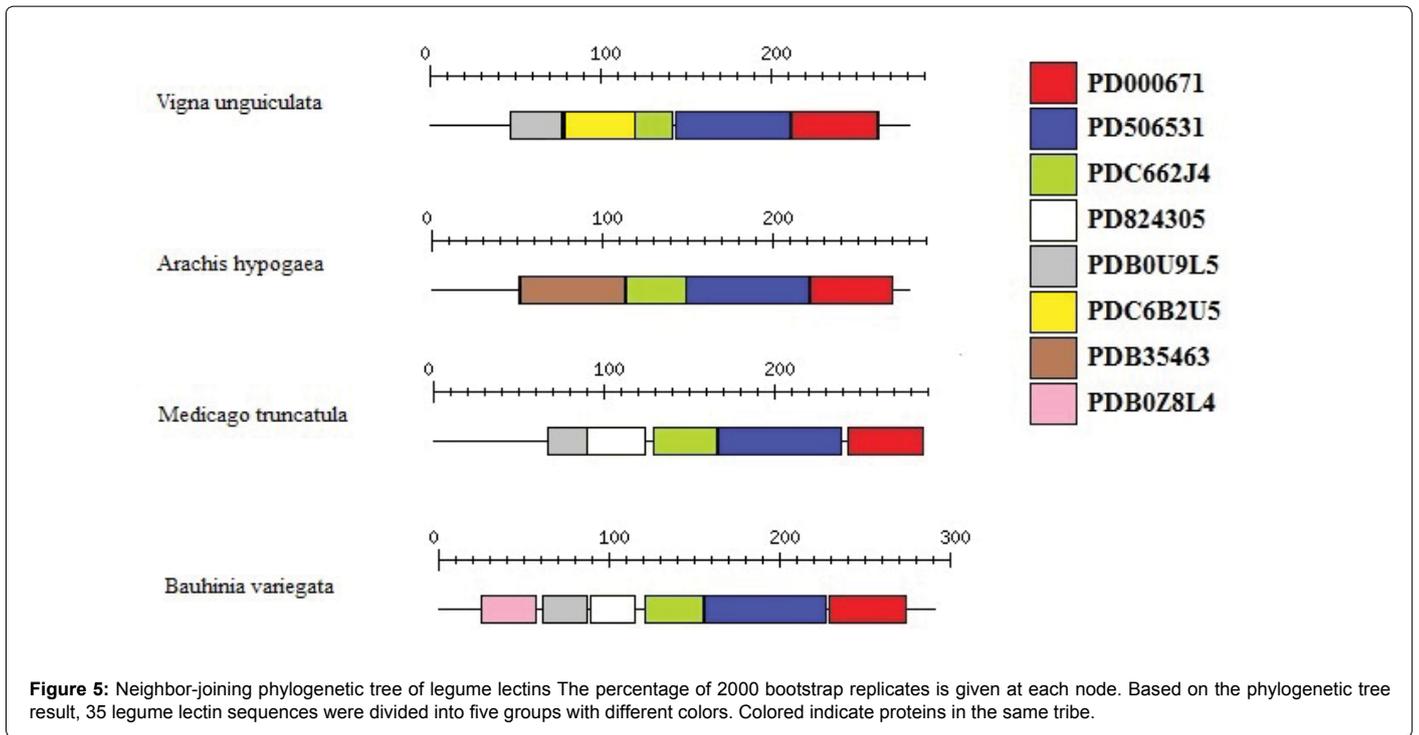

Figure 4: Tertiary structure prediction of VuL sequence. Identified functional domains by ProDom server are indicated by different colors as follows: PD000671 (Red), PD506531 (Blue), PDC66234 (Green), PDC6B2U5 (Yellow), PDB0U9L5 (Gray). Hydrophobicity is represented as a color gradient, with blue being the most hydrophilic, to white, to orange red for the most hydrophobic.
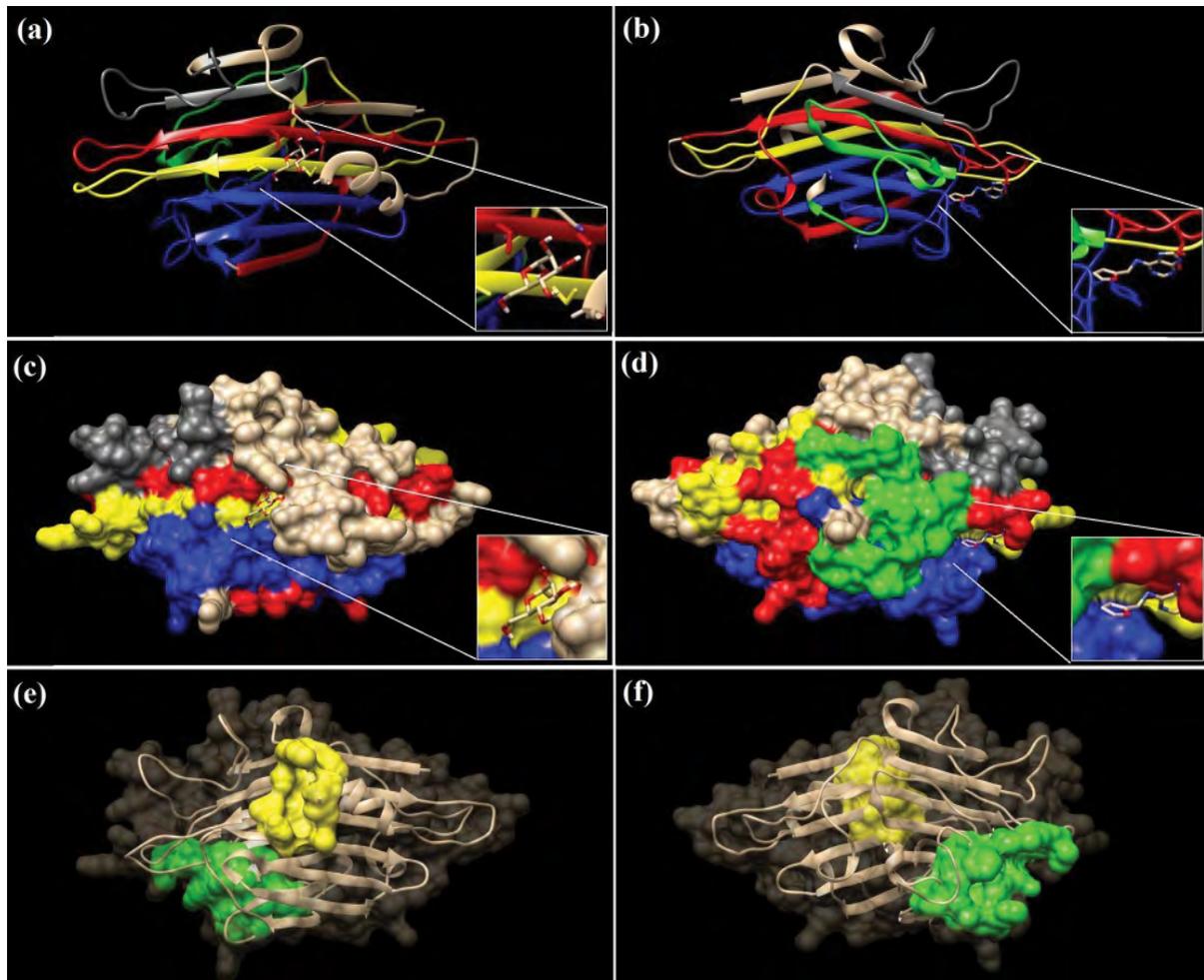
Citation: Moraes Filho RM, Rossiter JG, Cavalcanti Junior EA, Martins LSS (2017) *In Silico* Comparative Analysis of Legume Lectins. J Genet Genom 1 103.

Page 8 of 11

Figure 5: Neighbor-joining phylogenetic tree of legume lectins The percentage of 2000 bootstrap replicates is given at each node. Based on the phylogenetic tree result, 35 legume lectin sequences were divided into five groups with different colors. Colored indicate proteins in the same tribe.

Figure 6: Ramachandran plot of VuL obtained through MOLPROBITY server. (b) ProSA-web Z-score plot of VuL showing the Z value and (c) ProSA-web plot of VuL showing the energy graph of residue scores of a native protein structure.

Figure 7: Predicted interactions of VuL with Mannose (a and c) and kinetin (b and d), Different molecular pockets identified by DoGSitescorer server for mannose and glucose (e) and kinetin (f). Molecular pocket 3 (e) is showed in yellow and molecular pocket 1 is showed in green (f).

not common or totally absent in plants [3]. Lectins from *Hordeum vulgare* lectin and *Triticum aestivum* were reported to preferentially accumulate in nematode-infested roots. These proteins were found at the nematode feeding site and did not accumulate if the plant was inoculated with a nematode not specific for it [53]. In this study, we analyze structural and functional characteristics of 35 legume lectin proteins sequences from the NCBI database. Functional analysis revealed some sequences that had distinct physico-chemical properties but were similar at the amino acid level. The majority of the sequences were classified as hydrophilic with only six lectins from Group 1 being slightly hydrophobic. As it can be seen in Table 3, the isoelectric point for legume lectins was between 4.60 and 6.36 indicating their acidic character. The isoeletric point is the pH at which a protein carries no net charge. At a pH below their pI, proteins carry a positive net charge and above their pI they carry a negative net charge. The pI of a protein is also the pH at which the protein is least soluble, and therefore unstable [54]. In order to find patterns of conserved motifs, we ran the MEME SUITE tool. As shown in Table 2 and Figure 1, five conserved motifs were found. Motif 1, 2 and 3 were distributed in all sequences. Motifs 4 was absent only in *M. truncatula*, and Motif 5 was absent in *C.*

arietinum, M. truncatula and *P. rotundifolius.* With the exception of the motif 5, all other motifs exhibited at least one conserved glycine residue which suggests that this aminoacid may play a role in the molecular function. The analysis by the CELLO2GO server revealed that most of the evaluated legume lectins are located at the extracellular medium or associated with the plasmamembrane. The predicted molecular function of all proteins was associated with ion binding, kinase activity and enzyme regulator activities. In the phylogenetic analysis, the 35 analyzed sequences were divided in four groups. G1 was the largest and contained 21 members, being 20 of the Phaseoleae tribe and only one of the Robinieae tribe (*Robinia pseudoacacia*). G2 comprised 10 members of three tribes, namely, Phaseoleae, Sophoreae and Dalbergiae. G3 was comprised of only *B. purpurea* and *B. variegata*, of the Cercidae tribe, and G4 contained *M. truncatula* of the Trifolieae tribe and *C. arietinum* of the Cicerae tribe.

ProDom analysis revealed eight functional domains. As shown in Figure 3 and Table 5, domains PD000671, PD506531 and PDC662J4 were present in the four analyzed sequences. The first two are involved in Mannose/Glucose binding and the third in Cytokinin/Glucose binding. PDC6B2U5, was indicated as involved with Protein Serine/

Domain	Pfam classification	Gene Onthology
PD000671	Lectin LegB	Mannose/Glucose Binding
PD506531	Lectin LegB	Mannose/Glucose Binding
PDC662J4	Lectin LegB	Cytokinin/Glucose Binding
PD824305	Lectin LegB	Cytokinin/Glucose Binding
PDB0U9L5	Lectin LegB	Mannose/Glucose Binding
PDC6B2U5	Lectin LegB	Carbohydrate Binding, Protein Serine/Threonine kinase activity
PDB3563	Lectin LegB	Cytokinin/Mannose Binding
PDB0Z8L4	Lectin LegB	Carbohydrate Binding

Table 4: Functional domains of four selected species identified by ProDom server.

Ligand	Score	Global Energy	Interface area (Å²)	DoGSiteScorer Pocket	Interactive Domains
Mannose	2382	-26.96	271	3	PD000671, PD506531, PDB0U9L5
Glucose	2436	-22.26	277.20	3	PD000671, PD506531, PDB0U9L5
Kinetin	3146	-19.99	387	1	PD000671, PD506531, PDC662J4

Table 5: Results of VuL molecular docking analysis.

Threonine kinase activity, and was only present in *V. unguiculata* lectin (Representative of Group 1). The phythormone binding properties of legume lectin has been vastly studied. Crystalline Con A was reported to bind non-polar molecules such as growth factors or cytokinins and auxins and might function in regulation of cell-division or germination [55,56]. Lectins from *P. lunatus*, *D. biflorus*, *P. vulgaris* and *G. max* were reported to have adenine binding sites of varying affinities. *P. lunatus* lectins also bound to cytokinins besides adenine. These legume lectins have a conserved specific hydrophobic binding site completely independent of the carbohydrate binding site [5]. The knowledge on three-dimensional structures is of great importance to understand protein functions. Tertiary structures can be predicted from the amino acid sequences using different techniques and methods, including homology modeling. This methodology compares the structure of the query sequence with known protein sequences to predict their protein structure. A theoretic model of a representative legume lectin from *V. unguiculata* was generated based on homology modeling using known similar protein structures.

After validation and refinement, the selected model was used to predict and identify interaction sites with two monosaccharides (mannose and glucose) and a cytokinin (kinetin). *In silico* docking analysis confirm the interaction of the modeled legume lectin structure with the three molecules (Table 4). The analysis also shows that mannose and glucose interact with the lectin in the same molecular pocket. In the other hand, kinetin interact with the protein in an independent hydrophobic molecular pocket (Figures 3 and 5). The results of this comparative analyses shows that the evaluated lectin sequences are highly conserved throughout leguminous species. Functional analysis suggest that legume lectins could have additional functions aside their carbohydrate binding activities. The molecular pocket and interactions with the kinetin molecule also points to a possible regulatory function for this lectin. The recent increase in the processing power of computers has enabled the emergence of *in silico* experimentation, where research is conducted via computer simulations with models closely reflecting the real world. *In silico* or bioinformatics analysis, can play a vital role in the interpretation of genomic and proteomic data. This methodologies combine knowledge from many disciplines such as mathematics, statistics, computer sciences, physics, biology, and medicine and have been extensively used for predicting function and structure of proteins from its amino acid sequence [11,27-31]. This comparative study employed *in silico* methods to analyze functional and structural characteristics of legume lectin sequences from public databases. These findings can provide useful information into the molecular basis of these proteins' functions (Figure 7).

Author Contributions

Moraes Filho, RM conceived and designed the experiment, Rossiter, JG, Cavalcanti Junior, EA and Moraes Filho, RM collected and analyzed the data, Moraes Filho and Martins LSS wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Etzler ME (1985) Plant lectins: molecular and biological aspects. Annu Rev Plant Physiol 36: 209-234.

2. Sharon N (1993) Lectin-carbohydrate complexes of plants and animals: an atomic view. Trends Biochem Sci 18: 221-226.

3. Peumans WJ, Damme Van EJ (1995) Lectins as plant defense proteins. Plant Physiol 109: 347-352.

4. Peumans WJ, Van Damme EJ, Barre A, Rouge P (2001) Classification of plant lectins in families of structurally and evolutionary related proteins. Adv Exp Med Biol 491: 27-54.

5. Roberts DD, Goldstein IJ (1983) Adenine binding sites of the lectin from lima beans (Phaseolus lunatus). J Biol Chem 258: 13820-13824.

6. Roberts DD, Goldstein IJ (1983) Binding of hydrophobic ligands to plant lectins: titration with arylaminonapthalene sulfonates. Arch Biochem Biophys 223: 479-484.

7. Hirsch AM (1999) Role of lectins (and rhizobial exopolysaccharides) in legume nodulation. Current Opinion in Plant Biology 2: 320-326

8. Hamblin J, Kent SP (1973) Possible role of phytohemagglutinin in Phaseolus vulgaris L. Nat New Biol 245: 28-29.

9. Bohlool BB, Schmidt EL (1974) Lectins: a possible basis for specificity in the Rhizobium-legume root nodule symbiosis. Science 185: 269-271.

10. Dazzo FB, Hubbell HD (1975) Cross-reactive antigens and lectins as determinants of symbiotic specificity in the Rhizobium-clover association. Appl Microbiol 30: 1017-1033.

11. Moreira GMSG, Conceição FR, McBride AJA, Pinto LS (2013) Structure Predictions of Two Bauhinia variegata Lectins Reveal Patterns of C-Terminal Properties in Single Chain Legume Lectins. PLoS ONE 8: e81338.

12. Vijayan M, Chandra N (1999) Lectins. Curr Opin Struct Biol 9: 707-714.

13. Svensson C, Teneberg S, Nilsson CL, Kjellberg A, Schwarz FP, et al. (2002). High-resolution crystal structures of Erythrina cristagalli lectin in complex with lactose and 2 L-fucosyllactose and correlation with thermodynamic binding data. J Mol Biol 321: 69-83.

14. Rao KN, Suresh CG, Katre UV, Gaikwad SM, Khan MI (2004) Two orthorhombic crystal structures of a galactose-specific lectin from Artocarpus hirsuta in complex with methyl-a-D-galactose. Acta Cryst 60: 1404-1412.

15. Pérez A, Sarkar A, Rivet C, Breton A, Imberty A (2015) Glyco3D: A portal for structural glycosciences. Meth Mol Biol 1273: 241-258

16. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, et al. (2009) Essentials of glycobiology. 2nd edn. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

17. Loris R, Hamelryck T, Boukaert J, Wyns L (1998) Legume lectin structure. Biochim Biophys Acta 1383: 9-36.

18. Brinda KV, Mitra N, Surolia A, Vishveshwara S (2004) Determinants of quaternary association in legume lectins. Protein Sci 13: 1735-1749.

19. Bezerra WM, Paiva C, Moreira RDA, Grangeiro TB (2006) Establishment of a heterologous system for the expression of Canavalia brasiliensis lectin: a model for the study of protein splicing. Genetics and Molecular Research 5: 216-223.

20. Cunningham BA, Wang JL, Waxdal MJ, Edelman GM (1975) The covalent and three-dimensional structure of concanavalin A. II. Amino acid sequence of cyanogen bromide fragment F3. J Biol Chem 250: 1503-1512.

21. Wang JL, Cunningham BA, Waxdal MJ, Edelman GM (1975) The covalent and three-dimensional structural of concanavalin A. I. Amino acid sequence of cyanogen bromide fragments F1 and F2. J Biol Chem 250: 1490-1502.

22. Becker JW, Reeke GN, Wang JL, Cunningham BA, Edelman GM (1975) The covalent and three-dimensional structure of concanavalin A, III, Structure of the monomer and its interactions with metals and saccharides. J Biol Chem 250: 1513-1524.

23. Hardman KD, Ainsworth CF (1973). Binding nonpolar molecules by crystalline concanavalin A. Biochemistry 12: 4442-4448.

24. Araújo-Filho JH, Vasconcelos IM, Martins-Miranda AS, Gondim DMF, Oliveira JTA (2010) A ConA-like lectin from Dioclea guianensis Benth. Has antifungal activity against Colletotrichum gloeosporioides, unlike its homologues, ConM and ConA. J Agric Food Chem 58: 4090-4096.

25. Delatorre P, Rocha BAM, Simões RC, Pereira-Júnior FN, Silva HC, et al. (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.

26. Chatterjee A, Mandal DK (2003) Biochim Biophys Acta 1648:174-183.

27. Darabi M, Seddigh S (2015) Bioinformatic characterization of aspartic protease (AP) enzyme in seed plants. Plant Syst Evol 301: 2399-2417.

28. Feng BZ, Li PQ, Fu L, Yu XM (2015) Exploring laccase genes from plant pathogen genomes: a bioinformatic approach. Genet Mol Res 14: 14019-14036.

29. Han Y, Zheng QS, Wei YP, Chen J, Liu R, et al. (2015) In silico identification and analysis of phytoene synthase genes in plants. Genet Mol Res 14: 9412-9422.

30. Vatansever R, Filiz E, Ozyigit II (2015) In silico identification and comparative analysis of molybdenum (Mo) transporter genes in plants. Braz J Bot 39: 1-13

31. Moraes Filho RM, Martins LSS (2016) In silico comparative analysis of tylenchid nematode pectate lyases. Genet Mol Res 15: gmr15038402.

32. Hu B, Jin J, Guo AY, Zhang H, Luo J, et al. (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31: 1296-1297.

33. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook, pp: 571-607.

34. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, et al. (2014) CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS ONE 9: e99368.

35. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A (2015) The TOPCONS web server for combined membrane protein topology and signal peptide prediction. Nucleic Acids Res 43: 401-407.

36. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

37. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. (2009) MEME SUITE: tools for motif discovery and searching. Nucl Acids Res 37: 202-208.

38. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.

39. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10: 845-858.

40. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera - a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612.

41. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr 66: 12-21.

42. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35: W407-W410.

43. Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M (2012). Combining global and local measures for structure-based druggability predictions. J Chem Inf Model 52: 360-372.

44. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl Acids Res 33: W363-W367.

45. Söding J (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics 21: 951-960.

46. Jefferys BR, Kelley LA, Sternberg MJ (2010) Protein folding requires crowd control in a simulated cell. J Mol Biol 397: 1329-1338.

47. Chen VB, Davis IW, Richardson DC (2009) KING (Kinemage, Next Generation): a versatile interactive molecular and scientific visualization program. Protein Sci 18: 2403-2409.

48. Kim YH, Woloshuk CP, Cho EH, Bae JM, Song YS, et al. (2007) Cloning and functional expression of the gene encoding an inhibitor against A, flavus alpha amylase, novel lectin-like protein from L. purpureus (Dolicos lablab). Plant Cell Rep 26:395-405.

49. Lioi L, Galasso I, Lanave C, Daminati MG, Bollini R, et al. (2007) Evolutionary analysis of the APA genes in the Phaseolus genus: wild and cultivated bean species as sources of lectin-related resistance factors. Theor Appl Genet 115: 959-970

50. Dannenhoffer JM, Schulz A, Skaggs MI, Bostwick DE, Thompson GA (1997) Expression of the phloem lectin is developmentally linked to vascular differentiation in cucurbits. Planta 201: 405-414.

51. Van Damme EJM, Goossens K, Smeets K, Van Leuven F, Verhaert P, et al. (1995) The major tuber storage protein of Araceae species is a lectin. Characterization and molecular cloning of the lectin from Arum maculatum L. Plant Physiol 107: 1147-1158.

52. Komath SS, Kavitha M, Swamy MJ (2006) Beyond carbohydrate binding: new directions in plant lectin research. Org Biomol Chem 4: 973-988.

53. Oka Y, Chet I, Spiegel Y (1997) Accumulation of lectins in cereal roots invaded by the cereal cyst nematode Heterodera avenae. Physiol Mol Plant Pathol 51: 333-345.

54. Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA (2001) The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. Protein Sci 10: 1206-1215.

55. Hardman KD, Ainsworth CF (1976) Structure of the concanavalin A-methyl alpha-D-mannopyranoside complex at 6-A resolution. Biochemistry 15: 1120-1128.

56. Edelman GM, Wang JL (1978) Binding and functional properties of Concanavalin A and its derivatives J Biol Chem 253: 3016-3022.