Phytochemistry 139 (2017) 60–71

Contents lists available at ScienceDirect

Phytochemistry

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

cDNA cloning, molecular modeling and docking calculations of L-type lectins from *Swartzia simplex* var. *grandiflora* (Leguminosae, Papilionoideae), a member of the tribe Swartzieae

Paulo A.C. Maranhão [a], Claudener S. Teixeira [a], Bruno L. Sousa [b], Ito L. Barroso-Neto [c], José E. Monteiro-Júnior [d], Andreia V. Fernandes [e], Marcio V. Ramos [a], Ilka M. Vasconcelos [a], José F.C. Gonçalves [e], Bruno A.M. Rocha [a], Valder N. Freire [f], Thalles B. Grangeiro [d, *]

[a] Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Ceará (UFC), Fortaleza, Ceará, 60440-900, Brazil
[b] Faculdade de Filosofia Dom Aureliano Matos, Universidade Estadual do Ceará, Av. Dom Aureliano Matos, 2060, Limoeiro do Norte, CE, 62930-000, Brazil
[c] Departamento de Química Analítica e Físico-química, UFC, Fortaleza, Ceará, 60455-760, Brazil
[d] Departamento de Biologia, UFC, Fortaleza, Ceará, 60440-900, Brazil
[e] Laboratório de Fisiologia Vegetal e Bioquímica, Instituto Nacional de Pesquisas da Amazônia (MCTI-INPA), Manaus, Amazonas, 69067-375, Brazil
[f] Departamento de Física, UFC, Fortaleza, Ceará, 60440-760, Brazil

ARTICLE INFO

Article history:
Received 9 November 2016
Received in revised form
15 March 2017
Accepted 11 April 2017
Available online 14 April 2017

Keywords:
Swartzia simplex var. *grandiflora*
Leguminosae
cDNA cloning
Legume lectin

ABSTRACT

The genus *Swartzia* is a member of the tribe Swartzieae, whose genera constitute the living descendants of one of the early branches of the papilionoid legumes. Legume lectins comprise one of the main families of structurally and evolutionarily related carbohydrate-binding proteins of plant origin. However, these proteins have been poorly investigated in *Swartzia* and to date, only the lectin from *S. laevicarpa* seeds (SLL) has been purified. Moreover, no sequence information is known from lectins of any member of the tribe Swartzieae. In the present study, partial cDNA sequences encoding L-type lectins were obtained from developing seeds of *S. simplex* var. *grandiflora*. The amino acid sequences of the *S. simplex grandiflora* lectins (SSGLs) were only averagely related to the known primary structures of legume lectins, with sequence identities not greater than 50–52%. The SSGL sequences were more related to amino acid sequences of papilionoid lectins from members of the tribes Sophoreae and Dalbergieae and from the *Cladratis* and Vataireoid clades, which constitute with other taxa, the first branching lineages of the subfamily Papilionoideae. The three-dimensional structures of 2 representative SSGLs (SSGL-A and SSGL-E) were predicted by homology modeling using templates that exhibit the characteristic β-sandwich fold of the L-type lectins. Molecular docking calculations predicted that SSGL-A is able to interact with D-galactose, *N*-acetyl-D-galactosamine and α-lactose, whereas SSGL-E is probably a non-functional lectin due to 2 mutations in the carbohydrate-binding site. Using molecular dynamics simulations followed by density functional theory calculations, the binding free energies of the interaction of SSGL-A with GalNAc and α-lactose were estimated as −31.7 and −47.5 kcal/mol, respectively. These findings gave insights about the carbohydrate-binding specificity of SLL, which binds to immobilized lactose but is not retained in a matrix containing D-GalNAc as ligand.

1. Introduction

Lectins are a very heterogeneous group of carbohydrate-binding proteins. They are widespread in nature, occurring in all domains of life. They are defined as proteins that have at least one non-catalytic domain that binds reversibly to a specific mono- or oligosaccharide (Peumans and Van Damme, 1995). These proteins were first detected in the seeds of castor bean (*Ricinus communis*) at the end of the nineteenth century and in the last years, genome and transcriptome analyses have shown that lectin genes are ubiquitous in plants (Jiang et al., 2010). In plants, lectins play a role as defense proteins against the attack of pathogens, nematodes and insects,

* Corresponding author.
E-mail address: tbgrangeiro@gmail.com (T.B. Grangeiro).

http://dx.doi.org/10.1016/j.phytochem.2017.04.007
0031-9422/

and some of them have been intensively investigated as potential tools for crop protection (Macedo et al., 2015). Based on the amino acid sequence of their carbohydrate-recognition domain (CRD), plant lectins are currently classified into 12 families of structurally and evolutionarily related proteins (De Schutter and Van Damme, 2015). Among these families, the so called legume lectins (or L-type lectins) occupy a prominent position. More than one hundred L-type lectins have been isolated from different legume species and some of these proteins, such as concanavalin A (ConA) from jack bean (*Canavalia ensifomis*) and phytohemagglutinin (PHA) from common bean (*Phaseolus vulgaris*), have become important tools in biological and biomedical research (Sharon and Lis, 1990).

Swartzia Schreb. (Leguminosae, Papilionoideae) is a neotropical genus of trees and shrubs that contains approximately 180 species. They occur primarily in lowland rainforests from southern Mexico and the Caribbean islands to southern Brazil and Bolivia, and a high diversity is specially found in the Amazon basin (Torke and Schaal, 2008). *Swartzia* species have been investigated as a rich source of secondary metabolites with promising biological properties such as molluscicidal (Magalhães et al., 2003) and antimicrobial (Osawa et al., 1992; Favre-Godal et al., 2015) activities. Although the presence of hemagglutinating activity in seed extracts of *Swartzia* species has been reported (Cavalcanti and Coelho, 1990; Fernandes et al., 2011), lectins from this taxon have been poorly investigated in further details and to date, only a lactose-specific lectin from *S. laevicarpa* seeds has been purified to homogeneity and partially characterized (Fernandes et al., 2012). Owing to the lack of amino acid sequence information, the relationship between *Swartzia* lectins and other known legume lectins remains elusive. In the present study, partial cDNA sequences encoding L-type lectins from *S. simplex* var. *grandiflora* (Raddi) R.S. Cowan were obtained and the relationship of the deduced amino acid sequences with those of known legume lectins was investigated. Homology modeling and docking calculations were then used to gain insights about their functionalities.

2. Results and discussion

2.1. The proteins encoded by the 3′ RACE products and their relationship with known legume lectins

The partial cDNA sequences obtained by 3′ RACE encode polypeptides which have 240 (SSGL-A to SSGL-D) and 244 (SSGL-E and SSGL-F) amino acid residues (Fig. 1). The multiple alignment of these 6 sequences had 249 sites, including 14 sites with insertions/deletions (~5.6% of all aligned positions), 124 sites with conserved residues (~49.8%) and 111 sites with variable residues (~44.6%). To provide experimental evidence about the existence of the products encoded by these partial cDNA sequences, *S. simplex* var. *grandiflora* seed proteins were resolved by SDS-PAGE and the N-terminal sequence of the main protein band, which exhibited an apparent molecular mass of ~30 kDa (Fig. S10), was determined. A 19-residues sequence was obtained (TKFVDNQEDLLLQGDAAEV), in which the stretch QEDLLLQGDAAEV aligned with the first 13 amino acid residues deduced from the cDNA sequences (Fig. S11). Therefore, the partial cDNA clones encode proteins which are very closely related to the major lectin found in the *S. simplex* var. *grandiflora* mature seeds. Chromosome counts have shown that this is a diploid species with $2n = 2x = 26$ (Pinto et al., 2016). Moreover, the partial cDNA sequences reported in the present study were obtained from a single isolated tree. Therefore, rather than being the products of allelic variants of the same gene, the different mRNA sequences encoding the SSGLs are likely to come from the expression of individual members of a multigene family. In flowering plants, lectins belonging to different families are usually encoded in the same genome and each family often contains a number of members (Dang and Van Damme, 2016). In soybean (*Glycine max*), for example, 94 predicted genes were identified as belonging to the legume lectin family (Van Holle and Van Damme, 2015).

When the amino acid sequences deduced from the partial cDNAs were compared, it was possible to cluster them into 2 groups according to their overall identity (Fig. 1 and Table 1). The first group comprised the 4 sequences with 240 amino acid residues (SSGL-A to SSGL-D; 99.1% mean sequence identity) and the second group contained the 2 sequences with 244 residues (SSGL-E and SSGL-F; 99.5% sequence identity). The average sequence identity between the sequences of these 2 groups was approximately 53.2% (excluding all sites with gaps). Due to the high sequence identity within each group, one representative sequence from each one (SSGL-A and SSGL-E, respectively) was selected for further analysis.

RPS-BLAST searches against the CDD revealed that the primary structures of SSGL-A and SSGL-E contain each one a complete copy of the lectin_L-type (legume type) superfamily domain (CDD accession number: cl14058). The conserved domain found in both sequences (residues 1–221 in the primary structure of SSGL-A and residues 1–226 in the amino acid sequence of SSGL-E) belongs to the family lectin_legume_LecRK_Arcelin_ConA (CDD accession number: cd06899), which is a member of the superfamily cluster mentioned above. Furthermore, BLAST searches against the NCBI non-redundant protein sequences database showed that the primary structures of SSGL-A and SSGL-E exhibited significant alignments with CLAII (Van Damme et al., 1995), a lectin from the bark of *Cladratis kentukea*, and with lectins from *Sophora* species such as *S. alopecuroides* and *S. flavescens* (Tables S1 and S2). Sequence identity of SSGL-A and SSGL-E with CLAII was 52.4 and 49.7%, respectively. To further investigate the relationship of SSGLs with the known legume lectins so far characterized, a phylogenetic tree was generated using primary structures of lectins from species belonging to representative tribes of the subfamily Papilionoideae (Fig. S12). As shown in Fig. 2, 9 supported clades were recovered, corresponding to the tribes from which the sequences are derived. The amino acid sequences of SSGL-A and SSGL-E were recovered in a single cluster, which was sister to a clade containing 3 supported subclades. These subclades contained amino acid sequences of lectins from members of the tribe Sophoreae as well as from species of the *Cladratis* and *Pterocarpus* clades, respectively. *Swartzia* is the type genus of the tribe Swartzieae, which includes 7 other genera (*Ateleia*, *Bobgunnia*, *Bocoa*, *Candolleodendron*, *Cyathostegia*, *Fairchildia* and *Trischidium*), as currently circumscribed (Cardoso et al., 2013). It has been predicted that the tribe Swartzieae diverged from the other legume lineages 48.9 ± 2.8 million years ago (Lavin et al., 2005). Furthermore, molecular phylogenetic studies have consistently shown that members of the tribes Swartzieae, Sophoreae and Dalbergieae (tribe Dalbergieae includes the *Adesmia*, *Dalbergia* and *Pterocarpus* clades) as well as those of the *Cladratis* clade constitute, with other taxa, the first branching lineages of the subfamily Papilionoideae (Cardoso et al., 2013). Therefore, the finding that the amino acid sequences of SSGL-A and SSGL-E have at most 50–52% sequence identity with other known legume lectins, and that the legume lectins which are more closely related to them are those from species of *Cladrastis* and *Sophora*, are likely the reflection of the phylogeny and evolutionary history of the taxon *Swartzia*.

2.2. Closest structural neighbors and functional features of SSGL-A and SSGL-E

When the SSGL-A sequence was compared to proteins with known three-dimensional structures, the lectins from the

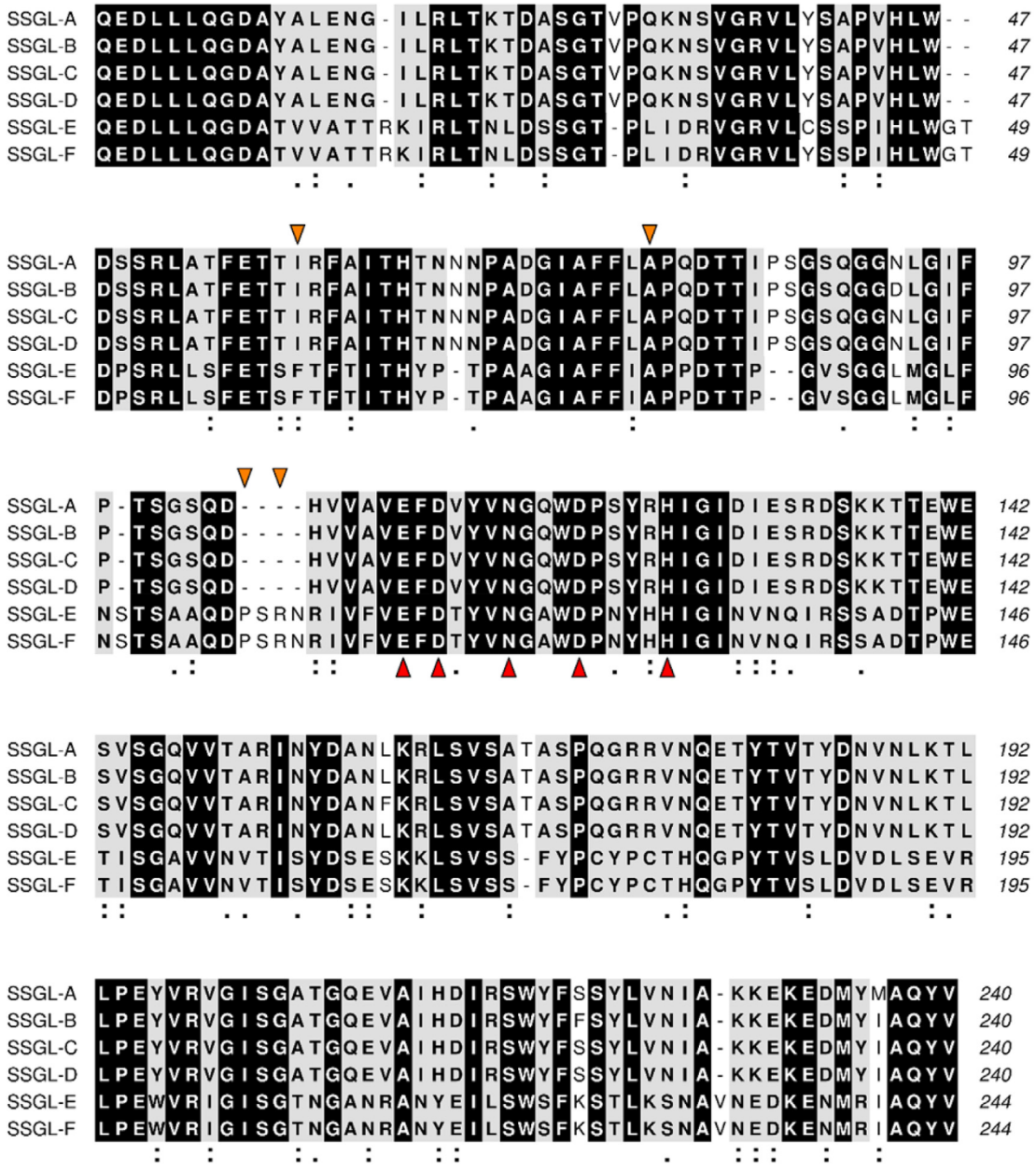

Fig. 1. Alignment of the amino acid sequences of SSGLs. Sites containing at each aligned position the same residue in the first 4 sequences and a different residue in the other 2 sequences are shaded in gray. Key residues that constitute the metal- and carbohydrate-binding sites of legume lectins are indicated by red and orange triangles, respectively. Sites containing residues with side chains that have strongly (:) or weakly (.) similar properties, scoring > 0.5 and ≤ 0.5 in the Gonnet PAM 250 matrix, respectively, are indicated. The alignment was rendered using the program ALINE (Bond and Schüttelkopf, 2009). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Pairwise sequence comparisons of the amino acid sequences of SSGLs, which are encoded by the cloned cDNA fragments. For each pair of compared sequences, the sequence identity (above the diagonal) and number of different amino acid residues (below the diagonal) between them are shown. These numbers were calculated based on the multiple sequence alignment shown in Fig. 1.

	SSGL-A	SSGL-B	SSGL-C	SSGL-D	SSGL-E	SSGL-F
SSGL-A	–	0.987	0.991	0.995	0.527	0.531
SSGL-B	3	–	0.987	0.991	0.531	0.536
SSGL-C	2	3	–	0.995	0.531	0.536
SSGL-D	1	2	1	–	0.531	0.536
SSGL-E	111	110	110	110	–	0.995
SSGL-F	110	109	109	109	1	–

papilionoid legumes *Platypodium elegans* (PELa), *Centrolobium tomentosum* (CTL), *Bowringia mildbraedii* (BMA), *Pterocarpus angolensis* (PAL) and *Maackia amurensis* (MAL) were recovered as the closest structural neighbors (Table S3). Searches against the PDB database recovered BMA, PELa, CTL and VML (*Vatairea macrocarpa* lectin) as the closest structural neighbors of SSGL-E (Table S4). These species are members of the tribes Dalbergieae (*P. elegans*, *C. tomentosum*, *P. angolensis*), Sophoreae (*M. amurensis*), Baphieae (*B. mildbraedii*) and the Vataireoid clade (*V. macrocarpa*). The tribe Baphieae and the Vataireoid clade have also been shown to represent, together with other clades, the early branching lineages of the papilionoid legumes (Cardoso et al., 2013).

To gain insights on the functional features of SSGL-A and SSGL-E,

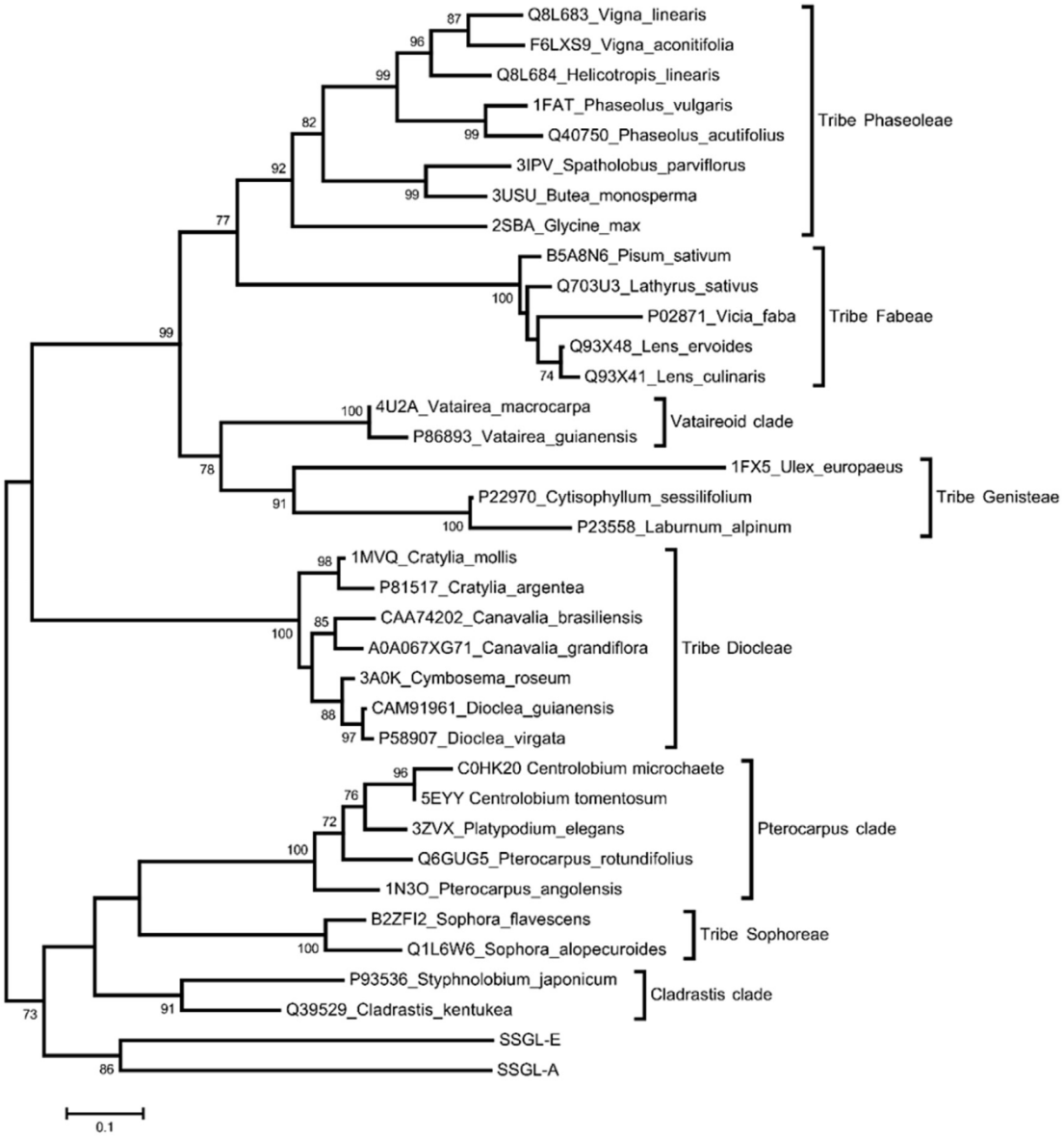

Fig. 2. NJ tree showing the phylogenetic relationship of SSGL-A and SSGL-E with representative legume lectins. The optimal tree with the sum of branch length = 6.94596515 is shown. The percentages of replicate trees in which the associated sequences clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1.27). The PDB codes or GenBank accession numbers of the amino acid sequences used are shown at the external nodes, next to the corresponding species names. The scale bar represents the branch lengths (the number of amino acid substitutions per site). The analysis involved 36 amino acid sequences. All positions containing gaps and missing data were eliminated and the final dataset had a total of 195 positions.

their amino acid sequences were compared to the primary structures of PELa (Benevides et al., 2012) and PAL (Loris et al., 2004) (Fig. 3). In PELa, 5 amino acid residues (Glu^{137}, Asp^{139}, Asn^{147}, Asp^{150} and His^{155}) constitute the metal-binding site for Ca^{2+} and Mn^{2+}. These residues are conserved in all legume lectins so far characterized and are required for carbohydrate-binding activity (Loris et al., 1998). These metal-binding site residues are also conserved in SSGL-A (Glu^{110}, Asp^{112}, Asn^{116}, Asp^{120} and His^{125}) and SSGL-E (Glu^{114}, Asp^{116}, Asn^{120}, Asp^{124} and His^{129}) (Fig. 3). The crystal structure of a PELa-trimannoside complex has revealed that 10 amino acid residues constitute its carbohydrate-binding site (CBS): 8 residues are involved in direct hydrogen bonds with the mannose units (Asp^{95}, Ala^{113}, Gly^{115}, Asn^{145}, Ser^{146}, Asn^{147}, Glu^{230} and Gln^{231}) whereas 2 other residues (Ala^{94} and Phe^{141}) maintain hydrophobic contacts with the sugar rings. Four out of these ten residues of PELa (Asp^{95}, Gly^{115}, Asn^{147} and Phe^{141}) are conserved in legume lectins and they are essential for carbohydrate binding (Loris et al., 1998). From the comparisons with PELa and PAL, 8 residues were identified in the sequences of SSGL-A (Ala^{71}, Asp^{72}, Gln^{90}, Gly^{92}, Tyr^{114}, Asn^{116}, Gln^{207} and Glu^{208}) and SSGL-E (Ala^{72}, Ala^{73}, Ser^{89}, Gly^{91}, Tyr^{118}, Asn^{120}, Ala^{210} and Asn^{211}) as putatively involved in their CBSs. The 4 key residues that are essential for carbohydrate binding are conserved in SSGL-A (Asp^{72}, Gly^{92}, Asn^{116} and Tyr^{114}), suggesting that it is a functional lectin. On the other hand, although 3 of these residues are conserved in SSGL-E (Gly^{91}, Asn^{120} and Tyr^{118}), one key aspartate (Asp^{95} in the PELa structure), whose side chain is involved in crucial hydrogen bonds with the ligand, is replaced by Ala (Ala^{73}). These findings suggested that SSGL-E is likely a non-functional lectin, which is unable to bind carbohydrate. To verify these assumptions, three-dimensional molecular models of SSGL-A and SSGL-E were build, validated and subjected to docking calculations with representative monosaccharides as potential ligands.

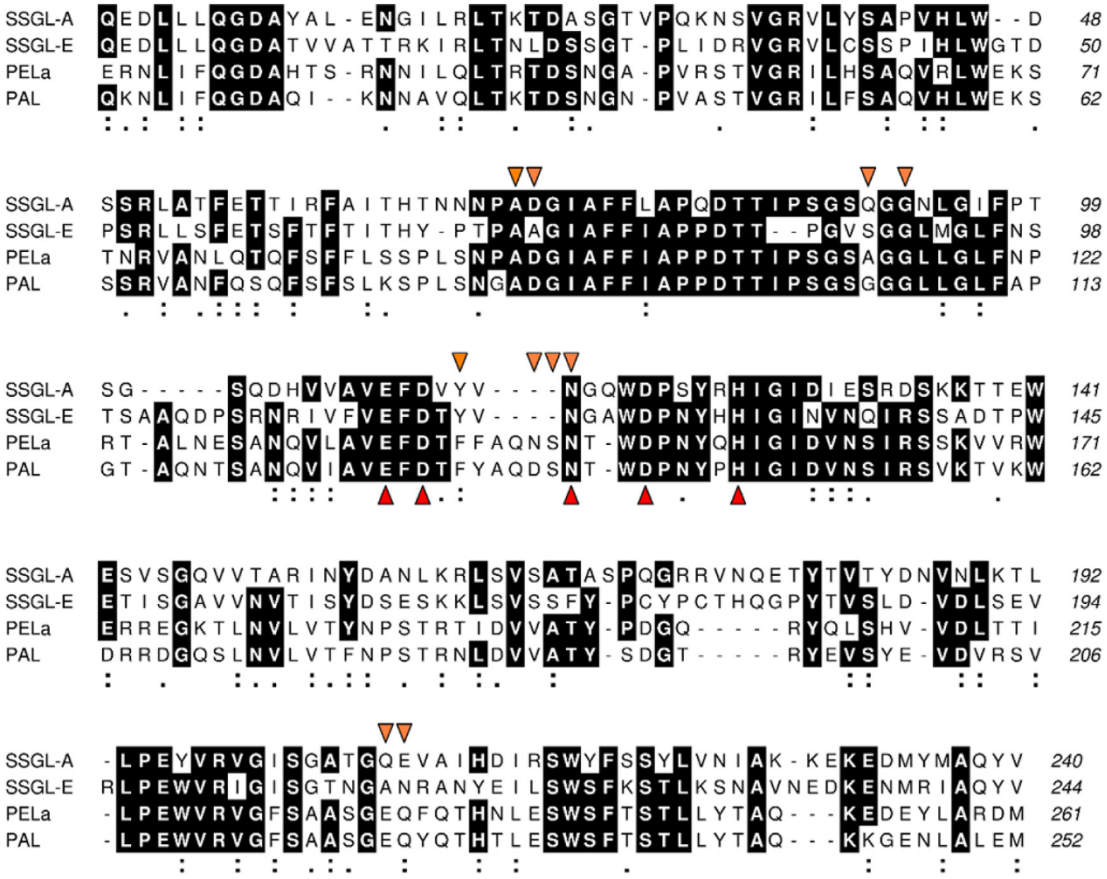

Fig. 3. Alignment of the amino acid sequences of SSGL-A and SSGL-E with the primary structures of PELa (PDB ID: 3ZVX**) and PAL (PDB ID:** 3Q8O**).** Sites containing the same amino acid residue in at least 3 sequences are shaded. PELa residues involved in metal and carbohydrate binding are indicated by red and orange triangles, respectively. Sites containing residues with side chains that have strongly (:) or weakly (.) similar properties, scoring > 0.5 and ≤ 0.5 in the Gonnet PAM 250 matrix, respectively, are indicated. The alignment was rendered using the program ALINE (Bond and Schüttelkopf, 2009). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

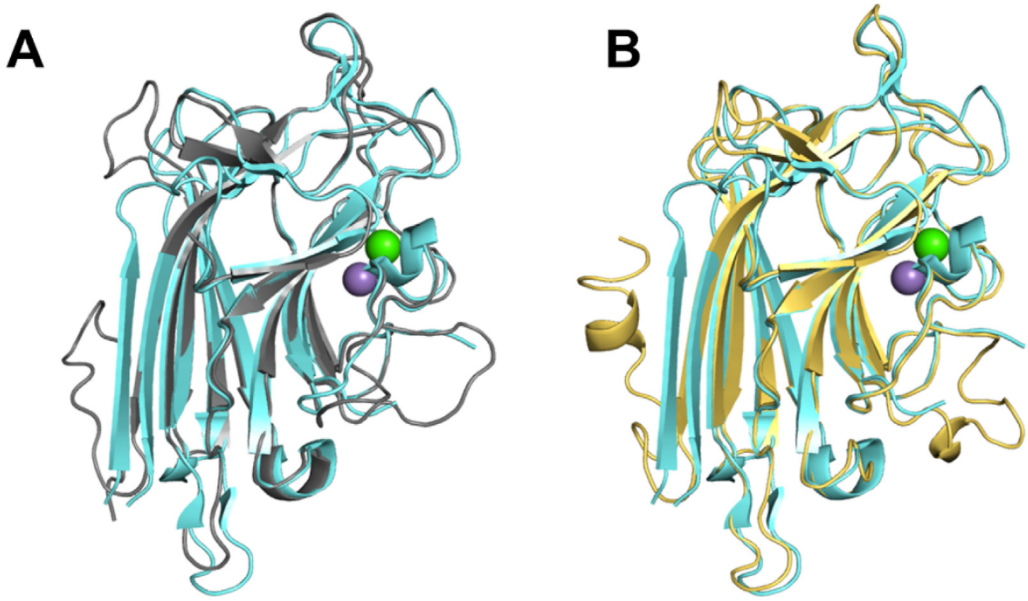

Fig. 4. Three-dimensional molecular models of SSGL-A and SSGL-E. Ribbon diagrams of the three-dimensional molecular models of SSGL-A (**A**; gray) and SSGL-E (**B**; yellow) superposed over the x-ray crystallographic structure of VML (cyan; PDB ID: 4U36). The carbohydrate-binding pocket is located on the right upper side of the images shown in **A** and **B**. Calcium (Ca^{2+}) and manganese (Mn^{2+}) ions are represented as green and purple spheres, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

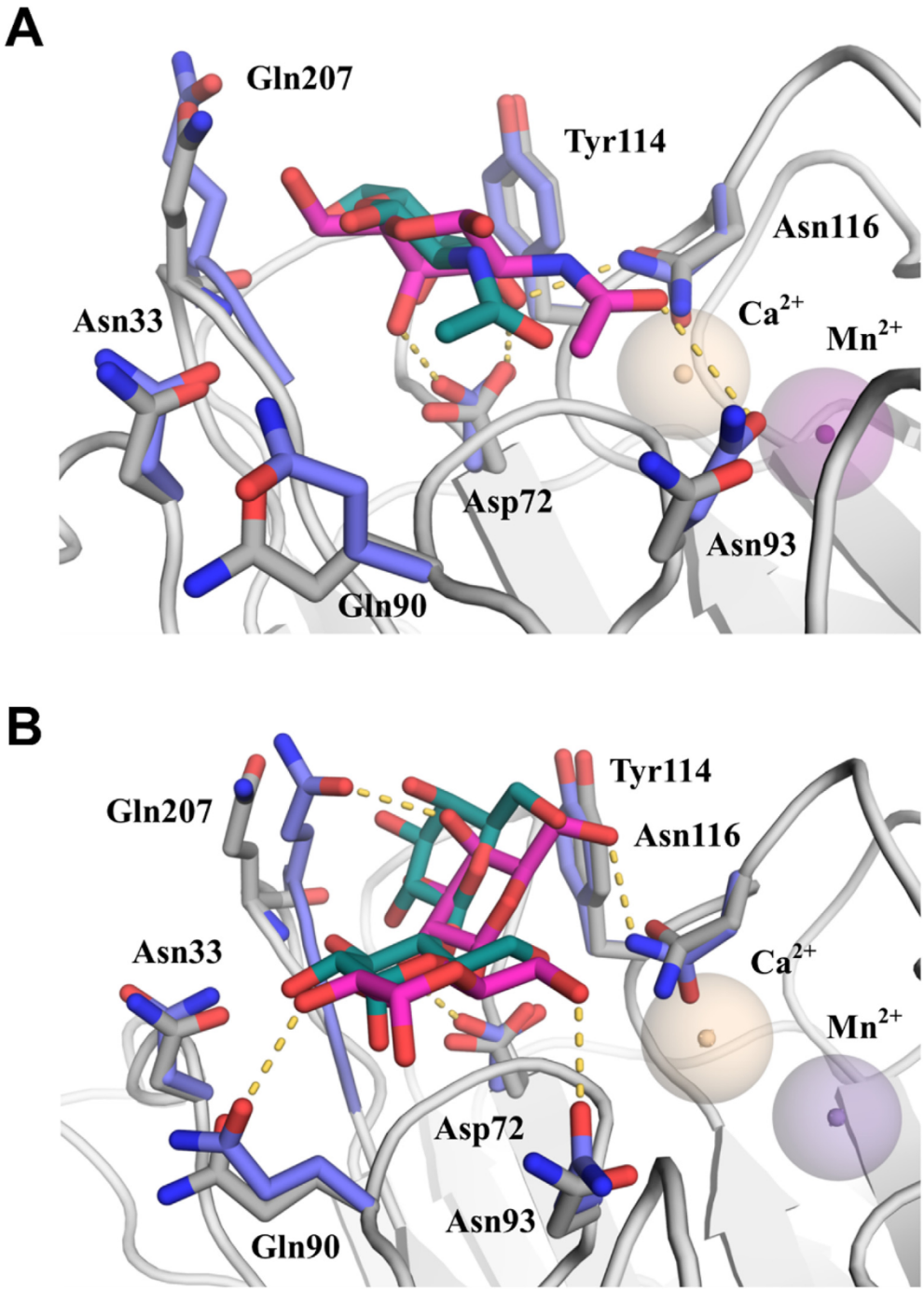

Fig. 5. Close-up view of D-GalNAc (A) and α-lactose (B) docked in the CBS of SSGL-A, which is shown in gray ribbon. The initial and final poses of the ligands during the MD simulations are colored pink and green, respectively, whereas the initial and final poses of the residues involved in carbohydrate coordination are represented as gray and purple sticks, respectively. Hydrogen bonds are shown as yellow dotted lines (distance cut-off = 3.5 Å) and N and O atoms are colored blue and red, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Assessment of the three-dimensional molecular models

PROSA2003 energy profiles of the SSGL-A and SSGL-E molecular models showed that most sequence-structure pairs had negative values (Fig. S15), which is an energetic feature expected for native protein folds (Sippl, 1993). Pair-wise structural comparisons between the SSGL-A and SSGL-E molecular models and the templates revealed RMSD values in the ranges of 0.59–0.89 Å (SSGL-A) and 11.79–20.41 Å (SSGL-E). These structural comparisons also showed TM-score values in the ranges of 0.90–0.92 (SSGL-A) and 0.17–0.36 (SSGL-E), indicating that models of correct topology (TM-score > 0.17) (Zhang and Skolnick, 2004) were predicted by the M4T method. PROCHECK and WHAT IF analyses (Table S5) showed that the local and global stereochemical parameters of the modeled structures had favorable values in comparison to the reference standards, suggesting a normal overall geometry. The QMEAN global raw scores confirmed the reliability of the modeled structures (Tables S6 and S7).

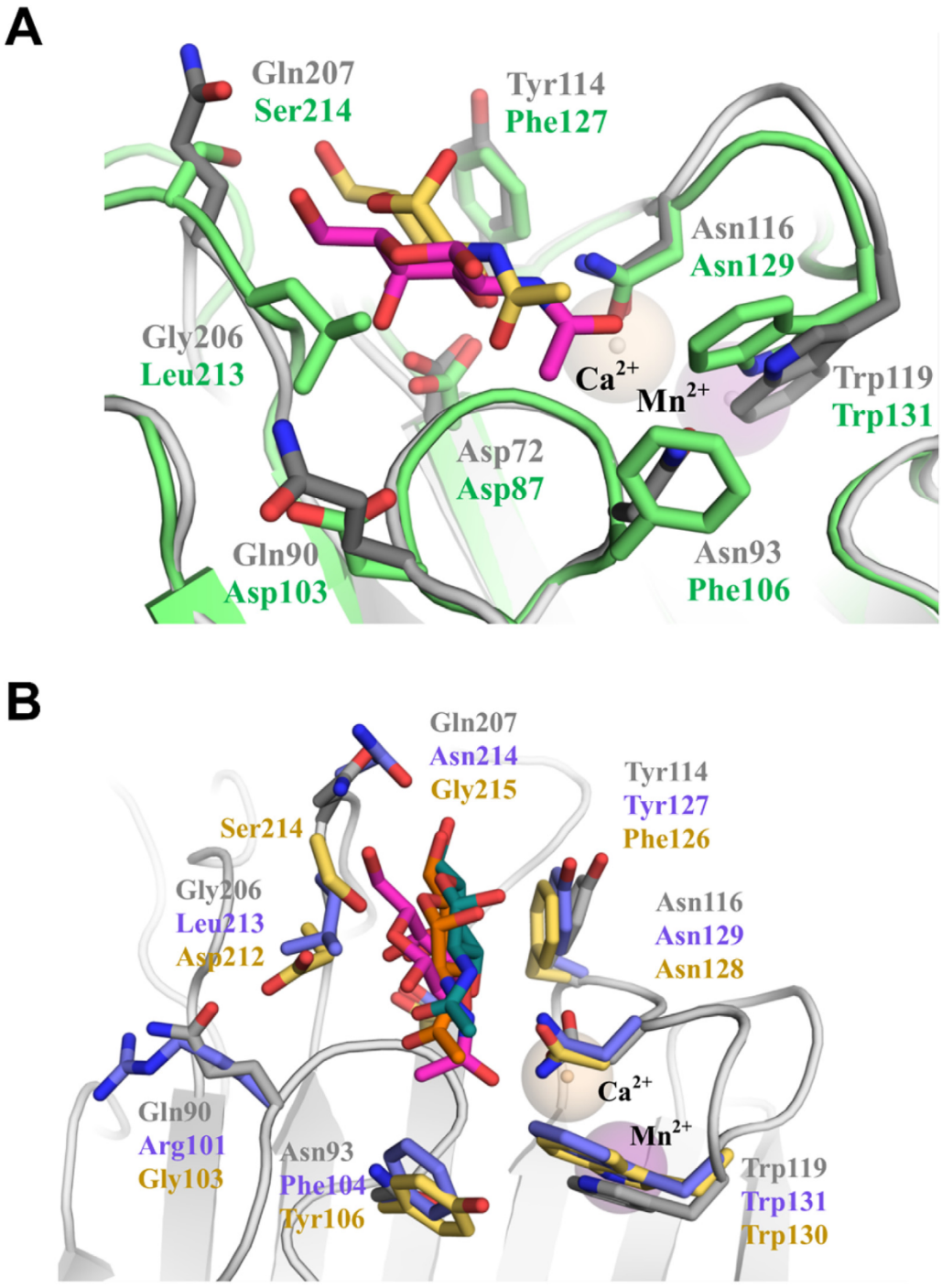

Fig. 6. Superposition of the CBS of SSGL-A, complexed with D-GalNAc, over the CBSs of representative GalNAc-binding legume lectins. (**A**) The three-dimensional molecular model of SSGL-A (gray) complexed with D-GalNAc (pink) was superposed over the x-ray crystallographic structure of VML (PDB ID: 4U36) complexed with the same monosaccharide (yellow). (**B**) The three-dimensional molecular model of SSGL-A (gray) complexed with D-GalNAc (pink) was superposed over the x-ray crystallographic structures of VVLB4 (PDB ID: 1N47; purple sticks) and WBAI (PDB ID: 2D3S; yellow sticks) complexed with the Tn antigen. The GalNAc residue bound to VVLB4 and WBAI are colored green and orange, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Moreover, the QMEAN z-scores of both models deviated less than 1 standard deviation from the mean score in similar sized high-quality protein structures in the reference dataset (Figs. S16 and S17). This showed that the global and local structural features of the tertiary models of SSGL-A and SSGL-E were comparable to those of experimental structures (Benkert et al., 2011). Altogether, these analyses indicated that the modeled structures were of good quality and reliable.

2.4. The predicted structures of SSGL-A and SSGL-E and the functionality of their CBSs

The modeled tertiary structures of both SSGL-A and SSGL-E (Fig. 4A and 4B) had the characteristic structural features of the L-type lectins, in which the CRD is composed of a β-sandwich of two major twisted antiparallel β-sheets. This β-sandwich is usually poor or devoid of any α-helical structure, comprising a major

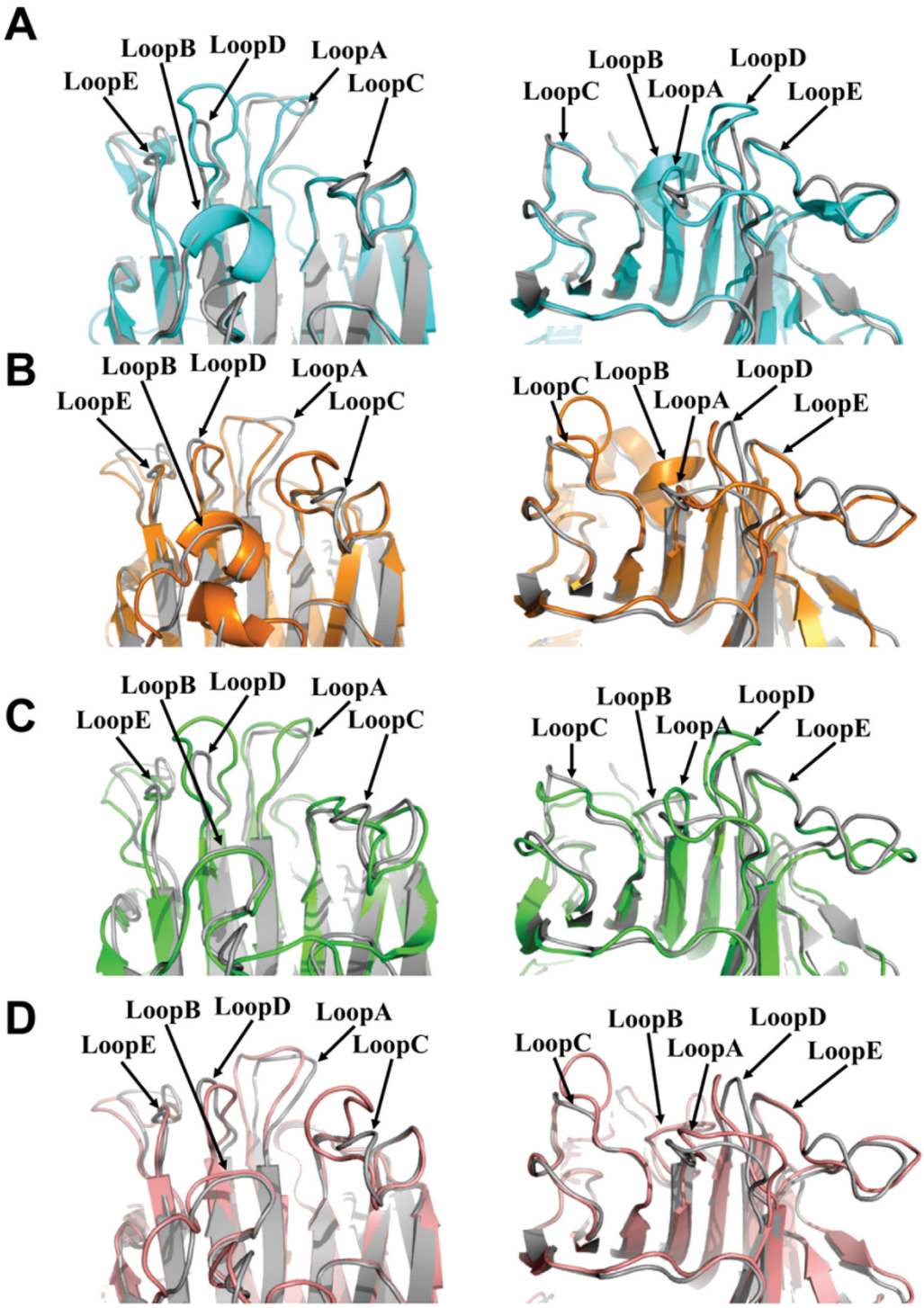

Fig. 7. Comparison of the loop regions that constitute the CBSs of SSGL-A and other legume lectins. The three-dimensional molecular model of SSGL-A (gray) was superposed over the x-ray crystallographic structures of VML (**A**; PDB ID: 4U36), PAL (**B**; PDB ID: 1Q8O), LPL (**C**; PDB ID: 3UJQ) and PELa (**D**; PDB ID: 3ZVX), and the 5 loops (A–E) that constitute their CBSs were identified and are indicated by arrows.

concave seven-stranded β-sheet and a minor convex six-stranded β-sheet, connected by short loops and β-bends (Loris et al., 1998). Molecular docking calculations revealed that, among the monosaccharides evaluated, the CBS of SSGL-A is able to properly accommodate solely D-Gal and D-GalNAc, presenting the same coordination observed for all legume lectins which display binding specificity towards Gal/GalNAc residues (Figs. S18A and S18B). The theoretical AutoDock Vina binding free energies of the top-ranked poses of D-Gal and D-GalNAc, docked in the CBS of SSGL-A, were −4.8 and −5.1 kcal/mol, respectively. In these protein-carbohydrate complexes, D-Gal and D-GalNac are stabilized in the CBS of SSGL-A through a network of hydrogen bonds involving 4 (Asp^{72}, Gly^{92}, Asn^{116} and Gln^{207}) and 5 (Asp^{72}, Gln^{90}, Gly^{92}, Asn^{116} and Gln^{207}) key residues, respectively (Figs. S18A and S18B and Table S8). In both complexes, the C3 and C4 hydroxyl groups of the monosaccharide interact with the O and N atoms of the side chains

of Asp[72] and Asn[116] and the main chain NH group of Gly[92]. On the opposite side of the sugar-binding pocket, the O atom of the pyranoside ring and the C1 and C6 hydroxyl groups interact with the O and N atoms of the side chain of Gln[207]. The main chain carbonyl and NH group of Gln[207] are also involved in these interactions (Table S8). In the interaction of SSGL-A with D-GalNAc, additional hydrogen bonds are formed between the *N*-acetyl O atom of the bound sugar and the main chain NH of Gly[92] and the main chain carbonyl of Gln[90] (Fig. S18B). This could explain the higher predicted affinity of SSGL-A for D-GalNAc (−5.1 kcal/mol) in comparison to that for D-Gal (−4.8 kcal/mol). These predictions also corroborate the previous observation that the hemagglutinating activity of SLL against rabbit erythrocytes is better inhibited by D-GalNAc in comparison to D-Gal (Fernandes et al., 2012). These authors also found that, although the hemagglutinating activity of SSL was inhibited by both D-GalNAc and lactose, SLL only bound to immobilized lactose. Indeed, the CBS of SSGL-A is also able to accommodate lactose (Fig. S18C and Table S9) with a predicted binding free energy = −6.6 kcal/mol. To assess the stability of D-GalNAc and lactose docked in the CBS of SSGL-A, molecular dynamics (MD) simulations were performed. RMSD plots revealed that bound D-GalNAc and lactose reached equilibrium after approximately 2500 and 600 ps of simulation, respectively (Fig. S19). Comparison between the initial and final poses of each accommodated ligand confirmed the essential role of the conserved triad Asp[72]-Gly[92]-Asn[116] for carbohydrate binding (Fig. 5 and Tables S10 and S11). Furthermore, to obtain more accurate predictions on the binding energies of each ligand to the CBS of SSGL-A, DFT-D calculations were performed, based on the stable poses obtained after the MD simulations. It has been shown that binding free energy calculations based on quantum mechanical (QM) methods have a good correlation with the experimental binding energies, allowing the binding affinity between a receptor and different ligands to be rank-ordered more properly. However, due to negligence of solvent effects, the QM interaction energies are much larger than the experimental ones (Chen et al., 2013). Therefore, the theoretical binding energies between SSGL-A and the two investigated ligands, based on the QM calculations, were estimated as −31.7 kcal/mol (D-GalNAc) and −47.5 kcal/mol (lactose), suggesting that SSGL-A likely exhibits greater binding affinity for lactose.

One noteworthy feature of the SSGL-A model is that its CBS is larger when compared to the CBS of VML, whose structure was used as template. In SSGL-A, the loop Thr[205]-Ala[210] is 3 residues shorter than the corresponding loop in VML (Ser[211]-Glu[219]), and this explains its larger sugar-binding pocket. In the complex VML-GalNAc, the sugar pyranose ring is stacked between the aromatic side chain of Phe[127] and the aliphatic side chain of Leu[213], which is located in the loop Ser[211]-Glu[219] (Sousa et al., 2015). Owing to the 3-residues deletion, Gly[206] of SSGL-A occupies the same position occupied by Leu[213] in VML and, as a consequence, the monosaccharide docked in the CBS of SSGL-A is slightly displaced towards the loop Thr[205]-Ala[210] (Fig. 6A). To further support the overall binding specificity of SSGL-A, as predicted using the molecular docking calculations and MD simulations, the D-GalNAc-SSGL-A complex was compared to the x-ray structures of the Gal/GalNAc-binding legume lectins VVLB4 (*Vicia villosa* isolectin B4) (Babino et al., 2003) and WBAI (basic agglutinin I from winged bean, *Psophocarpus tetragonolobus*) (Kulkarni et al., 2005). As shown in Fig. 6B, the orientation of GalNAc bound to SSGL-A is similar to the orientation of the GalNAc residue observed in the complexes between VVLB4 and WBAI with the Tn antigen (GalNAc-α-O-Ser/Thr), in which the acetamido group points towards the floor of the sugar-binding pocket. The most striking difference is the displacement of the galactopyranoside ring in the CBS of SSGL-A, which moves away from the side chain of Tyr[114], towards the loop Thr[205]-Ala[210] (Fig. 6A). It is noteworthy to mention that, although SSGL-A showed high overall sequence identity with the Man/Glc-specific lectins PAL (45.8%) and PELa (46.7%), remarkable differences are observed in the size and conformation of the long loop C region of SSGL-A in comparison to the equivalent region of these 2 proteins (Fig. 7B and 7D). On the other hand, when compared to VML and LPL, which are Gal/GalNAc-specific lectins, the size and conformation of SSGL-A loop C is very similar to the corresponding region of both lectins (Fig. 7A and 7C). The CBS of legume lectins is made up of 5 loops (A-E), which show great variation in size and sequence, and the long loop C is an important determinant of their carbohydrate-binding specificity (Abo et al., 2015; Soga et al., 2015). This comparative analysis shows that the modeled CBS of SSGL-A has the typical features of Gal/GalNAc-binding lectins.

In the case of SSGL-E, proper coordinates were not observed for any of the analyzed ligands, suggesting that it is very likely a non-functional lectin, which is unable to bind carbohydrate. In SSGL-E, the replacement of two key residues relative to VML were identified as the structural basis that could explain its non-functionality. Thus, the conserved Asp residue (Asp[72] in VML) of the invariant triad Gly-Asp-Asn, is replaced by an alanine residue (Ala[73] in SSGL-E), which disrupts important interactions for carbohydrate binding. Furthermore, the residue Leu[213] of VML, whose aliphatic side chain, together with the aromatic ring of Phe[127], helps to stabilize the monosaccharide conformation in the CBS through stacking forces, is replaced by a glycine residue (Gly[209] in SSGL-E), further disrupting key interactions in the CBS of SSGL-E (Fig. S20).

3. Conclusion

In conclusion, mRNAs encoding functional and non-functional L-type lectins are expressed in developing seeds of *S. simplex* var. *grandiflora*. Duplicate gene preservation by subfunctionalization suggests potential benefits for retention of the genes encoding the non-functional lectins.

4. Experimental

4.1. Plant material

Mature and developing seeds of *S. simplex* var. *grandiflora* (Raddi) R.S.Cowan were harvested from a tree growing at the Campus do Pici, UFC, Fortaleza, Ceará, Brazil. Voucher specimens (EAC 54159) were deposited at the Herbário Prisco Bezerra.

4.2. Plasmids, bacterial strains and reagents

The plasmid pGEM-T was purchased from Promega (Madison, WI, USA), whereas the *Escherichia coli* strain TOP10F′ was from Invitrogen (Carlsbad, CA, USA). All other reagents were of analytical grade.

4.3. RNA purification

Total RNA was purified from developing seeds according to the method described by Chang et al. (1993). The integrity of the RNA samples was checked by 1.5% agarose gel electrophoresis and the yield was estimated by measuring the absorbance at 260 nm (Sambrook et al., 1989). The purified total RNA was treated with RQ1 RNase-free DNase I (Promega) at 37 °C for 30 min (1 U of DNase I per μg of RNA) and used for cDNA synthesis.

4.4. 3′ RACE

Rapid amplification of cDNA 3′ ends (3′ RACE) was performed using the protocol described in the manual of the FirstChoice RLM-RACE Kit (Ambion Life Technologies, Carlsbad, CA, USA). To this end, DNA-free total RNA was reverse-transcribed to DNA using the ImProm-II Reverse Transcription System (Promega) and the 3′ RACE adapter (5′-GCGAGCACAGAATTAATACGACTCACTATAGG$(T)_{12}$VN-3′), following the protocol supplied by the enzyme's manufacturer. Next, the first-strand cDNA products were amplified by PCR using a gene-specific primer (5′-CARGARGAYYTNYTNYTNCARGGNGAYGC-3′) and the 3′ RACE outer primer (5′-GCGAGCACAGAATTAA-TACGACT-3′). The gene-specific primer was designed by referencing the N-terminal sequence of SLL (Fig. S1). Amplifications were performed in a final volume of 25 μL, which contained first-strand cDNA (980 ng), 1× reaction buffer (GE Healthcare Life Sciences, Piscataway, NJ, USA), 1.5 mM $MgCl_2$, 200 μM each dNTP, 0.5 μM each primer, and 2 U *Taq* DNA Polymerase (GE Healthcare). The reactions were performed in a PTC-200 thermocycler (MJ Research, USA) using the following cycling parameters: an initial denaturation step (3 min at 95 °C) followed by 32 cycles of 45 s at 95 °C, 45 s at 49 °C, and 1.5 min at 72 °C. After the last cycle, the reactions were further incubated for 8 min at 72 °C, quenched to 4 °C and analyzed by agarose gel electrophoresis.

4.5. Cloning of PCR products, DNA sequencing and sequence assembly

PCR products (Fig. S2) were ligated into the pGEM-T vector using T4 DNA ligase (Promega) at 4 °C for 16 h. Products from the ligation reactions were introduced into *E. coli* TOP10F′ cells by electroporation and the transformants were selected on LB agar containing 100 μg mL^{-1} carbenicillin, 30 μg mL^{-1} streptomycin 0.5 mM IPTG and 80 μg mL^{-1} X-Gal. Plasmid DNA was isolated from antibiotic-resistant colonies using the NucleoSpin Plasmid kit (Macherey-Nagel, Düren, Germany) and the presence of the inserts was confirmed by restriction digestion with *Pvu*II. The inserts of 10 clones were sequenced at the Macrogen Inc. (Seoul, South Korea) using the Sanger's dideoxy chain termination method. The sequences of both strands of each insert were determined using the primers T7 promoter (5′-TAATACGACTCACTATAGGG-3′) and SP6 (5′- ATTTAGGTGACACTATAG-3′). The Phred/Phrap/Consed package (Ewing et al., 1998) (Ewing and Green, 1998) (Gordon et al., 1998) was then used for removing vector and low quality sequences and contig assembly. Six unique cDNA contigs (GenBank accession numbers KX869901-KX869906) encoding 6 distinct polypeptide chains (named SSGL-A to SSGL-F) were generated (Figs. S3–S9).

4.6. Sequence analysis

Manipulation, editing and comparisons of DNA and amino acid sequences were routinely performed with the BioEdit 7.2.5 software package (Hall, 1999). Multiple alignments of DNA and amino acid sequences were performed using the program Clustal W (Thompson et al., 1994), as implemented in BioEdit 7.2.5. Searches for homologous proteins in public sequence databases were performed using BLAST (Altschul et al., 1997). The presence and delimitation of protein domains was performed by searching the NCBI Conserved Domain Database (CDD) through the CD-Search web service (Marchler-Bauer and Bryant, 2004).

4.7. Molecular evolutionary analysis

Amino acid sequences were aligned with ProbCons 1.12 (Do et al., 2005), available at the web server Phylogeny.fr (www.phylogeny.fr/index.cgi). The multiple alignment was used in the phylogenetic analysis, which was performed using the MEGA software, version 6.0 (Tamura et al., 2013). Pairwise evolutionary distances were computed using the Jones-Taylor-Thornton (JTT) method (Jones et al., 1992) and the trees were generated using the neighbor-joining (NJ) method (Saitou and Nei, 1987). The stability of the clades was assessed using the bootstrap method (Felsenstein, 1985).

4.8. Molecular modeling

Comparative protein structure modeling was performed by submitting the target sequences to the M4T server v. 3.0 (manaslu.aecom.yu.edu/M4T/) (Fernandez-Fuentes et al., 2007). The server performs the following main tasks in an automated manner: i) template search and selection, performed by the Multiple Template (MT) module; ii) target sequence to template structure(s) alignment, performed by the Multiple Mapping Method (MMM) module (Rai and Fiser, 2006); and iii) model building, performed by Modeller (Sali and Blundell, 1993). Therefore, the structures of *Lablab purpureus* (LPL; PDB ID: 3UJQ; 38.0 and 41.7% sequence identity with SSGL-A and SSGL-E, respectively) and *Vatairea macrocarpa* lectins (VML; PDB ID: 4U36; 45.5 and 47.2% sequence identity with SSGL-A and SSGL-E, respectively) were selected by the server and used as templates (Figs. S13 and S14). The geometric and stereochemical quality of the modeled structures was evaluated using the programs PROCHECK (Laskowski et al., 1993) and WHAT IF (Vriend, 1990). Model quality assessment was also performed using the QMEAN scoring function (Benkert et al., 2009).

4.9. Molecular docking calculations

A standard docking procedure was performed with AutoDock Vina version 1.1.2, which uses AutoDock Vina scoring function and the Iterated Local Search global optimizer algorithm in the docking calculations (Trott and Olson, 2010). Carbohydrate structures, which were used as potential ligands, were obtained from the PubChem Substance and Compound database (Wang et al., 2009) and were as follows: L-fucose (PubChem CID: 1706), D-galactose (PubChem CID: 6036), *N*-acetyl-D-galactosamine (PubChem CID: 84265), D-glucose (PubChem CID: 5793), *N*-acetyl-D-glucosamine (PubChem CID: 439174), D-mannose (PubChem CID: 18950) and α-lactose (PubChem CID: 84571). All of the torsional bonds of the ligands were free to rotate while the protein was held rigid. To prepare the protein and the ligands, polar hydrogen atoms were added using the AutoDock Tools version 1.5.6, and Kollman united atom partial charges (Morris et al., 1998) were assigned. The search space for the docking calculations, which was selected based on the oligosaccharides length, was defined by a 20 Å × 20 Å × 20 Å cube centered on the conserved CBS of SSGLs. Exhaustiveness was set to 15, and for all other parameters, default values were used. The ten top results, ranked according to the predicted binding affinity (expressed in kcal/mol), were analyzed. The solutions were first chosen based on the coordination of ligands co-crystallized with VML (Sousa et al., 2015) and the best predicted bound conformations were further ranked based on their theoretical binding energy values. Three-dimensional images of the ligand-protein complexes were prepared using the PyMol Molecular Graphics System, version 1.7.4 (Schrödinger, LLC).

4.10. Molecular dynamics (MD) simulations

The simulations were performed using the program NAMD

v2.10 (Nanoscale Molecular Dynamics) (Phillips et al., 2005), with the force fields CHARMM27 (Best et al., 2012) and CHARMM36 (Guvench et al., 2011), using the monomeric structure of SSGL-A and considering as starting coordinates for the carbohydrate ligands the best results from docking calculations. The software VMD (Visual Molecular Dynamics) was used for file preparation (Humphrey et al., 1996). The systems were solvated in cubic boxes of TIP3P water molecules with a margin of 10 Å along each dimension (Jorgensen et al., 1983). The net charge of the systems was neutralized by adding Na^+ or Cl^-, and periodic boundary conditions were applied to avoid edge effects. The nonbonded interactions were calculated using the switching method and a cut-off between 10 and 12 Å, along with particle-mesh Ewald summation, using a grid spacing of 1 Å for long range electrostatic interactions (Darden et al., 1993). The simulations were performed under NPT (constant normal pressure and normal temperature) conditions, using Langevin dynamics at a temperature of 300 K with a Langevin piston to maintain constant pressure at 1 bar. A time step of 2 fs was used with the SHAKE algorithm to fix bonds involving hydrogen atoms (Phillips et al., 2005).

Water molecules were initially minimized for 1000 steps (using the conjugate gradient energy minimization algorithm) and equilibrated for 200 ps, while the solute atoms were harmonically restrained. Afterwards, the complete systems were minimized during 10000 steps and then heated linearly from 100 K to 300 K over 200 ps. Finally, 5 ns MD simulations were performed for each system under NPT and the coordinates of all atoms were saved every 10 ps during the entire MD simulations. The ligand and protein stability during the MD simulations were measured by their deviation from the initial coordinates in terms of RMSD (Shen et al., 2013).

4.11. Binding energy calculations using a dispersion-corrected density functional theory (DFT-D) method

An imaginary sphere with radius of 5 Å was drawn from the coordinates of each carbohydrate docked in the CBS of the protein, as obtained after the 5 ns MD simulations. All residues near the ligands with at least one atom within the imaginary spheres were selected for the calculation of the interaction energy between the ligand and the CBS. First, the energy of the system formed by the protein fragment containing the CBS complexed with the carbohydrate was calculated ($E_{complex}$). Then, the energies of the fragment containing the CBS without the ligand (E_{model}) and of the isolated carbohydrate (E_{ligand}) were obtained. The interaction energy between the carbohydrate and the CBS (ΔE) was then calculated as follows: $\Delta E = E_{complex} - (E_{model} + E_{ligand})$.

The DFT simulations on the classically optimized structures were performed using the DMOL3 code (Delley, 2000), adopting a dispersion-corrected GGA+D exchange-correlation functional, which is adequate to describe systems where hydrogen bonds and dispersive forces, such as van der Waals interactions, are present. To model the dispersive forces, the scheme proposed by Tkatchenko and Scheffler (2009) was adopted. A Double Numerical Plus Polarization (DNP) basis set was selected to expand the Kohn-Sham orbitals of all electrons. The orbital cut-off for these computations was set to 3.7 Å and a total energy variation smaller than 10^{-6} Ha for two consecutive self-consistent field cycles was considered to achieve convergence. All interaction energy calculations were performed using the COSMO continuum solvation model (Klamt and Schüürmann, 1993) with dielectric constant (relative to the vacuum) $\varepsilon = 20$ (Vicatos et al., 2009).

Acknowledgements

This work was supported by research grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Projeto CNPq-Bionorte, grant number: 554307/2010-3), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.phytochem.2017.04.007.

References

Abo, H., Soga, K., Tanaka, A., Tateno, H., Hirabayashi, J., Yamamoto, K., 2015. Mutated leguminous lectin containing a heparin-binding like motif in a carbohydrate-binding loop specifically binds to heparin. PloS One 10, e0145834. http://dx.doi.org/10.1371/journal.pone.0145834.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Babino, A., Tello, D., Rojas, A., Bay, S., Osinaga, E., Alzari, P.M., 2003. The crystal structure of a plant lectin in complex with the Tn antigen. FEBS Lett. 536, 106–110.

Benevides, R.G., Ganne, G., Simões, R. da C., Schubert, V., Niemietz, M., Unverzagt, C., Chazalet, V., Breton, C., Varrot, A., Cavada, B.S., Imberty, A., 2012. A lectin from Platypodium elegans with unusual specificity and affinity for asymmetric complex N-glycans. J. Biol. Chem. 287, 26352–26364. http://dx.doi.org/10.1074/jbc.M112.375816.

Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343–350. http://dx.doi.org/10.1093/bioinformatics/btq662.

Benkert, P., Künzli, M., Schwede, T., 2009. QMEAN server for protein model quality estimation. Nucleic Acids Res. 37, W510–W514. http://dx.doi.org/10.1093/nar/gkp322.

Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., Mackerell, A.D., 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. J. Chem. Theory Comput. 8, 3257–3273. http://dx.doi.org/10.1021/ct300400x.

Bond, C.S., Schüttelkopf, A.W., 2009. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr. D Biol. Crystallogr. 65, 510–512. http://dx.doi.org/10.1107/S0907444909007835.

Cardoso, D., Pennington, R.T., de Queiroz, L.P., Boatwright, J.S., Van Wyk, B.-E., Wojciechowski, M.F., Lavin, M., 2013. Reconstructing the deep-branching relationships of the papilionoid legumes. South Afr. J. Bot. 89, 58–75. http://dx.doi.org/10.1016/j.sajb.2013.05.001.

Cavalcanti, M., Coelho, L., 1990. Isolation and partial purification of a lectin from Swartzia pickellii Killip (white jacaranda). Mem. Inst. Oswaldo Cruz 85, 371–372.

Chang, S., Puryear, J., Cairney, J., 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. 11, 113–116. http://dx.doi.org/10.1007/BF02670468.

Chen, M., Hiang, H., Lin, X., Chen, Y., Wang, M., Liu, J., 2013. DFT study of binding energies between acetohydroxyacid synthase and its sulfonylurea inhibitors: an application of quantum pseudoreceptor model. Commun. Comput. Chem. 1, 72–87. http://dx.doi.org/10.4208/cicc.2013.v1.n1.8.

Dang, L., Van Damme, E.J.M., 2016. Genome-wide identification and domain organization of lectin domains in cucumber. Plant Physiol. Biochem. 108, 165–176. http://dx.doi.org/10.1016/j.plaphy.2016.07.009.

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092. http://dx.doi.org/10.1063/1.464397.

De Schutter, K., Van Damme, E.J.M., 2015. Protein-carbohydrate interactions as part of plant defense and animal immunity. Molecules 20, 9029–9053. http://dx.doi.org/10.3390/molecules20059029.

Delley, B., 2000. From molecules to solids with the DMol3 approach. J. Chem. Phys. 113, 7756–7764. http://dx.doi.org/10.1063/1.1316015.

Do, C.B., Mahabhashyam, M.S.P., Brudno, M., Batzoglou, S., 2005. ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res. 15, 330–340. http://dx.doi.org/10.1101/gr.2821705.

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194.

Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8, 175–185.

Favre-Godal, Q., Dorsaz, S., Queiroz, E.F., Marcourt, L., Ebrahimi, S.N., Allard, P.-M., Voinesco, F., Hamburger, M., Gupta, M.P., Gindro, K., Sanglard, D., Wolfender, J.-L., 2015. Anti-Candida cassane-type diterpenoids from the root bark of Swartzia

simplex. J. Nat. Prod. 78, 2994–3004. http://dx.doi.org/10.1021/acs.jnatprod.5b00744.
Felsenstein, J., 1985. Confidence-limits on phylogenies - an approach using the bootstrap. Evolution 39, 783–791. http://dx.doi.org/10.2307/2408678.
Fernandes, A.V., Ramos, M.V., Goncalves, J.F.C., Maranhao, P.A.C., 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.O.M., Moreno, F.B., Pereira, J.O., de Carvalho Gonçalves, J.F., 2012. Purification and characterization of a lectin of the Swartzieae legume taxa. Protein Pept. Lett. 19, 1082–1088.
Fernandez-Fuentes, N., Madrid-Aliste, C.J., Rai, B.K., Fajardo, J.E., Fiser, A., 2007. M4T: a comparative protein structure modeling server. Nucleic Acids Res. 35, W363–W368. http://dx.doi.org/10.1093/nar/gkm341.
Gordon, D., Abajian, C., Green, P., 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8, 195–202.
Guvench, O., Mallajosyula, S.S., Raman, E.P., Hatcher, E., Vanommeslaeghe, K., Foster, T.J., Jamison, F.W., Mackerell, A.D., 2011. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. J. Chem. Theory Comput. 7, 3162–3180. http://dx.doi.org/10.1021/ct200328p.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph 14, 33–38.
Jiang, S.-Y., Ma, Z., Ramachandran, S., 2010. Evolutionary history and stress regulation of the lectin superfamily in higher plants. BMC Evol. Biol. 10, 79. http://dx.doi.org/10.1186/1471-2148-10-79.
Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.
Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926. http://dx.doi.org/10.1063/1.445869.
Klamt, A., Schüürmann, G., 1993. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc. Perkin Trans. 2, 799–805. http://dx.doi.org/10.1039/P29930000799.
Kulkarni, K.A., Sinha, S., Katiyar, S., Surolia, A., Vijayan, M., Suguna, K., 2005. Structural basis for the specificity of basic winged bean lectin for the Tn-antigen: a crystallographic, thermodynamic and modelling study. FEBS Lett. 579, 6775–6780. http://dx.doi.org/10.1016/j.febslet.2005.11.011.
Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291. http://dx.doi.org/10.1107/S0021889892009944.
Lavin, M., Herendeen, P.S., Wojciechowski, M.F., 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst. Biol. 54, 575–594. http://dx.doi.org/10.1080/10635150590947131.
Loris, R., Hamelryck, T., Bouckaert, J., Wyns, L., 1998. Legume lectin structure. Biochim. Biophys. Acta 1383, 9–36.
Loris, R., Van Walle, I., De Greve, H., Beeckmans, S., Deboeck, F., Wyns, L., Bouckaert, J., 2004. Structural basis of oligomannose recognition by the Pterocarpus angolensis seed lectin. J. Mol. Biol. 335, 1227–1240.
Macedo, M.L.R., Oliveira, C.F.R., Oliveira, C.T., 2015. Insecticidal activity of plant lectins and potential application in crop protection. Molecules 20, 2014–2033. http://dx.doi.org/10.3390/molecules20022014.
Magalhães, A.F., Tozzi, A.M.G., de, A., Santos, C.C., Serrano, D.R., Zanotti-Magalhães, E.M., Magalhães, E.G., Magalhães, L.A., 2003. Saponins from Swartzia langsdorffii: biological activities. Mem. Inst. Oswaldo Cruz 98, 713–718.
Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32, W327–W331. http://dx.doi.org/10.1093/nar/gkh454.
Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. 19, 1639–1662. http://dx.doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.
Osawa, K., Yasuda, H., Maruyama, T., Morita, H., Takeya, K., Itokawa, H., 1992. Isoflavanones from the heartwood of Swartzia polyphylla and their antibacterial activity against cariogenic bacteria. Chem. Pharm. Bull. (Tokyo) 40, 2970–2974.
Peumans, W.J., Van Damme, E.J., 1995. Lectins as plant defense proteins. Plant Physiol. 109, 347–352.
Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., Schulten, K., 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26, 1781–1802. http://dx.doi.org/10.1002/jcc.20289.
Pinto, R.B., de, F.M., Torke, B.M., Forni-Martins, E.R., 2016. Evidence for a conserved karyotype in Swartzia (Fabaceae, Papilionoideae): implications for the taxonomy and evolutionary diversification of a species-rich neotropical tree genus. Brittonia 68, 93–101. http://dx.doi.org/10.1007/s12228-015-9395-z.
Rai, B.K., Fiser, A., 2006. Multiple mapping method: a novel approach to the sequence-to-structure alignment problem in comparative protein structure modeling. Proteins 63, 644–661. http://dx.doi.org/10.1002/prot.20835.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815. http://dx.doi.org/10.1006/jmbi.1993.1626.
Sambrook, J., Fritsch, E., Maniatis, T., 1989. Molecular Cloning: a Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
Sharon, N., Lis, H., 1990. Legume lectins–a large family of homologous proteins. FASEB J. 4, 3198–3208.
Shen, J., Zhang, W., Fang, H., Perkins, R., Tong, W., Hong, H., 2013. Homology modeling, molecular docking, and molecular dynamics simulations elucidated α-fetoprotein binding modes. BMC Bioinforma. 14 (Suppl. 14), S6. http://dx.doi.org/10.1186/1471-2105-14-S14-S6.
Sippl, M.J., 1993. Recognition of errors in three-dimensional structures of proteins. Proteins 17, 355–362.
Soga, K., Abo, H., Qin, S.-Y., Kyoutou, T., Hiemori, K., Tateno, H., Matsumoto, N., Hirabayashi, J., Yamamoto, K., 2015. Mammalian cell surface display as a novel method for developing engineered lectins with novel characteristics. Biomolecules 5, 1540–1562. http://dx.doi.org/10.3390/biom5031540.
Sousa, B.L., Silva Filho, J.C., Kumar, P., Pereira, R.I., Łyskowski, A., Rocha, B.A.M., Delatorre, P., Bezerra, G.A., Nagano, C.S., Gruber, K., Cavada, B.S., 2015. High-resolution structure of a new Tn antigen-binding lectin from Vatairea macrocarpa and a comparative analysis of Tn-binding legume lectins. Int. J. Biochem. Cell Biol. 59, 103–110. http://dx.doi.org/10.1016/j.biocel.2014.12.002.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. http://dx.doi.org/10.1093/molbev/mst197.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Tkatchenko, A., Scheffler, M., 2009. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. Phys. Rev. Lett. 102, 073005. http://dx.doi.org/10.1103/PhysRevLett.102.073005.
Torke, B.M., Schaal, B.A., 2008. Molecular phylogenetics of the species-rich neotropical genus Swartzia (Leguminosae, Papilionoideae) and related genera of the swartzioid clade. Am. J. Bot. 95, 215–228.
Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. http://dx.doi.org/10.1002/jcc.21334.
Van Damme, E.J., Barre, A., Bemer, V., Rougé, P., Van Leuven, F., Peumans, W.J., 1995. A lectin and a lectin-related protein are the two most prominent proteins in the bark of yellow wood (Cladrastis lutea). Plant Mol. Biol. 29, 579–598.
Van Holle, S., Van Damme, E.J.M., 2015. Distribution and evolution of the lectin family in soybean (Glycine max). Molecules 20, 2868–2891. http://dx.doi.org/10.3390/molecules20022868.
Vicatos, S., Roca, M., Warshel, A., 2009. Effective approach for calculations of absolute stability of proteins using focused dielectric constants. Proteins 77, 670–684. http://dx.doi.org/10.1002/prot.22481.
Vriend, G., 1990. WHAT IF: a molecular modeling and drug design program. J. Mol. Graph 8, 52–56.
Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., Bryant, S.H., 2009. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. 37, W623–W633. http://dx.doi.org/10.1093/nar/gkp456.
Zhang, Y., Skolnick, J., 2004. Scoring function for automated assessment of protein structure template quality. Proteins 57, 702–710. http://dx.doi.org/10.1002/prot.20264.