

# Predictive structure and protein-ligand interface of novel lectin from rice bean (*Vigna umbellata*)

**Ankur Tripathi**

CSK HPKV: Himachal Pradesh Agricultural University

**Vipin Hallan**

IHBT: Institute of Himalayan Bioresource Technology CSIR

**Ritu Raj**

IHBT: Institute of Himalayan Bioresource Technology CSIR

**Rajan Katoch** (✉ [rajankatoch@yahoo.com](mailto:rajankatoch@yahoo.com))

CSK HPKV: Himachal Pradesh Agricultural University <https://orcid.org/0000-0001-6210-2916>

---

## Research Article

**Keywords:** Legume lectins, Rice bean lectin (RbL), Homology modeling, Molecular docking

**Posted Date:** May 7th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-491393/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Lectins are an important group of multivalent glycoproteins having the property of selectively recognizing and precipitating glycoconjugates. Although lectins have been reported from diverse biological sources, legume lectins is the best-characterized family of plant lectins. We have successfully cloned and sequenced the RbL ORF of 843bp from immature rice bean seeds (*Vigna umbellata*). We report the results of *in silico* analysis of novel lectin precursor of 280 amino acids from rice bean. BlastP analysis revealed more than 90% sequence similarity of RbL protein with *Vigna angularis* lectin and *Vigna aconitifolia* lectin. ProtParam analysis revealed acidic, stable and hydrophobic nature of RbL protein. Template based 3D structure of RbL protein was modeled using I-TASSER tool and validated as good quality model. Structural analysis revealed the presence of  $\beta$ -sandwich (Jelly roll fold or lectin fold) in modeled RbL structure. RbL protein was functionally annotated as a plant defense protein. Molecular docking was performed to analyze interactions of RbL protein with predicted ligands (N-acetyl-D-glucosamine  $\beta$ -galactose, Lactose and Adenine) and two selected ligands (Glucose and Mannose). Molecular dynamics (MD) simulations of RbL-ligand complexes confirmed robust hydrogen bonding interactions between ligands and RbL protein. The novel information generated in the study would be useful in exploring RbL protein for different biomedical and biotechnological applications.

# Introduction

Lectins constitute an important group of glycoproteins having ability to bind and precipitate glycoconjugates selectively and reversibly. These proteins are also known as hemagglutinins, for their selective binding with cell surface receptors on erythrocytes and make them agglutinate (Tsaneva and Van Damme 2020). This property differentiates lectins from other glycoproteins and is the easiest way of identification of lectins from different biological sources. Lectins have been identified from all three domains of life *viz.*, eukaryotes, prokaryotes and archea. Among the different groups of lectins, plant lectins are thoroughly investigated group. More than a century ago, Stillmark (1888) gave the first description of lectins while investigating the effect of castor bean extract (*Ricinus communis* L.) on red blood cells. Since the first insight of lectins, numerous plant lectins have been investigated for their biochemical and biological properties (Vandenborre et al. 2011).

Structurally, lectins have been classified into three categories, *viz.*, merolectins, hololectins and chimerolectins (Peumans and Van Damme 1995). These are also classified into different groups depending on their carbohydrate binding specificity (Van Damme et al. 1998). Based on sequence and structure of lectin motif, plant lectins have also been categorized into twelve families (Van Damme et al. 2008). Carbohydrate binding specificity enable lectins to exert multiple biological activities such as mitogenic, antibacterial, antifungal, antiviral, anticancer and insecticidal activities (Gautam et al. 2018); symbiotic interactions between host plants and microbes (De Hoff et al. 2009), plant hormone mediated growth and development (Lannoo et al. 2007; Lannoo and Van Damme 2010) and signaling (Jin et al. 2009).

Though lectins are ubiquitously distributed and associated with all life forms, the most investigated plant lectins belong to the family Leguminosae (Cavada et al. 2020) and most of them have been purified from mature seeds comprising up to 10% of total storage proteins (Laija et al. 2010). Apart from seeds, lectins have also been identified from vegetative tissues in lower amount. In seeds, lectins are synthesized as preprecursor molecule and sequestered in protein bodies during seed development and broken down during seed germination for providing essential amino acids (Van Damme et al. 2008). The inactive preprecursor polypeptide is converted into mature protein through distinct proteolytic processes such as C-terminal trimming of polypeptide.

Despite high structural and sequential similarities, legume lectins show variability in carbohydrate specificity and quaternary structures. The primary structure of legume lectins is made up from protomers of 250 to 300 amino acid residues (~30kDa). Legume lectins have higher proportion of  $\beta$ -sheets,  $\beta$ -turns and negligible amount of  $\alpha$ -helix in their secondary structure (Swamy et al. 1985), thus considered as  $\beta$ -sheet proteins. The 3D structure of legume lectins consist of flat six stranded anti-parallel  $\beta$ -sheets (Back face) and seven-stranded curved anti-parallel  $\beta$ -sheets (Front face), interconnected by different loops. The dome like structure represent the  $\beta$ -sandwich structure, related to Jelly roll fold also known as lectin fold with two hydrophobic cores, present between front and back  $\beta$ -sheets and between the curled loop and front  $\beta$ -sheets (Lagarda-Dias et al. 2017). Concavity in front  $\beta$ -sheets and second hydrophobic core is suited well to bind with carbohydrates moieties. Quaternary structure of legume lectins is characterized by an oligomeric structure in which monomers are assembled as homodimers (Canonical legume lectin dimer) or homotetrameric structure (dimers of dimers). Moreover, heterotetrameric structures have also been observed (Hamelryck et al. 1998). Different types of quaternary structures [Canonical, ECorL-type, GS4-type, DBL-type, ConA-type, PNA-type, GS1-type and DB58-type and Arcelin-5-type (monomeric) structures] are known for legume lectins (Brinda et al. 2005).

Legume lectins require divalent cations particularly  $\text{Ca}^{+2}$  and  $\text{Mn}^{+2}$  for their activity as they are essential for maintaining the stability of carbohydrate binding site. The metal binding site formed by the side chains of glutamic acid, aspartic acid and histidine amino acids in legume lectins is highly conserved (Rini 1995). The carbohydrate and metal-binding sites are localized in close vicinity at top of the front sheet in legume lectins structure (Cummings et al. 2017). Four loops A, B, C and D are responsible for the formation of carbohydrate binding sites (Sharma and Surolia 1997). The diversity in carbohydrate specificity of legume lectins is mainly due to variability in conformation and size of the D loop. To the some extent, the variability in carbohydrate binding is also specified by the loop C (Young and Oomen 1992). Variability in carbohydrate-binding specificities makes legume lectins a valuable tool in glycome research (Coelho et al. 2017). Legume lectins are the best model system to study protein- sugar interactions (Loris et al. 1998).

With the development of advance computational tools, the analysis of large amount genomic and proteomic data has become easy and less time consuming. Critical analysis of information embedded in structure of proteins is essential to characterize their physic-chemical properties and functions. Though reports on structural and functional characterization of various legumes lectins are available (Pinto et al.

2008; Moreira et al. 2013; Li et al. 2014; Filho et al. 2017; Osman and Konozy 2017), but there is no such report available on lectin from rice bean [*Vigna umbellata* (Thunb.) Ohwi and Ohashi]. Herein, we presented the *in silico* analysis of RbL protein for its structural and functional characterization. Molecular docking and MD stimulations of RbL protein-ligands complexes was also performed to study RbL protein-ligand interactions.

## Materials And Methods

In the present investigation, RbL ORF sequence which had been cloned and sequenced from immature rice bean seeds using degenerate primer set, was translated into 280 amino acids long polypeptide with ExPASy translate software (<http://web.expasy.org/translate/>) (Gasteiger et al. 2005). The deduced amino acid sequence was analyzed with different computational tools for structural and functional characterization of RbL protein.

### Primary and secondary structure analysis

Physicochemical parameters of deduced RbL sequence were analyzed using ProtParam server (<http://web.expasy.org/protparam>) (Gasteiger et al. 2005). PHD web server (Rost 1996) was used for secondary structure analysis of RbL protein. ProteinPredict (<https://www.predictprotein.org>) web server (Rost et al. 2004) was used to predict putative functions of RbL protein based on gene ontology. Conserved motif structure was analyzed using MEME SUITE tool (Bailey et al. 2009) with parameters including maximum number of motifs =7; minimum width of motif=6; and maximum width of motif = 50. The NCBI CDS online tool was used to analyze gene-encoded protein domains (Marchler-Bauer et al. 2015). We used Kyte and Doolittle (1982) method at ProtScale software (<http://web.expasy.org/protscale>) for hydrophobicity analysis of RbL protein and Signal IP 3.0 server for the prediction of signal peptide cleavage sites (Bendtsen et al. 2005). The putative *N*-glycosylation sites were predicted with NetNGly1.0 server (<http://www.cbs.dtu.dk/services/NetNGlyc/>) (Gupta and Brunak 2002).

### Sequence alignment, homology modeling, binding site prediction and molecular docking analysis

RbL protein sequence was aligned with other legume lectins sequences using Clustal W algorithm and formatted with ESPrpt2.2 (Gouet et al. 2003). RbL protein structure was modeled using I-TASSER modeling server (Roy et al. 2010). The model validation was done through the Zlab server (Anderson et al. 2005). Using ProSA-web interactive server, Z-score was also calculated to recognize errors in RbL protein structure (Wiederstein and Sippl 2007). GalaxySite server (<http://galaxy.seoklab.org>) (Heo et al. 2014) was used to predict binding ligands for RbL protein. Due to the absence of binding site knowledge, COACH (meta-server) was used to predict protein-ligand binding site in modeled RbL structure. COACH produced ligand-binding sites employing two complementary approaches, S-SITE and TM-SITE (Yang et al. 2013a, b).

CDOCKER component from Discovery Studio was adopted for molecular docking test which is a CHARMM based docking engine (Wu et al. 2003). The receptor is firm, while ligands are resilient throughout the docking. For each ligand, the interaction and CHARMM energy symbolize binding affinity (Brooks et al. 1983). Water particles are usually raised in a semi-flexible and firm docking because water particles might harm the protein-ligand complex structure. Water particles were released and hydrogen atoms were attached to protein. The RbL binding site spheres were labeled as the areas inside a radius of 10Å from the ligands geometric centroid. The structures of recognized hits were arranged and docked into RbL-binding pocket. Various poses for every selected molecule were produced and interpreted based on -CDOCKER interaction energy. The 2D interactions were generated from PLIP server (Salentin et al. 2015).

MD-simulations were conducted to investigate the binding stability of docked complexes of RbL protein with ligands adenine, glucose, β-galactose, lactose, mannose and N-acetyl-D-glucosamine executed in GROMACS 4.6.5 module using Gromos96 43a1 forcefield parameter (Hess et al. 1997; Van Der Spoel et al. 2005; Abraham et al. 2015). The PRODRG server was employed to produce ligands topologies which can be utilized in GROMACS (Schüttelkopf et al. 2004). The SPC water state was selected and the box was designated as cuboid. Periodical terminal limitations were implemented in all regions and Na<sup>+</sup>Cl<sup>-</sup> ions were affixed to compensate the system. The 50,000 moves of steepest descents energy minimization were carried to rest any steric disputes. Energy minimization was then ensued by 1000ps NPT equilibration. The particle mesh Ewald approach was applied to reckon the extensive range of electrostatic interplays and the bond length restrained by the LINCS algorithm (Hess et al. 1997). Parrinello-Rahman barostat and V-rescale thermostat were performed sequentially to sustain the steady pressure and heat (1 bar and 300 K). The 10 ns production run was conducted and the coordinates were kept. The generated trajectories were examined by outlining the root mean square deviation (RMSD), root mean square fluctuation (RMSF) and hydrogen bond (H-bond) for every frame.

## Result And Discussion

We have successfully cloned and sequenced RbL ORF of 843bp encoding 280 amino acids from immature rice bean beads using degenerate primers set. The nucleotide sequence of RbL cDNA has been submitted to the GenBank Database (The sequence has been submitted to the GenBank Database under accession No.MT043160). Herein, we report the *in silico* analysis of novel lectin precursor from rice bean (RbL). The results of the study have been discussed under the subsequent subheads.

### Primary and secondary structure analysis of RbL protein

ProtParam analysis revealed that RbL protein was composed of 280 amino acids (Supplementary Figure 1) of which Serine was the most abundant amino acid with 14.60% occurrence, followed by leucine (10.40%), threonine (7.50%) and glycine (7.10%) (Figure-1). Number of positive charged and negatively charged amino acid residues in RbL protein were 21 and 18, respectively. Similar to *Erythrina* lectin (EcorL) (Osman and Konozy 2017) and Winged bean lectin (WBL) (Kortt 1985), RbL protein had high

proportion of hydroxylic amino acids, aliphatic and acidic amino acids with low methionine content and no cysteine residues. In some lectins such as lima bean lectin (*Phaseolus lunatus*) cysteine residues is essential for both carbohydrates and metal ions binding activities (Roberts and Goldstein 1984). Several cysteine residues have been also been conserved in hevein lectin domain for maintaining the structure and folding (Aboitiz et al. 2004). Blast-P analysis of deduced RbL protein sequence showed 97.14% and 91.79% similarities with from Adzuki bean (*Vigna angularis*) and Moth bean (*Vigna aconitifolia*) lectins, respectively. The knowledge of pI (Isoelectric pH) of a protein is important for having an assessment of solubility, sub-cellular localization and its interactions. Cytoplasmic proteins have pI lower than physiological pH, while those in nucleus have pI more than the physiological pH (Nandi et al. 2005). Theoretical pI of RbL protein was 5.76 indicating negative charge on protein and acidic nature. The molecular weight of the deduced RbL protein was inferred to be 29833.57 Da. The half-life is a prediction of the time taken for half of the protein to disappear after its synthesis in cell. ProtParam software relies on "N end rule", which relates half-life of a protein to the identity of its N-terminal residue (Varshavsky 2011). The analysis revealed estimated half life of RbL protein was 30hrs.

The instability index value of 34.99 (Instability index < 40) revealed good stability of RbL protein. The high aliphatic index value (93.99) and the positive GRAVY value (0.145) revealed thermal stability and hydrophobic nature of putative RbL protein. Approximately 50% amino acids had positive hydrophobicity score (>0) which also confirmed hydrophobicity nature of RbL protein (Figure-2). Proteins with hydrophobicity score < 0 and > 0 are more likely to be globular (hydrophilic protein) and membranous (hydrophobic) in nature, respectively (Magdeldin et al. 2012).

Signal peptide of 15 to 20 amino acids length is usually present at the N-terminus of newly synthesized proteins. The presence of signal peptide determines their fate to enter secretory pathway and transportation to reside either inside certain organelles (Plant vacuole, or protein bodies), secreted from cell, or inserted into most cellular membranes (Tsaneva and Van Damme 2020). Most of constitutively expressed lectins are synthesized with signal peptide as a precursor. Signal peptide directs the transportation of proteins to ER lumen in order to confirm their entry into secretory pathway (Van Damme et al. 2008). During co-translational processing, signal peptide is cleaved off by a signal peptidase enzyme in ER lumen. Most of constitutively expressed legume lectins synthesized with a signal peptide and have destination to plant vacuole or extracellular space. Contrary to it, induced lectins are synthesized with no signal peptide, thus reside in nucleus or in cytoplasm (Lannoo and Van Damme 2010).

Signal IP 3.0 server predicted two signal peptide cleavage sites in RbL protein using two bioinformatics algorithms. The Hidden Markov Model (HMM) and Neural Network (NN) algorithms predicted the size of signal peptide of 24 and 26 amino acids, respectively (Supplementary Figure 2a and 2b). Signal peptide cleavage sites in *Vigna aconitifolia* (A<sub>24</sub>-A<sub>25</sub>) and *Vigna unguiculata* (A<sub>26</sub>-A<sub>27</sub>) lectin sequences have also been reported by Filho et al. (2017). The presence of signal peptide in RbL protein indicated that the protein enters into endomembrane system for transportation through membrane to or through cell membrane, to cellular compartments such as plant vacuole for storage. Signal peptide with its

hydrophobic core ( $\alpha$ -helical region) has the property of interacting with hydrophobic interior of membrane. Signal peptide with its hydrophobic core ( $\alpha$ -helical region) interacts with hydrophobic interior of plasma membrane and initiate the transport of protein across the membrane. A 13 amino acids (8 to 21) long hydrophobic region in signal peptide of RbL protein was predicted with Phobius server (<http://phobius.sbc.su.se/>) (Käll et al. 2007). The location of *Vigna aconitifolia* and *Vigna unguiculata* has been reported in plasma membrane and predicted stress response activity as one of their major biological functions (Filho et al. 2017). Since both legume lectins reveal high sequence similarity with RbL protein, therefore their functions could be assigned to the RbL protein. Moreover, the presence of conserved domain for lectin receptor kinase (LecRK) signifies the putative role of RbL protein in plant defense. Domain mapping of RbL sequence with NCBI CDS online tool revealed the presence of conserved lectin\_legume\_LecRK\_Arcelin\_ConA domain (domain architecture ID 10160902) (amino acid residues from 30 to 260). Moreover, dimer and tetramerization domains were found highly conserved in RbL protein which also indicated that protein would be of multimeric nature (Supplementary Figure 3a).

The identification of motifs in large genomic sequences is a cumbersome, however it become easier when motifs are large and have high sequence similarity (Ujinal et al. 2019). The analysis of conserved motifs in legume lectins including RbL protein was performed by MEME Suite tool ([meme-suite.org/tools/meme](http://meme-suite.org/tools/meme)). The suite function by looking for repeated, un-gapped sequence patterns in the provided sequences. The suite determines width and number of occurrences of each motif repeatedly in order to minimize 'e-value' of motif. The predicted motifs with a lower e-value have the probability of finding out an equally well conserved region in target sequence (Unger and Sussman 1993). Initially the suite tool was set with parameters for motif identification in the sequences with maximum number of motifs to be identified =7. In all 7 identified motifs, four were more conserved and functional motifs than other motifs. These four conserved regions showed low e-value, and the number of amino acid residues were also different than non-functional motif (Supplementary Figure 3b).

Legume lectins are predominantly  $\beta$ -sheet proteins for containing large proportion of  $\beta$ -pleated sheets and negligible amount of  $\alpha$ -helix (Swamy et al. 1985). PHD neural network server revealed 39.64%  $\beta$ -sheets, 51.43% random coil and 8.93%  $\alpha$ -helix in secondary structure of RbL protein. Since rice bean has dominance of the  $\beta$ -stands in the secondary structure, therefore belongs to the class of  $\beta$ -proteins.

### Multiple sequence analysis of RbL protein

To have better insight of conservedness in deduced RbL sequence, deduced RbL sequence was aligned with other legume lectin sequences using ClustalW algorithm and curated with ESPript 2.2 server (Figure-3). Highest sequence similarity of RbL protein was observed with, *Vigna angularis*, *Vigna aconitifolia* and *Vigna unguiculata* lectins. Genetically, rice bean also shares similarity in evolutionary pattern and in pod and seed characteristics with adzuki bean. Moreover, similarity in electrophoretic pattern also reveals a close relationship between rice bean and adzuki bean (Chandel et al. 1988). NetNGlyc 1.0 server predicted 4 putative glycosylation sites in deduced RbL sequence positioned at N<sup>5</sup>FS, N<sup>25</sup>ST, N<sup>106</sup>KS, N<sup>137</sup>AT (denoted by circle in Figure-3). The N-linked glycosylation is important for secreted proteins to maintain

correct folding state of soluble and membrane bound proteins (Haweker et al. 2010). The presence of four glycosylation sites revealed high stability of RbL protein as they probably increase the inter-subunit interactions in protein (Mitra et al. 2006).

Multiple sequence alignment of RbL protein and other legume lectins revealed conservedness in amino acid sequences of carbohydrate binding loops except loop D. Variability in loop D is responsible for different carbohydrate specificities of legume lectins (Cummings et al. 2017). Another important observation from sequence alignment was the conservation in amino acid residues which are responsible for metal ion binding (denoted by \$) and monosaccharide binding (denoted by #) in legume lectins. From the sequence alignment, the *N*-terminal (after signal peptide sequence) and C-terminal sequences in deduced RbL sequence were also identified.

Prosite analysis (<https://prosite.expasy.org/>) (Gasteiger et al. 2005) revealed two legume lectins signature motifs in deduced RbL sequence which were  $\beta$ -chain signature [VAVEFDT (147-153)] corresponds to N-terminal) (Aspartate in  $\beta$ -chain is conserved for metal binding) and  $\alpha$ -chain signature [LPQWVSGFS (227-236) corresponds to C-terminal]. The presence of these signatures revealed that RbL protein is probably a two chain lectin processed post-translationally during maturation in plant vacuole and interactions between two chains lead to formation of dimer which has also been observed for DBL, PHA-L, PHA-E, SBA lectins. The specific signature motifs in legume lectins determine and differentiate their dimerization state (Van Damme et al. 2008).

Multiple sequence alignment also revealed the presence of conserved leucine amino acid as a cleavage signal for the removal C-terminal peptide. The C-terminal sequence in RbL amino acid sequence was identified by following three rules (Moreira et al. 2013) which are (i) first amino acid of excised peptide is small or hydrophobic; (ii) cleavage occurs after an acid, polar or hydrophobic amino acid residue, but not after a basic amino acid and; (iii) cleavage spot is located 5 to 8 residues after a conserved leucine amino acid residue. By analyzing the sequence alignment of legume lectins, the C-terminal cleavage site in deduced RbL sequence was observed between aspartic acid and leucine residues (Asp<sup>270</sup> and Leu<sup>271</sup>) and further presence of the two leucine amino acid confers the stability of C-terminal  $\alpha$ -helix. The presence of C-terminal  $\alpha$ -helix has also been observed in the structure of DBL, SBA and BVL-II (Moreira et al. 2013). Further, hydrophobic nature of C-terminal peptide in deduced RbL sequence revealed its targeting to the vacuole (Chrispeels and Raikhel 1992). The presence of  $\alpha$ -helix at c-terminal confers stability to dimeric and tetrameric structure of legume lectins by getting sandwiched between  $\beta$ -sheets of two lectin monomers.

### **Homology modeling, prediction of ligand binding sites and molecular docking analysis of RbL protein**

Homology modeling provide an insight of structure function relationship of target protein through building a 3D structure using a template protein whose 3D structure is experimentally predicted. It produces high quality structural models when the similarity between the target and template protein is high, but sequences having similarity lesser than 20% identity produces unrelated structures. A 3D model

of RbL protein was generated using the I-TASSER modeling server. The top hit template was PHA-E lectin (*Phaseolus vulgaris* lectin) (PDBID: 5AVA) (80.71% sequence similarity with a coverage value of 0.93). The server modeled RBL protein including 280 amino acid residues using iterative threading assembly refinement as in Figure 4a. I-TASSER modeling begins from the templates classified by LOMETS from the Protein Data Bank. LOMETS, a threading approach, includes multiple threading applications. The estimated TM-score and RMSD for the best model is  $0.66 \pm 0.13$  and  $6.90 \pm 4.10\text{\AA}$ , with a -0.41 C-score. C-score is usually in the range of (-5, 2), where a more significant value means a model with tremendous confidence. In model validation, the chart color-coding black, dark grey, grey, light grey denote highly favored conformations  $\Delta \geq -2$ . White with black grid signifies favored conformations  $-2 > \Delta \geq -4$ . The White with grey grid renders unreliable conformations  $\Delta < -4$ . Highly favored measurements bestowed 209 residues (84.615%) as green crosses. Favored measurements are conferred as brown triangles 24 residues (9.717%). The non-favorable measurements were displayed as red circles 14 residues (5.668%) as shown in Figure 4b.

Similar to other legume lectins, the dome like structure also known as lectin fold, jelly roll fold and  $\beta$ -sandwich structure was also identified in RbL protein. The structure also showed six flat stranded anti-parallel  $\beta$ -sheets forming back face and seven-stranded curved anti-parallel  $\beta$ -sheets forming front face of structure. RbL consisted of 16- $\beta$  strands connected with turn and loops (50% of the total amino acid residues). Since, 3-D model of RbL protein showed concurrency with conserved legume lectins structure ( $\beta$  sandwich model), the RbL protein must have two hydrophobic cores, (i) between the front and back sheets and (ii) between the front sheet and three loops, forming metal binding site and carbohydrate recognition domain.

Three-dimensional structure of RbL protein visualized with UCSF Chimera software (Pettersen et al. 2004) to identify the presence of characteristic carbohydrate binding loops in RbL structure. ProSA-web (Protein Structure Analysis web) server predicts Z score which represents the score of input structure to the score of experimentally determined protein deposited in Protein Data Bank (PDB) (Berman et al. 2006). The positive Z-score value indicates instability of structure while zero and negative scores indicates ideal structure (Filho et al. 2017). The server computed z score of -7.10 for RbL structure (Large black dot in supplementary figure 4) which was within the acceptable range (-10 to 10) (Yadav et al. 2012). Moreover, the z score for RbL structure was approximately equal to the z score of *Vigna unguiculata* lectin structure (-7.64) (Filho et al. 2017) and Z score of template protein (-6.52) which suggests that the predicted RbL model is reliable and share close structural similarity with template protein.

RbL structure Analysis revealed the presence of four loops namely A, B, C and D connected by turns and loops which are responsible for carbohydrate binding specificity of legume lectins (Figure 5a). The conserved triad [(Asp(D)-Asn(N)-Gly(G)/Arg(R)] responsible for monosaccharides binding through hydrogen bonding was also identified. Further, aromatic amino acid (Tyrosine), which is required for stacking with the non-polar face of sugar, was also identified in Loop C (Cummings et al. 2017). The folding in RbL protein, constituted typically from  $\beta$ -sheets interlinked by turns and loops, appears to create a very rigid and strong structural scaffold. In legume lectins, changes in orientation of  $\beta$ -sheets

confer rigidity to structure and provide protection from proteolytic degradation (Van Damme et al. 1998). Since metal ions are required for the stabilization of carbohydrate binding site, Ca<sup>+2</sup> ion binding site formed by Glu<sup>150</sup>, Asp<sup>152</sup>, Asp<sup>160</sup>, His<sup>165</sup> amino acid side chains in RbL structure was also predicted (Figure 5b). The presence of cis-peptide bond between Ala<sup>111</sup> and Asp<sup>112</sup>, which is essential for stability of carbohydrate bonding site in legume lectins was also identified in deduced RbL sequence (Sharon and Lis 2002).

Proteins perform their biological functions by interacting with ligands. Identification of ligand binding sites is of paramount importance to characterize protein on the basis of their biological function (Heo et al. 2014). Since *in vivo* and *in vitro* elucidation of ligand binding sites in a protein 3D-structure a laborious process, time consuming and costly process, *in silico* prediction using bio-computational tools is the suitable way of predicting ligand binding sites (Roche et al. 2010). GalaxySite web server predicted  $\beta$ -galactose, *N*-acetyl glucosamine and lactose and adenine as putative binding ligands for RbL protein (Table-1). We used COACH server and CDOCKER engine to predict binding site and docking analysis of RbL protein with these ligands including two additionally selected ligands glucose and mannose. The structures of *N*-acetyl-D-glucosamine, lactose,  $\beta$ -galactose, mannose, glucose, and adenine with PubChem ID 439174, 6134, 439353, 18950, 5793 and 190 were retrieved from Pubchem in SDF format. We recorded ligands interaction energy based on -CDOCKER interaction energy from the docking report, as depicted in Table 2.

On interpreting, the interactions between RbL and selected ligands ( $\beta$ -galactose, *N*-acetyl-D-glucosamine, glucose, mannose, lactose and adenine). It is observed that the residues Asp<sup>112</sup>, Ala<sup>129</sup>, Asn<sup>156</sup>, and Leu<sup>241</sup> are involved in hydrogen bond interaction with the hydroxal group of the selected ligands except adenine (Figure 6). All the poses were adopted as springing subjects for the MD computations. The molecular dynamic simulations were analyzed based on RMSD, RMSF, and H-bond. RMSD values of protein-ligand complexes determined the structural changes from 0 to 10 ns. The overall RMSD values were between 0.20 to 4.50 nm and attained a steady state after the 4ns of simulation (Figure 7a). RMSF renders each atom's fluctuation throughout the simulation. RMSF was accounted for RbL protein-ligand complexes; the values fortified that the binding site residues bestowed small fluctuation. The average RMSF values were 0.10 to 0.26 nm for the binding site residues, as depicted in Figure 7b.

The intermolecular H-bonding within the protein and ligand performs a crucial task in protein-ligand complexes stabilization. The hydrogen bond network stability built between RbL protein and selected ligands was measured during the 10ns simulation run. The cumulative estimate of H-bonds in the complexes versus time is shown in Figure 8. The number of H-bonds with an average of 4 to 6 was established in RbL complexes throughout the simulation with  $\beta$ -galactose, *N*-acetyl-D-glucosamine, glucose, mannose, lactose and adenine. This indicates that the selected ligands have enduring and robust hydrogen bonding with RbL protein.

ProteinPredict web server revealed that RbL protein had carbohydrate binding as one of the important molecular functions (Table 3). Further, analysis also revealed that RbL protein has function in cellular

processes. This probably denotes the function of RbL protein as storage protein for providing essential amino acids during seed germination. Argot<sup>2.5</sup> web server also annotated RbL protein by identifying various GO terms among which the most significant terms were GO:0030246 (Carbohydrate binding); GO:0046872 (Metal ion binding); GO:0006952 (Defense response); GO:0043086 (Negative regulation of catalytic activity) GO:0050832 and GO:0042742 (Defense response to fungus and bacterium, respectively and GO:0031640 (Killing of cells of other organisms) (Supplementary Figure 5a and 5b).

Computational analysis of RbL structure also revealed the presence of adenine binding site which suggests its function in storage of adenine derived plant hormones in seeds and during stress as extracellular ATP (Adenine is one of the structural component of ATP) has been recognized as central signaling molecule in plant defense response (Delatorre et al. 2007; Cao et al. 2014; Choi et al. 2014). The interaction of RbL protein with ATP could be related to the presence of conserved receptor kinase domain in RbL protein. Moreover, RbL has specificity to bind with *N*-acetyl-D-glucosamine, a monomer of chitin, it is logical that RbL protein would have entomotoxic properties. Taken together, these results support the role of RbL protein in plant defense.

## Conclusions

For the first time, we carried out *in silico* analysis of novel lectin precursor from rice bean culminating into valuable information on structural and functional properties of RbL protein. ProtParam analysis revealed hydrophobic, acidic and stable nature of RbL protein with two putative signal peptide cleavage sites and a C-terminal domain. RbL protein was identified as  $\beta$ -sheet protein due to the dominance of  $\beta$ -sheets and the presence of jelly roll fold ( $\beta$ -sandwich structure) in 3D structure. Comparative sequence analysis showed high sequence similarity in the primary structure of RbL protein and other legume lectins. Molecular docking analysis unveiled the potential of *N*-acetyl-D-glucosamine  $\beta$ -galactose, glucose, mannose, lactose and adenine ligands to associate with the binding pocket of the RbL protein. Furthermore, molecular docking and dynamics simulation outcomes confirmed hydrogen bond interactions during the docking investigation for decent binding between ligands and its feasible binding site in RbL protein. Gene ontology analysis revealed that RbL protein has significant role in plant defense. Moreover, the presence of adenine binding site also supports the role of RbL protein in plant defense mechanisms. The current novel information on RbL protein would be valuable for conducting research on future implication of RbL protein in biomedical and biotechnological interventions such as glycan profiling, disease research and plant protection.

## Declarations

**Ethical Approval**—Not applicable

**Consent to Participate**- granted

**Consent to Publish**- Consent is granted for publishing paper in AB&B

## Authors Contributions:

ANKUR TRIPATHI— Generation plants under controlled conditions, experimentation, data analysis

VIPIN HALLAN- Laboratory aid, instrumentation, sequencing

RITU RAJ- Computational analysis

RAJAN KATOCH—Conceptualization, interpretation, experimental modeling, manuscript casting

**Funding-** Department of Biotechnology, GOI, New Delhi

**Competing Interests-** Not applicable

**Availability of data and materials-** Resources from our laboratory

## References

Aboitiz N, Vila-Perello M, Groves P et al (2004) NMR and modeling studies of protein-carbohydrate interactions: synthesis, three-dimensional structure, and recognition properties of a minimum hevein domain with binding affinity for chitooligosaccharides. *ChemBioChem* 5: 1245-1255.

<https://doi.10.1002/cbic.200400025>.

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2: 19–25.

<https://doi.org/10.1016/j.softx.2015.06.001>

Anderson RJ, Weng Z, Campbell RK, Jiang X (2005) Main-chain conformational tendencies of amino acids. *Proteins: Struct Funct Genet* 60: 679–689. <https://doi.org/10.1002/prot.20530>.

Bailey TL, Boden M, Buske FA et al (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37: W202-W208. <https://doi.org/10.1093/nar/gkp335>.

Bendtsen JD, Nielsen H, Von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795. <https://doi.org/10.1016/j.jmb.2004.05.028>.

Berman HM, Burley SK, Chiu W et al (2006) Outcome of a workshop on archiving structural models of biological macromolecules. *Structure* 14: 1211–1217. <https://doi.org/10.1016/j.str.2006.06.005>.

Brinda KV, Surolia A, Vishveshwara S (2005) Insights into the quaternary association of proteins through structure graphs: a case study of lectins. *Biochem J* 391: 1-15. <https://doi.10.1042/BJ20050434>.

Brooks BR, Bruccoleri RE, Olafson BD et al (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4: 187–217.

<https://doi.org/10.1002/jcc.540040211>.

- Cao Y, Tanaka K, Nguyen CT, Stacey G (2014) Extracellular ATP is a central signaling molecule in plant stress responses. *Curr Opin Plant Biol* 20: 82–87. <https://doi.org/10.1016/j.pbi.2014.04.009>
- Cavada BS, Osterne VJS, Oliveira MV et al (2020) Reviewing Mimosoideae lectins: A group of under explored legume lectins. *Int J Biol Macromol* 154: 159-165. <https://doi.10.1016/j.ijbiomac.2020.03.113>.
- Chandel KPS, Arora RK, Pant KC (1988) Rice bean- A potential grain legume. NBPGR Scientific Monograph No. 12. NBPGR, New Delhi, India.
- Choi J, Tanaka K, Cao Y, Qi Y, Qiu J, Liang Y (2014) Identification of a plant receptor for extracellular ATP. *Science* 343: 290–294. <https://doi.10.1126/science.343.6168.290>.
- Chrispeels MJ, Raikhel NV (1992) Short peptide domains target proteins to plant vacuoles. *Cell* 68: 613–616. [https://doi.org/10.1016/0092-8674\(92\)90134-X](https://doi.org/10.1016/0092-8674(92)90134-X)
- Coelho LCBB, Dos Santos Silva PM, De Menezes Lima VL et al (2017) Lectins, interconnecting proteins with biotechnological/pharmacological and therapeutic applications. *Evid Based Complement Alternat Med* 1594074. <https://doi.10.1155/2017/1594074>.
- Cummings RD, Etzler ME, Surolia A (2017) L-Type Lectins. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW and Etzler ME (eds), *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, New York
- De Hoff PL, Brill LM, Hirsch AM (2009) Plant lectins: The ties that bind in root symbiosis and plant defense. *Mol Genet Genom* 282: 1-15. <https://doi.10.1007/s00438-009-0460-8>
- Delatorre P, Rocha BA, Souza EP et al (2007) Structure of a lectin from *Canavalia gladiata* seeds: New structural insights for old molecules. *BMC Struct Biol* 7: 1-9. <https://doi.org/10.1186/1472-6807-7-52>.
- Filho RMM, Rossiter JG, Cavalcanti Junior EA, Martins LSS (2017) *In Silico* comparative analysis of legume lectins. *J Genet Genomes* 1: 1-11.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) *The proteomics protocols handbook*. Humana Press USA, pp, 571-607. <https://doi.org/10.1385/1-59259-890-0:571>
- Gautam AK, Srivastava N, Sharma B, Bhagyawant S (2018) Current scenario of legume lectins and their practical applications. *J Crop Sci Biotechnol* 21: 217-227. <https://doi.org/10.1007/s12892-018-0002-0>
- Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31: 3320-3323. <https://doi.org/10.1093/nar/gkg556>

- Gupta R, Brunak S (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* 310-322. [https://doi.org/10.1142/9789812799623\\_0029](https://doi.org/10.1142/9789812799623_0029)
- Hamelryck TW, Loris R, Bouckaert J, Wyns L (1998) Structural features of the legume lectins. *Trends Glycosci Glycotechnol* 10: 349–360. <https://doi.org/10.4052/tigg.10.349>
- Haweker H, Rips S, Koiwa H et al (2010) Pattern recognition receptors require N-glycosylation to mediate plant immunity. *J Biol Chem* 285: 4629–4636. <https://doi.org/10.1074/jbc.M109.063073>.
- Heo L, Shin WH, Lee MS, Seok C (2014) GalaxySite: Ligand-binding-site prediction by using molecular docking. *Nucleic Acids Res* 42: W210-W214. <https://doi.org/10.1093/nar/gku321>.
- Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A Linear Constraint Solver for molecular simulations. *J Comput Chem* 18: 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H)
- Jin H, Hong Z, Su W, Li J (2009) A plant-specific calreticulin is a key retention factor for a defective brassinosteroid receptor in the endoplasmic reticulum. *Proc Nat Acad Sci USA* 106: 13612–13617. <https://doi.org/10.1073/pnas.0906144106>
- Käll L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 35:W429-W432. <https://doi.org/10.1093/nar/gkm256>
- Kortt AA (1985) Characterization of the acidic lectins from winged bean seed (*Psophocarpus tetragonolobus* (L.)DC). *Arch Biochem Biophys* 236: 544-554.
- Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *Mol Biol* 1982; 157:105-132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0).
- Lagarda-Dias I, Guzman-Partida AM, Vazquez-Moreno L (2017) Legume lectins: Proteins with diverse applications. *Int J Mol Sci* 18. <https://doi.10.3390/ijms18061242>.
- Laija SN, Mahesh S, Smitha LS, Remani P (2010) Isolation and partial characterization of two plant lectins. *Curr Res J Biol Sci* 2: 232-237.
- Lannoo N, Van Damme EJM (2010) Nucleocytoplasmic plant lectins. *Biochim Biophys Acta* 1800: 190-201. <https://doi.10.1016/j.bbagen.2009.07.021>.
- Lannoo N, Vandenborre G, Miersch O, Smagghe G, Wasternack C, Peumans WJ (2007) The jasmonate-induced expression of the *Nicotiana tabacum* leaf lectin. *Plant Cell Physiol* 48: 1207-1218. <https://doi.org/10.1093/pcp/pcm090>.

- Li Y, Li T, Li J, Liu D, Yang J, Yang J, Zhang F, Sun S (2014) Molecular cloning, expression, and characterization of a *Sophora alopecuroides* lectin from *Escherichia coli*. *Acta Biochim Biophys Sin* 46: 749–756. <https://doi.10.1093/abbs/gmu060>.
- Loris R, Hamelryck T, Bouckaert J, Wyns L (1998) Legume lectin structure. *Biochim Biophys Acta* 1383: 9-36. [https://doi.10.1016/s0167-4838\(97\)00182-9](https://doi.10.1016/s0167-4838(97)00182-9).
- Magdeldin S, Yoshida Y, Li, H et al (2012) Murine colon proteome and characterization of the protein pathways. *BioData mining* 5: 11. <https://doi.10.1186/1756-0381-5-11>.
- Mitra N, Sinha S, Ramya TNC, Surolia A (2006) N-Linked oligosaccharides as outfitters for glycoprotein folding form and function. *Trends Biochem Sci* 31: 156-163. <https://doi.10.1016/j.tibs.2006.01.003>.
- 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. <https://doi.org/10.1371/journal.pone.0081338>
- Nandi S, Mehra N, Lynn AM, Bhattacharya A (2005) Comparison of theoretical proteomes: identification of COGs with conserved and variable pI within the multimodal pI distribution. *BMC Genom* 6: 1-12. <https://doi.org/10.1186/1471-2164-6-116>
- Osman M, Konozy E (2017) Insight into *Erythrina* Lectins: Properties, Structure and Proposed Physiological Significance. *The Open Bioact Compd J* 5: 57-71. <https://doi.10.2174/1874847301705010057>
- Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera- A visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612. <https://doi.org/10.1002/jcc.20084>
- Peumans WJ, Van Damme EJM (1995) Lectins as plant defense proteins. *Plant Physiol* 109: 347–352. <https://doi.10.1104/pp.109.2.347>
- Pinto LS, Nagano CS, Oliveira TM, Moura TR, Sampaio AH, Debray H, Pinto VP, Dellagostin OA, Cavada BS (2008) Purification and molecular cloning of a new galactose-specific lectin from *Bauhinia variegata* seeds. *J Biosci* 33: 355–363. <https://doi.10.1007/s12038-008-0055-2>.
- Rini JM (1995) Lectin structure. *Ann Rev Biophys Biomol Struct* 24: 551–577. <https://doi.org/10.1146/annurev.bb.24.060195.003003>
- Roberts DD, Goldstein IJ (1984) Isolation from lima bean lectin of a peptide containing a cysteine residue essential for carbohydrate binding activity. *J Biol Chem* 259: 909-914. [https://doi.org/10.1016/S0021-9258\(17\)43543-5](https://doi.org/10.1016/S0021-9258(17)43543-5)
- Roche DB, Tetchner SJ, McGuffin LJ (2010) The binding site distance test score: A robust method for the assessment of predicted protein binding sites. *Bioinformatics* 26: 2920-2921.

<https://doi.org/10.1093/bioinformatics/btq543>

Rost B (1996) PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol* 266: 525-539. [https://doi.org/10.1016/S0076-6879\(96\)66033-9](https://doi.org/10.1016/S0076-6879(96)66033-9)

Rost B, Yachdav G, Liu J (2004) The PredictProtein server. *Nucleic Acids Res* 1:W321-W326. <https://doi.org/10.1093/nar/gkh377>.

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725-738. <https://doi.org/10.1038/nprot.2010.5>

Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M (2015) PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Res* 43(W1): W443–W447. <https://doi.org/10.1093/nar/gkv315>.

Schüttelkopf AW, Van Aalten DMF (2004) PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60: 1355–1363. <https://doi.org/10.1107/S0907444904011679>.

Sharma V, Surolia A (1997) Analyses of carbohydrate recognition by legume lectins: Size of the combining site loops and their primary specificity. *J Mol Biol* 267: 433–445. <https://doi.10.1006/jmbi.1996.0863>.

Sharon N, Lis H (2002) How proteins bind carbohydrates: lessons from legume lectins. *J. Agric. Food Chem* 50: 6586–6591. <https://doi.10.1021/jf020190s>.

Stillmark H (1888) Über Ricin ein giftiges Ferment aus den Samen von *Ricinus communis* L. und einige anderen Euphorbiaceen. MD Thesis, University of Dorpat, Dorpat, Estonia

Swamy MJ, Sastry MVK, Surolia A (1985) Prediction and comparison of the secondary structure of legume lectins. *J Biosci* 9: 203–212.

Tsaneva M, Van Damme EJM (2020) 130 years of plant lectin research. *Glycoconj J* 37: 533–551. <https://doi.org/10.1007/s10719-020-09942-y>

Ujainwal M, Sahani PA, Singh N (2019) Comparative sequence and structural analysis of lectin protein in Chickpea (*Cicer arietinum* L.) and their relationship with fabaceae family. *Int J Proteomics Bioinformatics*; 4: 001-006.

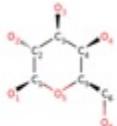
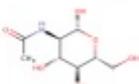
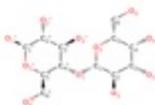
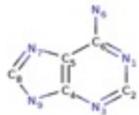
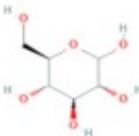
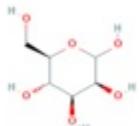
Unger R, Sussman JL (1993) The importance of short structural motifs in protein structure analysis. *J Comput Aided Mol Des* 7: 457-472. <https://doi.org/10.1007/BF02337561>

Van Damme EJM, Lannoo N, Peumans WJ (2008) Plant lectins. *Adv Bot Res* 2008; 48: 107-209. <http://hdl.handle.net/1854/LU-666112>

- Van Damme EJM, Peumans WJ, Barre A, Rouge P (1998) Plant Lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Cric Rev Plant Sci* 17: 575-692. <https://doi.org/10.1080/07352689891304276>
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26: 1701–1718. <https://doi.org/10.1002/jcc.20291>
- Vandenborre G, Smagghe G, Van Damme EJM (2011) Plant lectins as defense proteins against phytophagous insects. *Phytochemistry* 72: 1538-1550. <https://doi.10.1016/j.phytochem.2011.02.024>.
- Varshavsky A (2011) The N end rule pathway and regulation by proteolysis. *Protein Sci* 20:1298–1345. <https://doi.10.1002/pro.666>.
- 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. <https://doi.10.1093/nar/gkm290>.
- Wu G, Robertson DH, Brooks CL, Vieth M (2003) Detailed analysis of grid-based molecular docking: A case study of CDOCKER - A CHARMM-based MD docking algorithm. *J Comput Chem* 24: 1549–1562. <https://doi.10.1002/jcc.10306>.
- Yadav BS, Tripathi V, Kumar A, Khan MF, Barate A, Kumar A, Sharma B (2012) Molecular modeling and docking characterization of Dectin-1 (PAMP) receptor of *Bubalus bubalis*. *Exp Mol Pathol* 92: 7–12. <https://doi.10.1016/j.yexmp.2011.09.018>
- Yang J, Roy A, Zhang Y (2013a) BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res* 41(D1): <https://doi.org/10.1093/nar/gks966>
- Yang J, Roy A, Zhang Y (2013b) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* 29: 2588–2595. <https://doi.org/10.1093/bioinformatics/btt447>.
- Young NM, Oomen RP (1992) Analysis of sequence variation among legume lectins: A ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. *J Mol Biol* 228: 924–934. [https://doi.10.1016/0022-2836\(92\)90875-k](https://doi.10.1016/0022-2836(92)90875-k).

## Tables

**Table 1: Details of binding ligands for RbL protein**

Ligand name	Ligand structure	Templates for protein-ligand complex
$\beta$ -Galactose		1SBF_A, 1G9F_A, 1DBN_B, 1V6I_D
N-Acetyl-D-glucosamine		1AXZ_A, 2E7Q_A, 1HQL_A
Lactose		1GZC_A
Adenine		3UJO_B
Glucose*		-
Mannose*		-

(PDB Template ID: **1SBF\_A**: Soybean agglutinin (SBA); **1G9F\_A**: Soybean agglutinin in a complex with A biantennary blood group antigen analog; **1DBN\_B**: *Maackia amurensis* leucoagglutinin (lectin) with sialyllactose; **1V6I\_D**: Peanut lectin-lactose complex in acidic pH; **1AXZ\_A**: *Erythrina corallodendron* lectin in complex with D-galactose; **2E7Q\_A**: Crystal structure of basic winged bean lectin in complex with b blood group trisaccharide; **1HQL\_A**: The xenograft antigen in complex with the B4 isolectin of *Griffonia simplicifolia* lectin-1; **1GZC\_A**: High-Resolution crystal structure of *Erythrina cristagalli* lectin in complex with lactose; **3UJO\_B**: Galactose-specific seed lectin from *Dolichos lablab* in complex with adenine and galactose; \*: Manually selected ligand for docking analysis

**Table 2: CDOCKER interaction energy of selected ligands**

Ligands	-CDOCKER interaction energy
N-Acetyl-D-Glucosamine	25.45
Lactose	24.52
$\beta$ -galactose	24.39
Mannose	23.84
Glucose	23.49
Adenine	15.69

**Table 3: Molecular function prediction based on the gene ontology**

---

**Molecular function ontology**

<b>GO ID</b>	<b>GO Term</b>
<u>GO:0030246</u>	Carbohydrate binding
<u>GO:0048029</u>	Monosaccharide binding
<u>GO:0043167</u>	Ion binding
<u>GO:0046872</u>	Metal ion binding

**Biological process ontology**

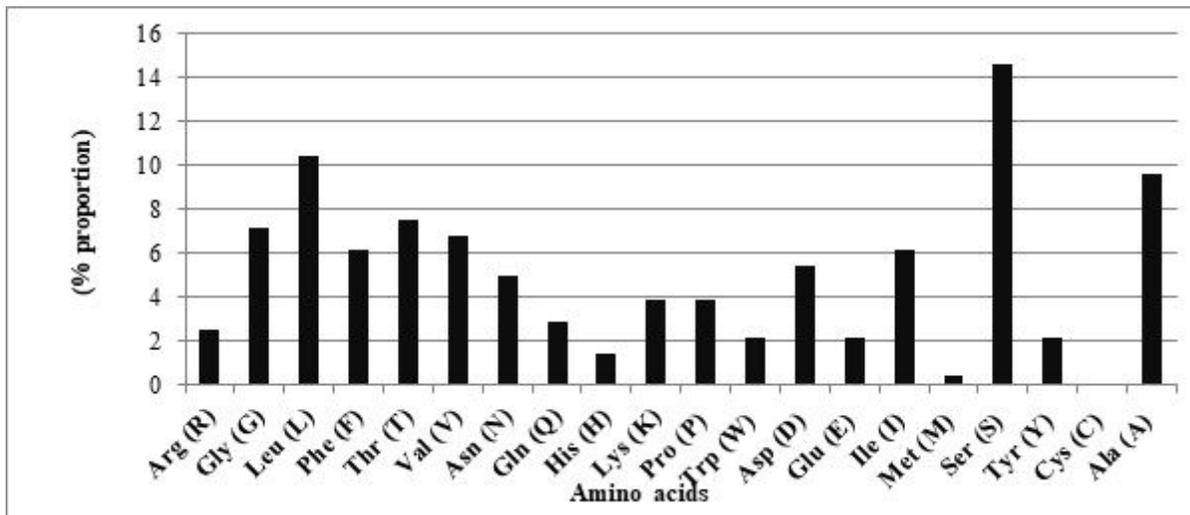
<u>GO:0009987</u>	Cellular process
<u>GO:0006952</u>	Defense response
<u>GO:0006950</u>	Response to stress
<u>GO:0050896</u>	Response to stimulus

**Cellular component ontology**

<u>GO:0005886</u>	Plasma membrane
<u>GO:0071944</u>	Extracellular matrix
<u>GO:0031224</u>	Integral component of membrane
<u>GO:0043229</u>	Intracellular organelle

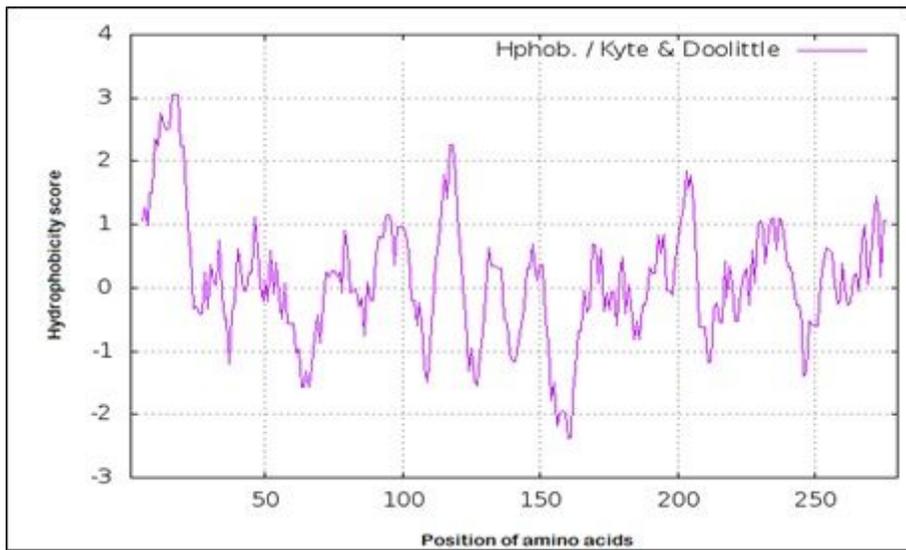
---

## Figures



**Figure 1**

Amino acid composition of RbL protein



**Figure 2**

Hydrophobicity analysis of RbL protein

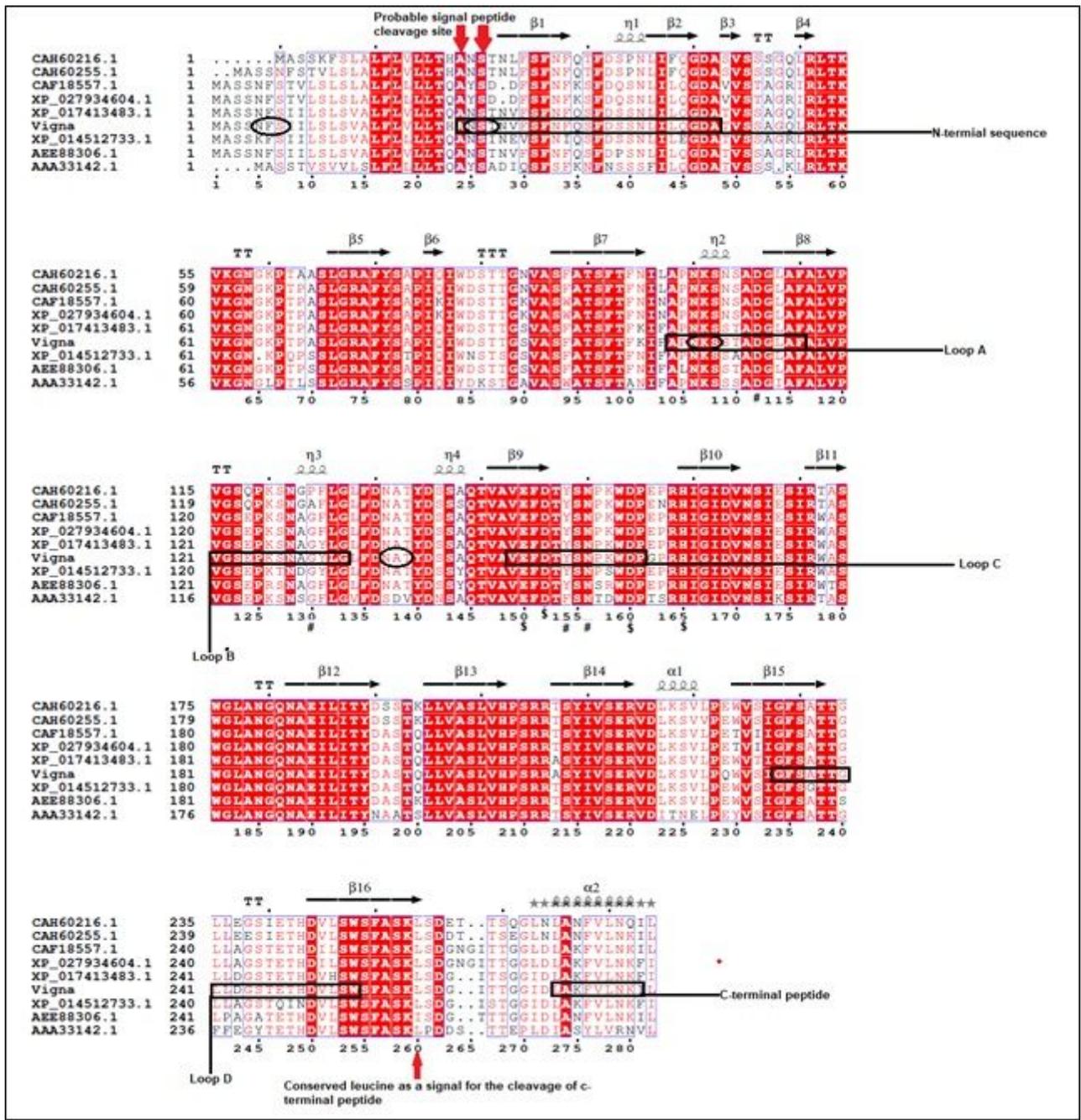
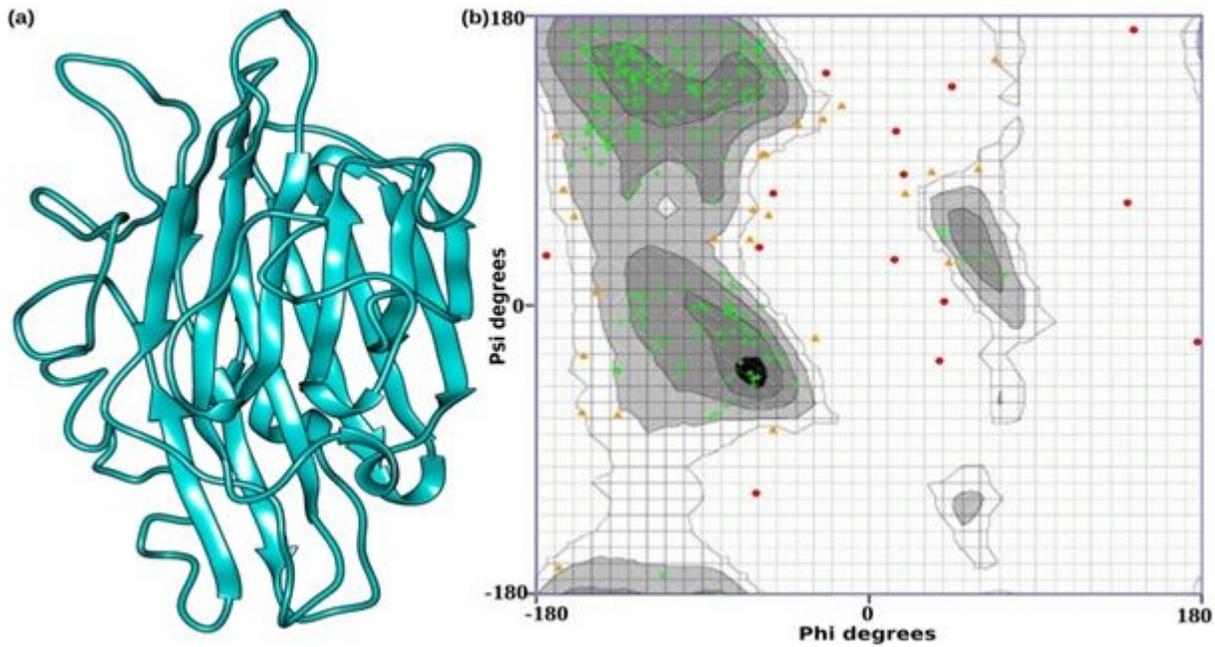


Figure 3

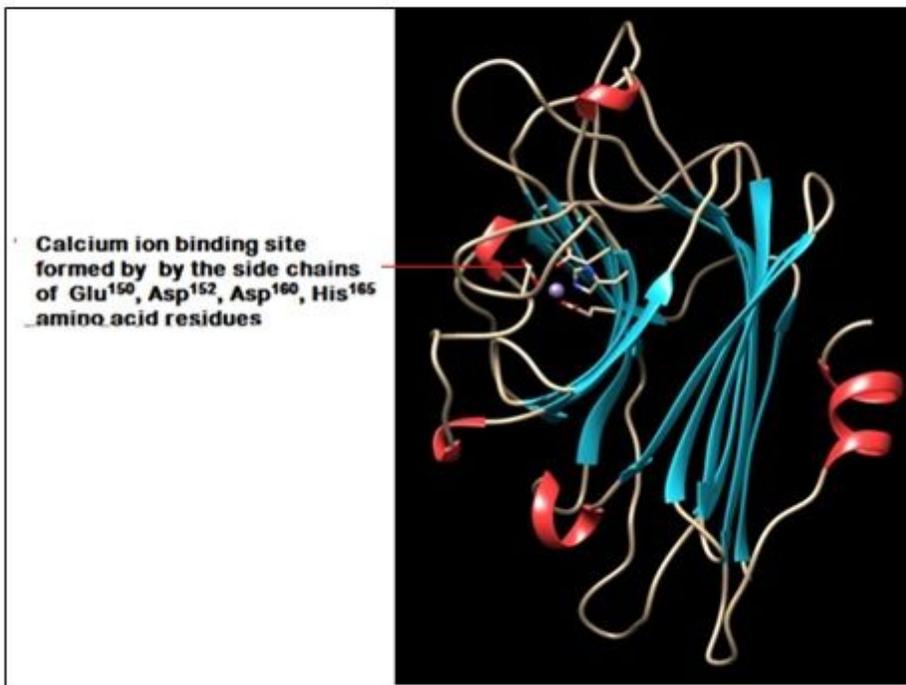
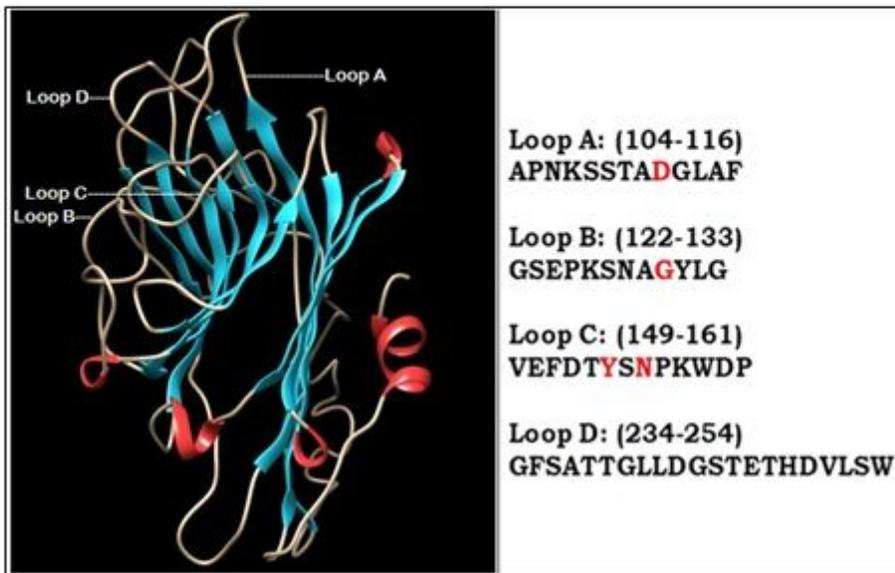
Multiple sequence alignment of deduced RbL sequence with other legume lectin sequences (XP\_017413483.1)- Lectin DB58-like [*Vigna angularis*]; (AEE88306.1) lectin [*Vigna aconitifolia*]; (XP\_014512733.1) lectin DB58-like [*Vigna radiata* var. *radiata*]; (XP\_027934604.1) lectin DB58-like [*Vigna unguiculata*]; (CAF18557.1) lectin precursor [*Vigna unguiculata*] (CAH60255.1) lectin precursor; [*Phaseolus parvulus*]; (CAH60216.1) lectin precursor [*Phaseolus filiformis*]; (AAA33142.1) lectin [*Vigna unguiculata* subsp. *cylindrica*] (Boxes: Amino acid residues involved in the formation of different loops, N-terminal and C-terminal sequences; Conserved amino acid residues for metal ion binding and monosaccharide binding are denoted by \$ and #, respectively; Circle represents putative N-linked

glycosylation sites. Upright and down arrows represent putative cleavage sites for c-terminal peptide and signal peptide, respectively.



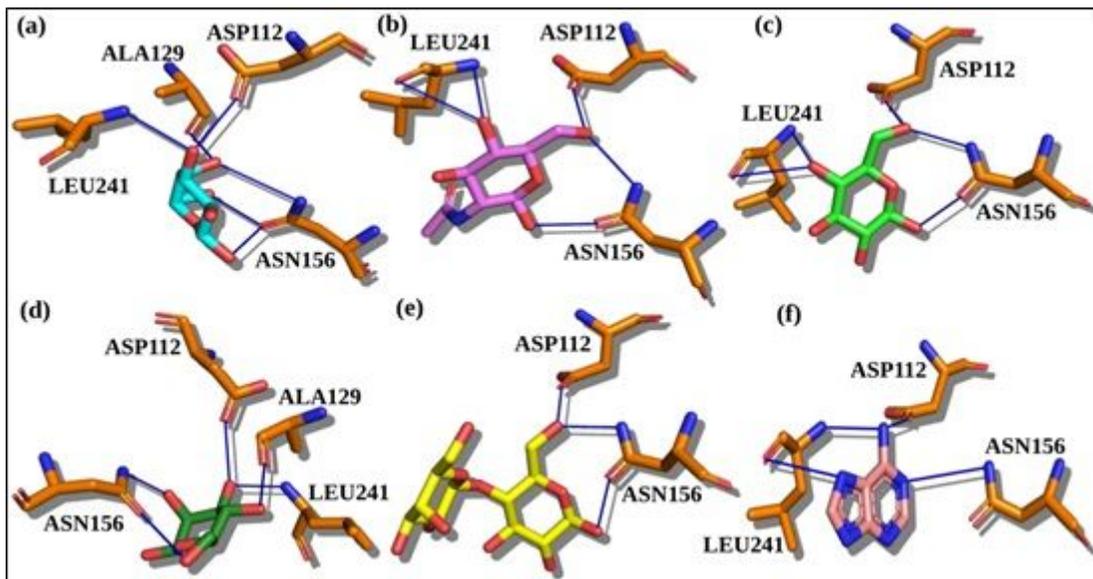
**Figure 4**

(a) The 3D structure of the protein is modeled from the I-TASSER server; (b) Ramachandran plot with the most favored (black, dark grey, grey, light grey) regions



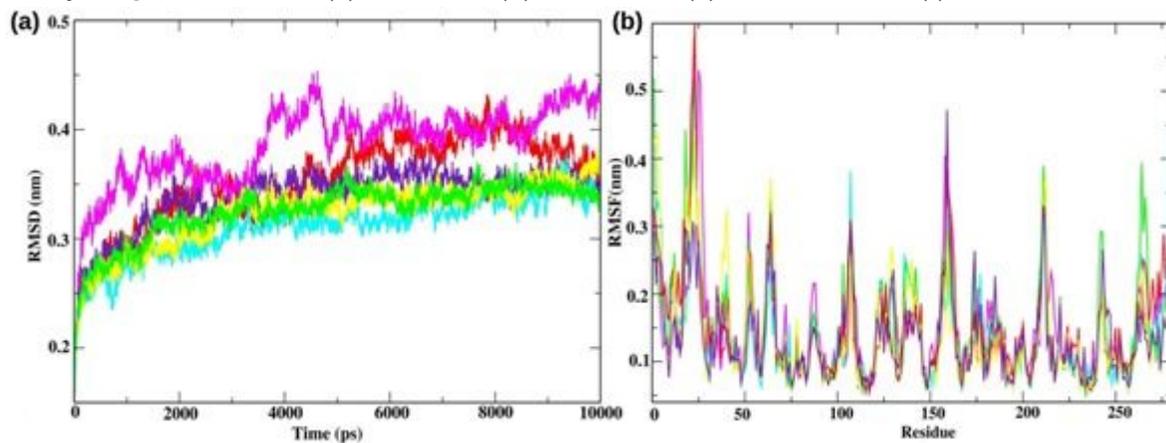
**Figure 5**

a: Carbohydrate binding loops in RbL protein b: Calcium binding site in RbL protein



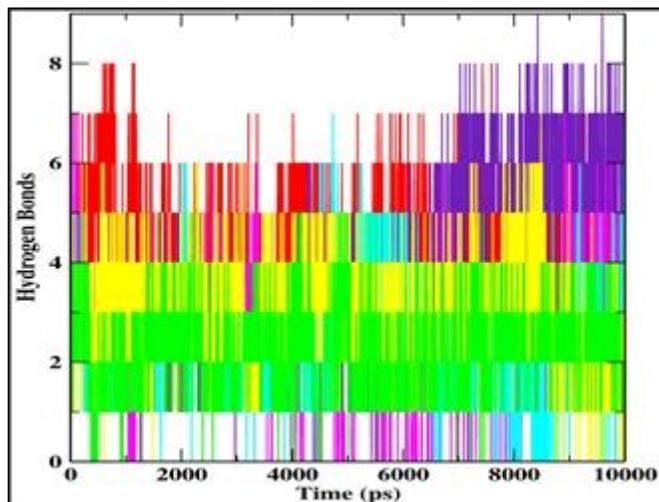
**Figure 6**

Graphical representation of the interactions between RbL and putative ligands (a)  $\beta$ -galactose (b) N-acetyl-D-glucosamine, (c) Glucose, (d) Mannose (e) Lactose, and (f) Adenine



**Figure 7**

Molecular dynamics simulations of RbL complexes (a) RMSD of Ca backbone over the 10 ns; (b) RMSF of Ca atoms over the 10 ns. The RbL complexes with  $\beta$ -galactose (red), N-acetyl-D-glucosamine (cyan), Glucose (green), Mannose (magenta), Lactose (indigo) and Adenine (yellow)



**Figure 8**

Graphical representation of number of hydrogen bond interactions formed during simulation in RbL complexes with  $\beta$ -galactose (red), N-acetyl-D-glucosamine (cyan), Glucose (green), Mannose (magenta), Lactose (indigo) and Adenine (yellow)

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [4.Supplementaryfigures.docx](#)