# Determination of Insecticidal potential of Lentil Lectin through Bio-computational Tool


## Abstract

Lectins from plants and animals are commonly used in direct defense against pathogens and in immune regulation. Lentil lectin (LL) is a metalloprotein, containing $Ca^{2+}$ and $Mn^{2+}$, which binds reversibly to polysaccharides and glycoconjugates containing glucose or mannose type sugars. The functional domain analysis of LL revealed metal binding and N-linked glycosylation site. The physico-chemical parameter like theoretical isoelectric point (pI) is 5.22, instability index is 27.51, aliphatic index is 78.33 and GRAVY value is -0.086. The secondary structure attributes of LL protein explain its structural behavior and biological activity. The present study using various bio-computational tools could, therefore, help in our understanding of LL protein structure and prove to potential candidate gene for generating transgenic for aphid resistance.



**Key words**: Lentil Lectins, Functional domain, Molecular docking, Molecular modelling


## Introduction:

Legume lectins are structurally and evolutionarily related to a well-defined group of lectins that were originally discovered in the seeds of legumes like jack bean, common bean, pea, peanut and soybean. Many of these lectins have been purified and characterized with respect to their structure, sugar-binding specificity and biological activity (Sharon &Lis, 1990; Van Damme*et al*., 1998; Peumans& Van Damme, 1999).Legume lectins exhibited fairly homogeneous molecular structure with ~30kDa subunits (Van Damme*et al*., 2007). These lectins shared extensive sequence homology and three dimensional structural similarities, but differed in carbohydrate specificity (Rouge *et al*., 1991). The monomer structure is characterized by presence of the "jelly roll" motif that is often associated with carbohydrate-binding activity. The "jelly roll" is characterized by the presence of three sets of antiparallel β-sheets. The sheets are connected by several loops of varying lengths (Loris *et al*. 1998; Vijayan and Chandra1999).Two metal ions, calcium and a transition metal, found in all the legume lectin structures are essential for the carbohydrate binding. ConA, pea, lentil and lathyrus lectins are dimers formed by side-by-side alignment of two monomers such that the two rear ß-sheets form a contiguous 12-stranded ß-sheet. The four different regions A, B, C and D, associated with the concave face of the seven-stranded curved ß-sheet at the top front-side of the subunit, form the binding site (Sharma &Surolia, 1997). The amino acid

UNDER PEER REVIEW

residues which bind $Ca^{2+}$ and $Mn^{2+}$ metal ions are highly conserved, while the residues which constitute the sugar-binding site are less conserved but exhibit similar properties (Lis& Sharon, 1998). The conserved Asp and Gly/Arg residues are present in loops A and B, whereas Asn and the hydrophobic residues (Phe/Tyr/ Trp/ Leu) are located in loop C. Size of the backbone of C loop determines carbohydrate specificity of the lectin. Aligned sequences of legume lectins showed 4 to 7 gaps in the binding loop D, indicating variation in the loop size (Sharma &Surolia, 1997) which contributes to broad specificity of legume lectins.

*Lens culinaris* lectin is a dimer composed of identical subunit with N- and C-terminal residues of threonine and alanine respectively. The present study was carried out to study the physico-chemical properties, secondary structure attributes and 3D model of LL protein. The insect receptor binding potential with LL protein was evaluated through docking study.

UNDER PEER REVIEW

**Material and Methods:**

**Sequence Retrieval and Analysis**

The amino acid sequence of lentil lectin (LL) (accession #KU382474.1) protein was retrieved from NCBI database. The functional domains of lectin were determined using the InterPro tool available on the EBI web page (www.ebi.ac.uk/interpro/). The physico-chemical properties like amino acid composition, pI, molecular weight, half-life and instability index were determined using Protparam (http://web.expasy.org/protparam/). Probability of proteindisorder was determined by the PrDOS (Protein disorder prediction server) tool (http://prdo s.hgc.jp).

**Structural analysis and homology-based modelling**

The secondary structure and solvent accessibility of LL was determined by the RaptorX protein structure server (http://raptorx.uchicago.edu/StructurePrediction/predict/). The 3D structure of the target protein LLwas generated using SWISS Model tool (https://swissmodel.expasy.org/) using Pea Lectin as a template through homology based modelling. The authenticity of the predicted models was further validated employing RAMPAGE tool (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php).

**Molecular docking and Active site mapping**

In order to accomplish the docking studies, *Acyrthosiphon pisum* membrane alanyl Amino peptidase-N (APN; Accession # DQ440823), was used as receptor for LL. ClusPro Docking server (http://cluspro.bu.edu/) was used and results were viewed through Discovery Studio 4.1 visualizer. The amino acid residues forming the cleft area involved in interacting with APN were also identified using ClusPro Docking server (http://cluspro.bu.edu/).

PDBSum tool (http://www.ebi.ac.uk/pdbs um) was used to further identify the amino acid residues binding with the ligand present on the lectin binding proteins which undergo glycosylation as a part of post translational modification for functional activation.

## RESULTS AND DISCUSSION

### Sequence analysis and characterization

The functional domain of LL protein sequence were defined using InterPro tool (Fig. 1). The residue annotation identifies metal binding sites of LL are $ASP^{151}$, $PHE^{153}$ and $ASP^{159}$. In case of legume lectins, the presence of metal ions i.e., $Mn^{2+}$ and $Ca^{2+}$ was documented to be very important (Sharon and Lis, 1990). This is signified by the evolutionarily conserved amino acid residues that bind to the metal ions. For example, Con A requires $Mn^{2+}$ and $Ca^{2+}$ for its activity (Hardman and Ainsworth, 1972). The residue annotation using InterPro tools also revealed that, $ASN^{135}$ of LL involved in N-linked glycosylation. The LL is characterized as acidic protein based on computed pI value 5.22 (pI<7). This protein play important molecular functions in mannose binding, carbohydrate binding and metal ion like $Ca^{+2}$ and $Mn^{+2}$ binding. The instability index of LL was 27.51, classifying it as a stable protein which is also justified with the result obtained from PrDOS tool. Two disordered regions were predicted in the protein sequence, of which the longest disordered region was found between $Gly^{264}$ to $Ala^{275}$ comprising 12 amino acid residues (Fig. 2). The aliphatic index of LL was 78.33. The Aliphatic index (Ai) of proteins determine its thermo stability under changing climatic conditions (Gupta *et al*. 2012). GRAVY indices for LL was -0.086, indicates the possibility of better interaction with water i.e. hydrophilic nature of the protein which is attributed to charged amino acid residues present in the protein sequence (25 negatively charged and 21 positively charged), suggesting that LL might be associated as extrinsically in plasma-membrane.

### Structural analysis and homology-based modelling

The secondary structure of LL generated with RaptorX predicted a total of 6% α helices, 41% β pleated sheets and 51% Coil (Fig. 2A). It also revealed solvent accessibility of this protein as 26% residues are buried into structure, 33% residues are exposed and 40% were medium (Fig. 2B). Most of legume lectin proteins contain about 40–50% b-sheet, 35–45 % β-turn and 0–10% α-helix, and thereby fall into a structurally distinct class of proteins. The β-sheet, α-helical and β-turn content predicted here agrees well with X-ray structure determination of Con A by Reeke*et al*. (1975). The 3D model of LL was generated by SWISSMODEL using pea lectin (ID - 2bqp.1) as template with identity of 89.70% and

UNDER PEER REVIEW

coverage was 85% (Fig. 3). The 3D model of APN with sequence identity 30.59% and coverage 89% generated by same server using APN1 from *Anopheles gambiae* as a template (ID-4wz9.1.A) (Fig. 4). The generated 3D models were further validated with RAMPAGE programme. The torsion angles ψ and ϕ were examined to access the reliability of the protein model. The result obtained in the validation of LL protein, 94.7% of the amino acid residues were found in the most favoured region, while 6.1% of amino acid residues were found in allowed region (Fig 5). The validation result for APN revealed that 93.5% of amino acid residues were found in most favoured region, while 4.7% and 0.2% of amino acid residues were present in allowed and outlier region (Fig 5).

**Molecular docking and active site mapping**

Molecular docking through ClusPro revealed positions of interaction between lentil lectin and APN receptor (Fig. 6). The amino acid residues name and positions involved in interaction are summarized in Table 1. The amino acid residue of both the protein??? involved in interchain H-bond, without selecting any residue in lectin protein and receptor, were evaluated using ClusPro server. Active site mapping for determining the residues involved in ligand binding i.e. XMM [(5-Bromo-4-Chloro-3-Indolyl)-A-D-Mannose] for lentil lectin protein is done by LigPlot using PDBSum tool (Fig. 6). The LIGPLOT displays all the interactions in term of hydrogen bond between the ligand and the residues of protein molecules (Table 2). A wide range of lectins, viz., GNA, Con A, PSA and ASA, exhibiting mannose or mannose/glucose sugar binding affinity, revealed profound anti-metabolic effects towards members of the homopteran insects both under in vitro (Habibi*et al*., 1993; Powell *et al*., 1993; Rahbe*et al*., 1995) and in planta conditions (Powell *et al*., 1995; Gatehouse *et al*.,1996; Rao*et al*., 1998). Most common sugar specificities expressed by lectins are towards Mannose, Mannose/Glucose, Mannose/Maltose, Fucose, Galactose/N-acetylgalactosamine, N-acetylglucosamine, sialic acid and complex glycan groups (Peumans and Van Damme, 1998). The nature of lectins and carbohydrates interaction has been detailed in a review by Del Carmen Fernandez-Alonso *et al*. (2012).

**Conclusion**

Lentil lectin protein sequence was retrieved from NCBI database and further subjected for *in silico* determination of insecticidal potential using various computational approaches. The structural attributes like secondary structure, physicochemical properties and the amino acid residues participating in the catalytic activity were determined using bio-computational tools. 3D model of the protein was generated via homology-based modelling

UNDER PEER REVIEW

which can help in understanding the structural and functional characteristics of the protein. The binding potential of LL with its probable ligand showed that this protein could serve as best candidate gene for generating transgenic against hemipterans. This model can serve as a template for characterizing various lectins from plant species since no templates are available for the same in the protein model database.

**References:**


1. Del Carmen Fernández-Alonso,M., Díaz, D., Berbis, M., A., Marcelo, F., Cañada J. and Jiménez-Barbero J., 2012. Protein-Carbohydrate Interactions Studied by NMR: From Molecular Recognition to Drug Design. *Current Protein and Peptide Science*, **13:**816-830.

   FAOSTAT, Food and Agriculture data (Statistics Division, Food and Agriculture Organization of United Nations, Rome, Italy) 2012 [http://faostate.fso.org/].
2. Gatehouse, A.M.R., Down, R.E., Powell, K.S., Sauvion, N., Rahbe, Y., Newell, C.A., Merryweather, A., Hamilton, W.D.O. and Gatehouse, J.A., 1996. Transgenic potato plants with enhanced resistance to peach-potato aphid *Myzuspersicae*. *Entomol. Exp. Appl*. **79:** 295-307.
3. Gupta, S.K., Rai, A.K., Kanwar, S.S., Sharma, T.R.,2012. Comparative analysis of zinc finger proteins 553 involved in plant disease resistance. *PLoS One***7(8):**e42578.
4. Habibi, J., Backus, E.A., and Czapla, T.H., 1993. Plant lectins affect survival of the potato leafhopper (Homoptera: Cicadellidae). *J. Econ. Entomol*. **86**: 945-951.
5. Hardman, K. D. and Ainsworth, C. F. (1972). Structure of concanavalin A at 2.4 Å resolution. *Biochemistry;***11**49:10-19.
6. Lis, H. and Sharon, N.,1998. Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev*. **98**: 637-674.
7. Loris, R., Hamelryck, T., Bouckaert, J., Wyns, L., 1998. Legume lectin structure. *Biochem. Biophys. Acta*. **1383**: 9-36.
8. MoraesFilho, R.M., Rossiter, J.G., Cavalcanti, Junior, E.A., Martins, L.S.S., 2017.*In Silico* Comparative Analysis of Legume Lectins. *J Genet Genom* 1:103.
9. Peumans, W.J. and VanDamme, E.J.M., 1998. Plant lectins; Specific tools for the identification, isolation and characterization of O-linked glycans. *Critic. Rev.Biochem& Mol. Biol*. **33**: 209-258.
10. Powell, K.,S., Gatehouse, A.M.R., Hilder, V.A., and Gatehouse, I.A., 1993. Antimetabolic effects of plant lectins and plant and fungal enzymes on the nymphal


UNDER PEER REVIEW

stages of 2 important rice pests, *Nilaparvatalugens* and *Nephotettixvirescens*. *EntomolExp Appl*. **66**: 119-126

11. Powell, K.S., Gatehouse, A.M.R., Hilder, V.A., and Gatehouse, J.A., 1995. Antifeedant effects of plant lectins and an enzyme on the adult stage of rice brown planthopper, *Nilaparvatalugens*. *Entomol. Exp. Appl*. **75**: 51-59.
12. Rahbe, Y., Sauvion, N., Febvay, G., Peumans, W.J.and Gatehouse, A.M.R., 1995. Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrthosiphon pisum*. *Entomol. Exp. Appl*. **76**: 143-155.
13. Rao, K.V., Rathore, K.S., Hodges, T.K., Fu, X., Stoger, E., Sudhakar, D., Williams, S., Christou, P., Bharathi, M., Bown, D.P., Powell, K.S., Spence, J., Gatehouse, A.M.R., and Gatehouse, J.A.,1998. Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper. *Plant J*. **15**: 469-477.
14. Reeke, G. N., Jr., Becker, J. W. and Edelman, B. M.,1975. The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. *J. Biol. Chem*., **250,** 1525-1547.
15. Rouge, P., Cambillu, C. and Bourne, Y., 1991. The three-dimensional structure of legume lectins. In: Lectin Reviews. Kilpatrick DC, Van Driessche E, Bog- Hansen TC (eds). Vol.1 Sigma Chemical co. St.Louis USA pp.143-159
16. Sharma, V. andSurolia, A., 1997. Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary activity. *J. Mol. Biol*. **267**: 433-445.
17. Sharon, N. and Lis, H., 1990. Legume lectins –a large family of homologous proteins. *The FASEB. J*. **4**:3198-3208.
18. Siddiqui S., Hasan S., and Salahuddin A., 1995. Isolation and characterization of cajanuscajan lectin. *Archives of biochemistry and biophysics*.,**319**;426-431.
19. Srinivas, V.R., Reddy, G.B., Ahmad, N., Swaminathan, C.P., Mitra, N. andSurolia, A., 2001. Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. *Biochem. Biophys. Acta*. **1527**: 102-111.
20. Van Damme, E.J.M., Allen, A.K. and Peumans, W.J., 1988. Related mannose- specific lectins from different species of the family Amaryllidaceae. *Physiol. Plant.***73**: 52-57.
21. Van Damme, E.J.M., Culerrier, R., Barre, A., Alvarez, R., Rouge, P. andPeumans, W.J., 2007. A novel family of lectins evolutionarily related to class V chitinases: an example of neofunctionalization in legumes. *Plant Physiol*. **144**: 662-672.

UNDER PEER REVIEW

22. Vijayan, M. and Chandra, N.,1999. Lectins. *Curr. Opin. Struct. Biol*. **9**: 707-714.

UNDER PEER REVIEW

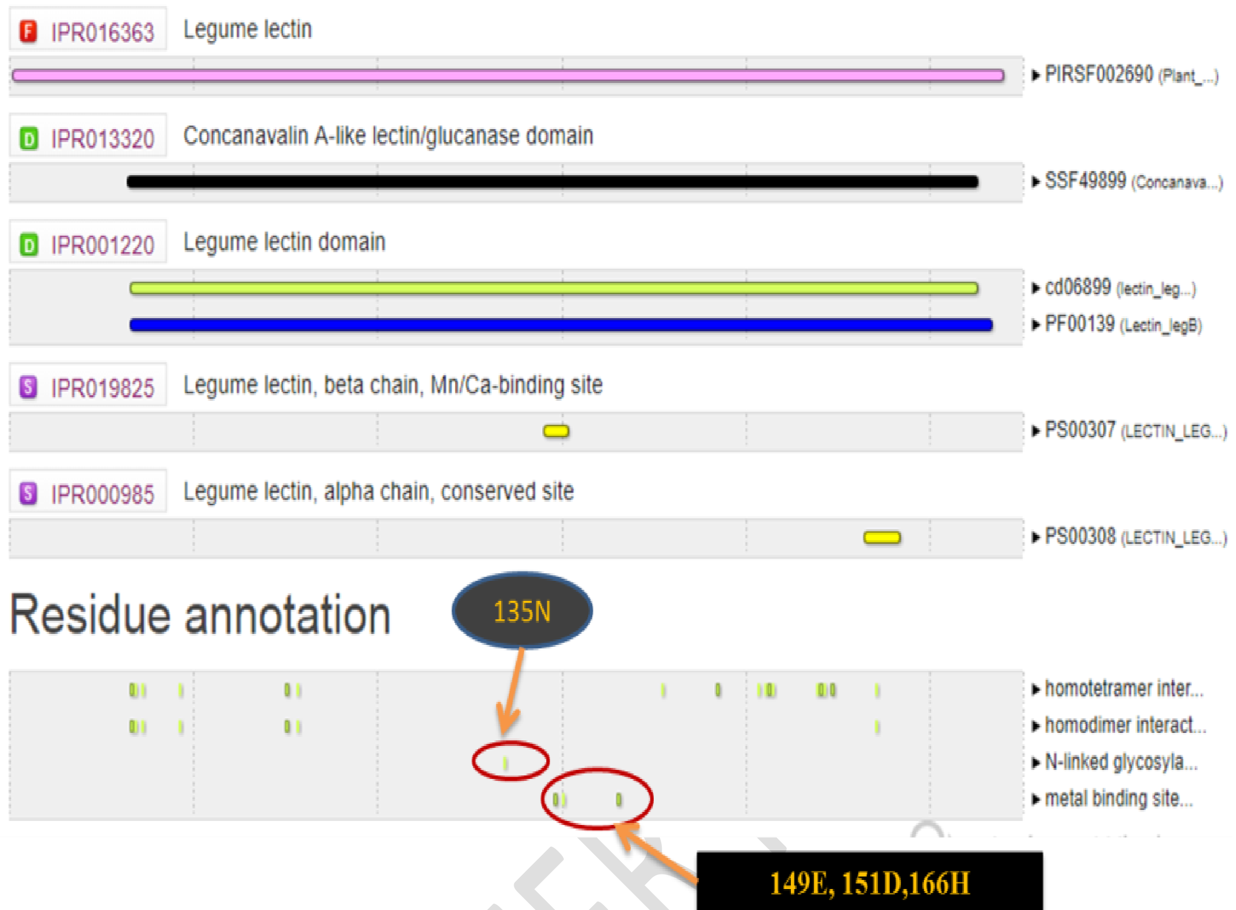


**Fig. 1 Functional domains analysis of LL protein sequence through InterPro**

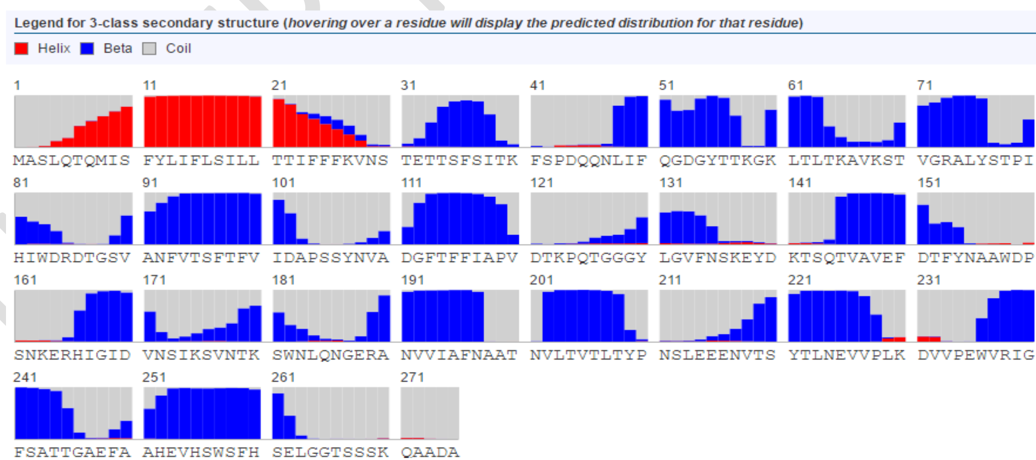


**Fig. 2A 3-class secondary structure displayed by individual amino acid residue of LL Protein**

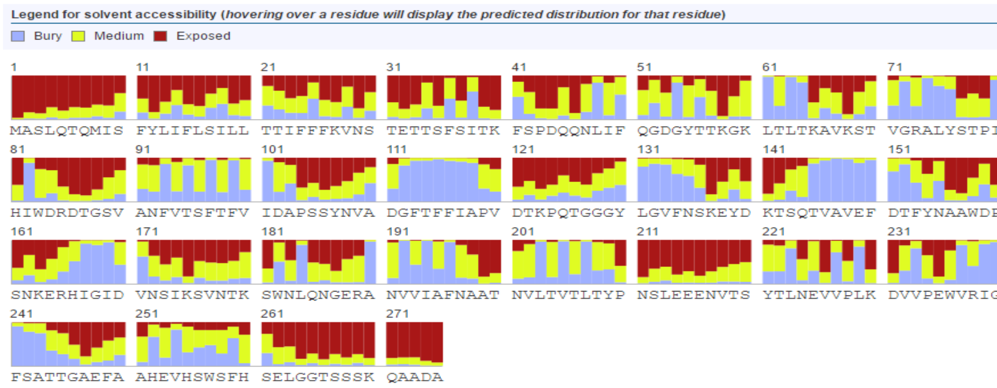


**Fig. 2B Solvent accessibility of individual amino acid residue of LL protein**

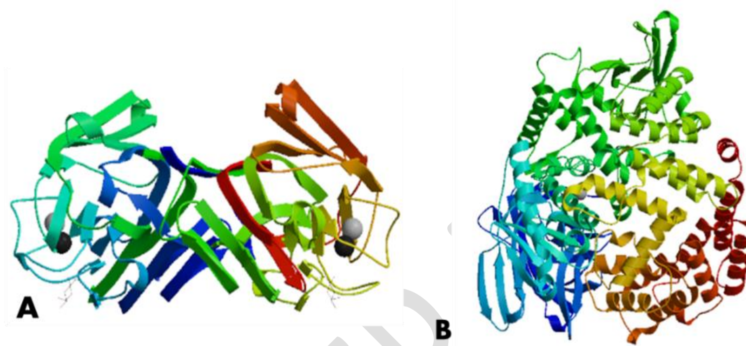


**Fig. 3 three dimensional model of (A)LL (B) APN generated via homology-based modeling using SWISS MODEL depicting various secondary structures—α helices, β pleated sheets and random coils**

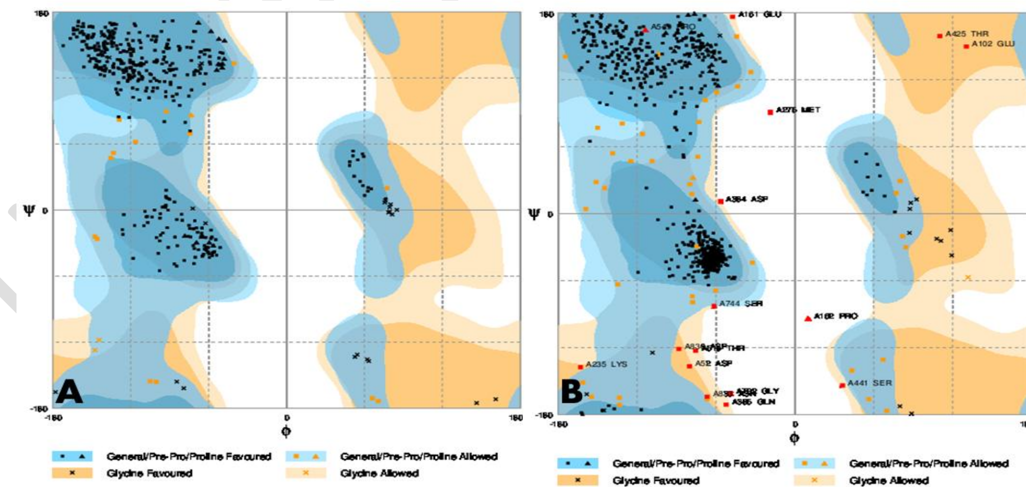


**Fig. 4 Validation of 3D predicted structure using RAMPAGE (A) LL (B) APN**

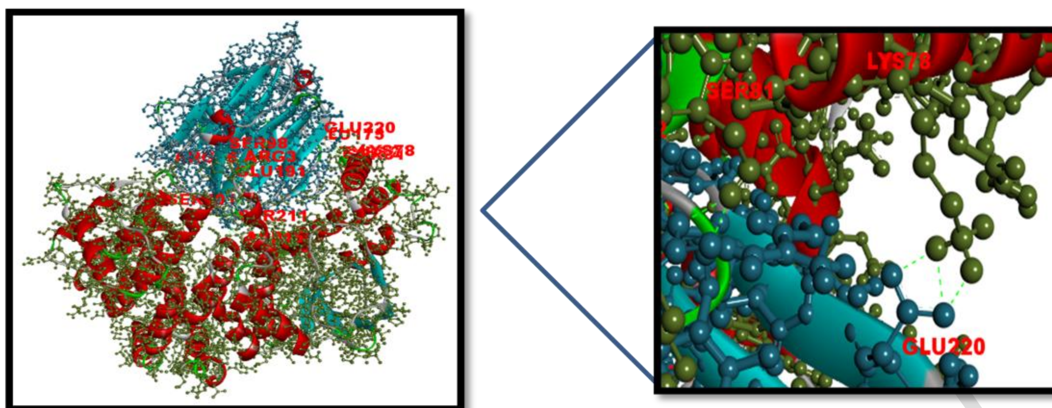


**Fig.5 Molecular docking of LL with APN, showing hydrogen bond (green dotted line) at interaction sites**

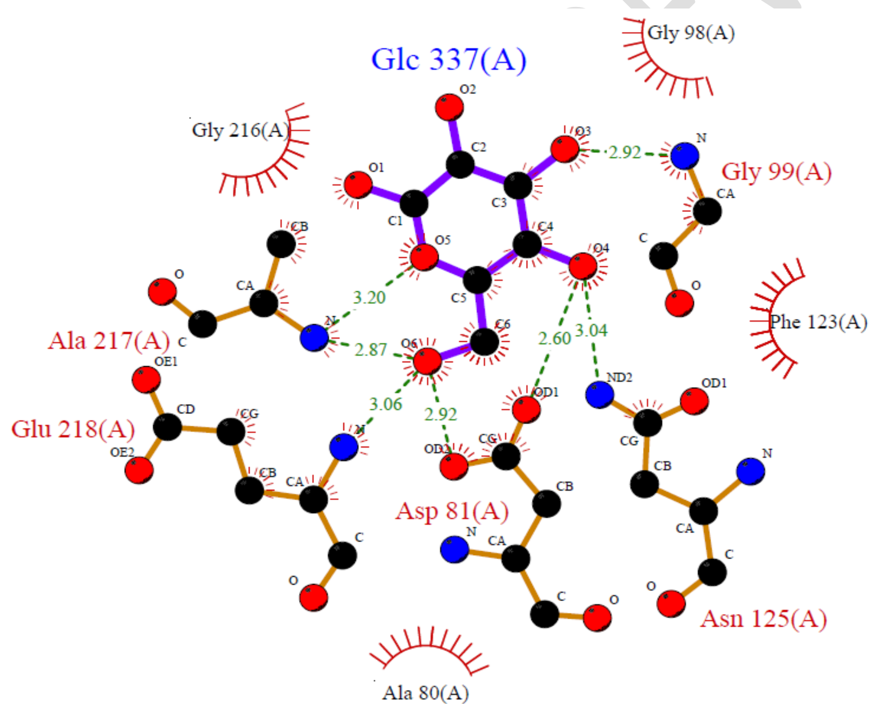


**Fig. 6 LigPlot of LL protein and its interaction with ligand XMM303 (A)**

| LL | Receptor<br>(Aphid gut) |
|---|---|
| GLU220 | LYS78 |

| GLU175 | TYR82 |
|---|---|
| GLU175 | SER81 |
| SER98 | ARG3 |
| GLU191 | ARG3 |
| ARG56 | SER401 |
| SER211 | ASP15 |

Table 1: Amino acid residues with their respective positions involved in interaction through H-bonding in molecular docking

| **Target protein (LL)** | | **Ligand (XMM)** | |
|---|---|---|---|
| Res. name | Res .position | Res. position | Distance ($A^0$) |
| ASP | 81 | 303 | 2.92 |
| GLY | 99 | 303 | 2.92 |
| ASN | 125 | 303 | 3.04 |
| ALA | 217 | 303 | 3.20 |
| ALA | 217 | 303 | 2.87 |
| GLU | 218 | 303 | 3.06 |

Table 2: Amino acid residue and their respective position involved in interaction with ligand XMM with H-bond length in LigPlot

UNDER PEER REVIEW