Computational Analysis to Study the Insecticidal Properties of Lectin Protein Through Docking Studies

Roma Chandra[1]*, Sapna Yadav[1]

[1]Department of Biotechnology, IILM University, Greater Noida, India.

ABSTRACT

Lectins belong to one of those protein families that exist in nature in abundance. Out of these, legume lectins are widely spread and display interesting antimicrobial and insecticidal properties. This is because, they recognize and bind to specific carbohydrate components without making any changes in the covalent assembly of recognized ligands, which makes them a suitable target candidate for useful applications in food and agriculture Moreover, numerous studies have demonstrated that the lectins found in legumes are poisonous to hemipterans. This study aims to elaborate on the insecticidal properties of *Cajanus cajan* (L.) Mill lectin (CCL) with receptor alanyl Aminopeptidase N (APN) from *Acyrthosiphon pisum* membrane and other insect origins through molecular modeling. The Protparam tool elaborated on the usefulness of physico-chemical parameters such as the aliphatic index, theoretical isoelectric point, and instability index along with the functional domains of the protein. The secondary structure of the CCL and APN was predicted to explain their structural and behavioral properties. Molecular docking was done between CCL as ligand and APN as receptor from different insect origins to obtain the cluster scores. To validate the current knowledge of the CCL protein structure and to demonstrate the prospective candidate gene for developing transgenics for improved aphid or other insect resistance, cluster scores with the lowest energy were evaluated using a variety of bio-computational tools.

Key Words: *Lectin, Docking, Molecular modeling, Acyrthosiphon pisum*

INTRODUCTION

Proteins or glycoproteins of a non-immune origin are lectins. They can reversibly attach to complex glycoconjugates' carbohydrate components while ensuring the covalent composition of all known glycosyl ligands [1]. Lectins can be detected through hemagglutination assays. Generally, they are categorized based on their structure, species location, and selectivity for carbohydrates. They play many physiological functions contingent on their characteristic properties and dispersal in tissues. The key function of lectin is to identify target molecules in different ways. This property renders it significant in a study that comprises purification, structural analysis, and application of these macromolecules to different extents like molecular or cell biology, immunology, medicine, clinical analysis, nanotechnology, and drug design or release. They can also be used as insecticidal agents in agriculture [2]. These proteins are found in all living species, especially bacteria, viruses, fungi, mammals, and plants. The lectins identified in legumes have received the most study and research among plant lectins [3]. These lectins are part of a very similar family of proteins as they are widely distributed in the seeds of legumes such as jack beans, common beans, peas, peanuts, and soybeans. The type of identified target, however, is greatly affected by the huge differences in their quaternary structures and carbohydrate specificities [4]. Lectins are a divergent group of carbohydrate-binding proteins of non-immune origin that are widely spread in all kingdoms of life.

Legume lectins show a homogeneous molecular structure of ~30kDa subunits. They share large sequence homology and tertiary structural similarities but vary according to their specificities for carbohydrates [5]. These lectins' tertiary structures are often made up of β-sheets connected by α helices, β turns, and bends. Between these β-sheets, in

Corresponding author: Roma Chandra
Address: Department of Biotechnology, IILM University, Greater Noida.
E-mail: ✉ roma.chandra@iilm.edu; romachandra2010@gmail.com
Received: 09 November 2022; **Revised:** 30 January 2023; **Accepted:** 03 February 2023

addition to the short loops, quaternary interfaces are formed. Loops connecting these β-sheets produce alternatively spliced chains that are typically free of helices [6]. High sequence and structural similarity were found in the monomers of legume lectin, with just minor variations in loop and strand lengths. The monomer structures exhibit a jellyroll motif with metal binding for divalent ions (Ca2+ and Mn2+) and a carbohydrate-binding domain (CRD). Yet, quaternary structures exhibit significant differences despite sharing a lot in common with primary, secondary, and tertiary structures. These differences affect how monomers interact with one another and whether or not proteins undergo modifications like glycosylation [7, 8]. All of the screened compounds 1–7 demonstrated outstanding in-silico% absorption, with the highest value obtained is 94.04%, according to the prediction study.

Pigeon pea (*Cajanus cajan* (L.) Mill sp.) is a crucial legume crop belonging to the family Leguminosae, which is mostly cultivated in warm-climate countries. It is a healthy food source that is high in nutrients for both people and domestic animals. Due to the seed's comparatively low fiber level and moderately high protein content, the grain offers great nutritional quality. Studies have shown that in a few villages in India, pigeon pea provides half of all proteins consumed [9]. Pigeon peas have the potential to fix nitrogen, making them an important component of sustainable cropping systems. Farmers recognize and value this tendency of pigeon peas to "replenish" the soil when planted after a grain crop [10]. The genome size of the pigeon pea, a diploid crop (2n = 22), is 808 Mb. Pigeon peas are mostly grown in India, Nepal, and Myanmar. India produces 2.86 million tonnes and has a 4.42 million hectares area among the three [11]. A comparable subunit containing threonine and alanine at the N- and C-termini of each subunit makes up the dimer of *C. cajan* lectin. The presence of a significant amount of acidic amino acids serves as the primary indicator of its amino acid composition [12]. The goal of the current work was to characterize the CCL protein and APN's secondary structure, 3D model, and physicochemical properties. The evaluation was done using molecular docking studies, the insect receptor's APN binding capability with the CCL protein.

MATERIALS AND METHODS

From the NCBI database, the *C. cajan* lectin amino acid sequence with accession number KU382473.1 was retrieved. InterPro tool was used to determine the functional areas of lectin. This tool is easily available on the EBI web page. Next, physicochemical properties of lectin such as its amino acid composition, molecular weight, pI value, instability index, half-life, etc. were determined using the Protparam tool, available on Expasy, and finally, the Protein disorder prediction server (PrDOS) tool was used to evaluate disorders in protein. RaptorX tool was used to predict the secondary level of protein structure as well as study the solvent accessibility of lectin. Tertiary structure prediction was also done using the SWISS Model tool through homology-based modeling using Pea Lectin as a template for the target protein. The predicted models were further confirmed using the RAMPAGE tool. To study protein target interaction studies, *Acyrthosiphon pisum* membrane alanyl Aminopeptidase N (APN) of accession number DQ440823.1 was considered as a target for binding CCL protein. Similar receptors from different origins were also used for molecular docking. Molecular docking was done using ClusPro docking. Result validation was done based on the models and their model scores.

RESULTS AND DISCUSSION

he InterPro tool discovered the CCL protein sequence's functional region represented in **Figure 1**. The metal-binding sites, GLU149, ASP151and HIS166 of the CCL protein sequence were identified based on its residue annotation. The presence of metal ions such as Ca^{2+} and Mn^{2+} was recognized to be very significant in legume lectins and has been demonstrated by evolutionarily conserved amino acids binding to these metal ions [13]. Also, depending on the residue annotation, it was revealed that ASN135 was involved in N-linked glycosylation. The CCL was described as an acidic and stable protein according to its pI value and instability index, which were 5.82 and 23.59 respectively. PrDOS tool predicted 2 disorderly regions in the whole protein sequence of CCL, the longest site being among Gly264 to Ala275 consisting of 12 amino acid residues (**Figure 2**). The Protparam tool predicted the physicochemical properties, showing that the protein consisted of 277 amino acids with a molecular weight of 30476.25. The projected half-life of CCL protein was found to be 30h in mammalian reticulocytes, 20h, and 10h in yeast and Escherichia coli respectively. Also, CCL protein was found thermally stable under different climatic conditions as its Aliphatic index (Ai) was counted as 80.80 [14]. Similarly, GRAVY indices of CCL were counted as -0.070, which showed the hydrophilic nature of a protein that refers to charged amino acid residues (25 negatively charged and 21 positively charged) present in the protein sequence, suggesting extrinsic association in the cell membrane.
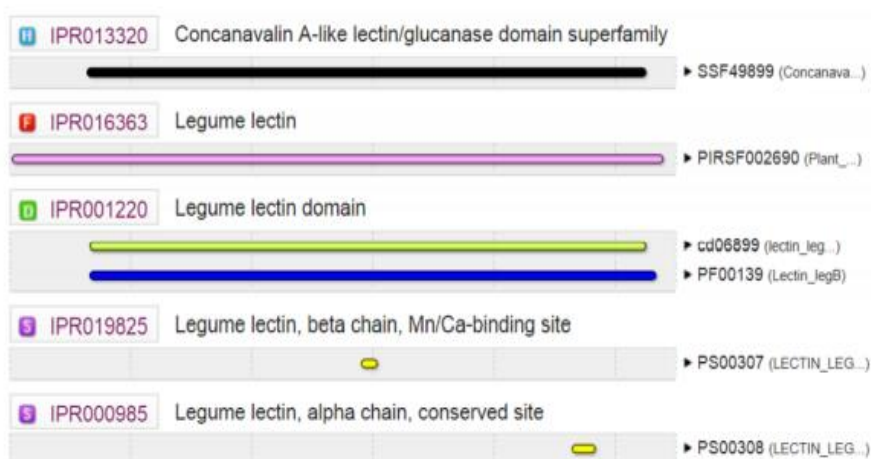
2

Figure 1. Results of InterPro tool for functional areas of lectin protein


Figure 2. Results of PrDOS tool for predicted disordered amino acid residues present in lectin protein

3

Our protein's secondary structure was predicted by the RaptorX tool to have a total of 7% alpha helices, 42% beta pleated sheet, and 50% coil **(Figure 3a)**. The protein's solvent accessibility was further demonstrated by the finding that 41% of the total residues were classified as medium, or neither exposed nor buried and that 27 percent of the total repeats were interred in the structure **(Figure 3b)**. Several studies have noted that most legume lectin proteins have 0-10% alpha helix, 40-50% beta sheets, and 35-45% turn, classifying them into distinct protein subclasses [15]. Furthermore, the predicted alpha helices, beta sheets, and turns in our work were extremely well supported by the X-ray structure of Con A as found by Reeke *et al*. [16].
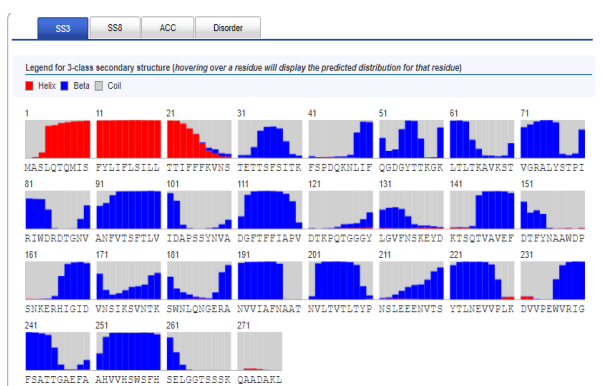
a)
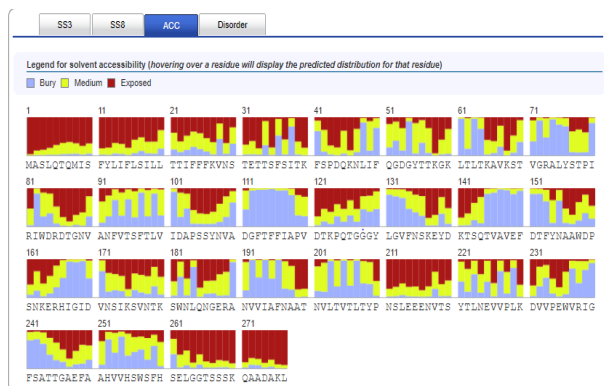
b)

Figure 3. a) Secondary level representation of lectin protein structure presented by individual amino acid residue, and b) Solvent accessibility of lectin protein presented by individual amino acid residue

Using the pea lectin template (PDB ID: 2bqp.1 A), which has an 88.41% sequence identity and an 85% coverage, SWISS-MODEL was used to create the three-dimensional model of CCL (**Figure 4a**). APN1 species *Anopheles gambiae* (PDB ID: 4wz9.1.A), which has a sequence homology of 30.59% and a coverage of 89%, was used as a template to create the 3D model of APN employing the same server (**Figure 4b**). Molecular docking was carried out by considering CCL as ligand and APN as receptor using the ClusPro docking server. A total of 13 receptors were used to carry out docking. These receptors were selected based on the sequence similarity of 30.59% and coverage of 89 % to our target APN receptor. All the selected receptors are listed below:

1. *Acyrthosiphon pisum* membrane alanyl Aminopeptidase N (APN; Accession # DQ440823.1)
2. 4kx7.1. A Glutamyl aminopeptidase Crystal structure of human aminopeptidase A
3. 6ea4.1. An endoplasmic reticulum (ER) aminopeptidase 2 ERAP2 bound to Aryl Sulfonamide Uncompetitive Inhibitor
4. 5ab0.1. An Endoplasmatic Reticulum Aminopeptidase 2 Crystal structure of aminopeptidase ERAP2 with ligand
5. 4wz9.1.AAGAP004809-PA APN1 from *Anopheles gambiae*
6. 5j6s.1. A ER aminopeptidase 2 Crystal structure of ER Aminopeptidase 2 (ERAP2) in complex with a hydroxamic derivative ligand
7. 5ab0.2. An ER Aminopeptidase 2 Crystal structure of aminopeptidase ERAP2 with ligand
8. 5cu5.1. An ER aminopeptidase 2 Crystal structure of ERAP2 without catalytic Zn(II) atom
9. 5cu5.2. An ER aminopeptidase 2 Crystal structure of ERAP2 without catalytic Zn(II) atom
10. 6mgq.1. An ER aminopeptidase 1 ERAP1 in the open conformation bound to 10mer phosphinic inhibitor DG014
11. 3se6.1. An ER aminopeptidase 2 Crystal structure of the human ER Aminopeptidase 2
12. 5j6s.2. An ER aminopeptidase 2 Crystal structure of ER Aminopeptidase 2 (ERAP2) in complex with a hydroxamic derivative ligand
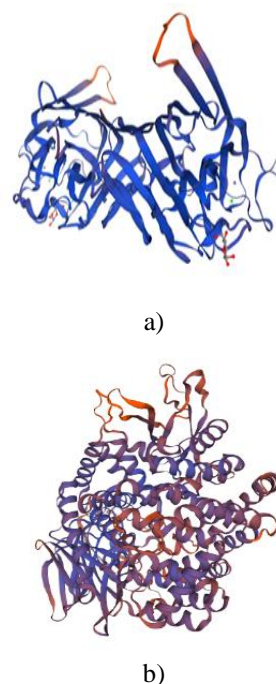13. 3mdj.1. An ER aminopeptidase 1 ER Aminopeptidase, ERAP1, Bound to the Zinc Aminopeptidase Inhibitor, Bestatin

a)

b)

Figure 4. a) 3D structure of lectin, and b) 3D structure of APN

Results obtained after molecular docking through the ClusPro docking server showed the interaction between lectin and different APN receptors (**Figure 5**). For every receptor, we obtained around 10 docked models each with their respective balanced coefficients. Among all the structures generated through molecular docking, the one with the lowest binding energy score was considered the most stable and is shown in **Table 1**. Several studies have documented that the more negative the binding energy, the more stable the ligand [17]. It is also believed that the binding energy is discharged during the engagement between the ligand and the target receptor, causing the complex's overall energy to decrease. The ligand's change from its minimal energy towards its coupled configuration with the protein is made up by the release of binding energy. As a result, when more binding energy is produced, the ligand has a larger probability of attaching to the receptor [18].
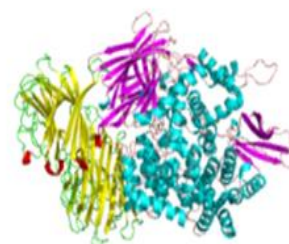

Figure 5. Molecular docking of lectin with APN

4

Table 1. Validation of docking results

S. No	Ligand	Receptor	Member	Representation	Weighted score
1	*C. cajan* lectin (CCL)	*Acyrthosiphon pisum* membrane alanyl Aminopeptidase N (APN)	26	Lowest energy	-671.3
2	*C. cajan* lectin (CCL)	4kx7.1. A Glutamyl aminopeptidase Crystal structure of human aminopeptidase A	51	Lowest energy	-617.1
3	*C. cajan* lectin (CCL)	6ea4.1. An ER aminopeptidase 2 ERAP2 bound to Aryl Sulfonamide Uncompetitive Inhibitor	26	Lowest energy	-671.3
4	*C. cajan* lectin (CCL)	5ab0.1. An ER aminopeptidase 2 Crystal structure of aminopeptidase ERAP2 with ligand	19	Lowest energy	-725.3
5	*C. cajan* lectin (CCL)	4wz9.1. An AGAP004809-PA APN1 from *Anopheles gambiae*	13	Lowest energy	-763.8
6	*C. cajan* lectin (CCL)	5j6s.1. An ER aminopeptidase 2 Crystal structure of ER Aminopeptidase 2 (ERAP2) in complex with a hydroxamic derivative ligand	38	Lowest energy	-737.2
7	*C. cajan* lectin (CCL)	5ab0.2. An ER aminopeptidase 2 Crystal structure of aminopeptidase ERAP2 with ligand	33	Lowest energy	-710.7
8	*C. cajan* lectin (CCL)	5cu5.1. An ER aminopeptidase 2 Crystal structure of ERAP2 without catalytic Zn(II) atom	52	Lowest energy	-748.7
9	*C. cajan* lectin (CCL)	5cu5.2. An ER aminopeptidase 2 Crystal structure of ERAP2 without catalytic Zn(II) atom	47	Lowest energy	-860.2
10	*C. cajan* lectin (CCL)	6mgq.1. An ER aminopeptidase 1 ERAP1 in the open conformation bound to 10mer phosphinic inhibitor DG014	33	Lowest energy	-793.7
11	*C. cajan* lectin (CCL)	3se6.1. An ER aminopeptidase 2 Crystal structure of the human ER Aminopeptidase 2	21	Lowest energy	-712
12	*C. cajan* lectin (CCL)	5j6s.2. An ER aminopeptidase 2 Crystal structure of ER Aminopeptidase 2 (ERAP2) in complex with a hydroxamic derivative ligand	15	Lowest energy	-678.1
13	*C. cajan* lectin (CCL)	3mdj.1. An ER aminopeptidase 1 ER Aminopeptidase, ERAP1, Bound to the Zinc Aminopeptidase Inhibitor, Bestatin	47	Lowest energy	-980.3

5

CONCLUSION

The lectin sequence information was obtained from NCBI and using a variety of bioinformatics methods, its insecticidal ability was further evaluated through in silico methods. The docking results demonstrated that lectin proteins can serve as the strongest candidate gene for creating transgenes against pests based on their binding affinity. Also, because there are no precedents for this in the polypeptide model database, these models can be used as a template for the identification of several lectins from plant species.

REFERENCES

[1] Irache JM, Durrer C, Duchene D, Ponchel G. In vitro study of lectin-latex conjugates for specific bioadhesions. J Control Release. 1994;31(2):181-8.

[2] Mishra A, Behura A, Mawatwal S, Kumar A, Naik L, Mohanty SS, et al. Structure-function and application of plant lectins in disease biology and immunity. Food Chem Toxicol. 2019;134:110827. doi:10.1016/j.fct.2019.110827

[3] Rajan K, Ankur T. Research advances and prospects of legume lectins. J Biosci. 2021;46(4):104. doi:10.1007/s12038-021-00225-8

[4] Tsaneva M, Van Damme EJ. 130 years of plant lectin research. Glycoconj J. 2020;37:533-51. doi:10.1007/s10719-020-09942-y

[5] Prajapat R, Singh P, Tiwari P, Mainkar P, Sahoo S, Rao A, et al. In silico analysis and molecular docking studies of *Cajanus cajan* lectin against aminopeptidase-N receptor from *Acyrthosiphon pisum*. Int J Curr Microbiol Appl Sci. 2018;7(06):959-67.

[6] Afzaal M, Saeed F, Aamir M, Usman I, Ashfaq I, Ikram A, et al. Encapsulating properties of legume proteins: recent updates & perspectives. Int J Food Prop. 2021;24(1):1603-14. doi:10.1080/10942912.2021.1987456

[7] Sharon N, Lis H. History of lectins: from hemagglutinins to biological recognition molecules. Glycobiology. 2004;14(11):53R-62R.

[8] Ambrosi M, Cameron NR, Davis BG. Lectins: tools for the molecular understanding of the glycocode. Org Biomol Chem. 2005;3(9):1593-608.

[9] Sarkar S, Panda S, Yadav KK, Kandasamy P. Pigeon pea (*Cajanus cajan*) an important food legume in Indian scenario – A review. Legum Res. 2018;43(5):601-10. doi:10.18805/LR-4021

[10] Sarkar S, Roy S, Ghosh SK. Development of marker-free transgenic pigeon pea (*Cajanus cajan*) expressing a pod borer insecticidal protein. Sci Rep. 2021;11(1):10543. doi:10.1038/s41598-021-90050-8

[11] FAO & WHO. Report 2022 – Pesticide residues in food - Joint FAO/WHO meeting on pesticide residues. Rome. 2023. doi:10.4060/cc4115en

[12] Siddiqui S, Hasan S, Salahuddin A. Isolation and characterization of *Cajanus cajan* lectin. Arch Biochem Biophys. 1995;319(2):426-31.

[13] Lagarda-Diaz I, Guzman-Partida AM, Vazquez-Moreno L. Legume lectins: proteins with diverse applications. Int J Mol Sci. 2017;18(6):1242. doi:10.3390/ijms18061242

[14] Han G, Qiao Z, Li Y, Wang C, Wang B. The roles of CCCH zinc-finger proteins in plant abiotic stress tolerance. Int J Mol Sci. 2021;22(15):8327. doi:10.3390/ ijms22158327

[15] Tripathi A, Hallan V, Kumar S, Raj R, Katoch R. Predictive structure and protein–ligand interface of novel lectin from rice bean (*Vigna umbellata*). Integr Biol. 2022;14(8-12):228-40. doi:10.1093/intbio/zyac018

[16] Reeke Jr GN, Becker JW, Edelman GM. The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. J Biol Chem. 1975;250(4):1525-47.

[17] Vasant OK, Chandrakant MA, Chandrashekhar KV, Babasaheb GV, Dnyandev KM. A review on molecular docking. Int Res J Pure Appl Chem. 2021;22(3):60-8.

[18] Yang C, Chen EA, Zhang Y. Protein-Ligand docking in the machine-learning era. Molecules. 2022;27(14):4568. doi:10.3390/molecules27144568

6