

Structure Prediction, Functional Characterization of Lectins and their Interaction with Tea Pest Receptor

Sathyapriya.P¹, Kalaivani.A², Arvinth.S³

¹Plant Physiology & Biotechnology Division, UPASI Tea Research Foundation, Tea Research Institute Valparai 642 127, Coimbatore District, Tamil Nadu

²Post doctoral fellow, Department of post-harvest science, Agricultural research organization, Volcani center Israel.

³Assistant professor in plant biology, PSG College of Arts and Science, Coimbatore, Tamil Nadu

ABSTRACT:- Tea is one of the most important agro-industrial crops sustaining the Indian economy. *Oligonychus coffeae* Nietner (Acari: Tetranychidae), the red spider mite (RSM), is a major pest of tea (*Camellia sinensis*) in most tea-producing countries. Nymphs and adults of RSM lacerate cells, producing minute characteristic reddish brown marks on the upper surface of mature leaves, which turn red in severe cases of infestation, resulting in crop loss. Molecular interaction of Zingiberaceae lectins with red spider mite symbiotic receptor protein was carried out. Initially genes of lectins were isolated through RNA isolation, cDNA synthesis and sequencing. ExPASy's Prot-param server and Cys_rec tool were used for physico-chemical and functional characterization of these proteins. Studies on secondary structure of these proteins were carried out by computational program SOPMA and hydrophobic cluster analysis was done. The three dimensional structure of proteins were modeled using I-TASSER and evaluated for best score models. Interaction between symbiotic receptor and lectins made complexes of two molecules. Surface accessibility test demonstrated significant masking of lectins in comparison to the free chaperonin molecule. Thus, these lectins cover up a certain surface area of the receptor, making that area sterically inaccessible to solvents or other molecules.

Keywords: tea, red spider mite, lectin, endosymbiont, *Buchnera* sp

1. INTRODUCTION

Pest/pathogen attack is an unavoidable phenomenon in plants as they do not have an immune response like humans. Being a monoculture crop tea provides a stable microclimate for a number of pests and disease. The perennial habit of this plant peculiar cultural conditions and humid climate of the tea growing areas are highly favorable for disease development. In India Red Spider Mite (RSM) was initially recorded in coffee by [1] in 1861, later in 1868, it was observed in tea. Infestation by RSM starts along midrib and veins initially and gradually spreads to the entire upper surface of leaves. In severe infestation, it damages the younger and older leaves and ultimately leads to defoliation and debilitation of the tea bush [2]. Pesticide application programme on tea in south India is guided by the recommendations of UPASI Tea Research Institute [3]. During the past one decade, the control of RSM has been achieved mainly by applying synthetic acaricides.

Long-term and continuous application of biocontrol agents were also made the pests to create the resistance towards them and it was proved in Bt toxins, where initially it was used against homopteran insect pests [4, 5]. Plants and insects are being interacted one another continuously in a complex way and plants developing number of strategies to overcome pest attack [6]. Plants synthesize a wide range of secondary metabolites which are toxic to herbivores. A few plant toxic substances exhibit more or less insecticidal properties viz., lectins, protease inhibitors, α -amylase inhibitors, ribosome-inactivating proteins, arcelins, canatoxins-like proteins and ureases[7].

The monocot mannose-binding lectins (MMBLs) have received much interest as well become an important tool in plant protection and plant biotechnology since their genes confer resistance against sucking insects and nematodes [8]. Unlike all other plant lectins, MMBLs are encoded by large families of closely related genes [9]. MMBLs are found to be very similar at the protein level; but differ in the processing and post-translational modifications of the primary translation products of their genes [10]. *Galanthus nivalis* lectin (GNA) was the first monocot mannose-binding lectin to be identified and it has been regarded for its insecticidal properties. Effect of GNA has well documented in vitro on the biology of Aphididae, Delphacidae and Cicadellidae [11, 12]. Using tobacco as a model system, GNA expression has been shown to adversely affect the population of *Myzus persicae* [8]. Subsequently, many studies have demonstrated deleterious effects of GNA expression in planta (potato, rice, wheat) towards different sap-sucking insects [13, 14].

Isolation of lectins with a wide range of specificities from the variety of the plant is the thirteenth area in the current integrated pest management [15]. This is believed to help in the crop resistance by introducing the novel genes against various plant-eating organisms. Hence, the identification of lectins with fungal/insecticidal activity from the new sources may reveal genes with potential characters for use in the genetic improvement of crops. This work aimed at molecular cloning of lectin genes and building of their structure through molecular modeling analyses. Additionally molecular interaction of the lectin and bacterial chaperonin were also provided.

2.0. MATERIALS AND METHODS

All the experiments were carried out at the United Planters Association of Southern India (UPASI) Tea Research Foundation, Tea Research Institute, Valparai, Tamil Nadu, India (latitude 10°30'N, longitude 27°0'S and altitude 1050 M) over a duration from 2011-2018. Molecular kits were purchased from QIAGEN and primers used for cDNA synthesis and gene amplification were from Sigma Chemicals Bangalore. PCR reaction components were purchased from MP biomedical, Bangalore. Gel extraction kit (QIAquick Gel Extraction kit, QIAGEN), transformation kit (InsTAclone™ PCR Cloning Kit, Thermo Scientific Company) was used for the DNA elution and cloning of DNA target into the vector respectively.

2.1. RNA Isolation

Tuber, rhizome and seeds of *Alpinia galanga* (Sitharathai), *Curcuma amada* (mango ginger), *Zingiber officinale* (ginger) and *Elettaria caramomum* (cardamom) were weighed (500 mg) individually and frozen with liquid nitrogen then powdered. Total RNA was isolated from Zingiberous plant tissues using the RNeasy mini Kit (RNeasy® Mini Handbook, Fourth edition, QIAGEN, April, 2006). To eliminate RNAase glassware, plastics and other materials used for RNA isolation were treated with 0.1% DEPC (diethyl pyrocarbonate) for overnight at 37°C, and then autoclaved for 15 min for eliminating residual DEPC before use as per instructions. Isolated RNA was quantified and confirmed formaldehyde (1%) agarose/EtBr gel electrophoresis.

2.2. Single strand cDNA synthesis

Synthesis of first strand cDNA was carried out using SMARTer® PCR cDNA Synthesis Kit (Clontech) synthesis as follows about 1µg of total RNA was mixed with 1µl of SMART IV Oligonucleotide (10µM) (5'-AAGCAGTGGTATCAACGCAGAGAGTGGCCATTACGGCGG-3'), 1 µl CDS III/3' PCR Primer (10µM) (5'-ATTCTAGAGGCCGAGGCGCCGACATG-d(T)₃₀VN-3') and incubated at 72°C for 2 min for first strand cDNA synthesis. Then tube was placed on ice for 2 min and 4.0 µl of 5X first-strand buffer, 1 µl of 20 mM DTT, 1 µl of 10 mM dNTP mix and 1 µl of Prime Script Reverse Transcriptase was added to the tube and mixed well by pipetting. Finally the tube was subjected to short spin and incubated at 42°C for 60 mins, placed immediately on ice to terminate first strand synthesis. First strand synthesized cDNA was stored at -20°C till further use.

2.3. PCR amplification of lectin genes

All PCR amplifications were carried out in the thermocycler (iCycler; Bio-Rad Laboratories, CA) using gene specific primers based on the conserved region. To obtain the coding region fragment, primers BamHI (5'-ACGGATCCATGGTCTACTACTTCATCTCT-3') and Hind III (5'-TAAAGCTTCAAGCAGCAC CGGTGCCA-3') were designed based on the database sequence (Genbank accession number: DQ525625). PCR reaction mixture was prepared with 2.5 µl of Buffer A (Tris with 15mM MgCl₂), 0.5 µl of MgCl₂, 1.0 µl of each primers BamHIFp (10 mM), Hind III Rp (10 mM) and dNTPs Mix (10 mM) taking 1.0 µl of DNA as template. 0.3 µl of TaqDNA polymerase (3U/µl) was used to initiate reaction. PCR was performed under the following condition: cDNA was denatured at 94°C for 5 min followed by 30 cycles of amplification (94°C for 45 s, 63°C for 30 s and 72°C for 1 min) and by 10 min at 72°C, 4°C holds. Resulted PCR product was then analyzed by 1.0 % agarose/EtBr gel.

2.4. Cloning and sequencing of lectin genes

The PCR fragment extracted and purified from the gel by using QIAquick® Gel Extraction kit protocol. (QIAquick Spin Handbook, QIAGEN, March, 2008). Purified DNA was cloned into the plasmid Vector, pTZ57R/T using InsTAclone™ PCR Cloning Kit (Appendix 1), Thermo Scientific by few modifications. The following components were added 5X ligation buffer (4 µl), ATP (0.3 µl), PEG (1.0 µl), DNA insert (5.0 µl), Vector pTZ57R/T (2.0 µl) and finally 1.0 µl of T4 DNA Ligase. Bacterial

transformation was carried out to clone the lectin genes for that *Escherichia coli* (E.coli) DH5 α bacteria cells were purposely treated with chloride salts and cold treatment to make them permeable to uptake the exogenous DNA (competency). 50 μ l of LB broth containing 5 μ l ligation mixture was streaked on different LB agar plates containing ampicillin, IPTG and X-gal containing plates were incubated overnight at 37°C to find effective transformation through blue/white further colony PCR. Positive colonies were selected and sequenced (Macrogen, Korea) for further analysis.

2.5. Physiochemical characterization of lectin sequences

Sequences were subjected to Vecscreen (<http://www.ncbi.nlm.nih.gov/tools/vecscreen/>) to remove the vector sequence (pTZ57R/T) then it was iterated to find similarity with other lectin available in the biological databases. Sequences were submitted to BLAST and a similarity search algorithm to retrieve the similar sequences based on their e-value, query coverage etc., and similar sequences were multiple aligned using clustal omega tool (<http://www.ebi.ac.uk/Tools/msa/clustalO/help/>) to identify their functional relationship with other manors- specific lectins from different plant families. Physiochemical parameters were calculated for the translated protein sequences using ExPASy's protparam server. CYS_REC was used to locate "SS bond" between the pair of cysteine residues, if present. The tool yields position of cysteines, total number of cysteine present and pattern, if present, of pairs in the protein sequence as output (<http://sunl.softberry.com/berry.phtml?topic>). SOPMA was employed for calculating the secondary structural features of zingiberous lectin sequences. Hydrophobic cluster analysis of lectin sequences were carried out using drawhca (<http://mobyli.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA#forms::HCA>). Phylogenetic tree was constructed using MEGA7 (<http://www.megasoftware.net/>) to see the evolutionary relationship with other associated proteins. The conserved domain of lectins was identified (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) in NCBI. Phylogenetic tree was constructed with Maximum likelihood statistical method with 500 replicates of bootstrapping. The tree was visualized in both original as well bootstrapped. Finally the sequences were submitted to both NCBI and PMDB for easy accession.

2.6. Model building and evaluation

The three dimensional structure of proteins were modeled using I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>). Quality of generating models was evaluated with PROCHECK by Ramachandran plot analysis. Stereo-chemical quality and accuracy of the selected models was further improved by subjecting it to energy minimization with the GROMOS 96 43B1 parameters set, implementation of the Swiss-PDB Viewer. Validation of generated models was further performed by VERIFY 3D and ERRAT programs. The process was applied for the analysis of Z scores and energy plots. The three dimensional structure of modeled proteins was analyzed using Deep View Swiss PDB viewer. Root Mean Square Deviation (RMSD) values were computed between the band of targets and template protein to visualize how much modeled protein deviates from the template protein structure.

2.7. Molecular docking of lectins with pest symbiotic receptor

CTAB method [16] was used to isolate genomic DNA of red spider mite and using universal 16S rRNA primers FD1 (ccgaattcgtcgacaacAGAGTTTGATCCTGGCTCAG) and RP1 (cccgggatccaagcttACGGTTACCTTGTTACGACTT) [17] endosymbiont sequence was amplified by commenting the PCR to denatured at 94°C for 3 min followed by 30 cycles of amplification (94°C for 30 s, 55°C for 30 s and 72°C for 1 min) and final elongation by 10 min at 72°C, 4°C holds. PCR product was sequenced and it was multiple aligned with available 16S rRNA gene sequences in the NCBI database. Phylogeny relationship was identified then the sequence was submitted. The three-dimensional coordinates of the bacterial chaperonin structure 1GRL, were obtained from the Protein Data Bank (rscb.org/PDB/). The coordinates of chain A were used for the docking purpose. A consensus oligosaccharide moiety, (GlcNAc)₂ (Man)₇, was obtained from the NMR solution structure of human Cd2 protein available in the Protein Data Bank under the code name 1GYA. The three dimensional co-ordinate structures of the proteins *Alpinia*, cardamom, ginger and mango ginger were obtained from the PDB with the following PDB IDs 1MSA, 1NPL, 1BWU and 3DZW respectively. The coordinates of the chain other than A of the above said PDB IDs were used for the docking purpose in order to avoid the confusion where 'A' chain of bacterial chaperonin is already chosen for docking purposes.

On closer inspection of the amino acid sequence of the Bacterial chaperonin protein, three asparagine-linked glycosylation signatures such as Asn 153, Ser 154, Asp 155, Glu 156 and Asn 229, Ile 230, Arg 231, Glu 232 and Asn 527, Asp 528, Ala 529, Ala 530 were found. However, our deep inspection on to the structure showed that the second and third signature remains buried into the quaternary structure of the bacterial chaperonin; hence, the first signature sequence was chosen as the

possible glycosylation site. This oligosaccharide moiety was covalently added to the Asn 153 - Glu 156 position of the bacterial chaperonin using the program Insight-II (Biosym Technologies) on a Dell optiplex work station. The resultant Protein Data Bank coordinates file was kept.

Structures of *Alpinia*, cardamom, ginger and Mango ginger proteins (ligands) was docked into the active site of its receptor Bacterial chaperonin using the program Gramm on a PC-LINUX work station. The experiment generated 4–5 good models, and the best model was chosen based on its energy rankings. The values of the accessible surface area for both the native protein (bacterial chaperonin) and the ligand-bound proteins were calculated using the homology module of the Insight-II program. The differences inaccessible surface area between native bacterial chaperonin and the docked complexes (*Alpinia*, Cardamom, Ginger, and Mango ginger proteins) was calculated for every residue using Insight-II. The schematics were rendered using the program Pymol.

3. RESULTS AND DISCUSSION

Initially, Mannose Binding Lectins (MBLs) were attracted because of their unique and exclusive specificities towards mannose hence these lectins have become very interesting tools in glycol conjugate research [18]. However, earlier discovery of their striking toxicity (eg, snowdrop lectin) to sap-sucking insects which has provided the hope for insect control [8]. In 2005 [19] initiated the cloning of zingiberaceae lectins by isolating the full-length cDNA of *Zingiber officinale* and this was the first report from this family. Different conventional procedures were initially tried but we could not able to get yield as the quality of the RNA, this might be due to the interference of secondary compounds present in the rhizome and seed. Finally, RNA was extracted using extraction kit and we could able to get quality product without DNA and protein contamination. Amount of RNA was considerably less but it is quite enough since 1µg is enough for further application. The integrity of the isolated RNA samples was analyzed by formaldehyde agarose (1%) gel electrophoresis shown in figure 1(A).

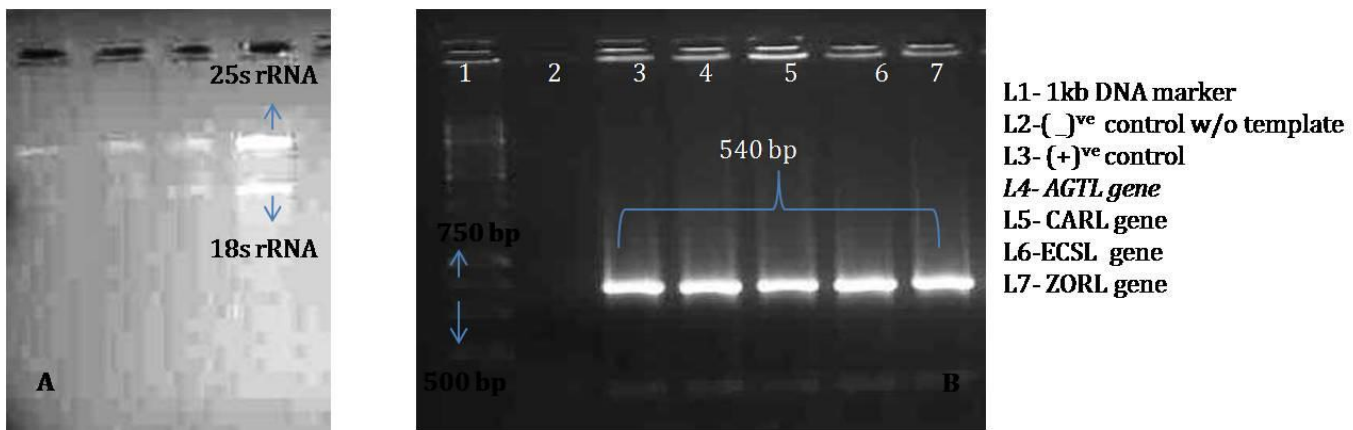


Figure -1. A: RNA isolation, B: PCR amplification of lectin gene

In present study first strand synthesis of cDNA was carried out with help of the primers (SMART IV and oligo 3’CDS) but the second step was carried out with *Bam*HI and *Hind*III primers figure 1(B) which is based on the conserved region of mannose binding lectin MMQDCNL. Resulted mannose sequences of *Alpinia galanga* tuber lectin (AGTL), *Curcuma amada* rhizome lectin (CARL), *Eletteria cardamom* seed lectin (ECSL) and *Zingiber officinale* rhizome lectin (ZORL) were submitted to NCBI database with following accession numbers KT313394, KT313395, KT313397 and KT313398 respectively. Garlic leaf lectin was used as positive control (Lane.3). No amplification was noticed in the negative control. Positive gene products were excised on the agarose gel and used to ligate into pTZ57R/T vector by T/A cloning. The recombinant molecule of pTZ57R/T containing gene was transferred to *E. coli* DH5α cells. Clones containing recombinant plasmids were selected by blue / white colony screening and white colonies were further subjected to colony PCR for confirmation. Colony PCR with universal M13 primers as well as gene specific primers showed amplification of the 740 and 540 bp respectively. Increased product size in colony PCR is due to additional base pairs of plasmids. Stable clones were taken for sequencing.

Lectins are usually considered as a heterogeneous group of proteins because of the apparent differences between individual lectins with respect to their molecular structure, sugar specificity, and biological activities. However, advances in the biochemistry, molecular cloning and structural analysis of lectins revealed the occurrence of seven families of structurally and evolutionarily related proteins [20]. Four of these families, namely the legume lectins, type 2 ribosome-inactivating proteins, chitin-binding lectins containing hevein domains and monocot mannose binding lectins are extended protein families and comprise most of the currently known lectins.

Two other families, namely the Cucurbitaceae phloem lectins and the amarantins are considered as small families as they comprise only a small number of lectins, which hitherto have been identified exclusively in Cucurbitaceae and *Amaranthus* sp., respectively. A third small lectin family, called the jacalin-related lectins, has originally been identified in Moraceae species like jack fruit (*Artocarpus integrifolia*) and other *Artocarpus* species, and osage orange (*Maclura pomifera*). All the *Artocarpus* lectins and the *Maclura pomifera* agglutinin have the same molecular structure, exhibit a very similar specificity and share a high sequence identity. Until recently, the so-called jacalin-related lectins had been found exclusively in a small taxonomic group, and therefore were considered as a minor lectin family. Hence, variation among the higher plant as well lack of molecular evolution sheds light to explore the lectin families.

Recombinant DNA technology involves the production of biologically important proteins in large when the availability of natural protein is limit. Heterologous protein expression is a handy way for obtaining the required protein in large quantity where the investment is a single gene instead of a large amount of plant material. A number of recombinant lectins have expressed using *Escherichia coli*, yeast, and *Pichia pastoris* as heterologous systems [21, 22]. Each of it has both advantage and dis-advantage in expressing the lectins since most of the lectins were glycosylated. Other than expression, the genes were simply analyzed for their characteristics. *Alocasia macrorrhiza* lectin could be an example in which the gene amplified through RACE and simple bioinformatics analysis was done [23]. As like in our present study we did not do tissue expression studies but sequence level characterization has done. Further, the sequences were translated *in-silico* into protein sequences and their three-dimensional structures were obtained through I-TASSER online structure modeling server.

Gene sequences of all our lectins were submitted to BLASTn and iterated (not shown). BLASTn results have shown matches with the previously reported mannose lectins. The maximum similarity was found with the other mannose-binding lectins from *Allium sativum*, *A. cepa* and *Annona squamosa*. Physiochemical characteristics of four lectin sequences were carried out with translated protein sequences *in-silico* using Expsy translate tool having a start and stop codon. The amino acid composition of the each translated lectin sequences were computed using the tool Protparam as shown in table-1. This tool also calculates different parameters viz, the theoretical molecular weight, theoretical pI, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Computed pI of AGTL – 4.76 is (pI<7) indicating its acidic nature, where as CARL-9.1 ECSL-8.84 and ZORL-7.6 are (pI>7) seems to be basic nature. Extension coefficient of lectins in M⁻¹ cm⁻¹ were seemed to be higher this might be due to the low percentage of trptophan and tyrosine residues. The range which is projected is assuming all pairs of Cys residues form cystines (eg. AGTL – 34045), and the other is assuming all Cys residues are reduced (AGTL-33920).

Table -1. The amino acid composition of lectins calculated by Protpram tool

Name of the proteins	Amino acid composition in %																			
	Alanine (A)	Arginine (R)	Asparagine (N)	Aspartic acid (D)	Cysteine (C)	Glutamine (Q)	Glutamic acid (E)	Glycine (G)	Histidine (H)	Isoleucine (I)	Leucine (L)	Lysine (K)	Methionine (M)	Phenylalanine (F)	Proline(P)	Serine (S)	Threonine (T)	Tryptophan (W)	Tyrosine (Y)	Valine (V)
AGTL	5.3 (8)	6 (9)	7.3 (11)	5.3 (8)	1.3 (2)	4 (6)	5.3 (8)	7.9 (12)	0	3.3 (5)	13.2 (20)	2 (3)	2.6 (4)	2 (3)	3.3 (5)	9.9 (15)	4.6 (7)	2.6 (4)	5.3 (8)	8.6 (13)

CARL	9.4 (17)	5.5 (10)	7.7 (14)	3.9 (7)	1.7 (3)	3.3 (6)	2.2 (3)	12.2 (22)	0	6.1 (11)	5 (9)	2.2 (4)	4.4 (8)	1.1 (2)	2.8 (5)	6.6 (12)	7.2 (13)	1.7 (3)	4.4 (8)	13.3 (24)
ECSL	9.4 (17)	5.5 (10)	7.7 (14)	3.9 (7)	1.7 (3)	3.3 (6)	2.2 (4)	11.6 (21)	0	5 (9)	6.1 (11)	2.2 (4)	4.4 (8)	1.7 (3)	2.8 (5)	7.2 (13)	7.2 (13)	1.7 (3)	4.4 (8)	12.2 (22)
ZORL	8.8 (17)	5.2 (10)	7.2 (14)	7.2 (9)	4.6 (4)	3.1 (6)	2.6 (5)	11.3 (22)	0	5.2 (10)	6.2 (12)	2.6 (5)	4.6 (9)	1 (2)	2.6 (5)	7.7 (15)	7.2 (14)	1.5 (3)	4.6 (9)	11.9 (23)

In our predictions, all four lectins have comparatively higher percent of aliphatic amino acids (alanine, valine, isoleucine, and leucine) than others (Table-1&2). The very high aliphatic index of all of our lectins infers that these proteins may be stable for a wide range of temperature. Instability index of all four lectins are smaller than 40 is predicted as stable. Table- 5 provides the list of predicted net hydrophobic residues in the lectins and can be taken as a positive factor for the increased thermo stability of the proteins [24]. All the proteins were found to be deficient in amino acid cysteine, and therefore lack the presence of disulfide linkages as also inferred from analysis of cys_rec result (http://www.softberry.com/berry.phtml?topic=cys_rec). In the absence of disulfide bond, extensive hydrogen bonding is believed to be responsible for the stability of these proteins. When expressing in percentage ZORL having 4.6% of Cysteine(C) residues than other three lectins and they are predicted to have di-sulfide bonds (Table-3). A positive score is predicted to be most probable positions of disulfide bonds. The very low GRAVY index of lectins infers that these proteins could result in a better interaction with water.

Table-2. Physicochemical parameters of four lectins

Name of the proteins	No of amino acids	pI	-R	+R	Extension Co-efficient	Instability index	Aliphatic index	GRAVY
AGTL	151	4.76	16	12	34045-33920	30.44	94.83	-0.152
CARL	181	9.1	10	14	28545-28420	17.23	90.94	0.161
ECSL	181	8.84	11	14	28545-28420	20.24	87.73	0.101
ZORL	194	7.6	14	15	30160-29910	20.81	87.37	0.072

In order to find homologous matches with the other proteins sequences, BlastP [25] search was done at NCBI. Most of them were annotated as mannose-binding lectins and proteins from various plants families including our own sequences. The first 15 (including the same sequence itself) hits with the highest similarity, which was all annotated as mannose-binding lectins from Alliaceae family. The conserved domain analysis performed reported that it belongs to B lectin super family (Figure-2). The secondary structure of our lectins was predicted by SOPMA methods of analysis program NPS@release 3.0 and the conclusion was got by secondary structure consensus prediction (Table-4).

Table-3. Presence of disulfide (ss) bond as predicted by Cys_Rec

Name of the proteins	CYS REC	Score
AGTL	Cys_31, Cys_55, Cys_26	-13.8, -38.3, -11.1
CARL	Cys_56, Cys_80	-21.9, -46.4
ECSL	Cys_28, Cys_59, Cys_83	-27.0, -28.1, -39.2
ZORL	Cys_35, Cys_60, Cys_90	-30.5, -24.7, 46.4

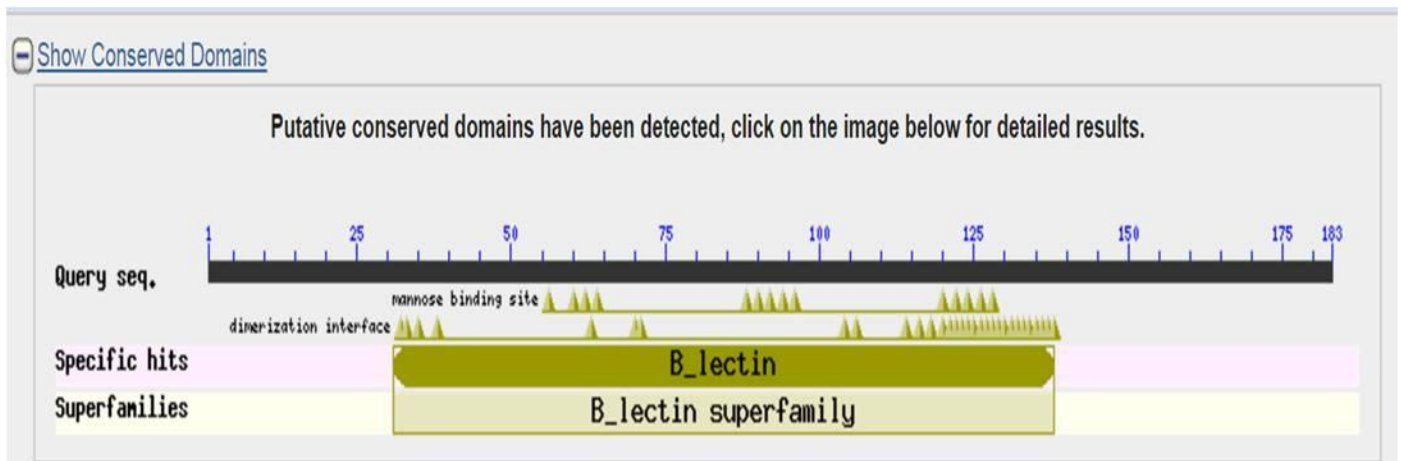


Figure-2. Conserved domain of Zingiberous lectin sequences

Table-4. Predicted secondary structures of lectins by SOPMA prediction method

Name of the proteins	Alpha helix (%)	3 ₁₀ helix (%)	Pi helix (%)	Beta bridge (%)	Extended strand (%)	Beta turn (%)	Random coil (%)	Other states (%)
AGTL	27.66	0	0	0	29.79	14.18	28.37	0
CARL	21.91	0	0	0	33.71	17.42	26.97	0
ECSL	21.35	0	0	0	29.78	16.85	32.02	0
ZORL	20.21	0	0	0	34.57	17.55	27.66	0

Table -5.Hydrophilic and hydrophobic residues content

Name of the proteins	Hydrophobic residues (%)	Hydrophilic residues (%)	Net hydrophobic residues content
AGTL	45.5	21.2	High
CARL	48.9	21.2	High
ECSL	52.1	22.6	Very high
ZORL	50.5	22	Very high

Hydrophobic cluster analysis (HCA) does not require powerful computer resources and can deal with distantly related proteins, even if no 3D data are available. In order to define hydrophobic clusters I, L, F, W, M, Y, V were considered as hydrophobic amino acids, whereas P was primarily considered as a breaker of these clusters, and A and C as mimetic, i.e. hydrophobic only in a hydrophobic environment [26]. Moreover, the hydrophobic residues were encircled, different symbols used for prolines (★), glycines (◆) which are often present in loops, and Cysteines (C) which may be involved in disulfide bonds. Additionally, Threonine (□) and Serine (▣) residues were also denoted. To simplify the use of this method HCA plot which plots the sequences on this pattern was written in Fortran 77, using color codes for amino acids.

HCA plots for all four lectins were presented in the figures (3-6). Plots of all four lectins share different shapes of clusters to infer about their corresponding secondary structures. Initial residues of our sequences (10-25 residues) are less or devoid of following amino acids(DENQHKR) is an indicator of completely buried secondary structures: mostly transmembrane helices (when the length is -20 residues) and more hardly, internal helices or long internal β-strands. The hydrophobic center of the four lectins namely *Alpinia*, mango ginger, cardamom, and ginger were centered around residue VYILM is long β- strand which is the center of the hydrophobic core of each protein. Additionally short, vertical hydrophobic dusters with very hydrophobic

residues throughout sequences and that are clearly separated from clusters by clear hydrophilic areas (loops) are another highly indicative of internal β strands.

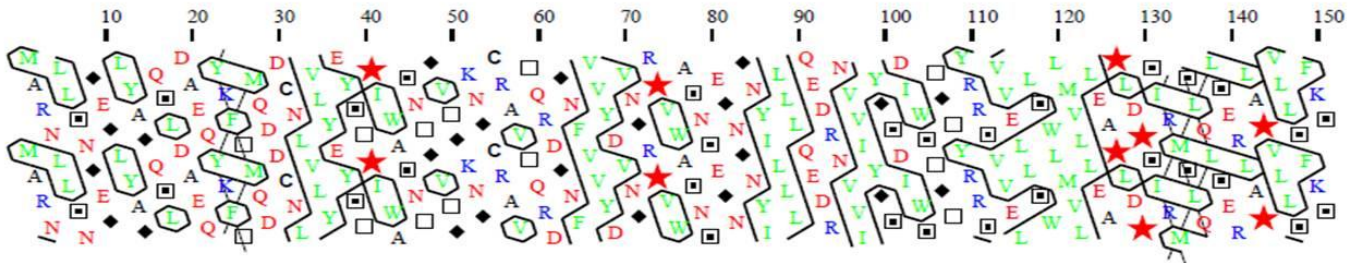


Figure.-3. Predicted HCA plot of *Alpinia galanga* lectin

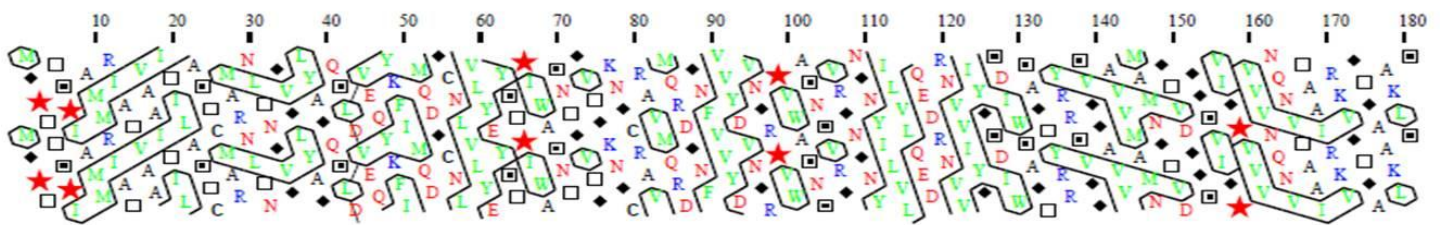


Figure-4. Predicted HCA plot of *Curcuma amada* lectin

Longer, horizontal clusters that are well separated by clear loops often denote amphiphilic α -helices. Mosaic (zig-zag) clusters that contain highly hydrophobic residues are often associated with β -strands particularly edge ones or an extended conformation they may not necessarily involved in a β -sheet. A clustering of alanine residues is sometimes observed on HCA plots. These clusters often correspond to α -helices.

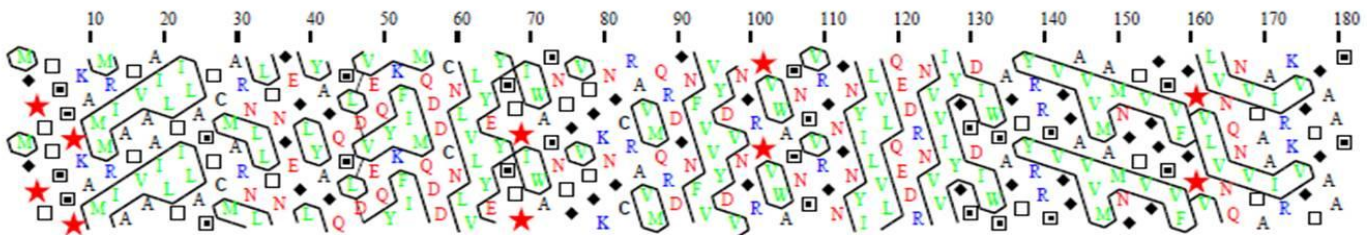


Figure-5. Predicted HCA plot of *Elettaria cardamomum* lectin

HCA plots are 2D-helical diagrams of protein sequences are routinely created by computer programs [27] which produce compact, monochrome or colored plots. Most of these programs feature automatic cluster contouring and division into structural segments at prolines. In our plots, amino acids were colored using green, black, blue and red for rapid identification. VILFWMY are drawn in green, ACGTS in black, P, D, E, N, Q in red and K, R in blue. The colored diagrams help in the rapid identification of deeply buried areas such as transmembrane helices where most of the residues appear in green or black, with the exception of proline (red). Due to their hydrophobic versatility, residues A, C, G, T and S can be interactively assigned to a hydrophobic cluster by comparison of with homologous proteins [28]. Proline, glycine, serine, and threonine are represented in special symbols (\star), (\blacklozenge), (\square) and (\square). In 1986 [29] reported the special characters of above amino acids in their review. According to that proline introduces the, largest constraints in the polypeptide chain and considered to be a systematic break in the clusters. Indeed, prolines often stop or distort helices and strands. Glycine has a large conformational flexibility. Serine and threonine are frequently encountered in loops but are also found in hydrophobic environments such as α -helices, where their hydroxyl group loses its hydrophilicity through hydrogen bonding with the carbonyl group of the main chain. In hydrophobic β -strands, threonine can sometimes replace more hydrophobic residues, probably due to the methyl group of its side-chain. Most of the lectin cDNAs which are cloned from plants were analyzed using HCA plots; *Calystegia* [30] and leguminosae lectins [31] for their structural conservation with another lectin.

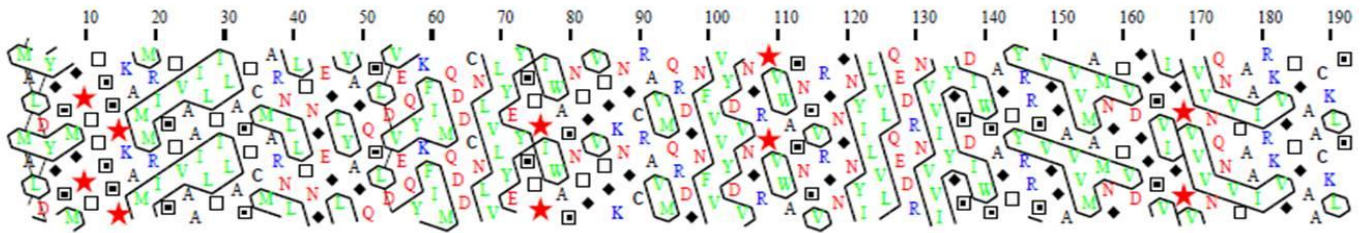


Figure-6. Predicted HCA plot of *Zingiber officinale* lectin

Since most of the identical hits (BlastP) were seen from Alliaceae members and also to know the relationship with other related sequences, multiple sequence alignment was built out with 13 MMBLs including AAQ55289 (*Typhonium divaricatum*-Araceae), AAA33346 (*Galanthus nivalis* - Amaryllidaceae), AAP04617 (*Amorphophallus konjac*- Araceae), AAQ18904 (*Amaryllia minuta* -Amaryllidaceae), AAB64237 (*Allium sativam*- Alliaceae), Q38759 (*Allium ampeloprosom* -Alliaceae), AAM28277 (*Ananas comosus* - Bromeliaceae), Q39728 (*Epipactis helleborine*- Orchidaceae),AAG10402 (*Crocus vernus* - Iridaceae),S62647(*Tulipa* hybridcultivar- Liliaceae);BAD67183 (*Dioscorea polystachya*- Dioscoreaceae), ALP82450 (*Alpinia galanga*- Zingiberaceae), ALP82451 (*Elettaria cardamomum*- Zingiberaceae), ALP82453 (*Curcuma amada*- Zingiberaceae), ALP82454 (*Zingiber officinale* -Zingiberaceae), ABF72850 (*Annona squamosa*- Annonaceae), ACA96191 (*Zingiber officinale* rhizome -Zingiberaceae) using clustalW revealed that our purified lectins are having conserved domains.

The consensus sequence motif **QXDXNXVXY** (where Q=Gln, D=Asp, N=Asn, V=Val, Y=Tyr, X=any another amino acid) is involved in alpha-D-mannose recognition. In addition to alignment analysis of our sequences with other 13 MMBLs from six different families showed that all the seven family lectins shared three common conserved binding sites (QDNVY) (not shown). These conserved domains conferred on them the capacity of specifically binding to carbohydrates, bioactivities, and functions as plant defense weapons. Hence, AGTL, CARL, ECSL, and ZORL are a member of the MMBL superfamily. Lectin sequences from all seven families were different in the amino acids length but they share similarity they may vary in their post-translation modification.

Phylogenetic tree of zingiberous lectin sequences was constructed along with amino acid sequences of other monocot families using MEGA7. Maximum Likelihood statistical method was used to draw a tree with bootstrap test and they have shown as tradition style respectively to visualize their relatedness easily (Figure -7). Zingiberous sequences get clustered with *Allium sativam* and *Annona squamosa* than other family lectins. This is correlating with BlastP iteration where our lectins are having similarity with *A. sativam* (96%), *A. ampeloprosom* (89%), other *allium* species (71- 85%), *A. squamosa* (96%). Other family lectins are seemed to be distantly related. *Dioscorea polystachya*, *Ananas comosus*, *Epipactis helleborine*, *Galanthus nivalis*, *Amaryllia minuta*, *Amorphophallus konjac* and *Typhonium divaricatum* are clustered in one group with the similarity of 52-65%. *Tulipa* hybrid cultivar and *Crocus vernus* is the other group with similarity 49%. By above functional characterization *Alpinia*, mango ginger, cardamom and ginger protein sequences were found to have many characters commonly possessed by mannose binding lectins. Phylogenetic analysis of lectins reveals they belong to the extended superfamily.

Bacterial symbionts play a prominent role in insect nutritional ecology by aiding in digestion of food or providing nutrients that are limited or lacking in the diet. Thereby, endo-symbionts open niches to their insect host that would otherwise be unavailable. The nutritional role of the primary (obligate) symbionts (*Buchnera aphidicola*) and the defensive role of various secondary (facultative) symbionts for their hosts have been particularly well studied in phloem-feeding aphids [32]. Thus, aphids have become a model system to study insect-bacteria interactions. Endosymbiosis is common in insects with more than 10% of insect species relying upon intracellular bacteria for their development and survival [33]. The symbiotic bacteria of aphids, *Buchnera aphidicola* live within large polyploid cells called bacteriocytes that are grouped into organ-like structures called bacteriomes located adjacent to the ovarioles. These are gram-negative, spherical or slightly oval cells which have not been cultivated outside aphid hosts. It has been demonstrated that *Buchnera* symbionts synthesize essential amino acids and other nutrients for their host aphids [34] and that removal of *Buchnera* via antibiotic or heat treatment results in retarded growth, sterility, and/or death in the host insect [35]. Due to their prevalence and importance to homopteran insects, *Buchnera* spp. are often referred to as primary symbionts. Experimental studies have provided some evidence that the nonessential amino acid glutamate is the major nitrogenous compound exported from the aphid and used by *Buchnera* for production of the essential, limiting amino acids [36, 37].

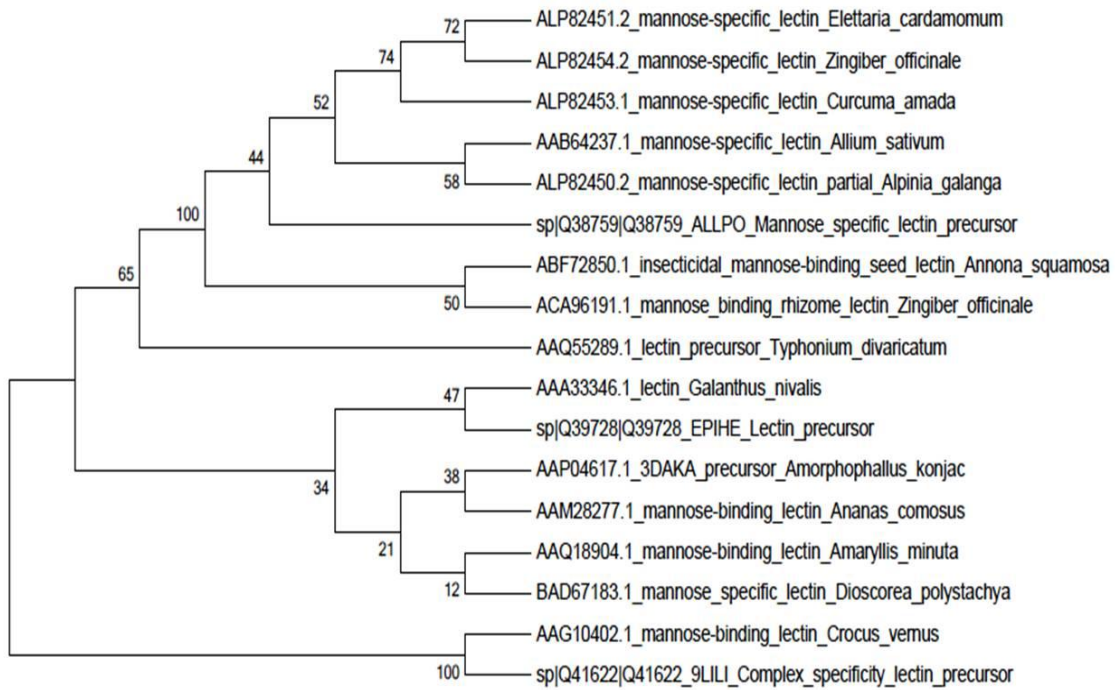


Figure -7. Phylogenetic tree of Maximum Likelihood - Bootstrapped method

In the present study, an attempt was made to identify the endosymbiont if it is present in the RSM pest. Aliquots of approximately hundred adult mites were collected from infected leaves in a saline solution and genomic DNA was isolated using modified CTAB method. By using this DNA as template, with help of universal eubacterial primer pairs (FD1 and RP1) that target the bacteria 16S rRNA gene was amplified [17]. The annealing temperatures of 55 °C for 30 sec with 30 cycles were used for the amplification. Resulted PCR amplicon was sequenced without further cloning. Sequenced gene was about 1504 bp and it was submitted to NCBI with accession number KX390632. Phylogenetic relationships were established between *Buchnera* sp. (*Oligonychus coffeae* N.) 16S rRNA gene, sequence and those of *Buchnera* sp. previously identified (figure -8).

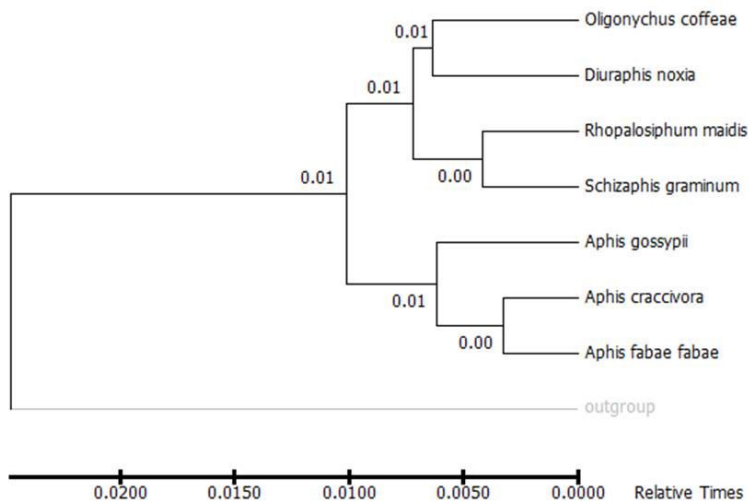


Figure -8. Phylogenetic tree constructed using the Neighbor-joining method. *Toxoptera aurantii* taken as out group

The reason to do this symbiont identification in Red Spider Mite (RSM) is to take it as one of the pest control measures. Through molecular identification, we could be able to identify gene fragment which has similarity with already available primary endosymbiont of aphids in the NCBI database. But they are sporadic or fewer data can be projected with supporting information on the presence of *Buchnera* species in mites. When comparing insects and mites, mites differ from insects in that they have only two main body parts: combined head and thorax section, and an abdomen. However, mites and insects are depending on the plants for their food supplements may believe to have similar abdominal behavior. Further, analysis of red spider mite internal morphology as well as characterization of their midgut through available experimental techniques might pave ways to prove our prediction of endosymbiont *Buchnera* sp.

Binding of toxins to receptors in the insect gut is an important step in toxicity. This also determines the specificity of a toxin to the insect. Garlic lectins bind to several proteins in the midgut of insects through glycan-mediated interactions [22]. Glycan binding ability of lectins has been established by several methods like inhibition of agglutination, enzyme-linked lectin adsorbent assay, surface plasmon resonance (SPR), isothermal titration calorimetry and glycan array [38, 39].

Translated protein sequences of all four lectins, as well as *Buchnera* sp., were submitted to I-TASSER a protein structure modeling server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>). Targeted sequences are first threaded through a representative PDB structure library with a pair-wise sequence identity cut-off of 70% to search for the possible folds by four simple variants of PPA methods, with different combinations of the hidden Markov model, PSIBLAST, profiles, the Needleman-Wunsch and Smith-Waterman alignment algorithms. The continuous fragments are then excised from the threading aligned regions which are used to reassemble full-length models while the threading unaligned regions (mainly loops) are built by *ab initio* modeling. The conformational space is searched by replica-exchange Monte Carlo simulations. The structure trajectories are clustered by SPICKER and the cluster centroids are obtained by the averaging the coordinates of all clustered structures.

To rule out the steric clashes on the centroid structures and to refine the models further, the fragment assembly simulation again, this starts from the cluster centroid of the first round simulation. Spatial restraints are extracted from the centroids and the PDB structures searched by the structure alignment program TM-align, which are used to guide the second round simulation. Finally, the structure decoys are clustered and the lowest energy structure in each cluster is selected, which has the C α atoms and the side-chain centers of mass specified [40].

For all four lectins the modeled structures were selected based on the C-score which is calculated based on the significance of threading template alignments and the convergence parameter r_s of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. The selected models were satisfied this condition, all models were resided in between the C-score range notably *Alpinia galanga* lectin (-2.02), *Curcuma amada* lectin (-2.90), *Elettaria cardamom* (-2.80) and *Zingiber officinale* lectin (-2.80). TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these qualities. The following figures -9 (A,a); (B,b); (C,c) and (D,d) showing the predicted 3D structures along with their ligand binding site of *Alpinia*, mango ginger, cardamom, and ginger respectively. Resulted models were validated for their quality and their stereo-chemical properties. The structures were submitted to PMDB database. Three-dimensional co-ordinates the lectins were selected from the PDB database based on the RMSD and TM score. The structures of the proteins *Alpinia*, mango ginger cardamom, ginger and were obtained from the PDB with the following PDB ids 1MSA, 3DZW, 1NPL and 1BWU respectively.

The docked bacterial receptor with four lectins and their complex is shown in figure -10A-D. The docked model, when analyzed critically, showed the interaction between the mannose molecules of the receptor with respective lectin residues through hydrogen bond were shown in tables(6-9). Schematic representation of residues interaction is depicted in the figure -10 a-d. Residues of each lectins interacting here satisfies the mannose binding domain ability with the bacterial receptor. Surface accessibility calculations done on this protein complex model determined the reduced surface accessibility in comparison to the free receptor molecule. The reduced accessibility was envisaged in the residues of the receptor that are responsible for the possible interaction with other potential molecules depicted in figure- 11 (a-d). In other words, the binding of individual lectins to the receptor covers up a certain area of the surface of the receptor, making that area sterically inaccessible to solvents or other molecules.

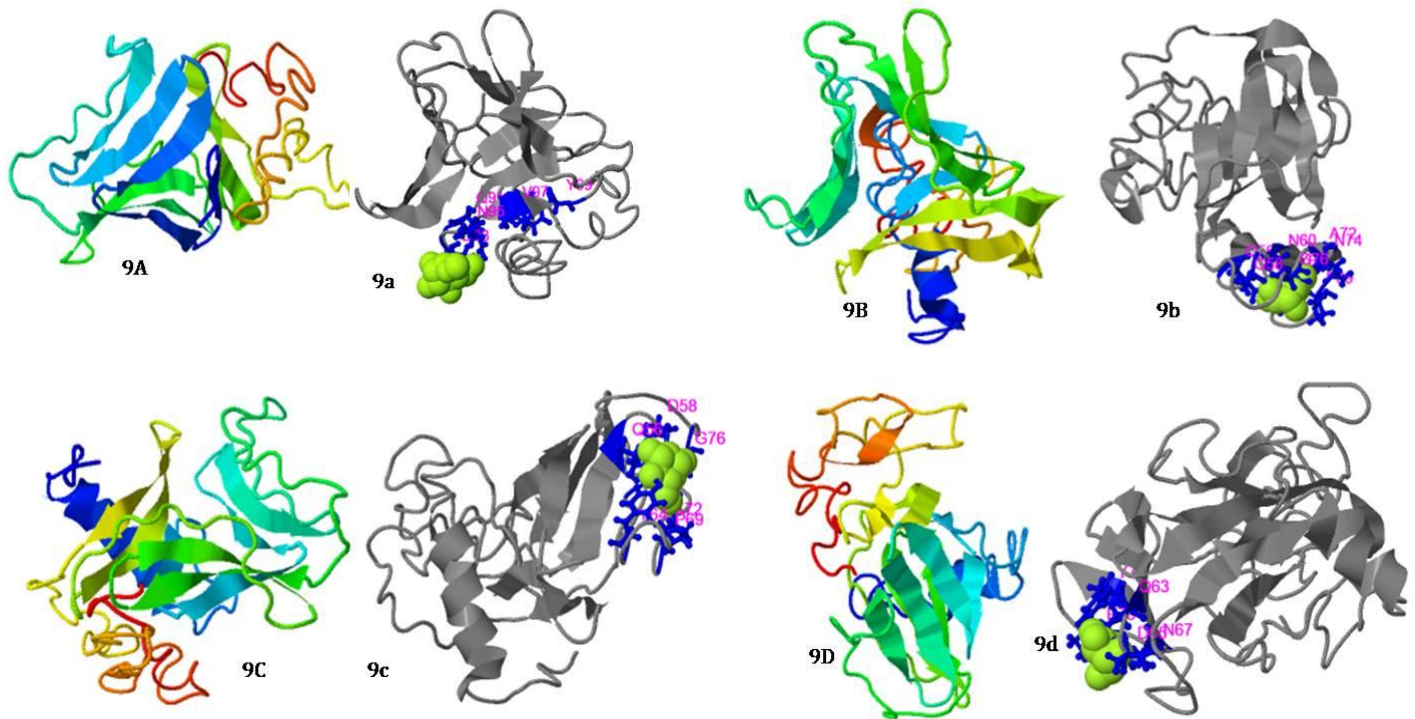


Figure -9. 3D structures along with their ligand bite site 9A & a- AGTL, 9B &b -CARL, 9C &c- ECSL and 9D &d- ZORL.

Table -6. The interaction of the mannose residues of oligosaccharide moiety with the receptor residues and the *Alpinia* residues

Mannose atoms	Receptor residues	Alpinia residues	Hydrogen bond length in (Å)
Man112	Gly211	-	1.6
Man113	Lys498	Arg92	
Man114	Ile175	-	

In the schematic representation, the oligosaccharide chain is depicted in the red colored stick conformation. The receptor is shown as a schematic and the green colored stick models are the residues interacting with mannose. Lectin is depicted as a schematic and its interacting residues are shown in cyan sticks. The hydrogen bond interactions are shown in yellow colored lined dots. The chaperonin is represented in green colored cartoon model whereas the lectins are represented as cyan colored cartoon model.

Interaction of garlic leaf lectin (ASAL) with endosymbiotic chaperonin (symbionin of *Buchnera aphidicola*) in *Lipaphis erysimi* was documented [41]. Lectins interact with the insect's intestinal epithelial cells or peritrophic membrane and/or glycosylated digestive enzymes. However, the exact mechanism of insecticidal action is unknown. Lectins exhibiting similar glycan specificity; might have different effects on closely related insects [7]. Garlic lectins show the glycan-mediated interaction with several putative receptors in the midgut of insects. Interactions with the digestive and molting enzymes are also carbohydrate mediated [22]. This shows that carbohydrate affinity of lectin plays a critical role not only other known biological functions but also in the insect specificity and insecticidal activity.

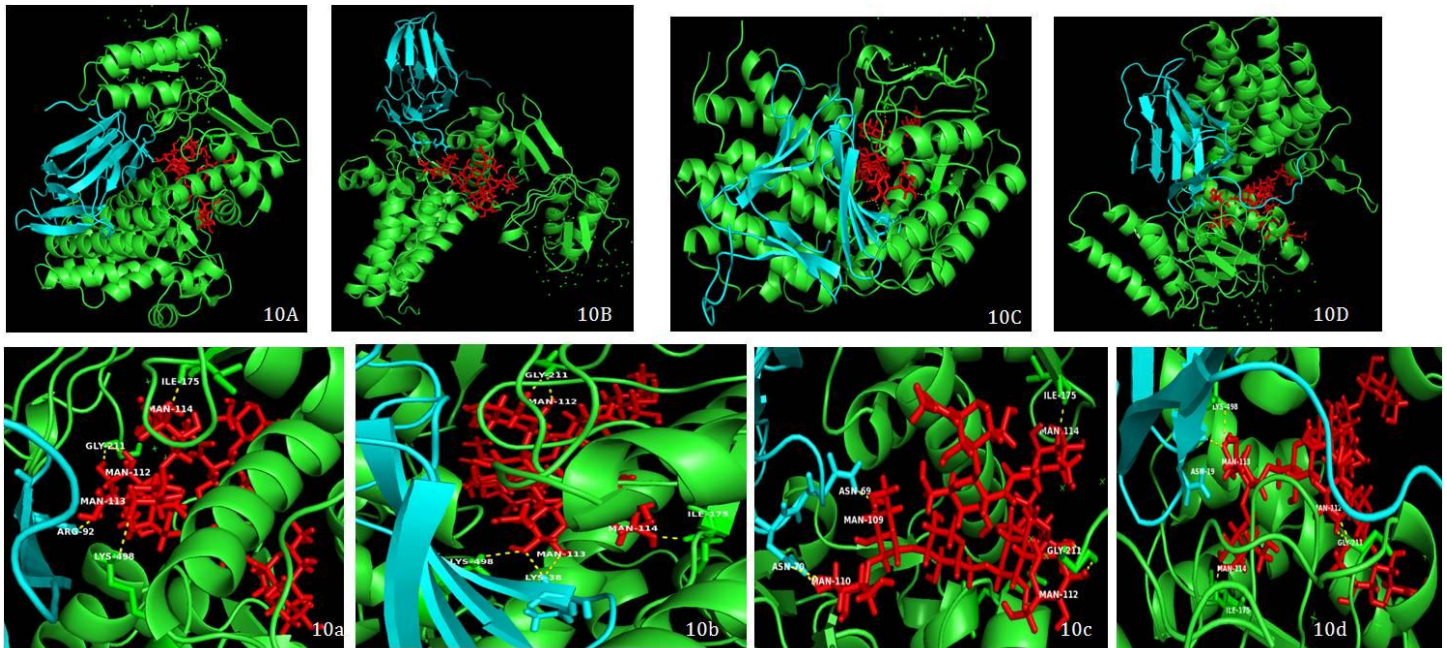


Figure-10. (A-D) Complex of Bacterial chaperonin (green) docked with mango ginger (cyan) through the oligosaccharide moiety (red) and (a-d) Molecular interactions of proteins 10A&a - AGTL, 10B &b - CARL, 10C &c - ECSL and 10D &d - ZORL.

Table -7.The interaction of the mannose residues of oligosaccharide moiety with the receptor residues and the mango ginger residues

Mannose atoms	Chaperonin residues	Mangoginger residues	Hydrogen bond length in (Å)
Man109	-	Asn69	1.7
Man110	-	Asn70	1.6, 2.6
Man112	Gly211	-	
Man113	Lys498	-	
Man114	Ile175	-	

Table- 8.The interaction of the mannose residues of oligosaccharide moiety with the receptor residues and the cardamom residues

Mannose atoms	Chaperonin residues	Cardamom residues	Hydrogen bond length in (Å)
Man112	Gly211	-	
Man113	Lys498	Lys 38	2.5, 2.8
Man114	Ile175	-	

Table -9.The interaction of the mannose residues of oligosaccharide moiety with the receptor residues and the ginger residues

Mannose atoms	Chaperonin residues	Ginger residues	Hydrogen bond length in (Å)
Man 112	Gly 211	-	
Man 113	Lys 498	Asn 19	2.3
Man 114	Ile 175	-	

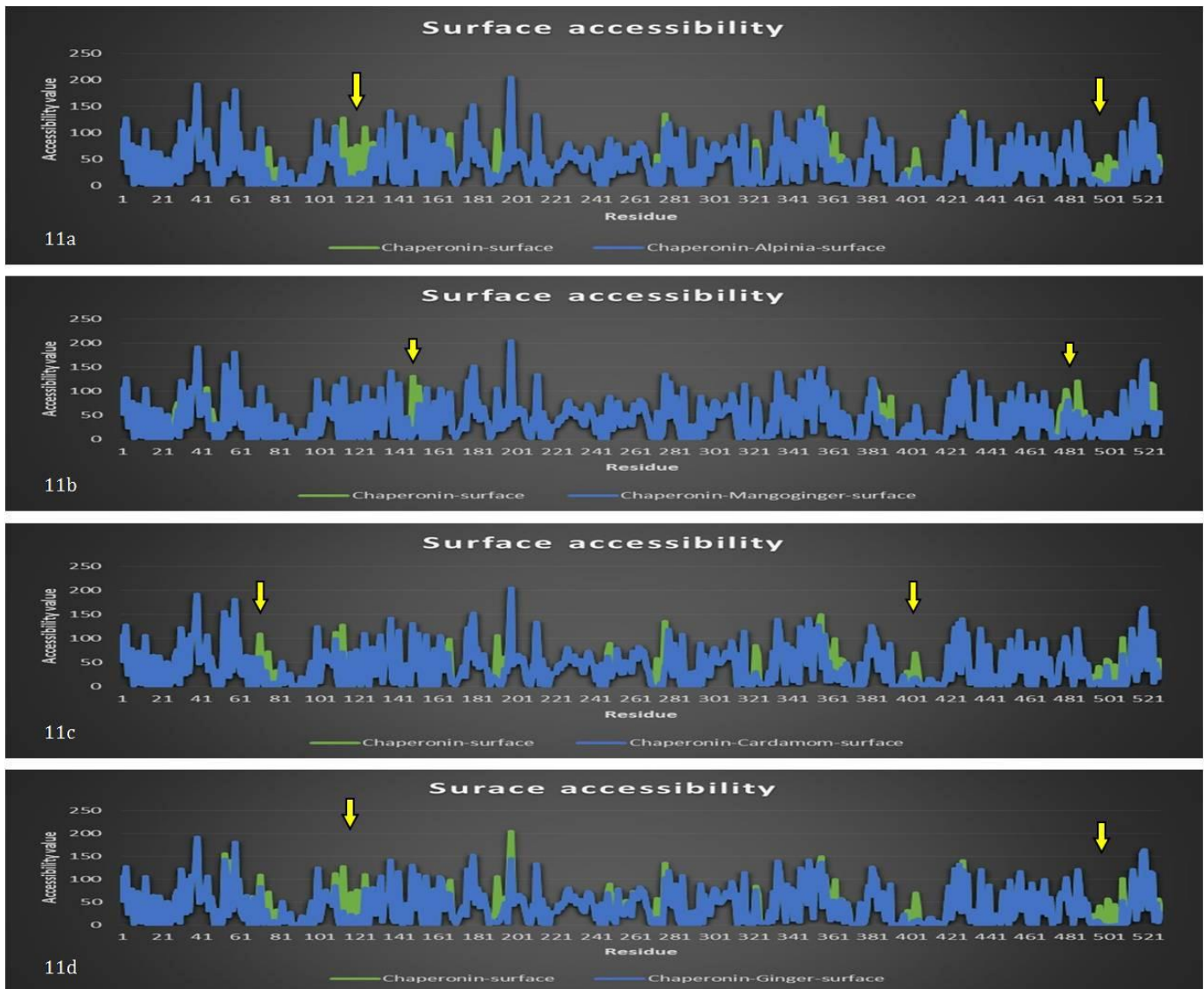


Figure - 11. Changes in the molecular surface accessibility of Chaperonin. Surface accessibility before (green) and after (blue) binding of lectin. The arrows show the region of maximum accessibility variations a-AGTL, b-CARL, c-ECSL and d-ZORL.

Study on the mechanism of insecticidal action of lectins was initiated nearly 25 years ago. In 1984, [39] proposed that lectins inhibit the nutrient absorption cause severe disruption of the epithelial cells in midgut by stimulating endocytosis. This results in disorganization and elongation of the striated brush border microvilli which leads to the swelling of epithelial cells and complete closure of the gut lumen. [42] reported the transportation of lectin across the midgut epithelial barrier after binding to the midgut glycoproteins and disruption in the microvilli of the striated brush border region. [43] reported the interaction of a lectin with glycosylated receptors on the surface of stomach epithelial cells which cause interference in normal metabolism and cell functions leading to the rapid feedback response on feeding behavior. These results suggest that each lectin has a different mode of action for different insects the cellular level. In lepidopteran insects, the garlic lectins interact with several glycosylated BBMVs proteins like Cadherin, APN, ALP, polycalins, and others in the midgut peritrophic membrane. Since most of them are reported as a receptor for β -endotoxins, there is a possibility that a few or all functions as a receptor for garlic lectins also. Binding of lectins to the midgut proteins affect the nutrient absorption. This might be due to the inhibition of digestive and other related enzymes. Further, garlic lectins are resistant to the midgut proteases and transported

to hemolymph. Where these lectins interact with cytochrome P450 and interfere with larval development. This results in severe growth retardation and premature death [22].

Not only garlic lectins, lectins from different plant families were also studied for their effectiveness against aphids, for instance, in 2005 [44] tested *Arum maculatum* tuber lectin of Araceae against two economically important sucking pests, *Lipaphis erysimi* and *Aphis craccivora*. They found two major receptor proteins of ATL (~40 kDa and ~35 kDa) from the brush border membrane vesicle (BBMV) of aphid guts. Due to the edible source of origin garlic lectins are supposed to be safe. The insecticidal action of garlic lectins is a multi-step process that involves the interaction with several glycosylated receptor proteins midgut of insects. This results in inhibition of nutrient absorption leading to mortality. Garlic lectins get accumulated into the hemolymph and ovarioles and interfere in development and reproduction.

4. CONCLUSION

All the effort on lectins characterization regards to their functions as well gene level analysis is to take them as a promising tool against pest and disease of crops. Earlier investigators have examined a number of bulb lectins *Allium* sp. and GNA have common features regarding their mannose-binding affinity, amino acid sequences, serological interactions and structural similarity which allow them to be categorized into a super-family of mannose-binding monocot lectins [45, 46, 47, 48]. The choice of a lectin to be used for crop protection must also depend on its non-toxicity towards human being and animals and definitely on the working mechanism. Although Zingiberous lectins have not explored on tea pest and disease control, detailed characterization of their properties and similarity with already available potential lectins will lead them to be considered in this concern. Moreover, the additional research on binding mechanism as well the expression level of our lectins through available experimental methods helps to consider them as the transgene for the control of different pests/pathogens in integrated pest management (IPM).

5. ACKNOWLEDGEMENT

We thank National Tea Research Foundation, Kolkata, India for funding the project on 'Lectin: an alternative biotic stress management system in tea' under which this work was carried out. Thanks are due to Dr. B. Radhakrishnan, Director for allowing us to carry out all the experiments. We extend our thanks to Dr. K. N. Chandrashekhara, former Senior Biotechnologist, Plant Physiology Division, for his support in designing the experiments as well his support in preparing the manuscript. We thank our laboratory colleagues for their kind support.

6. REFERENCES

- [1] J. Nietner, Observations on the enemies of the coffee tree in Ceylon, . Ceylon,Colombo, 1861. p. 31.
- [2] B. Radhakrishnan, Seasonal abundance of red spider mite, *Oligonychus coffeae* (Nietner) and the crop loss caused by it in tea, J. Plant .Crops, vol. **32** (Suppl), Jan2004, pp.354-356.
- [3] Babu and N. Muraleedharan, A note on the use of pesticide for the control of insect and mite pests of tea in south India. UPASI Tea Research Institute, 2010, pp23.
- [4] J. Ferre and J. Van Rie, Biochemistry and genetics of insect resistance to *Bacillus thuringiensis*, Annu. Rev. Entomol., vol. **47**, Mar2002, pp.501-533, doi:annurev.ento.47.091201.145234.
- [5] A.F. Janmaat and J. Myers, Rapid evolution and the cost of resistance to *Bacillus thuringiensis* in greenhouse populations of cabbage loopers, *Trichoplusia ni*, Proc Biol Sci., vol. **270**, Nov2003, pp. 2263-2270, DOI: 10.1098/rspb.2003.2497.
- [6] C.M. Smith and L.V. Boyko, The molecular bases of plant resistance and defense responses to aphid feeding; current status, Entomol Exp Appl., Vol. **122**(1), Dec2006, pp.1-16, doi: j.1570-7458.2006.00503.x.
- [7] C.R. Carlini and M.F. Grossi-de-Sa, Plant toxic proteins with insecticidal properties; a review on their potentialities as bio-insecticides, Toxicon, vol.**40** (11), Nov2002, pp.1515-1539, doi: 10.1016/s0041-0101(02)00240-4.
- [8] V.A Hilder, K.S Powell, A.M.R. Gatehouse, J.A. Gatehouse, L.N. Gatehouse, Y. Shi, W.D.O. Hamilton, A. Merryweather, C. Newell, J.C. Timans, W.J. Peumans, E.J.M Van Damme and D. Boulter, Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids, Transgenic Res., vol **4**, Jan1995,pp 18-25, doi: A:1008856703464.
- [9] E.J.M Van Damme, K. Smeets, F. Van Leuven, W.J. Peumans, Molecular cloning of mannose binding lectins from *Clivia miniata*, Plant Mol. Biol., vol **24**, Oct1994c, pp. 825-830,doi: https://doi.org/10.1007/BF00029865.

- [10] K. Smeets, E.J.M. Van Damme, W.J. Peumans, Comparative study of the post-translational processing of the mannose-binding lectins in the bulbs of garlic (*Allium sativum* L) and ramsons (*Allium ursinum* L), *Glycoconjugate J.*, vol.11, Nov1994, pp.309-320, doi:https://doi.org/10.1007/BF00731204.
- [11] K.S. Powell, J. Spance, M. Bharathi, J.A. Gatehouse, A.M.R. Gatehouse, Immunohistochemical and development studies to elucidate the mechanism of action of the snowdrop lectin on the rice brown planthopper, *Nilaparvata lugens* (Stal), *J Insect Physiol.*, vol.67, Jul1998, pp.529-539, doi: 10.1016/S0022-1910(98)00054-7.
- [12] K.S Powell, Antimetabolic effects of plant lectins towards nymphal stages of the planthoppers *Tarophagous proserpina* and *Nilaparvata lugens*, *Entomol Exp Appl.*, vol.99, Sep2001, pp. 71-77, https://doi.org/10.1046/j.1570-7458.2001.00803.x.
- [13] E. Stoger, S. Williams, P. Christou, R.E. Down and J.A. Gatehouse, Expression of the insecticidal lectin from snowdrop (*Galanthus nivalis* agglutinin, GNA) in transgenic wheat plants, effects predation by the grain aphids *Sitobion avenae*, *Mol Breed.*, vol. 5(1), Jan1999, pp. 65-73, doi:https://doi.org/10.1023/A:1009616413886.
- [14] X. Foissac, N.T. Loc, P. Christou, A.M.R. Gatehouse, and J.A. Gatehouse, Resistance to green leafhopper (*Nephotettix virescens*) and brown planthopper (*Nilaparvata lugens*) in transgenic rice expressing snowdrop lectin (*Galanthus nivalis* agglutinin, GNA), *J Insect Physiol.*, vol. 46(4), Apr2000, pp.573-583, doi:http://dx.doi.org/10.1016/S0022-1910(99)00143-2.
- [15] K. Michiels, E.J.M. Van Damme, G. Smagghe, Plant- insect interactions, what can we learn from plant lectins?, *Arch Insect Biochem Physiol.*, vol. 73, Apr2010, pp. 193-212, doi: 10.1002/arch.20351.
- [16] M.A. Saghai-Marroof, R.A. Jorgensen, R.W. Allard, Ribosomal DNA spacer-length polymorphisms in barley, Mendelian inheritance, chromosomal location and population dynamics, *Proc Natl Acad Sci U S A*, vol.81, Dec1984, pp. 8014-8018, https://doi.org/10.1073/pnas.81.24.8014.
- [17] W.G. Weisburg, S.M. Barns, D.A. Pelletier and D.J. Lane, 16S Ribosomal DNA Amplification for Phylogenetic Study, *J Bacteriol.*, vol.173(2), Oct1991, pp. 697-703, doi: 10.1128/jb.173.2.697-703.
- [18] Haselbeck, E. Schickaneder, H. Vondereltz and W. Hosel, Structural characterization of glycoprotein carbohydrate chains by using digoxigenin-labeled lectins on blots, *Anal. Biochem.*, vol.191, Nov1990, pp. 25-30, doi: https://doi.org/10.1016/0003-2697(90)90381-1.
- [19] Z. Chen, G. Kai, X. Liu, J. Lin, X. Sun and K. Tang, cDNA cloning and characterization of a mannose-binding lectin from *Zingiber officinale* Roscoe (ginger) rhizomes, *J Biosci.*, 30, Mar2005, pp. 213-220, doi:https://doi.org/10.1007/BF02703701.
- [20] E.J.M. Van Damme, W.J. Peumans, A. Pusztai and S. Bardocz, *Handbook of plant lectins, properties and biomedical applications*, Bognor Regis, UK, John Wiley and Sons, 1998,
- [21] N.P. Trung, E. Fitches and J.A. Gatehouse, A fusion protein containing a lepidopteran-specific toxin from the South Indian red scorpion (*Mesobuthus tamulus*) and snowdrop lectin shows oral toxicity to target insects, *BMC Biotechnol.*, vol. 6, Mar2006, pp.1810-1186, doi: 10.1186/1472-6750-6-18.
- [22] S.K. Upadhyay, S. Saurabh, P. Rai, R. Singh, K. Chandrashekar, P.C. Verma, P.K. Singh and R. Tuli, SUMO fusion facilitates expression and purification of garlic leaf lectin but modifies some of its properties, *J Biotechnol.*, vol.146, Mar.2010, pp. 1-8, doi: 10.1016/j.jbiotec.2010.01.013.
- [23] Y.R Zhu, J. Wang, B.Q. Huang, X.W. Hou, Molecular Cloning of a Lectin cDNA from *Alocasia macrorrhiza* and Prediction of Its Characteristics, *Plant Physiol Mol Biol.*, vol.32 (6), Dec 2006, pp. 634-642.
- [24] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel and A. Bairoch, ExPASy—the proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Res.*, vol.31, Jul2003, pp. 3784-3788, doi: 10.1093/nar/gkg563.
- [25] S.F. Altschul, W. Gish, W. Miller, E.W. Meyers and D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.*, vol. 215(3), Oct1990, pp. 403- 410, doi: 10.1016/S0022-2836(05)80360-2.
- [26] C. Gaboriaud, V. Bissery, T. Benchetrit and J.P. Mornon, Hydrophobic cluster analysis, an efficient new way to compare and analyse amino acid sequences, *FEBS Letters*, vol. 224(1), Nov1987, pp. 149-155, doi:https://doi.org/10.1016/0014-5793(87)80439-8.
- [27] B. Hennsdat, E. Raimbaud, V. Tran, J.P. Mornon, HCABAND, a computer program for the 2-D-helical representation of protein sequences, *Comput Appl Biosci.*, 6, Jan1990, pp.3-5, doi: 10.1093/bioinformatics/6.1.3.
- [28] K. Lemesle-Varloot, B. Henrissat, C. Gaboriaud, V. Bissery, A. Morgat and J.P. Mornon, Hydrophobic cluster analysis, procedures to derive structural and functional information from 2-D-representation of protein sequences, *Biochimie*, vol. 72(8), Aug 1990, pp. 555-574, doi: https://doi.org/10.1016/0300-9084(90)90120-6.
- [29] J.F Leszczynski and G.D Rose, Loops in globular proteins, a novel category of secondary structure, *Science*, vol. 244, Nov1986, pp. 849-855, doi: 10.1126/science.3775366.

- [30] E.J.M. Van Damme, A. Barre, P. Verhaert, P. Rouge, W.J. Peumans, Molecular cloning of mitogenic mannose/maltose-specific rhizome lectin from *Calystegia sepium*, *FEBS letters*, vol. **397**, Nov1996, pp.352-356, doi:https://doi.org/10.1016/S0014-5793(96)01211-2.
- [31] P. Rougé and L. Varloot, Structural homologies between leguminosae lectins as revealed by the hydrophobic cluster analysis (HCA) method, *Biochem Syst Ecol.*, vol.**18** (6), Oct1990, pp. 419-427, doi: 10.1016/0305-1978(90)90087-V.
- [32] K.M. Oliver, A.H. Smith and J.A. Russell, Defensive symbiosis in the real world – advancing ecological studies of heritable, protective bacteria in aphids and beyond, *Funct Ecol.*, vol. **28**, Jun2014, pp.341–355, doi:https://doi.org/10.1111/1365-2435.12133.
- [33] P. Baumann, N.A. Moran, L. Baumann and editors, Bacteriocyte-associated endosymbionts of insects, In; Dworkin M, editor, *The prokaryotes* [online], New York, Springer, 2000.
- [34] Douglas, Nutritional interactions in insect-microbial symbioses; aphids and their symbiotic bacteria *Buchnera*, *Annu Rev Entomol.*, vol. **43**, Jan1998, pp. 17–37, doi:https://doi.org/10.1146/annurev.ento.43.1.17.
- [35] Ohtaka and H. Ishikawa, Effects of heat treatment on the symbiotic system of an aphid mycetocyte, *Symbiosis*, vol.**11**, 1991, pp.19-30.
- [36] T. Sasaki and H. Ishikawa, Production of essential amino acids from glutamate by mycetocyte symbionts of the pea aphid, *Acyrtosiphon pisum*, *J. Insect Physiol.*, vol. **41**, Jan1995, pp. 41–46, doi:https://doi.org/10.1016/0022-1910(94)00080-Z.
- [37] L.F. Whitehead and A.E. Douglas, A metabolic study of *Buchnera*, the intracellular bacterial symbionts of the pea aphid *Acyrtosiphon pisum*, *J Gen Microbiol.*, vol. **139**, Apr1993, pp. 821–826, doi: 10.1099/00221287-139-4-821.
- [38] T.K. Dam, P.K. Bachhawat, G. Rani and A. Suriola, Garlic (*Allium sativum*) lectins bind to high mannose oligosaccharide chains, *J Biol Chem.*, vol. **273**, Mar1998, pp. 5528-5535, doi: 10.1074/jbc.273.10.5528.
- [39] K. Bachhawat, C.J. Thomas, B. Amutha, M.V. Krishnasastri, M.I. Khan and A. Surolia, On the stringent requirement of mannose substitution in mannooligosaccharides for the recognition by garlic (*Allium sativum*) lectin, *J Biol Chem.*, vol. **276**, Feb2001, pp. 5541–5546, doi: 10.1074/jbc.M009533200.
- [40] Y. Zhang, I-TASSER server for protein 3D structure prediction, *BMC bioinformatics*, vol. **40**(9), Jan2008, pp. 1-8, doi: https://doi.org/10.1186/1471-2105-9-40.
- [41] S. Banerjee, D. Hess, P. Majumder, D. Roy and S. Das, The Interactions of *Allium sativum* Leaf Agglutinin with a Chaperonin Group of Unique Receptor Protein Isolated from a Bacterial Endosymbiont of the Mustard Aphid, *The J Biol Chem.*, vol. **279**(22), Mar2004, pp.23782–23789, doi: 10.1074/jbc.M401405200.
- [42] A.M.R. Gatehouse, F.M. Dewey, J. Dove, K.A. Fenton and A. Pusztai, Effect of seed lectins from *Phaseolus vulgaris* on the development of larvae of *Callosobruchus maculatus*, mechanism of toxicity, *J Sci Food Agric.*, vol. **35**(4), Apr1984, pp. 373-380, doi: https://doi.org/10.1002/jsfa.2740350402.
- [43] N. Sauvion, C. Nardon, G. Febvay, A.M.R. Gatehouse, Y. Rahbe', Binding of the insecticidal lectin Concanavalin A in pea aphid, *Acyrtosiphon pisum* (Harris) and induced effects on the structure of midgut epithelial cells, *J Insect Physiol.*, vol. **50**(12), Dec2004, pp.1137- 1150, doi: 10.1016/j.jinsphys.2004.10.006.
- [44] P. Majumder, H.A. Mondal and S. Das, Insecticidal activity of *Arum maculatum* tuber lectin and its binding to the glycosylated insect gut receptors, *J. Agric. Food Chem.*, vol.**53**(17), Jul2005, 6725-6729, doi: 10.1021/jf051155z.
- [45] E.J.M. Van Damme, K. Smeets, S. Torrekens, F. Vanleuven, I.J. Goldstein and W.J. Peumans, The Closely Related Homomeric and Heterodimeric Mannose-Binding Lectins from Garlic Are Encoded by One-Domain and 2-Domain Lectin Genes, Respectively, *Eur J Biochem.*, vol.**206**, Jun1992, pp. 413-420, https://doi.org/10.1111/j.1432-1033.1992.tb16941.x.
- [46] K. Smeets, E.J.M. Van Damme, F. Van Leuven, W.J. Peumans, Isolation and characterization of lectins and lectin-alliinase complexes from bulbs of garlic (*Allium sativum*) and ramsons (*Allium ursinum*), *Glycoconjugate J.*, vol. **14** (3), Mar1997, pp.331–334, doi: 10.1023/a:1018570628180.
- [47] E.J.M. Van Damme, K. Smeets, I. Engelborghs, H. Aelbers, J. Balzarini, A. Pusztai, F. Van Leuven, I.J. Goldstein, W.J. Peumans, Cloning and characterization of the lectin cDNA clones from onion, shallot and leek, *Plant Mol Biol.*, vol.**23**, Oct1993,365–376, doi:https://doi.org/10.1007/BF00029011.
- [48] N.R. Chandra, G. Ramachandraiah, K. Bachhawat, T.K. Dam, A. Surolia and M. Vijayan, Crystal structure of a dimeric mannose-specific agglutinin from garlic; quaternary association and carbohydrate specificity, *J Mol Biol.*, vol. **285**, Jan1999, pp. 1157–1168, doi: 10.1006/jmbi.1998.2353.