CHAPTER 5

Purification and physiochemical characterization of lectin domain (BGAF_Lec) of β-glucosidase aggregating factor (BGAF) from Sorghum bicolor

5.1 Summary

A protein containing c-terminal lectin domain of BGAF from *Sorghum bicolor* was cloned, expressed in *E. coli* strain and purified using Ni-NTA affinity chromatography. The purified protein is named BGAF_Lec is a monomer of 18 kDa and gel filtration analysis shows that it exists as a monomer in solution. BGAF_Lec is stable in pH range 5-9 and despite being a recombinant protein is thermally stable up to 30 minutes at 50°C. The hemagglutination activity of BGAF_Lec is strongly inhibited by D-fructose and fetuin and weakly with mannose and N-Acetyl-D-galactosamine. Fluorescence spectroscopy showed intrinsic fluorescence of purified BGAF_Lec was quenched by different carbohydrates such as D-fructose, mannose, N-Acetyl-D-galactosamine, with D-fructose showing highest effect with a dissociation constant value Kd of $2.1X10^{-3}$. Circular dichroism studies showed changes in CD signal upon addition of D-fructose, but without any significant change in the secondary structure. BLAST analysis showed homology of BGAF_Lec with other well-known β-glucosidase aggregating factors and some mannose-specific jacalin-related lectins (JRLs).

5.2 Introduction

Lectins are carbohydrate-binding proteins of non-immune origin spread across nature, especially in the plant kingdom. In plants, they have been isolated from seeds, roots, leaves, bark, rhizomes and tubers. Plant lectins have at least one non-catalytic carbohydrate-binding domain, which binds specifically to mono or oligosaccharides without altering their structure (Lis and Sharon, 1998). Plant lectins are involved in diverse functions by their recognition capacity such as phagocytosis and binding of microorganisms to host cells, intracellular transport of glycoproteins, endocytosis, cellular migration and adhesion (Lis et al., 1998, Wragg et al., 1999, Vijayan et al., 1999). Many plants such as jacalin and conA have been used to differentiate between normal and malignant tissues. Plant lectins agglutinate different type of red blood cells, this capacity of them is usually utilized to identify them.

Plant lectins are broadly divided into five classes based on their three-dimensional structural fold, such as legume lectins, monocot mannose-binding lectins, β-trefoil lectin, a lectin with hevein domain and β-prism I lectin (Peumans et.al., 2001). Among all the class of lectins, legume lectins are the most studied and they have helped in gaining insight on protein-glycan interactions, which led to the development of many applications of plant lectins (Peumans and Van Damme, 1998). Among other classes, in recent years there has been many studies on β-prism I lectin or popularly known as Jacalin related lectins or JRLs. Jacalin related lectins derived their name from the first member of this family isolated from the seeds of jackfruit (Moreira & Ainouz, 1981). Also, Jacalin from was the first one from this family to be studied through X-Ray crystallography (Shankarnarayanan et al., 1996). Lectins from this family have been found to carry out numerous functions, such as inhibition of HIV via binding with the specific region of HIV, ability to recognize particular immunoglobulin AI from human serum, stimulation of different T cell function and specific binding with tumor associated T-antigen (Corbeau et al., 1994). In plants, especially in monocot plants, jacalin are also present in the form of monocot chimeric jacalin, where they are induced by the jasmonate and contain another domain called dirigent domain. Monocot chimeric jacalin share similarities with monocot mannose-binding lectins and classical jacalin, but are also different from them (Qing-Hu Ma, 2014).

In this chapter, we have discussed cloning, expression, purification and physiochemical analysis of a lectin domain.

5.3 Material

All the Molecular biology grade reagents were purchased commercially. Mannose, galactose, N-Acetyl-D-glucosamine, and N-Acetyl-D-galactosamine, were obtained from Himedia Laboratories, Mumbai, Maharashtra, India. Fetuin and other chemicals required were procured from Sigma-Aldrich. Protein molecular weight markers were acquired from Thermo Fisher Scientific, Framingham, MA, USA. All media like Luria-Bertani (LB), Tryptone, Peptone, Yeast Extract, were purchased from HIMEDIA. Plasmid and Gel Extraction kit were purchased from Qiagen. The cDNA coding for beta-glucosidase aggregating factor (BGAF) from Sorghum (Gene Bank id: DQ866804.1) was synthesized artificially and obtained from Gene script Inc, USA.

Bacterial Strains and Vectors: BL21 (DE3) pLysS *E. coli* strain was obtained from Invitrogen Life Technologies, Grant Island, NY, USA. Bacterial strains of *E. coli*, *P. aeruginosa* and *S.aureus,* were procured from Microbial type culture collection, IMTECH, Chandigarh, India. pETM30a and pET28a was a gift from Dr. C. Ratna Prabha and Dr. Ashima Bhardhwaj respectively. pETM10 was obtained from EMBL, Heidelberg, Germany.

5.4 Methods

5.4.1 Preparation of chemical competent cells

The competent cells for transformation were prepared chemically as described in section 4.4.1.1 of chapter 4.

5.4.2 Transformation of *E. coli* cells

Transformation of *E. coli* cells with plasmid DNA was done as described in section 4.4.1.2 of chapter 4.

5.4.3 Plasmid DNA isolation

Plasmid DNA from *E. coli* cells was isolated as described in section 4.4.1.3 of chapter 4.

5.4.4 Agarose gel electrophoresis

DNA, both linear and circular was analyzed using gel electrophoresis method described in section 4.4.1.4 of chapter 4.

5.4.5 Restriction Digestion

Restriction digestion of PCR product and vector DNA was carried out according to protocol mentioned in section 4.4.1.5 of chapter 4.

5.4.6 Agarose Gel Extraction

DNA from agarose gel was extracted according to the protocol described in section 4.4.1.6 of chapter 4.

5.4.7 Ligation Reaction

Ligation of restricted digested PCR product into vector DNA was done according to manual provided by the manufacturer of T4 DNA ligase enzyme. Different ratios of insert DNA to plasmid DNA were tried.

5.4.8 Cloning of BGAF_Lec

The gene for BGAF_Lec was amplified from the full-length gene of BGAF (GeneBank id DQ866804.1) cloned in pUC57 vector, using the forward primer (5'-GGCGAATTCATGGGAGGTTCCGGA-3') and reverse primer (5'-GGCGCGGCCGCGCAGTCTTGAAACATT-3') with restriction sites 5` EcoRI and 3` NotI. The PCR amplification was performed in a DNA thermal cycler (Applied Biosystems) programmed for 4 minutes at 94°C and 35 cycles of 30s at 94°C, 45s at 55°C, 1 minute at 72°C and finally reaction mixture was kept for 10 minutes at 72°C. After

amplification, PCR product was double-digested (EcoRI, NotI) and ligated to similarly digested pET-series of an expression vector for obtaining N-terminal or C-terminal His-tagged fused protein. Positive clones were confirmed by restriction digestion, gel shifting and finally through nucleotide sequencing to check proper orientation and correct sequence stored at -20°C.

5.4.9 Expression trials of BGAF_Lec

E. coli expression strains BL21(DE3) & BL21(DE3) PlysS were transformed with vectors carrying the cloned gene and positive clones selected by culturing on 1.5% LB agar plates containing suitable antibiotic marker selection A single colony of the cells with clones was used to innoculate 4ml of LB media with suitable antibiotic and cultured overnight at 37°C. A 0.2ml aliquot of this culture was then used to inoculate 20ml of LB media supplemented with the antibiotic and incubated at 37°C till optical density (OD) reading of the media at 600 nm reached 0.6 and expression induced with varying concentrations of IPTG at different temperature and different lengths of time to get the optimal expression. The bacteria were harvested by centrifugation at 10000 x *g* for 15 minutes at 4 °C, the supernatant removed, and the pellet was either snap frozen in liquid nitrogen and stored at -80°C or it was lysed. For lysis, the pellet was suspended in lysis buffer (50 mM Tris-HCl pH 7.8, 200 mM NaCl, 20 mM imidazole, 5% glycerol, 0.1% Triton X 100 & 1mM PMSF) and sonication was performed to lyse cells. Finally, the sample was centrifuged at 12000 X g at 4°C for 20 minutes, and supernatant and pellet were separated and stored for further analysis.

5.4.10 Purification of BGAF_Lec

Purification was achieved by affinity chromatography using the hexahistidine tag fused to protein. 3ml Ni-NTA agarose affinity matrix was packed in a column, column equilibrated with lysis buffer and supernatant from bacteria lysis was loaded onto the column. The column was washed with lysis buffer containing 50mM imidazole to remove the nonspecifically bound proteins, and elution was achieved with the lysis buffer containing 300mM imidazole and samples analyzed on SDS-PAGE.

5.4.11 Molecular mass determination of BGAF_Lec

5.4.11.1 SDS-PAGE analysis of BGAF_Lec

The relative molecular mass of the purified protein was obtained by SDS-PAGE performed according to method described in the section 2.5.2.1 of chapter 2 of this thesis.

5.4.11.2 Gel filtration studies of BGAF_Lec

Gel-filtration of BGAF_Lec was carried out according to method described in section 2.5.2.2 of chapter 2 of this thesis.

5.4.12 Hemagglutination assay and carbohydrate specificity of BGAF_Lec

Hemagglutination and carbohydrate inhibition studies of BGAF_Lec was carried out according to method describe in section 2.5.5 of chapter 2 of this thesis.

5.4.13 Effect of pH and Temperature on hemagglutination activity of BGAF_Lec

Effect of temperature and pH on lectin activity of BGAF_Lec was studied according to method described in section 2.5.6 of chapter 2 of this thesis.

5.4.14 Carbohydrate binding analysis of BGAF_Lec using fluorescence spectroscopy

The binding of different carbohydrate which was showing inhibition of hemagglutination activity of lectin was studied by titrating the protein against carbohydrate solutions and monitoring the change in fluorescence. Titrations were performed by adding aliquots (1µl at each time) of test sugars from a 1M stock in 100mM Tris-HCl (pH 7.5) added to 1ml of purified protein (0.1mg/ml) in the same buffer. Samples were mixed thoroughly with pipetting, making sure that no bubble formation occurred, corrections made to compensate the dilution effect upon addition of sugar to lectin and at the highest concentration of the added sugar, the volume change was less than 5 % of the solution in the cuvette. Each data point was an average of three independent sets of experiments with

standard deviation (SD) less than 5%. The following equation was used to determine the association constant (Ka) (Chipman *et al.,* 1967).

$\log [C]f = -\log[Ka] + \log [(F0-FC)/(FC-F\infty)]$ (Eq. 4.1)

From the ordinate intercept of the double reciprocal plot of F0/(F0-FC) versus 1/[C], where F0 and FC are the fluorescence intensities of the free protein and of the protein at a sugar concentration [C], F∞, the fluorescence intensity upon saturation of all the sugar binding sites is obtained. In the plot of log[(F0-FC)/(FC-F∞)] versus log[C], the abscissa intercept yielded the Kd value (the dissociation constant) for the protein-sugar interactions, which is the reciprocal of Ka (the association constant). The free energy changes of the Association (ΔG) were calculated by using the equation:

$-\Delta G = RT \ln(Ka)$ (Eq. 4.2)

Where R is a constant and T is the temperature at which the reaction was carried out.

5.4.15 Determination of Secondary structure elements of BGAF_Lec via circular dichroism spectroscopy

To record the CD spectra, a Jasco J-810 (Jasco corp., Tokyo, Japan) spectropolarimeter equipped with Peltier thermostat was used. Two concentrations of purified protein (5μM and 25μM) in 50mM sodium phosphate buffer (pH 7.5) were placed in a 2mm path length rectangular quartz cell and spectra recorded at a scan speed of 20nm/min with a response time of 4s, a slit width of 2nm was adjusted in both, the far-UV region (250 to 200 nm) and the near-UV region (300 to 250nm), respectively. Analysis of the CD spectrum was done using three different methods, namely CDSSTR, CONTINLL and SELCON3 (Sreerama *et al.,* 1999; Johnson, W.C., 1999; Provencher & Glockner, 1981) and employing the software routines available at CDPro. A basis set containing 43 proteins was used as a reference for fitting the experimental spectrum. To investigate the effect of carbohydrate binding on the secondary and tertiary structure, the spectra were recorded with and without 1 mM D-fructose.

5.4.16 Antimicrobial assay of BGAF_Lec

To check the effect of lectin on the growth of three bacterial species (*E. coli*, *S. aureus,* and *P. aeruginosa*) antimicrobial assay was carried out in 96 well plates. A colony of bacteria was inoculated in 5ml LB and incubated at 37°C till the OD_{600} reached 0.1 (10^5-10^6 cells/ml). This culture was used as inoculum for the antibacterial assay. Purified lectin (1mg/ml) in PBS was used at two final concentrations of 0.25mg/ml and 0.5mg/ml. 2µl of was added to each well in a 96 well plate and incubated at 37°C for 16 hours. After 16 hours the O.D was measured at 600 nm on a Bio-Rad Elisa plate reader. The assay was repeated three times.

5.4.17 Crystallization trials of BGAF_Lec

The initial crystallization trials of lectin were carried out using Screen 1 (Crystal Screen TM-HR2-110) and Screen 2 (Crystal ScreenTM-HR2-112), Index and Index HT crystallization buffers from Hampton Inc using the micro batch method. Besides crystallization screens a homemade screen of different PEGs and variations in their concentration and conditions were also screened. The different buffer conditions were set up in the micro batch 60 wells plate acquired from griener biologics. The plates were filled with a total of 8ml of oil which was a mixture of paraffin and silicone oil in the ratio of 1:1. Purified lectin was dialyzed against 50mM Tris-HCl (pH 7.5) and 50mM NaCl and different concentration (2mg/ml to 5mg/ml) was used to set up the crystallization. The protein and different buffers were added to the wells of plates in a different ratio of (1:1,1:2, 1:3) at different protein concentration generating a wide range of different conditions for crystallization.

5.4.18 Bioinformatics analysis of BGAF_Lec

The homology search to find out the homologs protein sequences was done using BLAST tool present in the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Protein BLAST or pBLAST was used, and query sequence was searched against Non-redundant protein sequences from Viridiplantae. BLAST search was carried out using BLOSUM 62 matrix with expected values of 10. MUSCLE and ClustalW program present in MEGA

Suite were used to carry out multiple sequence alignments to find out the conserved stretches of sequences. Molecular modeling and model refinement were done using I-tasser server (Yang *et al.,* 2015) and modrefiner tool (Xu & Zhang, 2011), respectively. Chimera was used for molecular visualization and superposing the structures (Pettersen *et al.,* 2004). Molecular docking of the ligand, in the putative binding site, was done using Autodock tool (ADT) (Morris et al., 2009). Intermediary steps, such as pdbqt files for protein and ligands preparation and grid box creation were completed using Graphical User Interface program ADT. ADT assigned polar hydrogens, united atom Kollman charges, solvation parameters and fragmental volumes to the protein. AutoGrid was used for the preparation of the grid map using a grid box. The grid size was set to 32 × 28 × 26 XYZ points, respectively, and grid center was designated at dimensions (x, y, and z): 26.223, 90.443 and 22.613, respectively. Autodock employed iterated local search global optimizer and during the docking procedure, both, the protein and ligands were considered as rigid. The pose with the lowest energy of binding or binding affinity was extracted. Pymol was used for viewing hydrogen bonding of ligand with protein. The coordinates of N-Acetyl-D-galactosamine and mannose were downloaded from ZINC database (http://zinc.docking.org/)

5.5 Results

5.5.1 Cloning of BGAF_Lec

The BGAF_Lec fragment of ~ 420 bp was obtained by PCR amplification (Figure 5.1 A). The full-length gene for BGAF (GeneBank id DQ866804.1), cloned into pUC57 was used as a template. PCR amplified fragment was purified, and gel eluted for further use. The purified amplicon was digested with EcoRI and NotI and cloned into three different vectors pET28a, pETM10 and pET30a vectors (Table 5.1). Positive clones of transformed DH5α cells were isolated from kanamycin LB plates. The recombinants were assessed by gel shift assay and restriction digestion (Figureure5.1 B, 5.1 C). The clones that showed positive results in gel shift assay, and restriction digestion were cultured, plasmid DNA isolated and subjected to sequencing. The sequence obtained for ~420 bp (Figure 5.2 A) was translated to protein sequence (Figure 5.2 B), and it is 100% similar to the desired sequence.

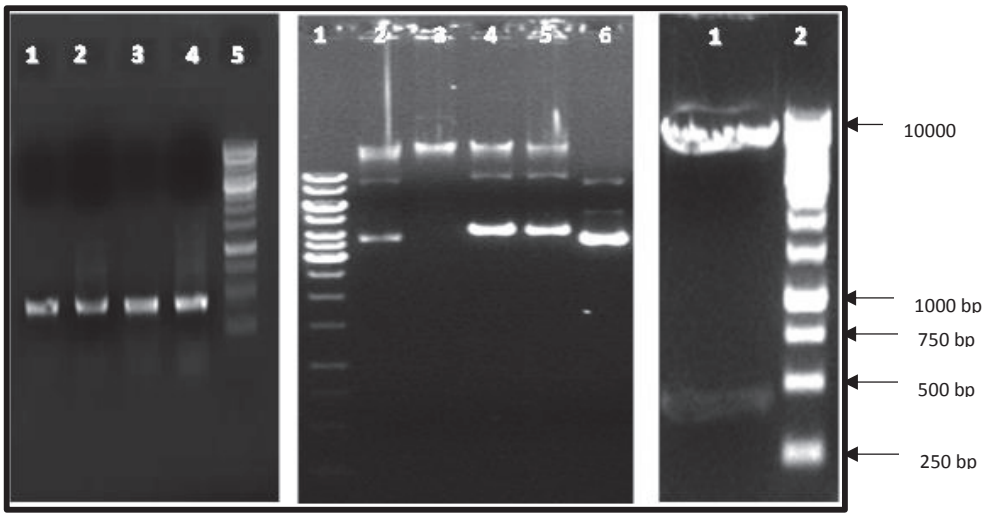


Figure 5.1 Agarose Gel (1.0%) profile of BGAF_Lec clone (A) Lane 1-4 shows the ~450 bp amplicon obtained by PCR. **(B)** Plasmid DNA isolated from the positive clones of transformed DH5α and lane 4 & 5 shows retarded mobility **(C)** Restriction digestion of clone showing retarded mobility. Lane 1 show released gene from the cloned plasmid.

Table: 5.1 Cloning of BGAF_Lec in different vectors

Vector	Clone	Sequencing
pETM10	Yes	Yes
pET28a	Yes	Yes
pET30a	Yes	Yes

(A)

```
ATGGGAGGTTCCGGAGGATCAGCGATGGACATTACCGAAGCTCCACGTAGACTGGAGTCAATCACTGTTTATGCC
GGTGTTGTGTTGGATAGCATTGCATTTAGTTACATCGACAACTCTGGCCAAAAAAGATCCGCAGGACGTTGGGGTG
GACCAGGAGGTGGCGGACCTCATACAATTCAGCTGGGTGAATCAGAGGTCATCACGGAAGTTAGCGGCACATTTG
GAACGTATTACAATGCTACCACTATTACATCTATCAAATTCGTCACCAACCTGAATAAGACTTACGGACCATGGGG
TGGAGGACAGGGAGCAAGTTTCACTATTCCTGTGCAACCAGGTTCTGCTATCGTGGGCTTTTTCGTCCGTGGAGCG
ACCTATCTGCAAGCTATTGGTGTGTACGTCAGAACTTTGTAA
```

(B)

```
BGAF        MAKLQVTPAAAFTEYKELNFQGLYLHHMFWNRPKANQARIIENKAPLGIGATVVNNWEVY
Lectin      ------------------------------------------------------------

BGAF        DGPGENAKLVARAQGSHIYAGKWANSFSLVFVDERFSGSTLEVMGIVVETGEWAVVGGTG
Lectin      ------------------------------------------------------------

BGAF        QFAMANGVISKRLHKSTSDGNIIELTIRAFCPPLKGTMYPVIKVGPFGGSGGSAMDITEA
Lectin      ----------------------------------------------MGGSGGSAMDITEA
                                                           :************

BGAF        PRRLESITVYAGVVLDSIAFSYIDNSGQKRSAGRWGGPGGGGPHTIQLGESEVITEVSGT
Lectin      PRRLESITVYAGVVLDSIAFSYIDNSGQKRSAGRWGGPGGGGPHTIQLGESEVITEVSGT
            ************************************************************

BGAF        FGTYYNATTITSIKFVTNLNKTYGPWGGGQGASFTIPVQPGSAIVGFFVRGATYLQAIGV
Lectin      FGTYYNATTITSIKFVTNLNKTYGPWGGGQGASFTIPVQPGSAIVGFFVRGATYLQAIGV
            ************************************************************

BGAF        YVRTL
Lectin      YVRTL
            *****
```

Figure 5.2 DNA and protein sequence of BGAF_Lec F(A) Nucleotide sequence of 420 bp and (B) Translated protein sequence of BGAF_Lec showing 140 amino acids.

5.5.2 Expression trials of BGAF_Lec

Initially expression trials of lectin cloned in pETM10 and pET28a vectors showed no detectable expression of lectin on SDS-PAGE. Expression trials in autoinduction media at different temperature also yielded the similar results. But expression trials of lectin cloned in pET30a vector in LB and autoinduction media showed significant expression of lectin at the correct size in SDS-PAGE gel analysis (Figure 5.3). Summary of expression trials at different conditions are detailed in Table 5.2

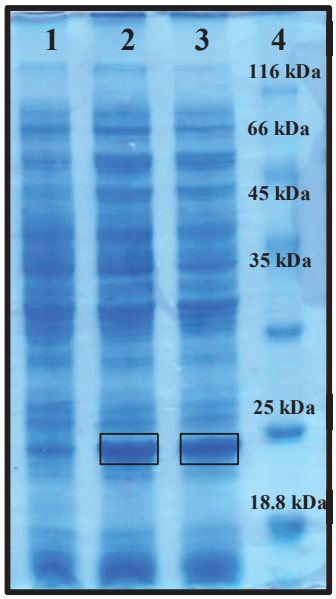


Figure 5.3 SDS-PAGE analysis of Expression trials of BGAF_Lec. Gene for BGAF_Lec cloned in pET30a vector in BL21 (DE3) PlysS strain. Lane 1 shows samples where no IPTG added, and it is used as a control. Lane 2,3 shows induction by IPTG at 0.2 mM, 0.4 mM, respectively at 24°C. Lane 4 shows molecular weight markers.

Table 5.2 Summary of expression trials of BGAF_Lec in different pET vectors. Gene for BGAF_Lec cloned in three different pET series of vectors with a combination of two different strains. Expression levels were assayed by visual inspection on Coomassie stained SDS-PAGE gel. + indicates that expression was detected. – indicates that no expression was detected.

***E. coli* Strain** / **Vectors**		**LB media**								
		0.2 mM IPTG			**0.4 mM IPTG**			**0.6 mM IPTG**		
		20°C	**24°C**	**37°C**	**20°C**	**24°C**	**37°C**	**20°C**	**24°C**	**37°C**
BL21(DE3)	pETM10	-	-	-	-	-	-	-	-	-
	pET28a	-	-	-	-	-	-	-	-	-
	pET30a	-	-	-	-	-	-	-	-	-
BL21(DE3) pLysS	pETM10	-	-	-	-	-	-	-	-	-
	pET28a	-	-	-	-	-	-	-	-	-
	pET30a	+	+	+	+	+	+	+	+	+

5.5.3 Purification of BGAF_Lec

Purification of His-tagged lectin was carried out through a nickel affinity column. After purification samples from each step of purification process was analyzed on SDS-PAGE gel (Figure 5.4). On the gel, His-tagged BGAF_Lec protein was found to be purified.

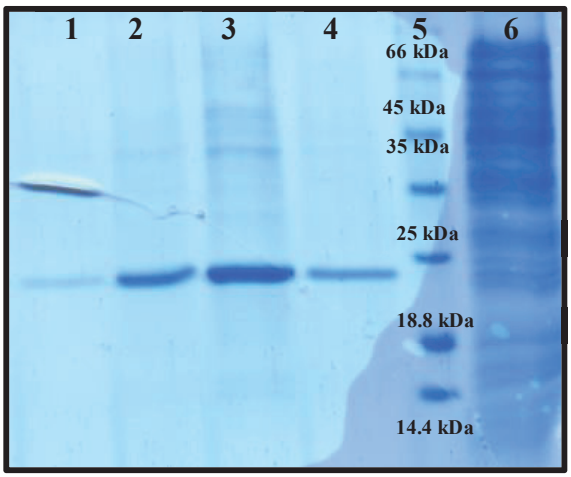


Figure 5.4 SDS-PAGE analysis of Purification process of BGAF_Lec Lane 1-4 shows fractions of purified BGAF_Lec eluted from the column with buffer containing 300 mM imidazole. Lane 6 Shows sample collected after washing the column with washing buffer. Lane 5 shows Molecular weight markers.

5.5.4 Hemagglutination and carbohydrate specificity of BGAF_Lec

Lectin showed hemagglutination activity with 1% rabbit erythrocytes. The hemagglutination activity was inhibited strongly by 0.5mM D-fructose. Mannose and N-Acetyl-D-glucosamine also showed inhibition but at a higher concentration of 20 and 50mM, respectively. The hemagglutination activity was also inhibited by fetuin with a minimum inhibitory concentration (MIC) of 0.1 % (Table 5.3)

Table 5.3 Inhibition of lectin activity of BGAF_Lec

Carbohydrates	Minimum Inhibitory Concentration (MIC) mM
Mannose	20
Galactose	NI
Lactose	NI
Glucose	NI
Fructose	0.5
N-Acetyl-D-galactosamine	50
N-Acetyl-D-glucosamine	NI
Methyl- α-D-mannopyranoside	NI
Methyl- α-D-glucopyaranoside	NI
Fetuin	0.1%

5.5.5 pH and temperature stability of BGAF_Lec

The hemagglutination activity of BGAF_Lec was retained between pH 5-9 (Figure 5.6) and was stable until 50°C. Heating about 50° C leads to unfolding and subsequent precipitation **of protein.**

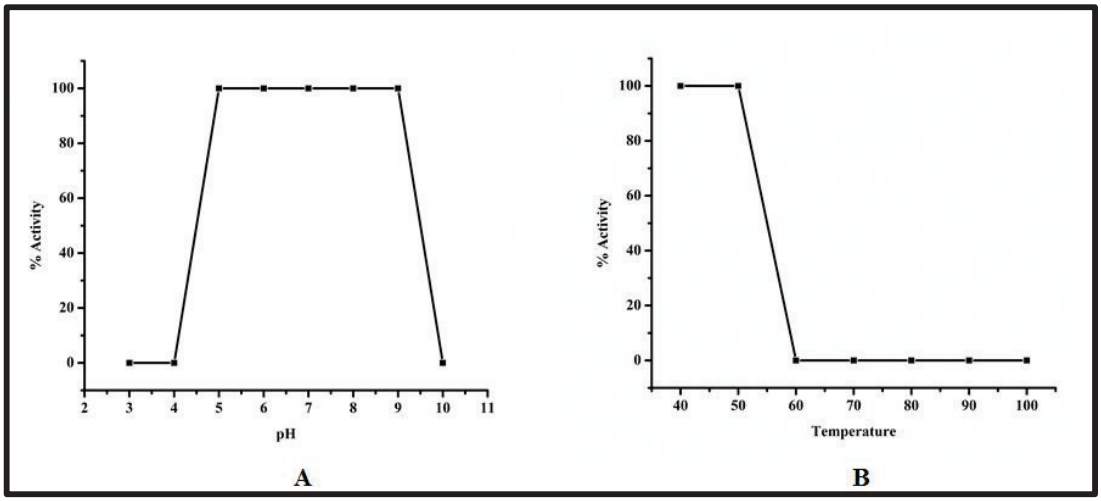


Figure 5.5 Effect of pH and temperature on hemagglutination activity of BGAF_Lec. pH stability (A) and temperature stability (B) of BGAF_Lec.

5.5.5 Carbohydrate binding analysis of BGAF_Lec by Fluorescence spectroscopy

Upon excitation at 280nm lectin gave an emission spectrum centered at 338 nm. Titration of BGAF_Lec with sugars resulted in quenching of the lectin fluorescence without any change in emission maxima (Figure 5.6, 5.7). Titration of BGAF_Lec with N-Acetyl-D-galactosamine and Methyl- α-D-glucopyaranoside resulted in a decrease in the emission intensity by 53% and 62%, respectively. In Figure 5.6A spectrum one corresponds to the protein alone and spectra 2-10 correspond to those recorded in the presence of increasing concentrations of N-Acetyl-D-galactosamine. In Figure 5.7A spectrum one corresponds to the protein alone and spectra 2-10 correspond to those recorded in the presence of increasing concentration of Methyl- α-D-glucopyaranoside. Figure 5.6 B (N-Acetyl-D-galactosamine) and figure 5.7 B (Methyl-α-D-glucopyaranoside) shows a plot of the change in fluorescence intensity, ΔF (= Fo – F) as a function of the added ligand concentration, which represents the binding curve for the titration. Double logarithmic plots are shown in figure 5.6 C and 5.7 C, respectively were used to determine the dissociation constant of lectin for respective carbohydrates. From X intercept of these plots, the dissociation constant Kd was determined as $3.5X10^{-3}$ and $5.3X10^{-3}$, respectively. Reciprocal values of dissociation constant were considered as association constant The slope of these graphs were found to be ~ 1, indicating each lectin subunit binds one saccharide molecule. The fluorescence titration data for the interaction of other carbohydrates were also analyzed in similar manner and association constant and corresponding ΔG values for binding are listed in Table 5.4

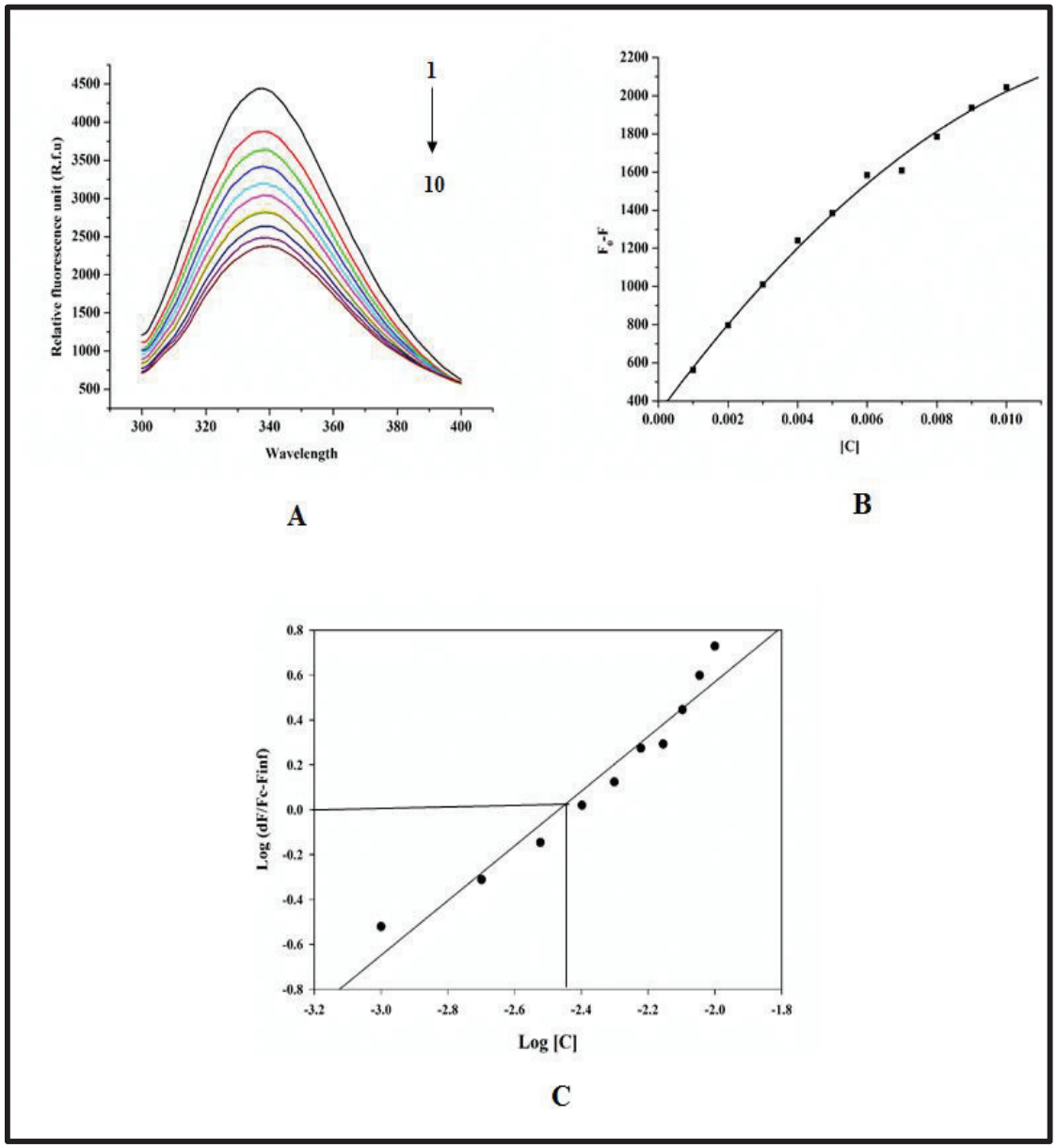


Figure 5.6 Fluorescence spectra of BGAF_Lec in absence and presence of GalNac. (A) Represents fluorescence quenching of lectin on the addition of N-Acetyl-D-galactosamine to the protein solution. (B) The plot of (Fo-F) vs. [C] shows saturation of binding site with increasing concentration of N-Acetyl-D-galactosamine (C) Double logarithmic plot for calculating binding affinity of N-Acetyl-D-galactosamine with BGAF_Lec.

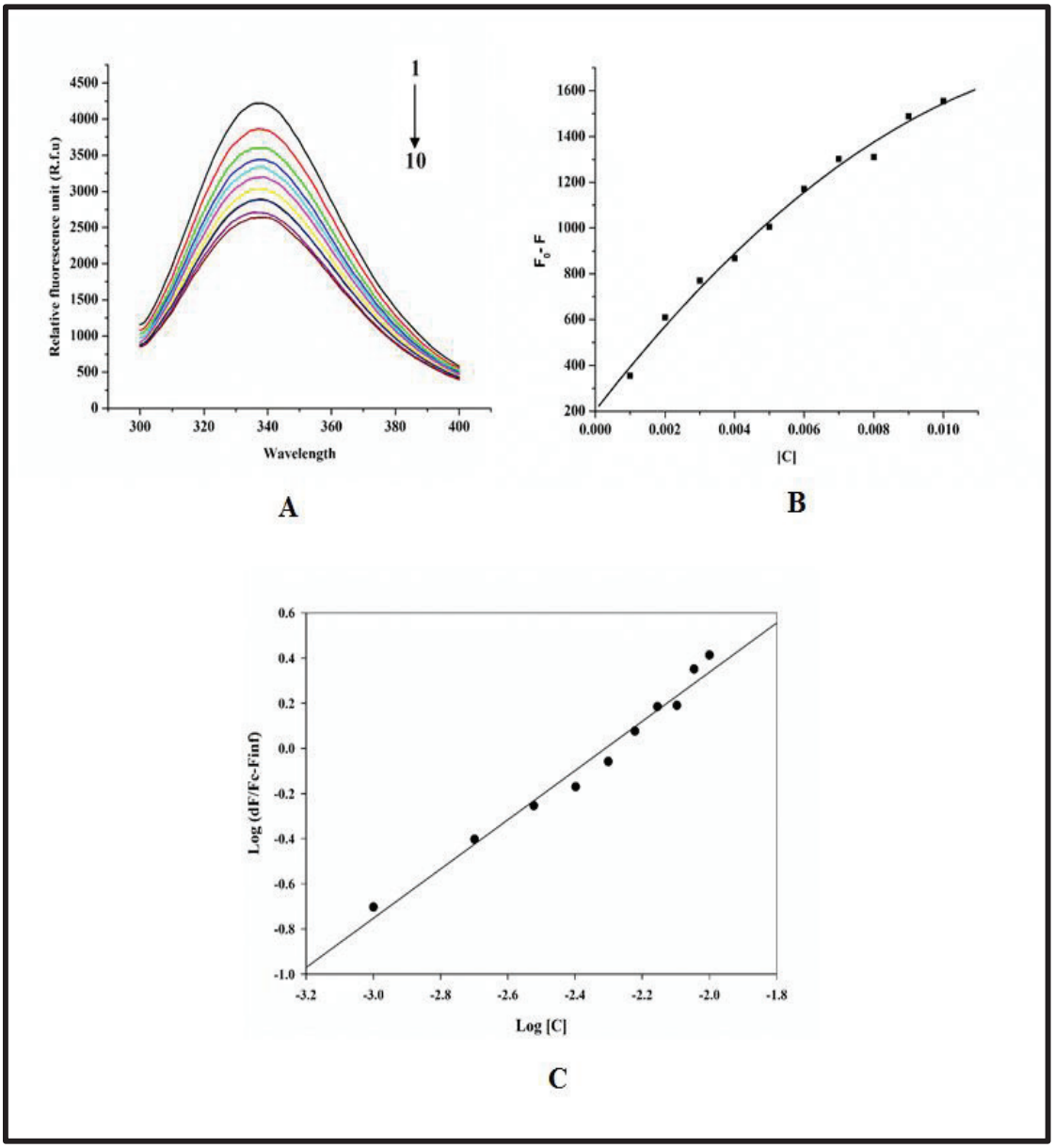


Figure 5.7 Fluorescence spectra of BGAF_Lec in absence and presence of Methyl-α-D-glucopyaranoside (A) Represents fluorescence quenching of lectin on the addition of Methyl-α-D-glucopyaranoside to the protein solution. (B) The plot of (Fo-F) vs. [C] shows saturation of binding site with increasing concentration of Methyl-α-D-glucopyaranoside (C) Double logarithmic plot for calculating binding affinity of Methyl-α-D-glucopyaranoside with BGAF_Lec.

Table 5.4 Dissociation constants, Association constants and corresponding Gibb`s free energy values for BGAF_Lec with different carbohydrates.

Carbohydrate	Dissociation Constant (Kd)	Association Constant (Kb= 1/Kd) M^{-1}	ΔG°b (Kcal.mol^{-1})
N-Acetyl-D-galactosamine	$3.5X10^{-3}$	$2.8571X10^{2}$	-14.41
Methyl-α-D-glucopyaranoside	$5.3X10^{-3}$	$1.886X10^{2}$	-13.36
Fructose	$2.1X10^{-3}$	$4.76X10^{2}$	-15.72
Mannose	$6.3X10^{-3}$	$1.587X10^{2}$	-12.90

5.5.6 Secondary structure analysis of BGAF_Lec with Circular Dichroism (CD) spectroscopy

Circular dichroism spectra of lectin in native state (Solid line) and in the presence of D-fructose (dotted line) are shown in figure 5.8 A. The far-UV spectrum of native lectin is characterized by a minimum near 219nm with zero crossing near 203nm. In the presence of N-Acetyl-D-galactosamine, there was no change in minimum, but the intensity was decreased suggesting no change in the secondary structure in the presence of D-fructose.

The analysis of near-UV CD spectra provided an understanding of the tertiary structure of a protein by giving insight into the interaction of aromatic amino acid side chains with peptide bonds. Analysis of near-UV CD spectra (Figure 5.8 B) of lectin in the absence and presence of N-Acetyl-D-galactose amine showed Λ_{max} at 253nm and 285nm, respectively. The peak at 253nm signified the presence of phenylalanine and the peak at 285nm represented the presence of tyrosine, thereby implying that the structure of BGAF_Lec is compact and rigid.

To derive information on the content of secondary structure elements of lectin, the far-UV CD spectrum was analyzed by three different methods, namely CDSSTR, CONTINLL and SELCON3 and the results obtained from these analyses are given in Table 5.5. To check the effect of D-fructose on the secondary structure of BGAF_Lec, far UV spectrum collected in the presence of D-Fructose was also subjected to analysis with programs mentioned above in CDPro with similar parameters and data obtained from these analyses suggested, N-Acetyl-D- galactosamine did not change secondary structure elements of BGAF_Lec.

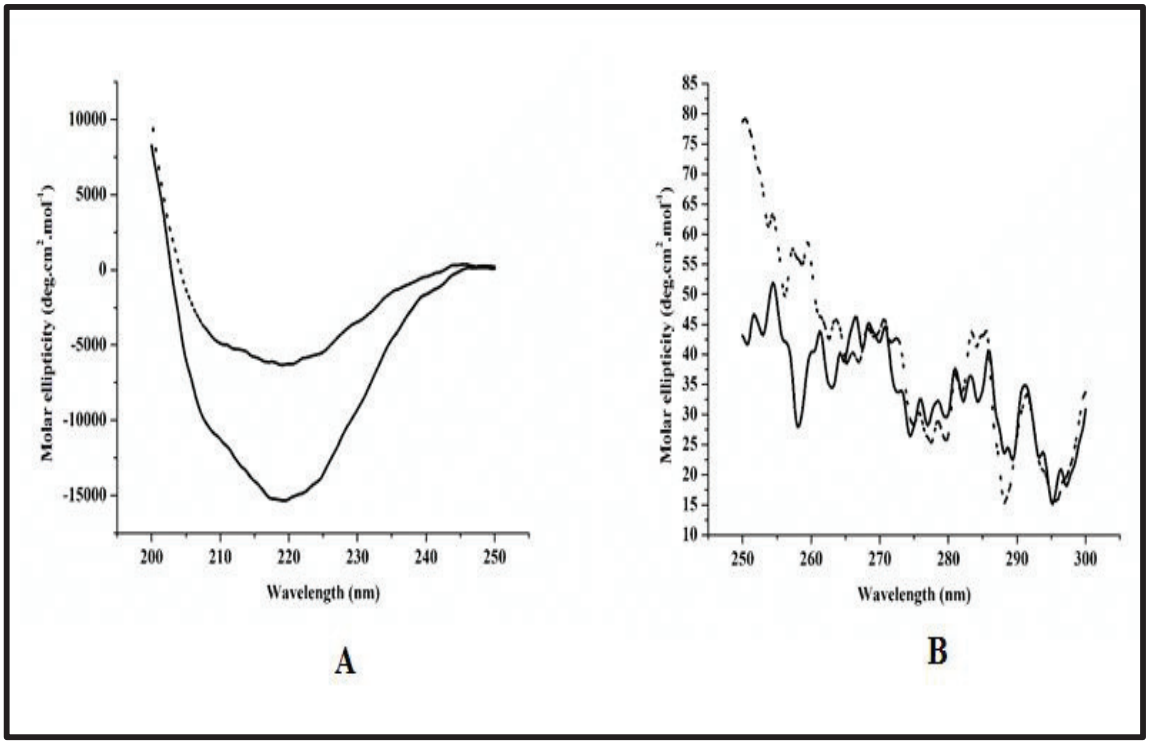


Figure 5.8 CD analysis of BGAF_Lec in absence and presence of carbohydrate. (A) Far-UV spectra of BGAF_Lec (B) Near UV spectra of BGAF_Lec. The solid line represents BGAF_Lec in the buffer, and the dotted line represents BGAF_Lec in complex with D-Fructose.

Table 5.5 Calculated secondary structure elements of BGAF_Lec using CDPro suite and prediction by PSIPRED

From CD Spectral Analysis								
Method	**α%**		**β%**		**Turns**		**Unordered**	
	Without Fructose	**With Fructose**	**Without Fructose**	**With Fructose**	**Without Fructose**	**With Fructose**	**Without Fructose**	**With Fructose**
CDSSTR	7	6	35	37	26	24	31	32
CONTINLL	3.3	4.9	41.4	39.6	21.1	22	34.2	33.5
SELCON3	9.2	9.2	46.3	46	11.0	10.9	34.1	33.9
Average	**6.5**	**6.7**	**40.9**	**40.86**	**19.36**	**18.96**	**33.1**	**33.1**
From Theoretical prediction								
PSIPRED	0.0		50		50		-	

5.5.7 Antimicrobial activity of BGAF_Lec

To measure the activity of BGAF_Lec on different strains of bacteria, their growth in terms of O.D was monitored in presence and absence of lectin, and it did not inhibit the growth of this microorganism completely (Figure 5.9)

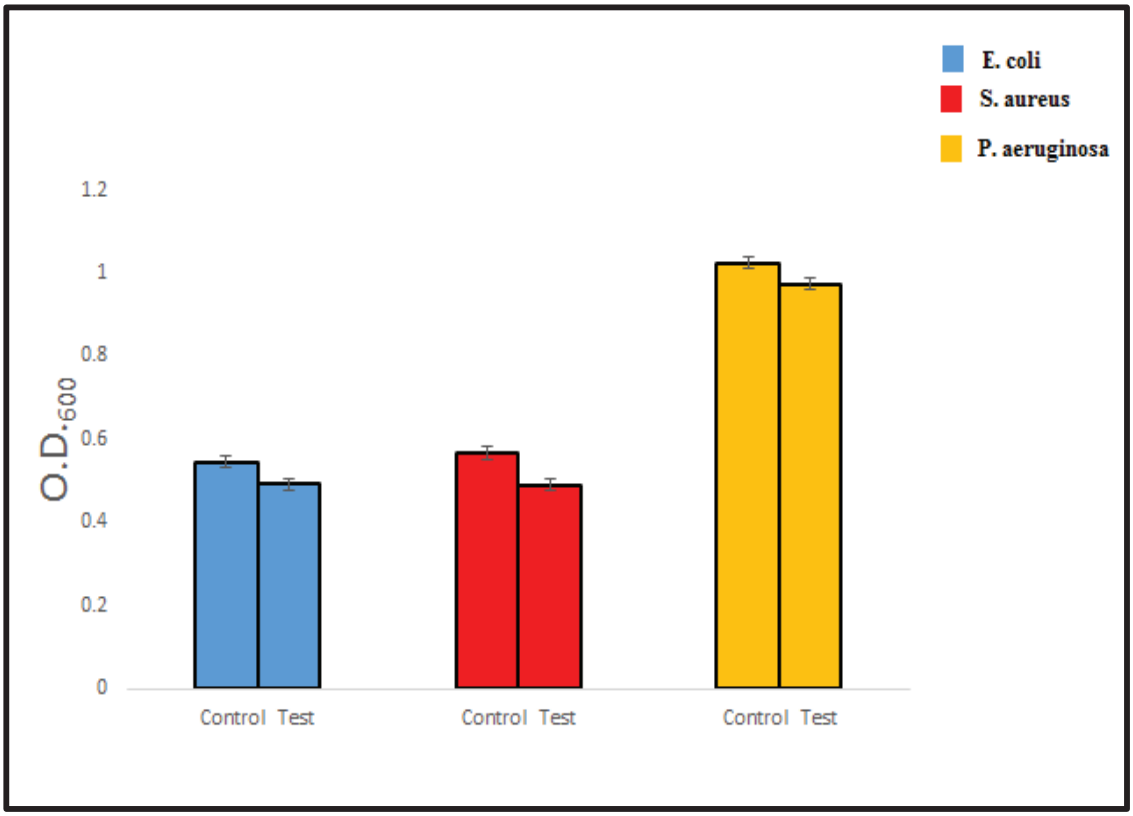
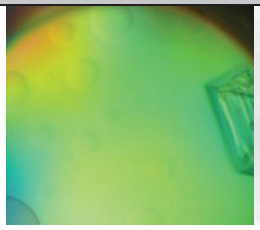
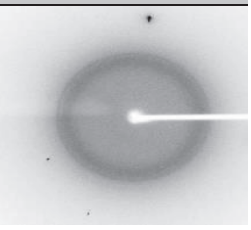
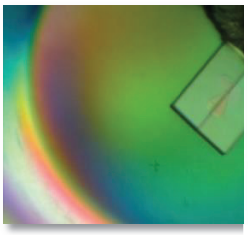
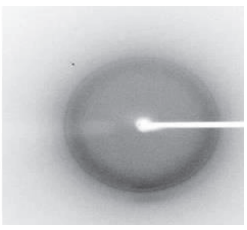


Figure 5.9 Antibacterial activity of BGAF_Lec. Activity of BGAF_Lec against three strains of bacteria with control

5.5.8 Crystallization trials of BGAF_Lec

For extensive crystallization trials, BGAF_Lec was purified by Ni-NTA affinity chromatography (up to 5 mg/ml) and used for conducting trials in micro-batch plates under oil conditions. Crystals were obtained in only two conditions (Table 5.). Upon diffraction, the crystals were found to be of salts instead of protein.

Table 5.6 Crystallization trials of BGAF_Lec

Name	Crystallization Condition	Image of Crystal	Diffraction Image
BGAF_Lec_C1	0.1 M Sodium acetate trihydrate pH 4.5, 3.0 M Sodium Chloride		
BGAF_Lec_C2	0.05 M Potassium phosphate monobasic, 20% w/v PEG 8000		

5.5.9 Bioinformatics analysis of BGAF_Lec

5.5.9.1 Sequence analysis of BGAF_Lec

BLAST analysis using a sequence of lectin as a query found most of the hits with a beta-glucosidase aggregating factor or jasmonate-induced proteins or hypothetical proteins. Not a single hit from classical lectins belong to jacalin class was obtained. The nearest homolog of BGAF_Lec from a different plant other than sorghum was jacalin-related lectin from *Oryza sativa* with 63% sequence identity and E-score of $6e^{-51}$ (Table 5.7).

Since BLAST output did not show homology between BGAF_Lec and classical jacalin-like lectins, PDB database was searched using lectin as a query. Total of 82 structure of Jacalin-like lectins from various plants. The nearest homolog of BGAF_Lec in PDB data was mannose-specific lectin from *Helianthus tuberosus* with 37% identity and E-value of $2.58e^{-14}$ (Table 5.7).

Table 5.7 BLAST analysis of BGAF_Lec

Accession	Description	Query coverage	E-Value	Max Identity %
ABI24164.1	Beta-glucosidase aggregating factor (*Sorghum bicolor*)	100	8e-94	99
EAY82651.1	Hypothetical protein Osl 37872 (*Oryza sativa*)	100	6e-51	63
XP_002461766.1	Hypothetical protein SORBIDRAFT	100	3e-61	73
NP_001141303.1	Uncharacterized protein (*Zea mays*)	100	1e-45	63
ABJ97445.1	Beta-glucosidase aggregating factor 1 (*Zea ma ys*)	92	2e-44	62
AAR20919.1	Jasmonate-induced protein (*Triticum aerstivum*)	100	3e-47	61

Table 5.8 Homology search for BGAF_Lec in PDB database

PDB ID	Uniprot Id	Description	E-Value	Max Identity %
1C3K	Q9ZQY5	Crystal structure of *Helianthus tuberosus* lectin	2.58e-14	37
1ZGR	P83304	Crystal structure of *Parkia platycephala* seed lectin	3.92e-14	39
1X1V	Q8L5H4	Structure of banana lectin - Methyl-Alpha-Mannose-complex	4.6e-14	35
5AV7	P93114	Crystal structure of Calsepa lectin in complex with bisected glycan	5.18e-10	31

5.5.5.9 Multiple seqeunce alignment and Phylogenetic analysis of BGAF_Lec

A multiple sequence alignment carried out with sequences of the lectin domains from other known BGAF and classical Jacalin lectin obtained from PDB database reflects overall similarities. The multiple alignments showed that the second half (C-terminal side) the sequence is more conserved in the different members compared to first half (N-terminal). The four different loops which play a vital role in carbohydrate recognition are showed in the multiple sequence alignment as L1, L2, L3 and interacting or IL loop (Figure 5.10).

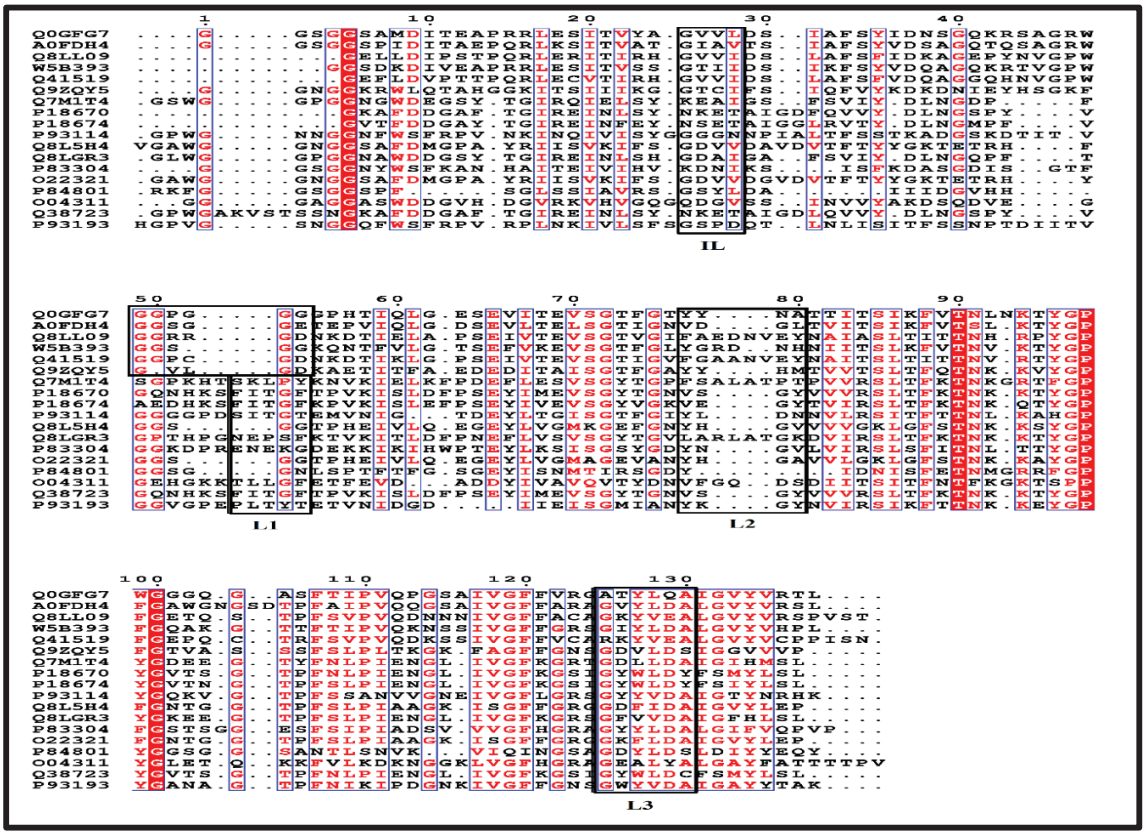


Figure 5.10 Multiple sequence alignment of BGAF_Lec: MSA of BGAF_Lec with sequences of lectin domain from other known BGAF and classical JRLs. Residues in the red show conservation of that particular residue among all the sequence. Residues in light red means one or more of the sequence have a similar substitution of amino acids. Black box shows amino acids constitute different loops L1, L2, L3, and IL, which forms the binding site in JRLs.

Based on the multiple sequence alignment, a phylogenetic analysis was also carried out to gain insight into the interrelatedness of the various sequences. The phylogenetic trees were constructed using three widely used algorithms in MEGA Suite and are showed in Figure 5.11. All three program indicated a similar branching pattern, grouping all sequence into two major clusters, named cluster 1 and 2. The first cluster has an identifiable sub-cluster which is made up of sequences from lectin domain of BGAF from different plants such as sorghum, maize. The other part of cluster one is made up of lectins from monocotyledonous plants. Cluster 2 is made up sequence from all the dicotyledonous plants and a sequence from red algae. Cluster 2 also has a very distinct subgroup consists of all the Moraceae family lectins.

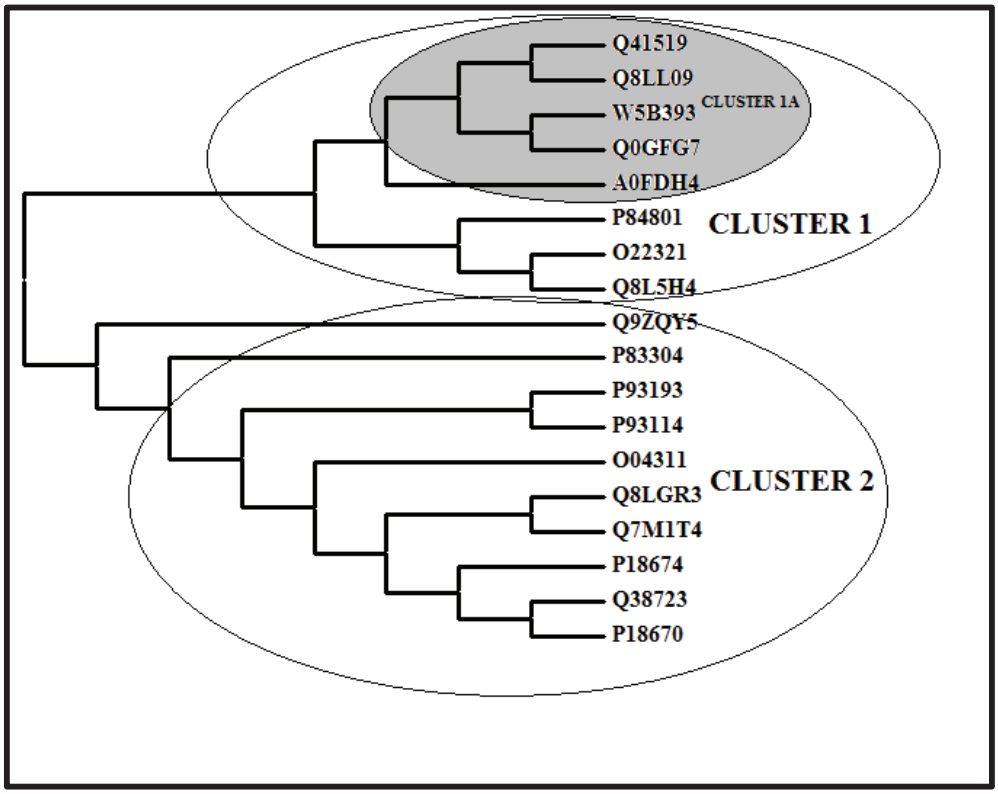


Figure 5.11 Phylogenetic analysis of BGAF_Lec Phylogenetic analysis of BGAF_Lec with other known JRLs and lectin domains of other BGAF forms two major clusters; Cluster 1 having all the monocotyledons mannose binding JRLS sequences. Cluster 1 also has a sub-cluster 1A which contains all the sequences of lectin domains from different BGAF. Cluster 2 contains sequence from dicotyledons plants and sequence from red algae.

5.5.9.3 Modelling and Secondary structure prediction of BGAF_Lec

The three-dimensional structure of BGAF_Lec was predicted by I-TASSER server with C-score of 1.01. C-score is a confidence score for estimating the quality of predicted model by I-TASSER server. C-score is typically in the range of [-5,2], where C-score of higher value signifies a model with high confidence. The modeled three-dimensional structure was generated using the structure of known lectin from Jerusalem artichoke (*Helianthus tuberosus*) (PDB Id: 1C3M). The generated model consisted of 9 β- strands and loops which are represented in different colors and each strand are numbered from N-terminal to C-terminal (Figure 5.12 A). Loops which are very important for the formation of carbohydrate binding site are named as L1, L2, L3 and IL, where IL stands for interacting loop (Figure 5.12 B). Shape and size of these loops play very important role in carbohydrate specificity. The secondary structure of BGAF_Lec was predicted on the basis of its sequence by PSIPRED server. The output obtained from PSIPRED server is matching with the modeling output suggesting BGAF_Lec consists of only nine β-strands and loops (Figure 5.12 C).

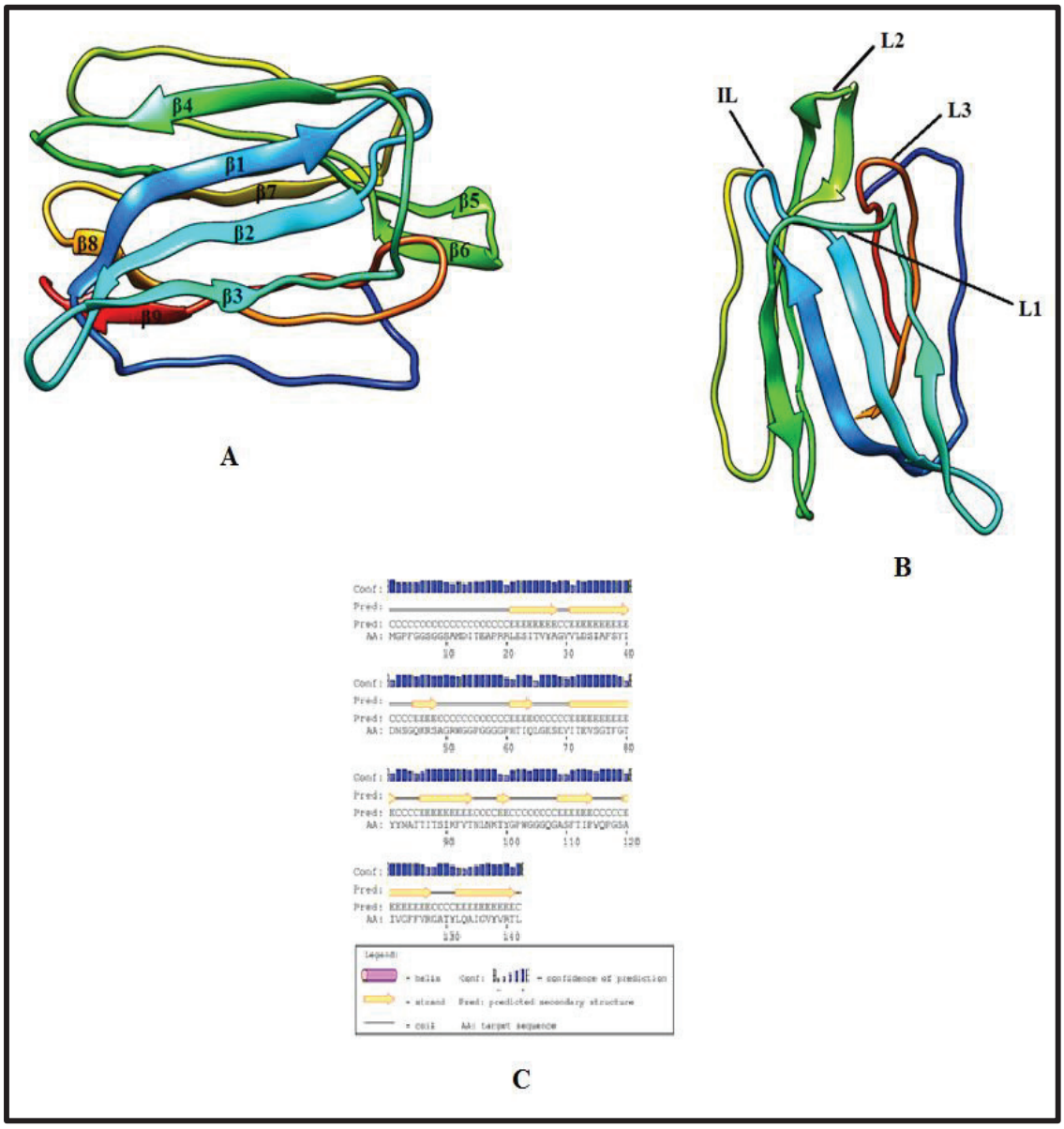


Figure 5.12 Homology modeling and secondary structure prediction of BGAF_Lec (A) Homology model of BGAF_Lec predicted by I-TASSER server with nine β-sheets arranged in 3 Greek key shapes. (B) Four loops namely L1, L2, L3, and IL, which forms the carbohydrate binding site. (C) Secondary structure predicted by PSIPRED Server.

A structural homolog of Lectin was obtained from DALI Server. BGAF_Lec showed the highest homology with banana lectin (PDB Id: 3MIT) with r.m.s.d. Values of 0.7 and Z-score of 24.3. The homologs of lectin and their respective r.m.s.d values are shown in Table 5.8

Table 5.9 Structural homolog of BGAF_Lec from DALI server.

PDB ID	Chain	RMSD Values	% Identity
3MIT	A	0.7	35
4PIF	C	0.9	36
2BMZ	B	0.7	35
3MIU	A	0.7	35
1X1V	A	0.7	35

5.5.9.4 Molecular docking analysis of BGAF_Lec

To find the amino acid involved in binding of D-fructose and GalNAc, we performed a molecular docking experiment using Autodock Tools. A homology model of BGAF_Lec generated by I-TASSER server was used. The coordinates of D-fructose and GalNAc was obtained from ZINC database. Since, most of the jacalin-related lectins have conserved binding site and previous studies showed Amino acid GGVLD was involved in carbohydrate binding (Kittur et al., 2009), this information was used to define the binding site. Results from these studies with D-fructose showed that Valine present at position 26th forms hydrogen bond with C6 of D-fructose. Besides valine 26, other residues such as Gly 25, Val 27, Asp 29, Gly 54, and Tyr 127 present in binding pocket interact with D-fructose via hydrophobic interactions and the calculated binding energy for this interaction was -4.0 Kcal/mol. In case of interaction of GalNac with lectin the orientation of GalNac is such that no hydrogen bond formation takes place but only hydrophobic interactions between different amino acids of binding site and GalNac stabilize the interaction with binding energy of - 2.9 Kcal/mol (Figure 5.13)

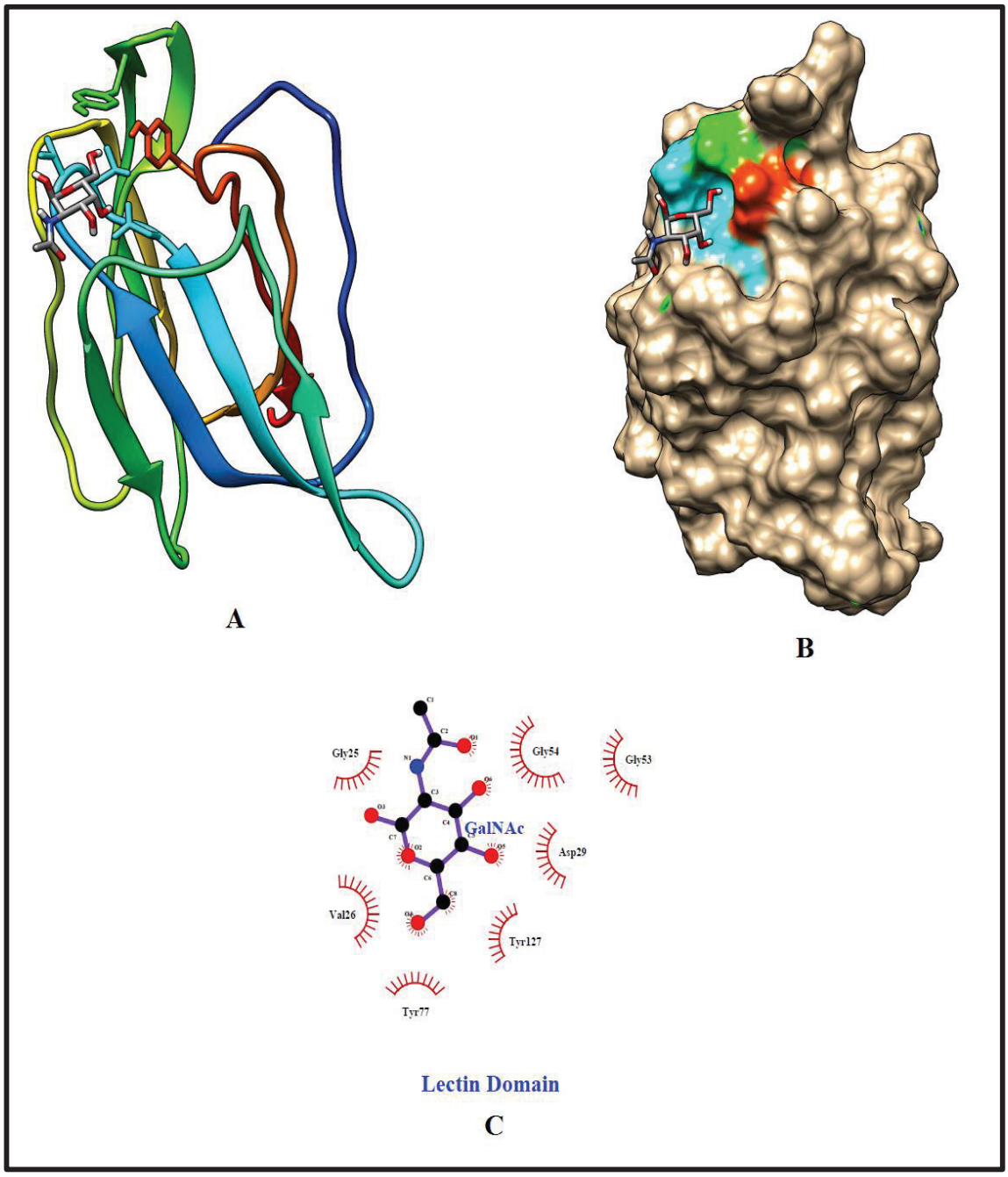


Figure 5.13 Molecular docking studies of BGAF_Lec with GalNac (A) Graphical representation of binding of GalNAc with BGAF_Lec. (B). The surface of lectin and pocket where GalNac binds with BGAF_Lec. (C) Graphical representation of hydrophobic interactions between GalNac and amino acid residues of lectin

5.6 Discussion

In this study, we have cloned and purified a lectin named BGAF_lec (Jacalin-like) which forms the c-terminal domain of β-Glucosidase aggregating factor from *Sorghum bicolor*. The gene for BGAF_Lec was amplified from the full-length gene of BGAF clone. Heterologous expression host strains of *E. coli* transformed with clones of the gene in pETM10 and pET28a vectors, in did not show any expression compared to control samples. To overcome this problem, BGAF_Lec gene was recloned in pET30a vector which did show significant expression of BGAF_Lec compared to control samples. BGAF_Lec was purified using Ni-NTA affinity purification technique. The final yield of purified protein from 200ml of LB media was 2.5 mg. Since all the three vectors used in this study are from the pBR322 origin, the lack of expression in two vectors (pET-28a and pETM10) and abundance of expression in pET-30a is unknown. One of the possible reason could be a lack of stability of first two vectors in *E. coli* host, which might lead to omission of a vector containing the gene of expression out of *E. coli* cells. This leads to propagation of *E. coli* cells without target vector and in the end lack of expression (Fakruddin et al., 2013).

The purified BGAF_Lec showed hemagglutination activity, and it was inhibited by D-fructose at very low concentration (0.5 mM), it was also inhibited by other carbohydrates such as N-Acetyl-D-galactosamine (50mM), mannose (20 mM), and fetuin (0.1%), which are similar to the results obtained in earlier studies (Kittur et al., 2009). To the best of our knowledge, this is the first report of jacalin-like lectin showing a strong affinity with D-fructose. In fact, an in-depth search of PDB database did not show any jacalin structure with D-fructose.

Results obtained from fluorescence spectroscopy studies contradicted the trends obtained in inhibition studies (Table). In fluorescence spectroscopy studies all the carbohydrate tested showed an almost similar binding affinity towards BGAF_Lec, while in the case of inhibition assay the D-fructose (0.5 mM) showed much higher affinity compared to other carbohydrates such as GalNac (50 mM).

Table 5.10 Comparison of Kd values and MIC (mM) values of different carbohydrate for BGAF_Lec

Carbohydrate	Kd	MIC (mM)
D-Fructose	$2.1X10^{-3}$	0.5
N-Acetyl-D-galactosamine	$3.5X10^{-3}$	50
Methyl-α-D-glucopyaranoside	$5.3X10^{-3}$	NI
Mannose	$6.3X10^{-3}$	20

One of the other contradictory trends in interaction of BGAF_Lec with GalNac was lack of change in tryptophan environment of BGAF_Lec. In Previous chapter we had shown when GalNac interacted with full-length BGAF, there was a change in the ʎmax leading to blue shift in fluorescence signal. While the addition of GalNac to BGAF_Lec did not elicit such change in ʎmax suggesting no huge change in tryptophan environment.

Secondary structure analysis through CD analysis showed BGAF_Lec contained mostly β-strand and loops which are consistent with all the known jacalin from other plants. Upon addition of sugar, the intensity of CD signal changes drastically but secondary structure elements remains same suggesting there is change in energy landscape upon ligand binding, but it did not alter the structure of the protein (Figure. 5.8)

Apart from biochemical and biophysical studies, bioinformatics was also used to analyze the obtained data. One of the important aspects of jacalin-like lectin is their sequence-structure correlation. Multiple sequence alignments of BGAF_Lec with other lectin domain of BGAFs from different plants and jacalin-like lectins from other plants showed conservation across plant species leading to their recognition as jacalin-related lectins. All Jacalin related lectins are having a β-prism-I fold, made of three β-sheets arranged into a prism. Sequence mapping according to secondary structure elements indicated that the amino acids in strand regions are highly conserved, while the amino acids

in loop regions are highly variable in their sequence and length helping generate diversity in carbohydrate binding. In-depth analysis of sequence-structure relationship revealed conserved residues lie most in 1st and 3rd sheet regions and of the β-prism and loop L3 (c-terminal of sequence), while the variable region, which is predominantly in first half of the sequence (N-terminal of sequence) and loop IL, L1, and L2 are major factor generating major diversity of carbohydrate recognition as seen in (Figure 5.10).

Another interesting variation in different JRLs is their quaternary association. Analysis of PDB database for the structure of different jacalin reveals a wide variety of quaternary association. The crystal structure of heltuba showed octameric arrangement on the other hand crystal structure of jacalin showed tetrameric arrangements. The diversity in the quaternary association has been attributed to the spread of exposed hydrophobic amino acid on the surface of different jacalin. The presence of hydrophobic residues on the surface which are exposed to outside environment on a subunit needs to be protected. Oligomerization helps in burying these amino acids and render stability to the overall structure and dictating the mode of quaternary association. To understand this phenomenon, hydrophobic surface in a subunit of jacalin, which is buried on the quaternary association, is shown in figure 5.15 Analysis of BGAF_Lec surface for the presence of exposed hydrophobic residues were done using Deep view- Swiss PDB Viewer (Guex and Peitsch, 1997). In this analysis, a hydrophobic residue having a solvent accessible surface area > 30 $Å^2$ was considered. The analysis showed less number of hydrophobic residues exposed on the surface of BGAF_Lec compared to the crystal structure of heltuba and jacalin (Figure 5.14). The

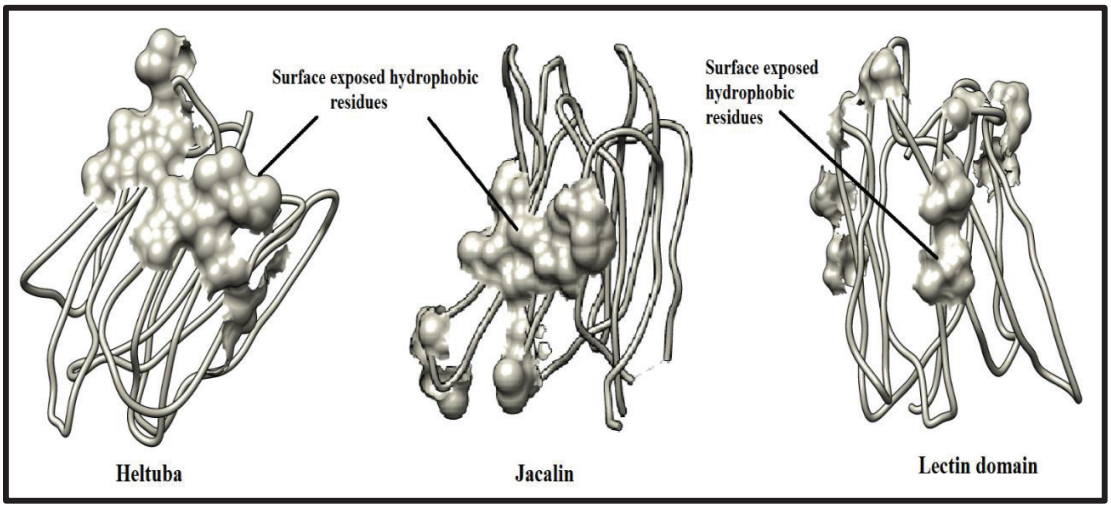


Figure 5.14 Analysis of surface exposed hydrophobic residues. Surface exposed hydrophobic residues on the surface of Heltuba, Jacalin, and BGAF_Lec, that have a solvent accessibility of > 30Å

less number of exposed hydrophobic residues on the surface of BGAF_Lec might be the reason for it to remain as a monomer in solution (Table 5.9).

Table 5.11 List of hydrophobic residues having solvent accessible surface areas >30 A°

Protein	Surface exposed hydrophobic residues with solvent accessibility of > 30Å
Heltuba	5_L, 17_I, 25_F, 46_V, 51_A, 57_A, 69_F, 71_A,75_M, 89_V, 92_P, 96_V, 101_F, 103_L, 104_P, 105_L, 114_F, 127_V, 128_V, 129_V, 130_P.
Jacalin	22_I, 25_L, 29_P, 30_F, 33_P, 39_L, 40_P, 44_V, 50_F, 51_P, 54_F, 64_P, 67_A, 68_L, 69_A,71_P, 73_P, 90_P, 99_F, 101_L, 102_P, 107_L, 121_A, 122_I, 124_V, 126_M, 127_A, 128_L.
BGAF_Lec	7_A, 8_M, 10_I, 17_L, 20_I, 22_V, 26_V, 40_G, 56_P, 59_I, 61_L, 110_P, 135_V, 138_L

It is well-established fact that loops which connect the strands play very important role in carbohydrate recognition and binding. Many crystal structures of JRLs with specificity towards galactose suggest that residues in L1 form a hydrophobic lid to the binding pockets (Raval et al., 2004) (Figure 5.16). In mannose binding JRLs this hydrophobic lead is absent, which allows binding of oligosaccharides. The lectin domain of all the BGAF considered for this study also lack the hydrophobic lead, which suggests evolutionary they belong to mannose binding JRLs. This observation is supported by the phylogenetic analysis where lectin domain of all BGAF are clustered with other monocot mannose-binding JRLs. In the L2 presence of aromatic amino acid in both mannose and galactose-specific JRLs has been observed (Figure 5.16). It is believed that size of the loop and orientation did not affect the binding of monosaccharides in anyway, but the spatial position of aromatic amino acid and orientation of loop can strongly influence the oligosaccharide binding. Superposition of BGAF_Lec with mannose and galactose-binding lectins and full-length BGAF demonstrate the above-mentioned observation very well. In Figure 5.16, it shows alignments of aromatic amino acid of banana lectin, jacalin, and BGAF_Lec, but in the case of BGAF, the orientation of aromatic amino acid changes substantially and also the orientation of loop is different from other known lectins. These changes in L2 might be the underlying cause of the difference in specificity of full-length BGAF and lectin domain of BGAF.

L3 plays a very important role in determining the lectin activity by providing the conformation and housing the aspartic acid involved in interactions with both galactose and mannose binding JRLSs. The high sequence conservation in this loop provides signature motif GXXXD to JRLs. BGAF_Lec lacks this signature motif in L3, and yet able to maintain the conformation like other well-known JRLs.

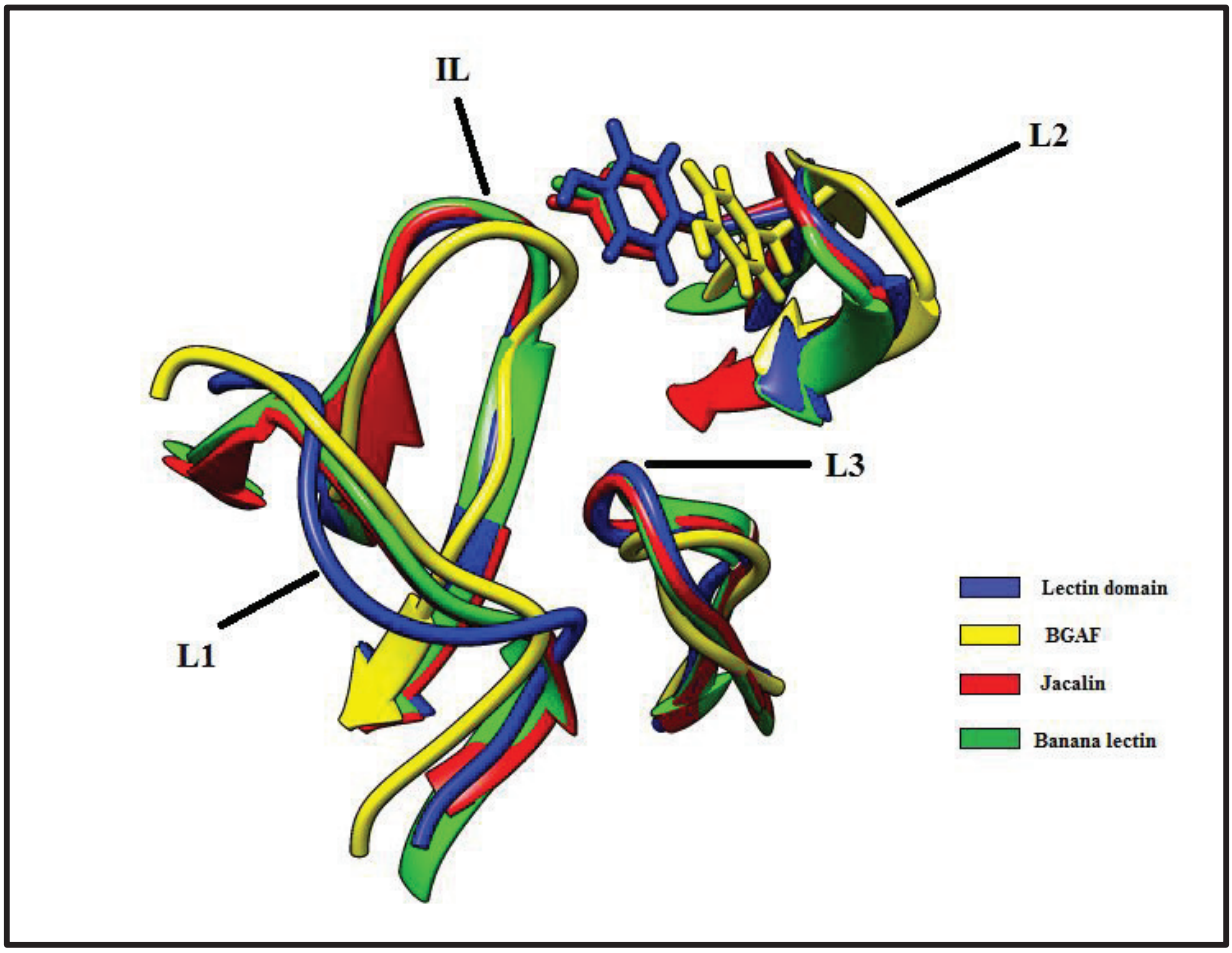


Figure 5.15 Superposition of loop region of BGAF_Lec Superposition of loop region of BGAF and BGAF_Lec with loop region of Jacalin and banana lectin.

5.7 Conclusion

Gene coding for lectin domain of BGAF from *Sorghum bicolor* was cloned into a bacterial expression vector. The BGAF_Lec was obtained in purified form via Ni-NTA affinity chromatography with final concentration up to 5mg/mL. Purified BGAF_Lec showed hemagglutination activity, and it was inhibited by D-fructose, mannose, N-Acetyl-D-galactosamine, and fetuin. The protein was stable up to 50°C and in pH range 5-9. The fluorescence spectroscopy studies showed that D-fructose has a similar affinity towards BGAF_Lec as GalNac, contrary to the observation of inhibition of hemagglutination assay showed. Fluorescence spectroscopy also revealed affinity of BGAF_Lec for Methyl-α-D-

glucopyaranoside, which was not detected by hemagglutination inhibition assay. Secondary structure analysis with CD spectroscopy revealed a change in CD signal upon addition of carbohydrate, but it did not affect the secondary structural elements of protein. Crystallization trials of BGAF_Lec with different buffer condition did not yield any protein crystals. Bioinformatics analysis showed BGAF_Lec belongs to mannose-specific JRLs. Molecular docking study showed that BGAF_Lec interacts with D-fructose via hydrogen bonds, and with GalNac through hydrophobic interactions.