CHAPTER 4

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

4.1 Summary

A protein containing N-terminal dirigent domain and c-terminal lectin domain from *Sorghum bicolor* was cloned, expressed in *E. coli* strain and purified using Ni-NTA affinity chromatography. The purified protein is known as a beta-glucosidase aggregating factor (BGAF) is a monomer of 38kDa and gel filtration analysis showed that it exists as a monomer in solution. It is stable in pH range 5-6 and 8-9 and despite being a recombinant protein, it is thermally stable up to 30 minutes at 60°C. The hemagglutination activity of BGAF is strongly inhibited by N-Acetyl-D-galactosamine and fetuin. Mannose, fructose, and Methyl α-D- mannopyranoside also inhibits hemagglutination activity of BGAF but at a much higher concentration. Fluorescence spectroscopy shows that the intrinsic fluorescence of purified BGAF was quenched by different carbohydrates and N-Acetyl-D-galactosamine has highest effect leading to blue-shift effect and binding dissociation value Kd of $6.3X10^{-5}$. Circular dichroism studies showed BGAF is a mixture of α- helices and β-strands and in the presence of N-Acetyl-D-galactosamine the secondary structure changes and percentage of β-strand increases by up to 10%. BLAST analysis showed that it is homologous to other very well-known β-glucosidase aggregating factors and share 67% of sequence similarity with BGAF from maize.

4.2 Introduction

Over the course of evolution emergence of multi-domain proteins have played a vital role in a single protein exhibiting large numbers of functions. They may have different binding partners, catalytic activity, may exist and become a part of a regulatory or structural network where multiple interactions (protein-protein, protein-RNA, protein-carbohydrates, etc.) are important. β-Glucosidase (E.C 3.2.1.21) is an enzyme which catalyzes the hydrolysis of aryl or alkyl β-D-glucosides and β-linked oligosaccharides and releases glucose and an aglycone. These enzymes occur all across the living kingdom. The natural substrates of β-glucosidase in plants are cyanogenic and hydroxamic acid β-glucosides of plant secondary metabolites and β-linked oligosaccharides released from the digestion of plant cell wall during germination (Niemeyer, M., 1988; Cuevas et al., 1992 & Leah et al., 1995). Most β-glucosidase exhibit broad specificity towards aglycone moiety of their substrate but very narrow specificity towards glycone moiety, but most of the β-glucosidase from plants have a narrow specificity towards aglycone portion as well. The aglycones play a very important role in plant defense, growth, and development. Cyanogenic β-glucosides upon hydrolysis releases respiratory poison HCN thus plays a very important role in plant defense against pathogen and herbivores. The shoots of young sorghum plants contain large amounts of cyanogenic glucoside dhurrin (p-hydroxy-(s)-mandelonitrile-pD-glucopyranoside) and hydrolysis of dhurrin by β-glucosidase cleaves it into the glycone and aglycone moieties. This aglycone part is further processed by α-hydroxynitrile lyase and releases the hydrocyanic acid and HCN. Hydrocyanic acid has long been considered to play an important part in plant defense (Hruska, 1988; Hughes et al., 1982).

In certain verities of maize, β-glucosidase exists as part of large quaternary aggregates (Esen et al., 2000). Initially, these verities were thought as null mutant for of β-glucosidase. The β -glucosidase zymograms of such genotypes are devoid of enzymatic activity, but further studies showed the presence of β-glucosidase activity in solution and identification of a 60-kDa polypeptide reacting specifically with anti β -glucosidase sera on immunoblots (Esen et al., 2000). The reason for these phenomena was the presence of a β-glucosidase aggregating factor (BGAF) which specifically interacts with β-glucosidase and prevents it from entering the zymogram. Further, Bioinformatics analysis revealed

presence of BGAF like the sequence in other plants belonging to grass family (Kittur et al., 2007; Blanchard et al., 2000; Esen et al., 2000).

BGAF belongs to jasmonic acid-induced protein family and is a modular protein consisting of two domains; the N-Terminal dirigent or disease response domain and C-terminal jacalin-like lectin domain. (Kittur et al., 2007). Pfam database showed at least 72 sequences having an N-terminal dirigent domain and C-terminal lectin-like domain, and all these 72 sequences are distributed in grass family poacea.

Sequence comparison between cDNA of β-glucosidase from maize and sorghum shows 70% identity. Despite sharing high sequence identity, they differ in their capacity to bind with BGAF (Kittur et al., 2009; Kittur et al., 2007). Both isozymes of β-glucosidase in maize binds to BGAF but β-glucosidase from sorghum does not bind to BGAF and the reasons for this lack of binding are yet to be answered. It is thought that by aggregating the β-glucosidase, BGAF prevents their degradation by proteases released by pests and studies show that the lectin domain of BGAF may be responsible for aggregation of β-glucosidase (Kittur et al., 2007). The role of dirigent domain in aggregation and the interactions between β-glucosidase and lectin domain, are not well understood and need to be answered and to understand this phenomenon, we have characterized BGAF from *Sorghum bicolor* in this study.

4.3 Material

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 Corporation, Bangalore, Karnataka, India. 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 BGAF was synthesized and obtained from genescript Inc, USA. Auto-induction media was prepared according to studier method (Studier F.W., 2005).

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, MSU Baroda, and Dr. Ashima Bhardhwaj, IIAR, respectively and pETM10 was obtained from EMBL, Heidelberg, Germany.

4.4 Methods

4.4.1 General Cloning protocols

4.4.1.1 Preparation of competent cells

A single colony of the desired *E. coli* strains maintained on LB plate was inoculated into 10ml of LB media and cultured overnight at 37^0C. 50ml of LB media was inoculated with 0.5ml of this inoculum and cultured until an OD of 0.5 was obtained at 600nm. The cells in the culture were chilled on ice for 30 minutes were centrifuged gently (3,000 X g) at 4^0C and the pellet was re-suspended in half the culture volume containing 0.1M $CaCl_2$ additionally and was incubated on ice for 15 minutes and then centrifuged at 3,000 X g at 4^0C. The pellet was finally resuspended in 2.5ml of 0.1 M $CaCl_2$ and adding 15% ice cold glycerol (final volume) and cells were aliquoted and stored at -80^0C

4.4.1.2 Transformation of *E. coli* cells

For transformation, either plasmid DNA (50-100 ng) or the ligation buffer mix was added to 100µL of competent *E. coli* cells and incubated on ice for 30 minutes. Subsequently, the cells were subjected to heat shock at 42^0C for 1 minute and then cooled on ice for 5-10 minutes. 1ml of LB medium was added to it, and the cells were cultured at 37^0C for 1 hour and then plated on LB-agar plates containing antibiotics as per vector and strain and incubated at 37^0C for 12 to 18 hours. Individual colonies were picked and screened for transformation.

4.4.1.3 DNA isolation

Discrete colonies obtained on plated LB media were inoculated in 10ml of LB media containing desired antibiotics and cultured at 37^0C for 20 hours. The cells were harvested by centrifugation at 8,000 X g for 5 minutes. The cell pellet was resuspended in 200µL P1 solution (50 mM Tris-HCl pH 8.0, 10 mM EDTA and 100 µg/mL RNase) and the cells were lysed using 200µl P2 solution (0.2 N NaOH and 1% SDS). Following lysis 200µL P3 solution (3 M potassium acetate, pH 5.5) was added and resulting suspension subjected to centrifuge at 13,500 X g for 20 minutes. The supernatant separated, and the DNA present in the supernatant was precipitated by adding 0.8 volumes of isopropanol and centrifuged at 13,500 X *g* for 15 minutes. The pellet obtained was washed with 70% ethanol to remove excess salts and air dried. The DNA was re-suspended in sterile double distilled water or Tris-HCl buffer and analyzed by agarose gel electrophoresis.

4.4.1.4 Agarose gel electrophoresis

The extracted DNA was analyzed by agarose gel electrophoresis. Depending on the percentage of the gel, agarose was weighed and added to the required volume of 1 X TAE, heated in a microwave oven and cooled to 55°C. Ethidium bromide was added to final concentration of 0.5µg/ml, mixed well and poured into the gel trough and allowed to solidify. Samples were mixed with sample buffer and loaded into the well, the gel electrophoresis conducted for 45-60 minutes in 1X TAE buffer at 100V and finally, DNA was visualized under a UV-trans illuminator.

4.4.1.5 Restriction Digestion

Restriction digestion of DNA and plasmid was performed according to the instructions provided by the manufacturer (Thermo Fisher Scientific, Framingham, MA, USA.). The enzymes EcoRI and NotI were used in double digests with compatible buffer O as described in the manual provided by the manufacturer.

4.4.1.6 Agarose Gel Extraction

DNA was purified from a band excised from an agarose gel using the QIAquick Gel Extraction Kit from Qiagen (Hilden, Germany). The protocol provided by the manufacturer was used. For these experiments, the samples were analyzed using the 1.2% low melting agarose (Sigma-Aldrich, USA) which helps in increasing the quantity and purity of extracted DNA. The increase in quality and quantity is because it is easier to remove the low melting agarose from the samples compared to normal agarose, which remains as trace with the extracted DNA.

4.4.1.7 Ligation Reaction

Insert and plasmid DNA were combined in a ratio of approximately 3 moles insert to 1 mole plasmid, and placed on ice. 10X T4 ligase buffer (Roche, Basel, Switzerland) was thawed on ice, and diluted with sterile deionized water to make 1X and added to a concentration of 0.35U/µl to the above reaction mixture. This mixture was then incubated at various temperatures for various time durations to derive the best condition for this particular experiment

4.4.1.8 Cloning of BGAF in pET vectors for heterologous expression

Gene encoding the full-length protein of BGAF (GeneBank id DQ866804.1) was synthesized and obtained from Genescript inc., USA. The synthesized gene was codon-optimized for expression in heterologous *E. coli* host and two restriction sites 5` EcoRI and 3` Not I was introduced in gene sequence to sub-clone it in expression vectors. The plasmid (pUC57) carrying the full-length gene for BGAF was double digested (EcoRI, NotI) and excised full-length gene was 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 sequence similarity. Once the clone was confirmed to be of BGAF and in the right orientation, it was transformed to BL21(DE3) pLysS strain for expression trials.

4.4.2 Heterologous expression of BGAF in E. coli strains

Two *E. coli* strains BL21(DE3) & BL21(DE3) pLysS were first transformed with vectors carrying the gene for BGAF and plated on 1.5% LB agar containing Kanamycin (50µg/ml) and kanamycin (50µg/ml) plus chloramphenicol (34µg/ml), respectively and incubated at 37°C overnight. Next day a single colony from the plate was used to inoculate 4ml of LB media containing kanamycin (50µg/ml) plus chloramphenicol (34µg/ml). This culture was grown overnight at 37°C, and a 0.2 ml aliquot of this culture was then used to inoculate 20ml of LB media with kanamycin (40µg/ml) plus chloramphenicol (20µg/ml). This culture was grown at 37°C with shaking at 220 rpm for 3 hours or till O.D reached 0.6 and then the culture was induced for gene expression by using different concentrations of IPTG (0.2 to 0.6 mM) at different temperature (20°C, 24°C and 37°C) for expression of BGAF. The gene expression was also optimized using autoinduction media (Studier F.W., 2000). The cells were harvested by subjecting the culture to centrifuge at 10000 X *g* for 15 minutes at 4 °C and collecting the pellet which was lysed using lysis buffer (50 mM Tris-HCl pH 8.5, 200 mM NaCl, 20 mM imidazole, 5% glycerol, 0.1% Triton X 100 & 1mM PMSF). Sonication was performed on lysis mixture incubated in an ice bath, using 5X30-second pulses of sonication with 15 seconds cooling time and finally the sample was subjected to centrifuge at 12000 X *g* at 4°C for 20 minutes, and supernatant and pellet were stored at -20°C.

4.4.3 Affinity purification of BGAF

Purification was achieved by exploiting the hexahistidine tag fused to protein. 3ml nickel affinity matrix, Ni-NTA acquired from Sigma-Aldrich was packed in a column, and the column was equilibrated in lysis buffer and cell lysate was loaded onto the column. The column was washed with lysis buffer plus imidazole (50 mM) to remove the nonspecifically bound protein, and final elution was achieved with the lysis buffer containing 300 mM imidazole.

4.4.4 Molecular mass determination of BGAF

4.4.4.1 Sodium dodecyl sulfate - Polyacrylamide gel electrophoresis (SDS-PAGE) of BGAF

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.

4.4.4.2 Gel filtration studies of BGAF

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

4.4.5 Peptide mass fingerprinting for protein identification

The excised band from the SDS-PAGE gel was destained and in-gel digested with trypsin. The extracted trypsinized peptides were then subjected to MALDI-TOF/MS/MS using AB SCIEX QSTAR© elite LC-MS/MS system (AB SCIEX, Framingham, MA, USA) at Sandor Proteomics, Hyderabad, India. The peptides fragment peak list generated by 4000 Series Explorer software was searched against the MASCOT search engine (http://www.matrixscience.com) using the protein pilot software to identify them.

4.4.6 Hemagglutination activity and carbohydrate specificity of BGAF

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

4.4.7 Effect of Temperature and pH on lectin activity of BGAF

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

4.4.8 Carbohydrate binding analysis of BGAF using fluorescence spectroscopy

The binding association of different carbohydrate with positive inhibition of hemagglutination activity was further analyzed by fluorescence spectroscopy. Titrations

were performed by adding aliquots (1µl at each time) of test sugars from stock (1M) in 100mM Tris-HCl (pH 8.5) to 1ml of purified protein (0.1mg/ml) in the same buffer. The temperature was maintained at 30°C throughout the experiment and samples were mixed thoroughly with pipetting, making sure that no bubble formation occurred. Corrections were also made for dilution effects and at the highest concentration of the 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).

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

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 protein solution and the protein at a sugar concentration [C] respectively, 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:

$$-\Delta G = RT \ln(Ka) \quad \text{(Eq. 4.2)}$$

4.4.9 Determination of Secondary structure elements of BGAF via circular dichroism spectroscopy

To record the CD spectra, a Jasco J-810 (Jasco corp., Tokyo, Japan) polarimeter equipped with Peltier thermostat was used. Two concentrations of purified protein (5µM and 25µM) in 50mM sodium phosphate buffer (pH 8.5) were placed in a 2mm path length rectangular quartz cell and spectra recorded at a scan speed of 20nm/minute with a response time of 4s and 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 in the absence as well as in the presence of 1mM galactose also separately. The structural changes that occurred in the protein during the thermal unfolding of the protein were determined by recording the CD spectra between 25°C and 90°C, at a scan rate of 1°C/min and by monitoring the changes in ellipticity (Θ) at 214nm.

4.4.10 Crystallization trials and data collection of BGAF

The initial crystallization trials 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. Besides crystallization screens a homemade screen of different series of PEGs were also used. The different buffer conditions were set up in the micro batch 60 wells plate acquired from griener biologics. The plates were filled with total of 8ml of oil mixture of paraffin and silicone oil in ratio of 1:1 and the protein and different buffers were added to the wells of plates in different ratio of (1:1,1:2, 1:3) at different protein concentration generating a wide range of different conditions for crystallization.

4.4.11 Antimicrobial assay of BGAF

To check the effect of protein on the growth of three bacterial species (*E. coli*, *S. aureus,* and *P. aeruginosa*) antimicrobial assay was carried out in 96 well plates. 5ml LB was inoculated with a colony of the bacteria cultured on LB plates and incubated with shaking at 37°C till the OD_{600} reached 0.1 (10^5-10^6 cells/mL). This culture were used as inoculum for the antibacterial assay. Purified protein (1mg/ml) in PBS was used in two final concentrations of 0.25 mg/mL and 0.5 mg/mL. The purified protein was diluted in Muller Hinton broth (MHB) and 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 3 times.

4.4.12 Bioinformatics analysis of BGAF

The homology search to find out the homologous 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 and 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, and auto dock 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/)

4.5 Results

4.5.1 Gene synthesis and cloning of BGAF gene in pET vectors

The full-length gene of BGAF from *Sorghum bicolor* was obtained from Genescript Inc., U.S.A in pUC57 vector and the 918 bp fragment obtained was cloned in three different pET series of vectors (Table 4.1). Positive clones of transformed DH5α *E. coli* cells were selected by the gel shift assay and restriction digestion (Figure. 4.2) and the gene sequence from isolated plasmids in positive clones (Lane 3, Figure. 4.2) sequenced. The sequence obtained for 918 bp matched exactly with the synthesized gene exactly (Figure. 1).

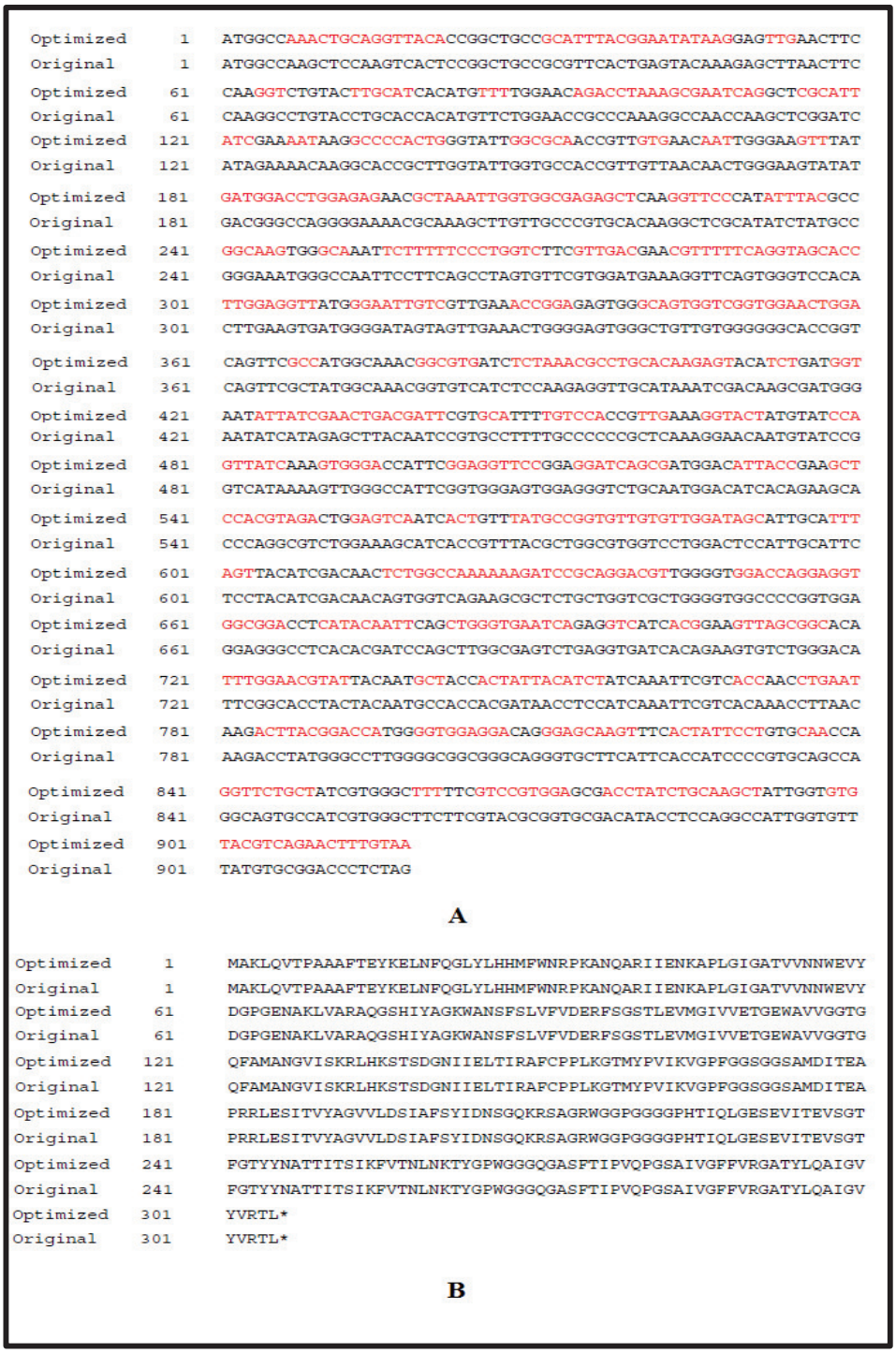
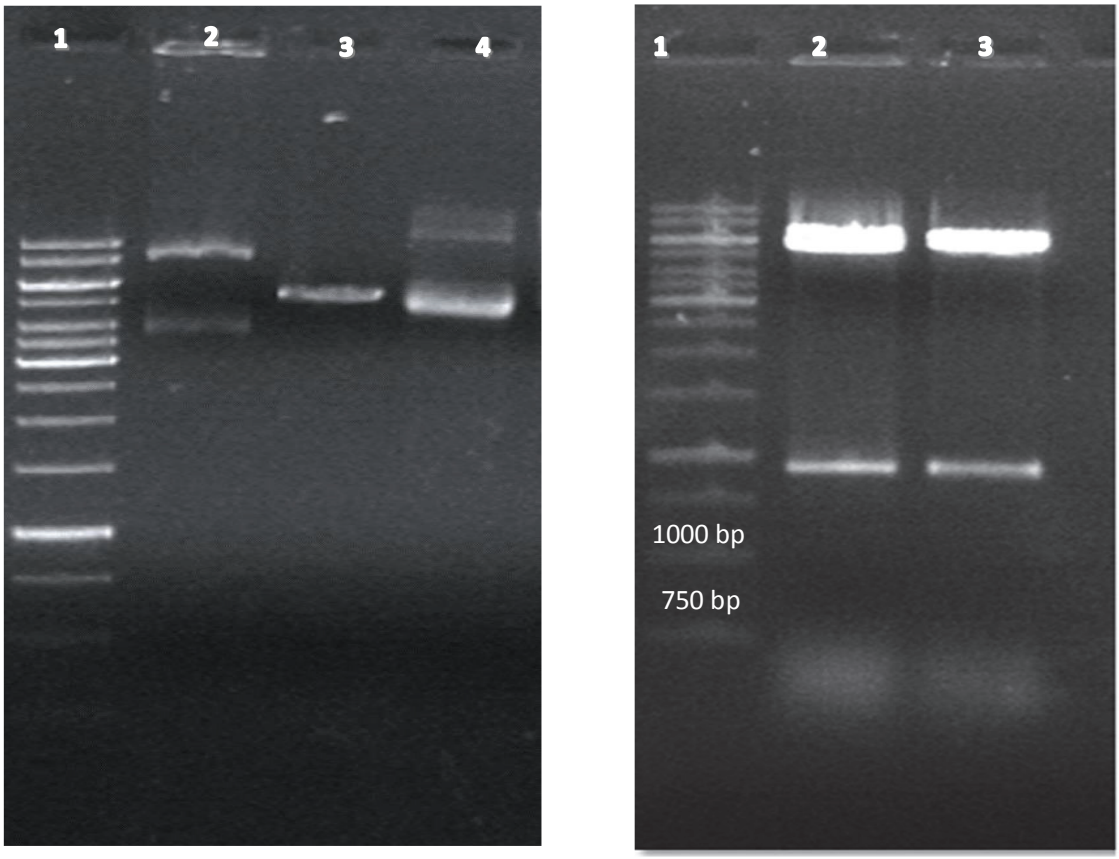


Figure 4.1 DNA and Protein sequence of BGAF (A) Codon optimized DNA sequence of BGAF. Change in codons is shown in red. (B) Translated protein sequences from Original BGAF gene sequence and optimized gene sequence



(A) (B)

Figure 4.2 Agarose gel profile of BGAF clone. (A) Plasmid DNA isolated from the positive clones of transformed DH5α cells and lane 2 and 3 showed retarded mobility. (B) Lane 2 and 3 shows insert released from the clone with retarded mobility.

4.5.2 Expression trials of BGAF

Initially expression trials of BGAF cloned in pETM10, and pET28 vectors were carried out in LB medium with different concentration IPTG and varied temperatures. After induction with IPTG and appropriate incubation at different temperatures, cells were lysed. SDS-PAGE analysis of lysed samples was carried out, and results showed no detectable expression of BGAF on SDS-PAGE. Expression trials in autoinduction media at different temperature also yielded the similar results. But expression trials of BGAF cloned in pET30a vector in LB and autoinduction media showed significant expression of BGAF protein at correct size in SDS-PAGE gel analysis (Figure 4.3) and summary of expression trials at different conditions are detailed in Table 4.1

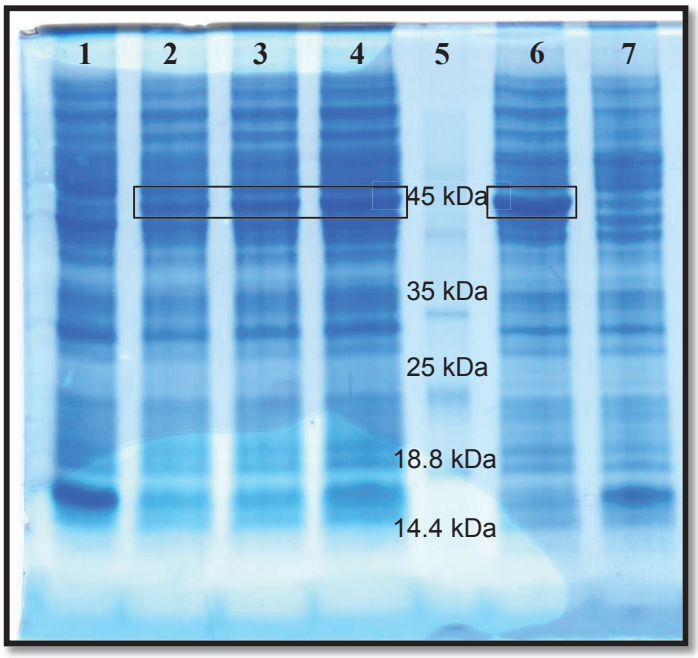


Figure 4.3 SDS-PAGE analysis of Expression trials of BGAF BGAF gene cloned in pET30a vector in BL21 (DE3) PlysS strain. Lane 1 shows samples where no IPTG is added, and it is used as a control. Lane 2-4 shows induction by IPTG at 0.2mM, 0.4mM and 0.6mM, respectively at 24°C. Lane 5 shows molecular weight markers. Lane 6 shows sample from autoinduction media after overnight growth at 24°C and lane 7 shows sample which was taken before induction takes place (after 5hour growth) in autoinduction media.

Table 4.1 Summary of expression trials of BGAF in different pET vectors. BGAF gene 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**									**Autoinduction 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	24°C	37°C
BL21(DE3)	pETM10	-	-	-	-	-	-	-	-	-	-	-
	pET28a	-	-	-	-	-	-	-	-	-	-	-
	pET30a	-	-	-	-	-	-	-	-	-	+	+
BL21(DE3) pLysS	pETM10	-	-	-	-	-	-	-	-	-	-	-
	pET28a	-	-	-	-	-	-	-	-	-	-	-
	pET30a	+	+	+	+	+	+	+	+	+	+	+

4.5.3 Affinity purification of BGAF

Purification of His-tagged BGAF was carried out through a nickel affinity column and eluted samples analyzed on SDS-PAGE gel, and the purified BGAF found as a single band corresponding to its molecular weight of 36 kDa (Figure 4.4)

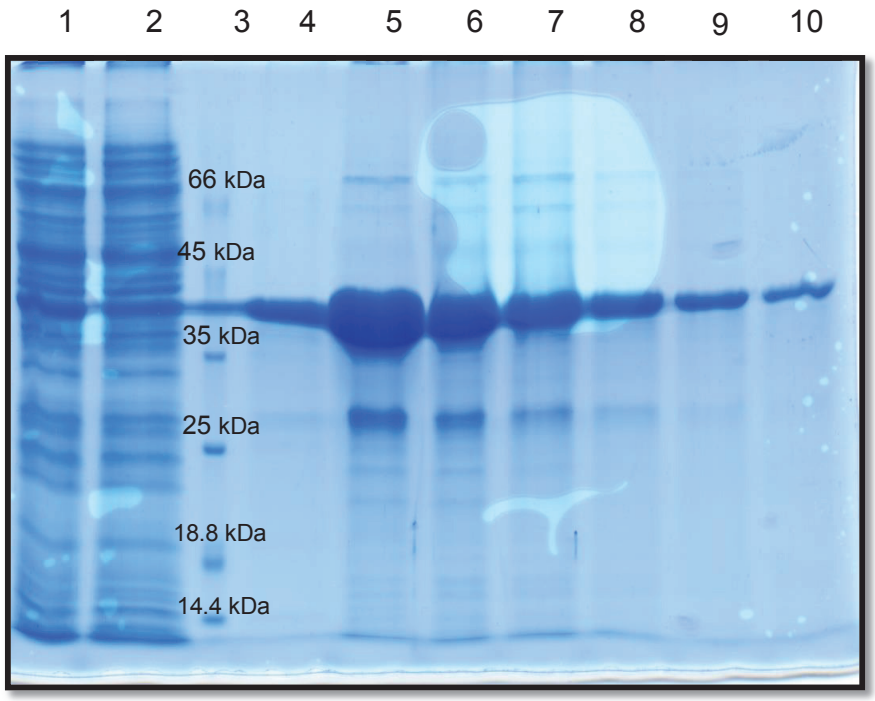


Figure 4.4 SDS-PAGE analysis of purification process of BGAF. Lane 1 shows sample from flow through. Lane 2 Shows sample collected after washing the column with washing buffer. Lane 3 shows Molecular weight markers. Lane 4-10 shows fractions of purified BGAF eluted from the column with buffer containing 300 mM imidazole.

4.5.4 MALDI-TOF analysis for protein identification

To confirm the identity of purified protein, MALDI-TOF analysis was carried out on gel-band cut from the SDS-PAGE gel (Figure 4.4, lane 8) and the spectra obtained from MALDI-TOF the analysis of spectra was carried out with Mascot search analysis. The results showed the purified protein was a Beta-glucosidase aggregating factor from *Sorghum bicolor* with a score of 105, which was statistically significant (Figure 4.5).

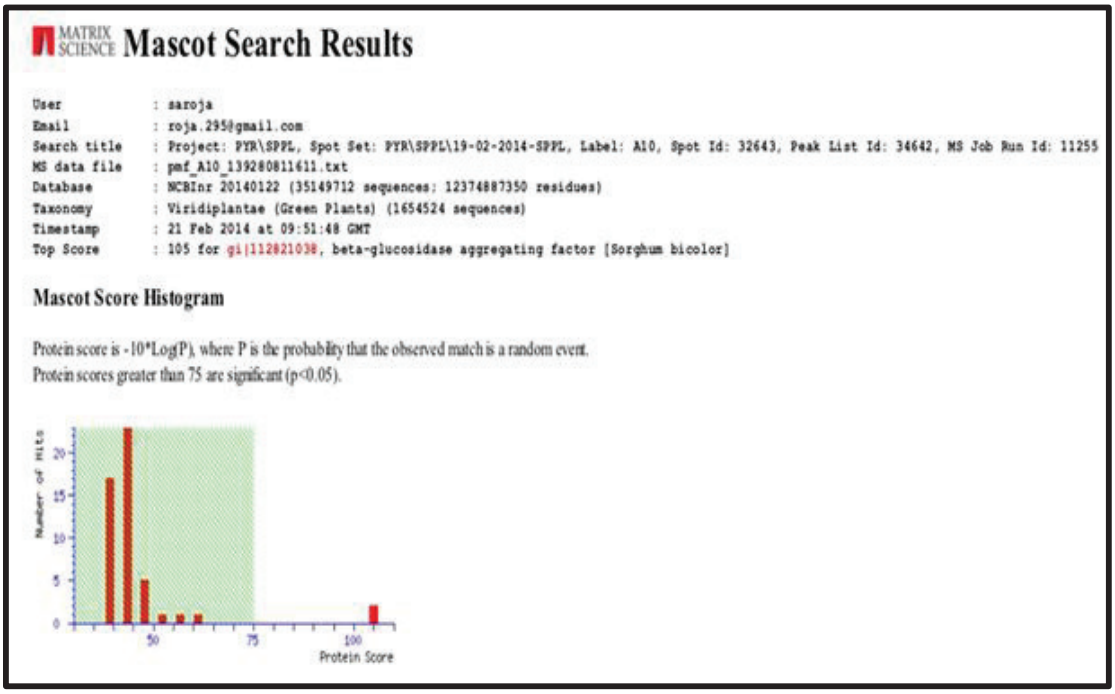


Figure 4.5 MALDI-TOF analysis for identification of BGAF MALDI-TOF analysis showed purified protein to be a beta-glucosidase aggregating factor from Sorghum bicolor with a score of 105.

4.5.5 Hemagglutination activity and Carbohydrate specificity of BGAF

BGAF showed hemagglutination activity with 1% rabbit erythrocytes. The hemagglutination activity of BGAF was inhibited strongly by N-Acetyl-D-galactosamine (1 mM). Fructose, Mannose, Methyl-α-D-mannopyranoside also showed inhibition but at a higher concentration. The hemagglutination activity was also inhibited by fetuin with a minimum inhibitory concentration (MIC) of 0.1 % (Table 4.2)

Table 4.2 Inhibition of hemagglutination activity of BGAF. Inhibition of hemagglutination activity of BGAF by different carbohydrates and all the concentrations are mentioned in mM. (NI stands for no inhibition).

Carbohydrates	Minimum Inhibitory Concentration (MIC) mM
Mannose	75
Galactose	NI
Lactose	NI
Glucose	NI
Fructose	80
N-Acetyl-D-galactosamine	1
N-Acetyl-D-glucosamine	NI
Methyl- α-D-mannopyranoside	70
Methyl- α-D-glucopyaranoside	NI
Fetuin	0.1%

4.5.6 Effect of pH and temperature on hemagglutination activity of BGAF

The hemagglutination activity of BGAF was retained between pH 5-6 and 8-9 (Figure 4.6) and was stable until 60°C and heating about 60^0C lead to unfolding and subsequent precipitation of BGAF.

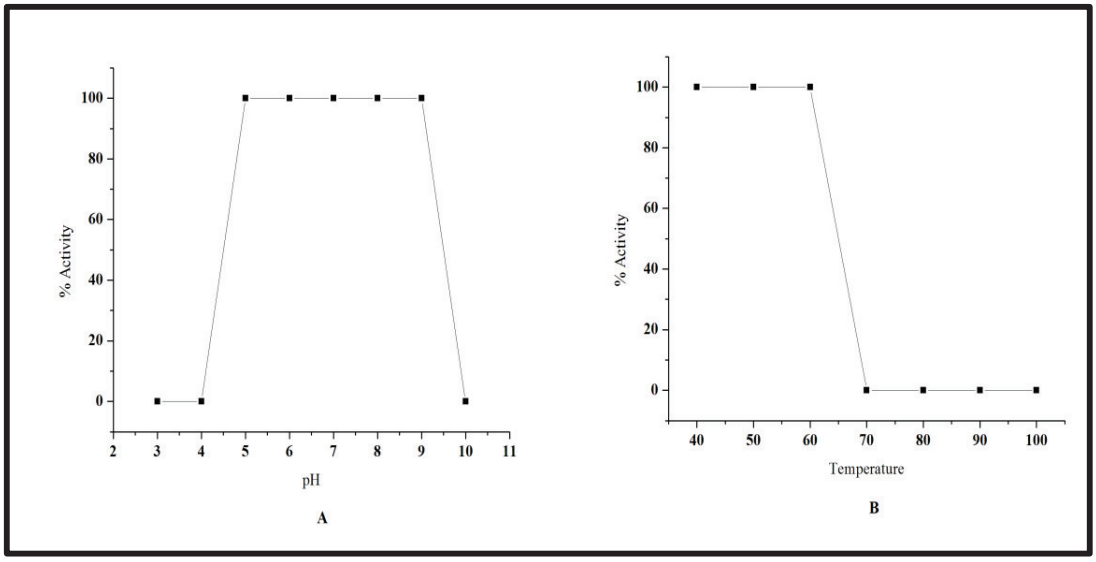


Figure 4.6 Effect of pH and temperature on hemagglutination activity of BGAF Graphical representation of effect of pH (A) and temperature (B) on hemagglutination activity of BGAF.

4.5.8 Carbohydrate binding analysis of BGAF through fluorescence spectroscopy

Upon excitation at 280nm BGAF gave an emission spectrum centered at 342nm. Titration of BGAF with sugars resulted in quenching of the lectin fluorescence without any change in emission maxima except N-Acetyl-D-galactosamine7. Titration of BGAF with N-Acetyl-D-galactosamine resulted in a decrease in the emission intensity by 30-50% along with 8nm blue shift in the emission maximum (Figure 4.7 A). Double logarithmic plots are shown in Figure 4.7 C, and 4.8 C were used to determine the dissociation constant of BGAF for respective carbohydrates. The slope of these graphs was found to be ~ 1 (1.14), indicating each lectin subunit binds one saccharide molecule. The fluorescence titration data for the interaction of other carbohydrates were also analyzed in a similar manner, and association constant corresponding ΔG values for binding are described (Table 4.3).

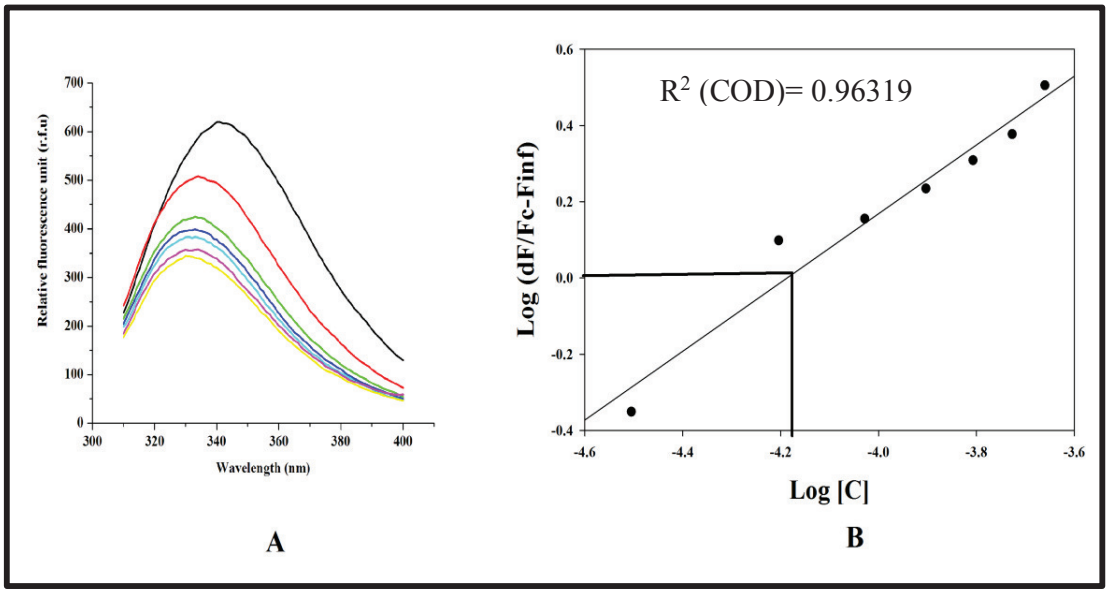


Figure 4.7 Fluorescence spectra of BGAF in absence and presence of GalNac (A) Fluorescence spectra of BGAF in the absence (Black) and after addition of defined aliquots from the stock of N-Acetyl-D-galactosamine. (B) Double logarithmic plot for binding of N-Acetyl-D-galactosamine with BGAF

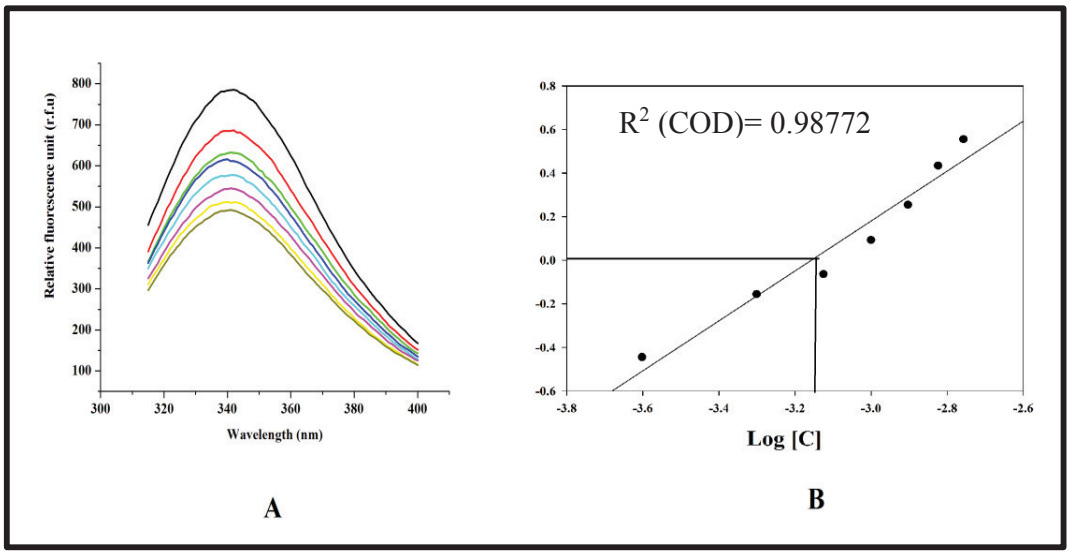


Figure 4.8 Fluorescence spectra of BGAF in absence and presence of Methyl-α-D-mannopyranoside (A) Fluorescence spectra of BGAF in the absence (Black) and after addition of defined aliquots from the stock of Methyl-α-D-mannopyranoside (B) Double logarithmic plot for binding off with Methyl-α-D-mannopyranoside BGAF. The plots were obtained using Chipman analysis.

Table 4.3 Dissociation constants, Association constants and corresponding Gibb`s free energy values for BGAF with different carbohydrates

Carbohydrate	Dissociation Constant (Kd)	Association Constant (Kb= 1/Kd) M^{-1}	ΔG°b (Kcal.mol-1)
N-Acetyl-D-galactosamine	$6.3X10^{-5}$	$1.5873X10^{4}$	-24.66
Methyl-α-D-mannopyranoside	$6.9X10^{-4}$	$1.449X10^{3}$	-18.54
Galactose	$5.4X10^{-3}$	$1.8518X10^{2}$	-13.31
Mannose	$6.3X10^{-3}$	$1.587X10^{2}$	-12.90

Inhibition studies of hemagglutination assay with different sugars suggested fructose inhibits the activity of BGAF at 80mM, but fluorescence spectroscopy with fructose did not show binding of fructose with BGAF. Galactose did not inhibit hemagglutination activity of BGAF, but fluorescence spectroscopy studies of BGAF with galactose shows spontaneous binding of galactose with BGAF. These results suggest a limitation of hemagglutination assay to determine the various binding partner of lectins.

4.5.9 Determination of Secondary structure elements of BGAF via circular dichroism spectroscopy

Circular dichroism spectra of BGAF in native state (Solid line) and in the presence of N-Acetyl-D-galactosamine (dotted line) are shown in Figure 4.9A. The far-UV spectrum of native BGAF is characterized by a minimum near 212nm with zero crossing near 203nm. In the presence of N-Acetyl-D-galactosamine, the minimum near 212nm shifted towards 218nm with a decrease in intensity. These changes in spectra suggested a change in secondary structure of the protein, after addition of N-Acetyl-D-galactosamine.

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 4.9 B) of BGAF 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 is compact and rigid.

Data from CD melt studies indicated that the protein was stable until 65°C but beyond 65°C, it unfolded and precipitated (Figure 4.9 C). These findings are by the observation of temperature stability assay; wherein BGAF showed hemagglutination activity till 60°C.

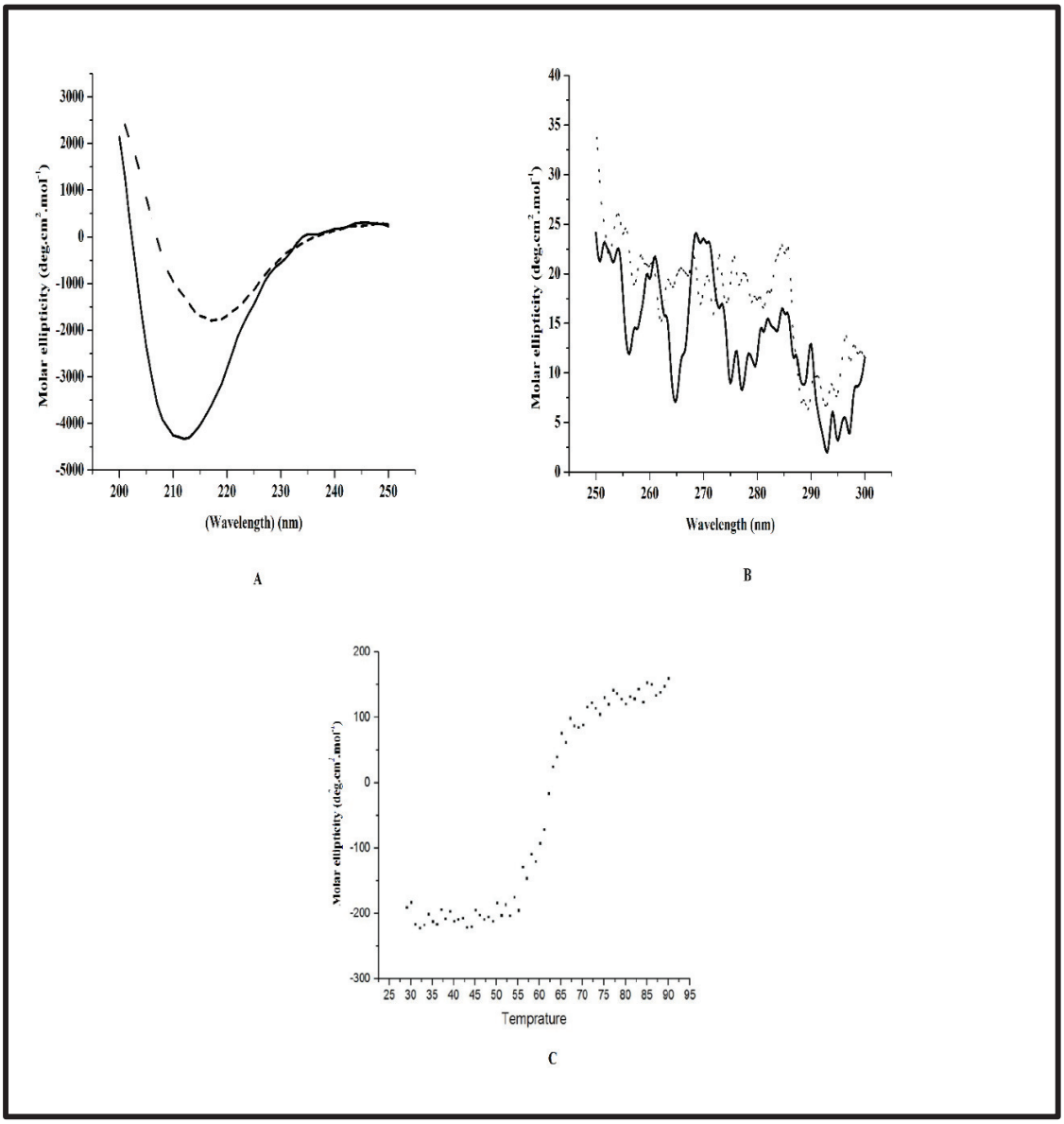


Figure 4.9 Circular Dichroism (CD) spectra of BGAF in absence and presence of GalNac CD and CD melt analysis of BGAF in absence and presence of carbohydrate. (A) Far-UV spectra of BGAF (B) Near UV spectra of BGAF. The solid line represents BGAF in the buffer and the dotted line represents BGAF in complex with N-Acetyl-D-galactosamine. (C) CD melt analysis of BGAF monitored at 218 nm as a function of temperature.

To derive information on the content of secondary structure elements of BGAF, 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 4.5. To check the effect of N-Acetyl-D-galactosamine on the secondary structure of BGAF, far UV spectrum collected in the presence of galactose was also subjected to analysis with above-mentioned programs in CDPro with similar parameters. The results obtained are given in Table 4.5. Data obtained from these analyses suggested, N-Acetyl-D- galactosamine conferred a significant change in secondary structure of BGAF, which is consistent with spectral analysis

Table: 4.4 Calculated secondary structure elements of BGAF using CDPro suite and prediction by PSIPRED

From CD Spectral Analysis								
Method	α%		β%		Turns		Unordered	
	Without GalNac	With GalNac	Without GalNac	With GalNac	Without GalNac	With GalNac	Without GalNac	With GalNac
CDSSTR	27	12	27	33	22	24	23	31
CONTINLL	15.1	6.5	26.6	34.4	16.7	24.2	41.7	33.5
SELCON3	24.7	10.6	28.9	30.2	20.0	22.4	29.4	37.4
Average	22.2	8.7	27.5	32.5	19.56	23.53	31.36	33.96
From Theoretical prediction								
PSIPRED	0.0		45.72		54.28		-	

4.5.10 Antimicrobial activity of BGAF

To measure the activity of BGAF on different strains of bacteria, their growth in terms of O.D was monitored in presence and absence of BGAF. BGAF inhibits the growth of *E. coli* completely. In case of *S. aureus,* there was a moderate decrease in growth whereas in *P. Aeruginosa* no change in growth pattern in presence of BGAF was observed (Figure 4.10)

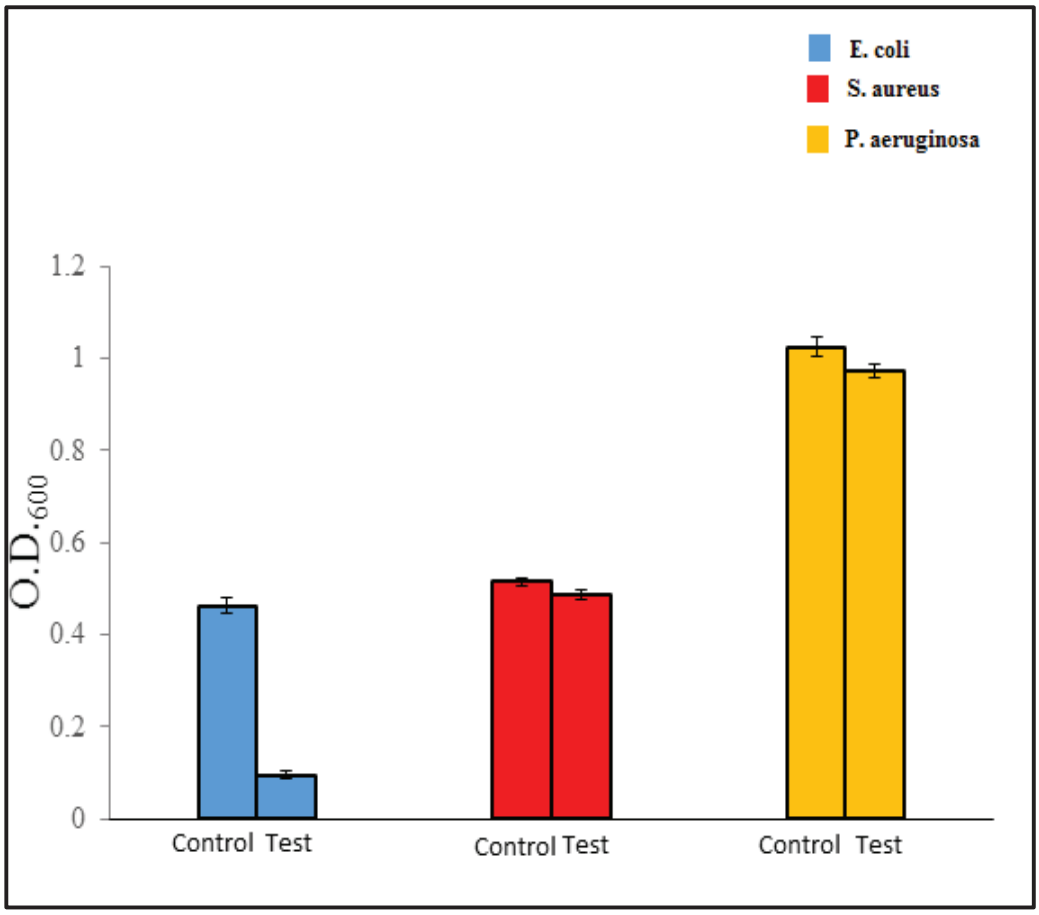
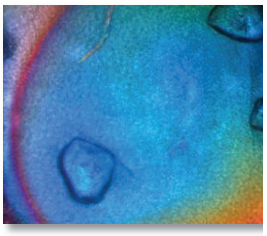
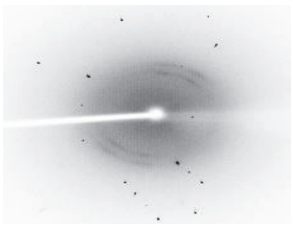
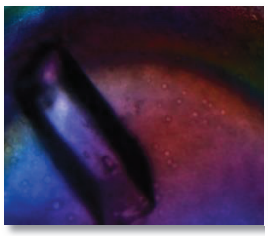
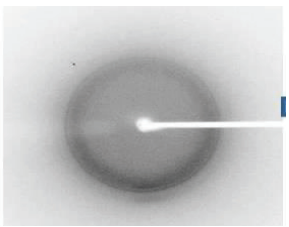
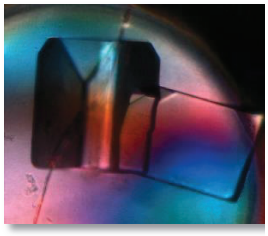
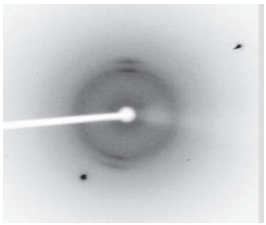


Figure 4.10 Antimicrobial activity of BGAF Three different strains of bacterial were tested for antimicrobial activity of BGAF.

4.5.11 Crystallization trials of BGAF

BGAF was purified in high concentration by Ni-NTA affinity chromatography (up to 25 mg/mL). And crystallized using various conditions and the observations are in Table 4.6. All the crystals obtained during trials turned out to be salt crystals when tested in the X-Ray beam. Crystallization conditions were further optimized by varying pH, concentration of protein and precipitants, and incubation temperatures, but no improvement in crystals was obtained.

Table 4.5 Crystallization trials of BGAF

Name	Crystallization Condition	Image of Crystal	Diffraction Image
BGAF_C1	0.5 M Sodium Chloride, 0.1 M Sodium Citrate tibasic dehydrate pH 5.6, 2% v/v Ethylene imine polymer		
BGAF_C2	0.1 M Sodium phosphate monobasic monohydrate, 0.1 M Potassium phosphate monobasic, 0.1 M MES monohydrate pH 6.5, 2.0 M Sodium chloride		
BGAF_C3	0.1 M HEPES pH 7.5, 4.3 M Sodium chloride		

4.5.12 Bioinformatics analysis of BGAF

Sequence analysis: The BLAST analysis using sorghum BGAF as query sequence found most of the hits with a beta-glucosidase aggregating factor or jasmonate-induced proteins from different plants. The nearest homolog of sorghum BGAF is another beta-glucosidase aggregating factor from maize with 68% sequence identity (Table 4.7).

Table 4.6 BLAST analysis of BGAF

Accession	Description	Query coverage	E-Value	Max Identity
ABI24164.1	Beta-glucosidase aggregating factor (*Sorghum bicolor*)	100	0.0	100
XP_002461766.1	Hypothetical protein SORBIDRAFT	99	2e-148	74
NP_001141303.1	Uncharacterized protein (*Zea mays*)	100	2e-131	68
ABJ97445.1	Beta-glucosidase aggregating factor 1 (*Zea mays*)	100	2e-130	67
XP_008651439.1	Predicted jasmonate-induced protein isoform X1 (*Zea mays*)	100	9e-130	68
EAY82651.1	Hypothetical protein (*Oryza sataiva* indica group)	99	8e-123	61
AAR20919.1	Jasmonate-induced protein (*Triticum aerstivum*)	100	1e-118	61

4.5.12.1 Multiple Sequence Alignment (MSA)

The multiple sequence alignment of BGAF was carried out with other sequences of BGAF present in other plants. Both the N-terminal dirigent domain and C-terminal lectin domain showed conservation among all the sequence, especially glycine residues (Figure 4.11).

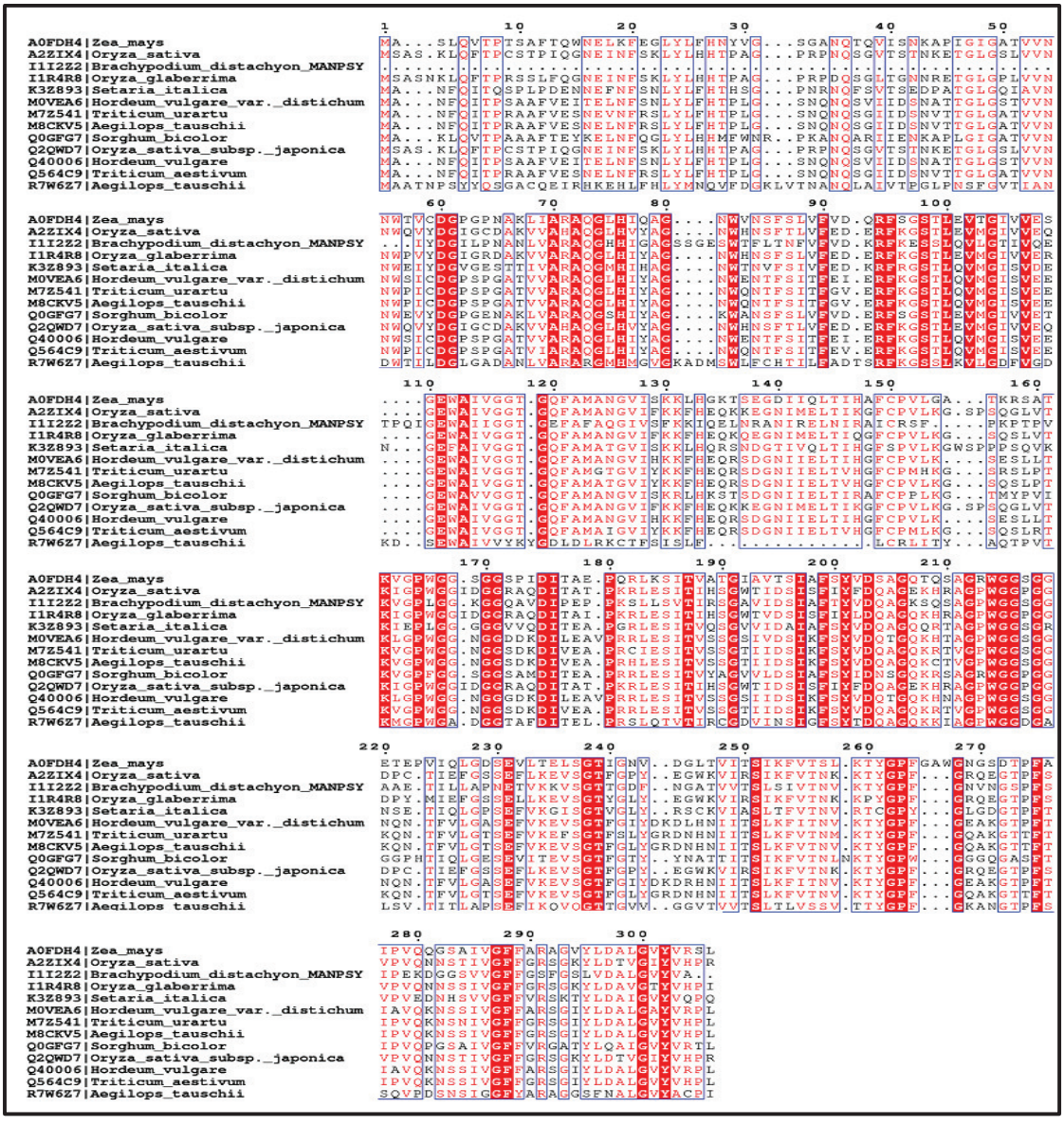


Figure 4.11 Multiple sequence analysis (MSA) of BGAF. BGAF with other beta-glucosidase aggregating factor from different plants. Residues in red show conservation of that particular residue among all the sequence. Residues in light red means one or more of the sequence have similar substitution of amino acids

4.5.12.2 Phylogenetic analysis

The phylogenetic analysis was carried out among BGAF, and homologous lectins from plants and sequences were collected from the UniProt database. Total of 13 sequences were selected for this analysis. The phylogenetic (Figure 4.12) analysis of these sequences formed two major clusters. 1 The cluster contains BGAF sequences from *Bracypodium distachyon* and *Aegilops tauschii,* while rest of the other BGAFs are clustered with BGAF from sorghum and maize.

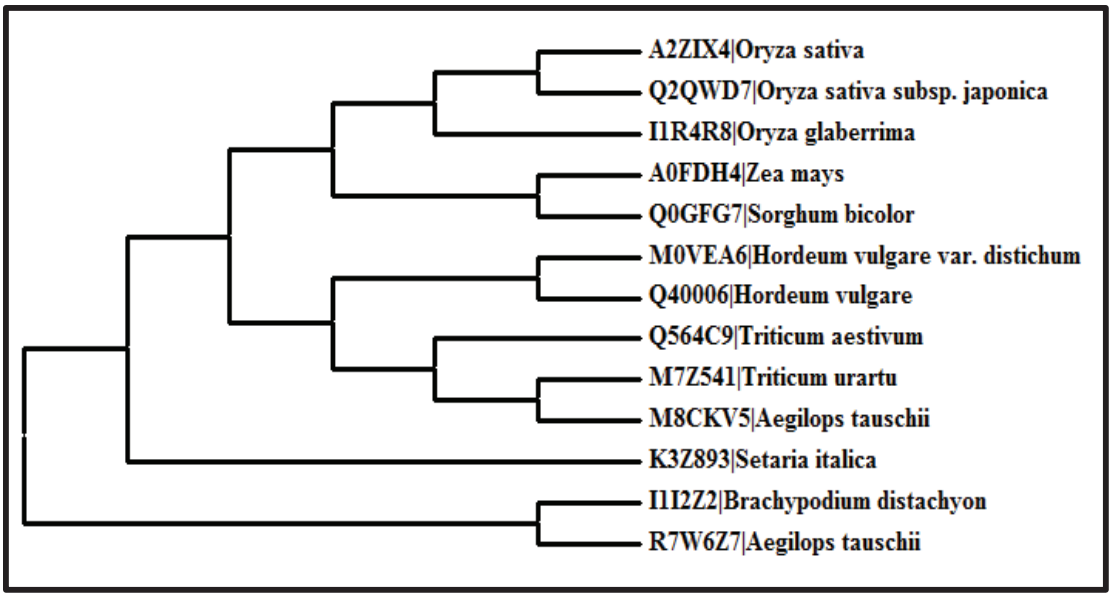


Figure 4.12 Phylogenetic analysis of BGAF. Dendogram of Sorghum BGAF with other BGAF from plants, showing a close relationship with maize BGAF.

4.5.12.3 Modeling and Secondary structure prediction of BGAF

Three-dimensional structure of BGAF protein is still elusive. To overcome this problem, the accurate three-dimensional structure of BGAF was predicted by I-TASSER server with C-score of -1.5. The modelled three-dimensional structure was generated using known lectins structures from PDB as templates. The generated model consisted 18 β-strands and loops which are represented in different colors, and each strand is numbered from N-terminal to C-terminal (Figure 4.13). The N-terminal dirigent domain and C-terminal lectin domain are connected by a stretch of 20 amino acids linker region. This linker region connects strand 8 (dirigent domain) to strand 9 (lectin domain). The Number of β-strands (18) is consistent with the secondary structure predicted by PSIPRED server based on sequence (Figure 4.14).

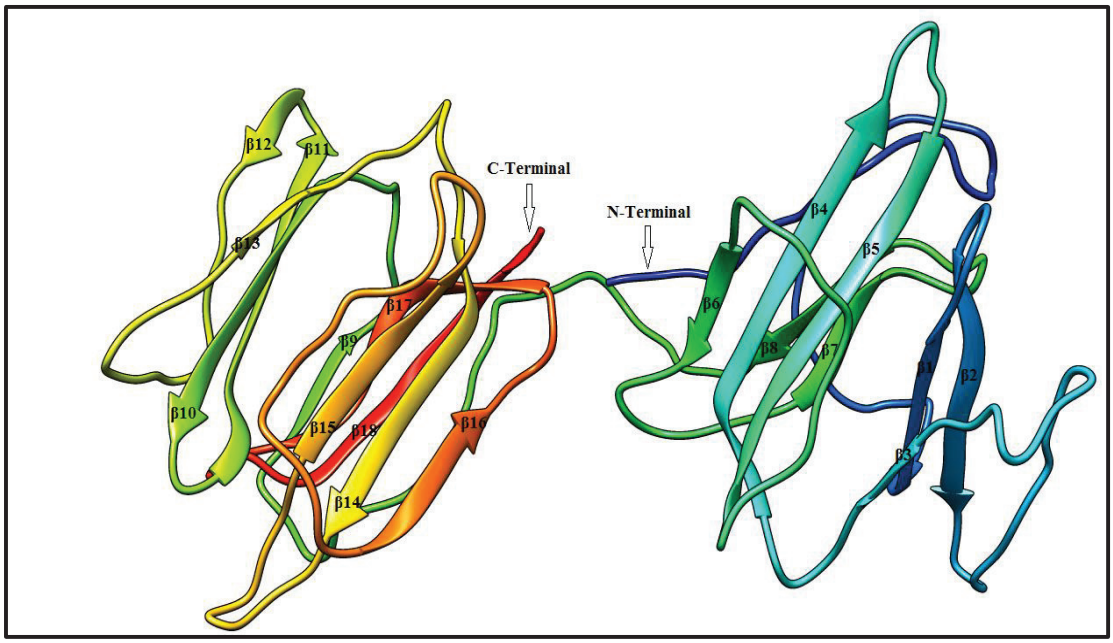


Figure 4.13 Homology modelling of BGAF. The structure of BGAF predicted using I-TASSER server, showing jacalin-like fold with 18 β-sheets joined by a linker region.

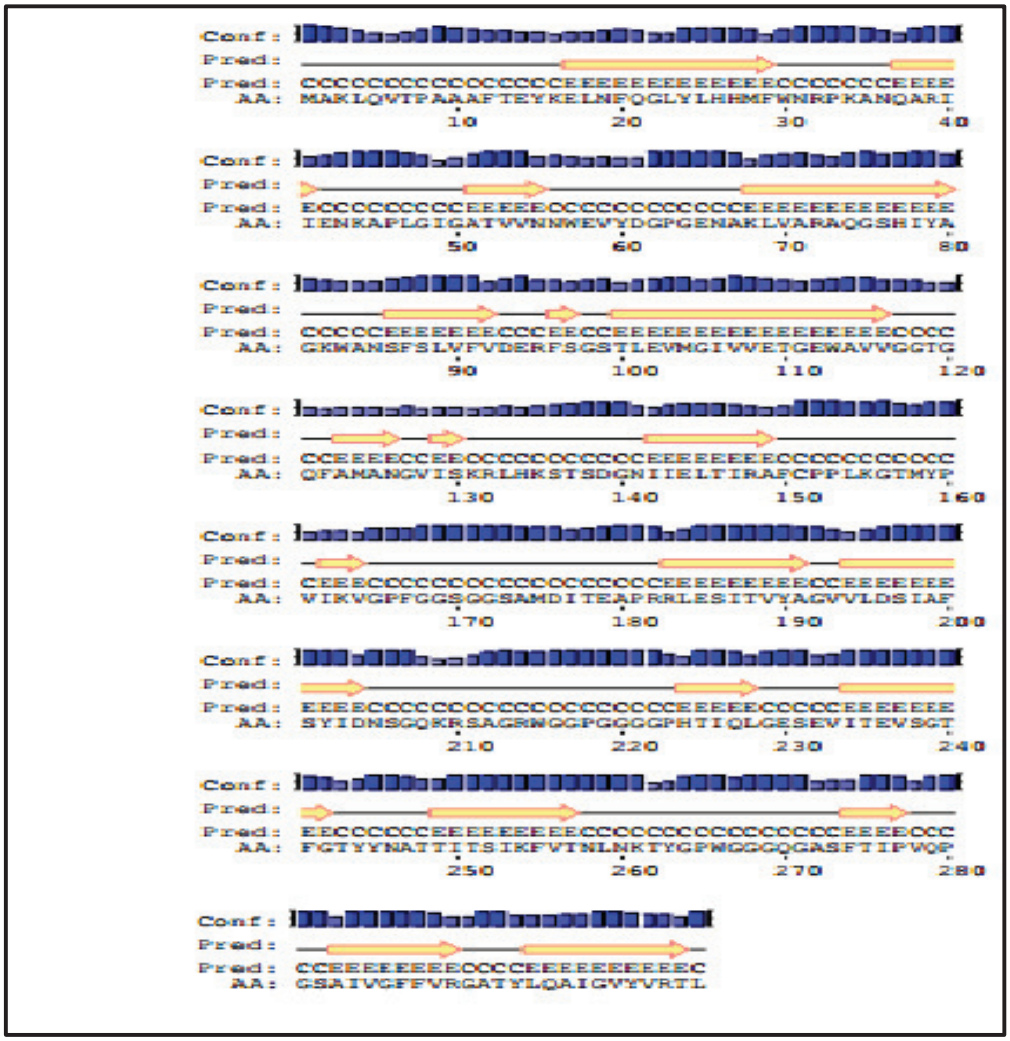


Figure 4.14 Secondary structure prediction of BGAF. The secondary structure predicted by PSIPRED server mainly consisted of β-sheets (Yellow). The blue bar shows the confidence level for individual amino acids.

DALI server was used to find out the structural homologs of BGAF and BGAF showed homology with Jasmonate-induced protein (2JZ4) with an r.m.s.d values of 2.1 and Z-score of 25.9. The other homologs of BGAF and its r.m.s.d values are listed in Table 4.8.

Table 4.7 Structural homologs of BGAF from DALI Server

PDB ID	Chain	RMSD Values	% Identity	Protein name
2JZ4	A	2.1	15	Jasmonate inducible protein
3P8S	B	3.8	26	Jacalin
1J4T	C	2.8	26	ARTOCARPIN
3MIV	B	2.3	31	Lectin
1XXR	A	4.0	26	Mannose-binding lectin

4.5.12.4 Molecular Docking Studies of BGAF

To find out which amino acids are involved in carbohydrate binding, docking studies of BGAF with different carbohydrates were carried out. 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 showed that Glycine present at position 168 and 169 forms a hydrogen bond with C2 of N-Acetyl-D-galactosamine. Besides, glycine, arginine present at position 290 also forms a hydrogen bond with carbohydrate (Figure 4.15). The binding energy for this interaction was -5.2Kcal/mol.

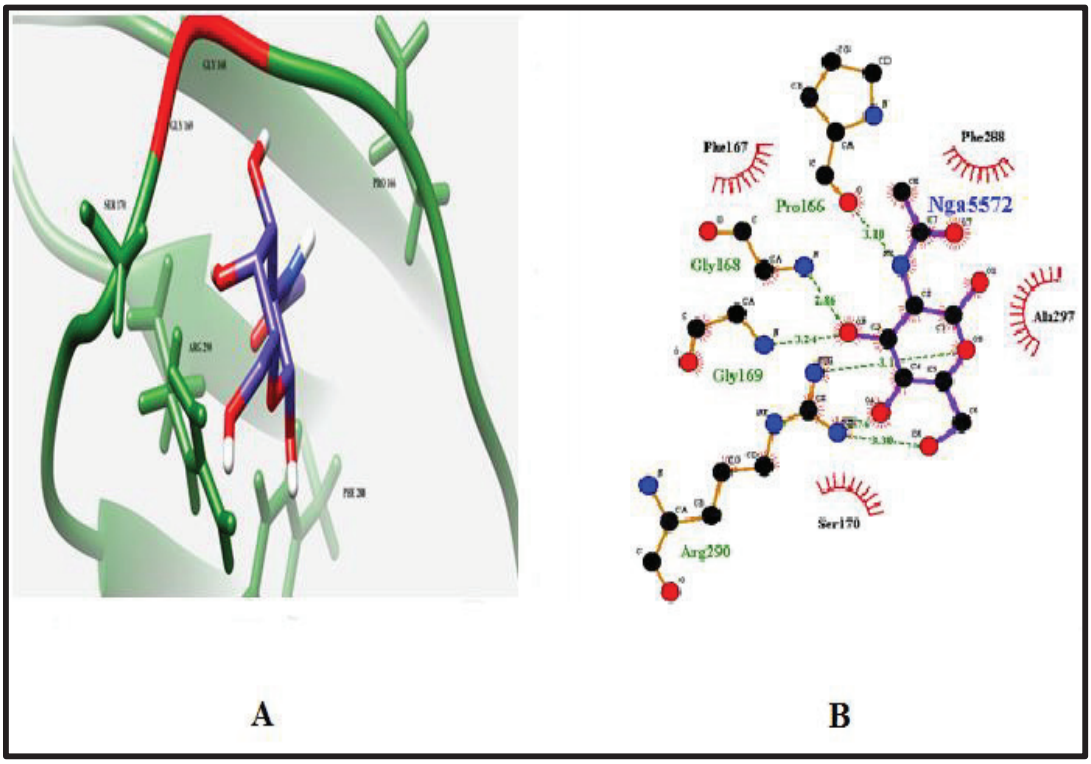


Figure 4.15 Molecular docking of BGAF with GalNac. (A) Docking of GalNac with cavity surrounding GG..GVVLD binding site. (B) Hydrogen bond formation between GalNac and Gly 168, Gly 169 and Arg 290.

4.6 Discussion

Beta-Glucosidase Aggregating Factor is a chimeric protein with two distinct domains, N-terminal dirigent domain, and C-terminal JRL domain, with known roles in plant defense. Till, date only two BGAF from *Zea mays* and *Sorghum bicolor* have been characterized in detail with respect to its function of aggregation and carbohydrate binding properties. BGAF from maize aggregates the β-glucosidase enzyme, on the other BGAF from sorghum, lack this ability despite showing almost 70% sequence homology. One of the other major difference between maize and sorghum BGAF is their carbohydrate specificity. Maize BGAF is strongly inhibited by GalNac, galactose, and Methyl-α-D-galactospyranoside, similarly, sorghum BGAF is strongly inhibited by GalNac, but it does not show any inhibition by galactose or Methyl-α-D-galactospyranoside (Kittur et al., 2007). The reason behind this lack of activity by sorghum BGAF is not known, and we have tried to characterize this protein in terms of its biochemical and biophysical properties.

In this study, we have cloned and purified a β-Glucosidase aggregating factor consisting of the N-terminal dirigent domain and C-terminal lectin domain. The gene for BGAF was synthesized with codon optimization. Codon usage is one of the most important factors in heterologous expression in prokaryotes (Lithwick et al., 2003; Ikemura. T, 1981). In general, the more codons that a gene contains, which are preferentially used in the host, the most likely outcome would be an expression of protein at a reasonable level. Low expression levels are seen for the genes with rare codons appears in a cluster or at the N-terminal part (Kane. J. F., 1995). The most common solution to overcome this problem is to alter the rare codon present in the target gene with codon which is most frequently used in host system without altering the final amino acid sequence. There have been several reports, showing 10 to 20 folds increase in protein expression after codon optimization (William et al., 1988; Perlak et al., 1991; Kotula et al., 1991).

For Initial expression trials of BGAF, cloned in pETM10 and pET28a vectors in heterologous expression host strains of *E. coli* did not show any expression compared to control samples. To overcome this problem, BGAF gene was recloned in pET30a vector which did show significant expression of BGAF compared to control samples. BGAF

protein was purified using Ni-NTA affinity purification technique. The final yield of purified protein from 200ml of autoinduction media was 20mg/ml. 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 interest 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 showed hemagglutination activity, and it was inhibited by N-Acetyl-D-galactosamine, mannose, Methyl-α-D-mannopyranoside and fructose, which are consistent with previous studies (kittur et al., 2009). The result obtained from fluorescence spectroscopy studies shows the affinity of each carbohydrate with BGAF. N-Acetyl-D-galactosamine has the highest affinity towards BGAF, and these results are in line with the results obtained from hemagglutination assay.

One of the most interesting features of the interaction of N-Acetyl-D-galactosamine with BGAF is a change in the ʎmax upon addition of sugar. The decrease in the ʎmax is known as the blue-shift effect. Blue-shift occurs when the fluorophore, in this case, amino acid tryptophan, tyrosine, and phenylalanine, place in a less polar solvent (Lakowicz et al., 2006). Blue-shift with the addition of sugar suggested a change in the protein structure and increased the stability of protein upon ligand binding. These observations are in line with the observation we found in Circular dichroism spectroscopy where on the addition of carbohydrate, secondary structure elements of BGAF changes drastically.

Secondary structure analysis showed that BGAF is a mixture of α-helical and β-strand proteins, and there is a huge change in secondary structure upon ligand binding (Figure 4.9 A). All the jacalin-like lectins are predominantly consist of β- sheets with little or no presence of α-helical component (Singh et al., 2005). Since very little or no information is available on dirigent domain structure, it is believed that dirigent domain may contain a mixture of both helical and sheet structure and upon ligand binding, to lectin domain, there is a change in secondary structure of full-length BGAF protein. Thermal unfolding studies using circular dichroism showed that secondary structure of BGAF is

practically unaltered up to 55°C, but after this point, it suddenly starts unfolding and about 60°C it is in the unfolded state completely (Figure 4.9 C). This is correlated with temperature stability studies of BGAF, upon heating above 60°C lead to loss of hemagglutination activity of BGAF.

There have been reports showing the anti-bacterial activity of many GalNac specific lectins. Many studies in past have shown that component of bacterial cell wall, such as LPS, peptidoglycans, and O-antigen on the surface function as a recognition site for lectins and plays a role in the antibacterial activity of lectins (Klafke et al., 2013). In gram-negative bacteria, the LPS plays an important part with lectin interaction. From our results, we can infer that *E. coli* have in it LPS composition, one of the sugars for which BGAF showed specificity. The sugar core of *E. coli* is composed of the common sugars such as mannose, galactose, glucose and N-Acetyl-D-glucosamine (Klafke et al., 2013). Out of these sugar for two sugars, mannose and galactose are specific for BGAF. The concentration needed for bacterial growth inhibition by BGAF is similar to *Halichondria okadai* lectin, which is a GalNAc specific lectin (Kawsar et al., 2010).

It has been well proven that amount of charged amino acid like glutamate, aspartate, arginine and lysine confers high solubility to protein in an aqueous environment, but it also makes it difficult to crystallize that protein (Li and Lawson 1995). BGAF sequence is consisting of 13 lysine, 12 arginines, 15 glutamate/glutamine residues and the presence of these residues in the sequence of BGAF making it highly soluble and difficult to crystallize. Another possible reason for the difficulty in protein crystallization is highly flexible or unfolded region of BGAF, which might interfere in crystallization. These might be the possible reasons for no three-dimensional structural information on dirigent domain till data.

Apart from biochemical and biophysical studies, bioinformatics was also used to analyze the obtained data. Phylogenetic analysis of BGAF with other homologous BGAF and lectins from plant formed four distinct clades. These 4 clades include 2 clades of plant lectins, 1 clade of BGAF from sorghum and maize and one clade of other BGAFs. The BGAF from sorghum was distributed separately with maize BGAF. Data from this study suggest BGAF from sorghum is a link between lectins and other BGAF. Structural

homology search with DALI server showed similarity with different classes of lectins. These results are very natural, since the homology model was based on only lectin domain since no structural information is available about the dirigent domain.

4.7 Conclusion

The full-length gene coding for β-glucosidase aggregating factor (BGAF) was cloned into a bacterial expression vector. The BGAF was obtained in purified form via Ni-NTA affinity chromatography with final concentration up to 20mg/mL. Purified BGAF showed hemagglutination activity, and it was inhibited by mannose, N-Acetyl-D-galactosamine, Methyl-α-D-Mannopyranoside, and fructose. The protein was stable up to 60°C and in pH range 5-9. The fluorescence spectroscopy studies showed that GalNac has a very strong effect on BGAF structure, and it stabilizes it. Fluorescence spectroscopy also revealed affinity of BGAF for galactose which was not detected by hemagglutination inhibition assay. Secondary structure analysis with CD spectroscopy revealed protein is a mixture of α-helices and β-sheets, but upon carbohydrate binding, it has more β-sheets compared to holoprotein. Thermal melt with CD spectroscopy showed loss of secondary structure around 60°C, which was consistent with temperature stability assay. Crystallization trials of BGAF with different buffer condition did not yield any protein crystals. Bioinformatics analysis showed BGAF is an evolutionary link between lectins and other BGAF in plants. Molecular docking study showed that GalNac interacts with two glycines of putative binding site GG.. GVVLD.