Chapter 5

Bioinformatics analysis, cloning, expression, purification and characterization of Mucus binding protein (MubP) and Fibronectin binding protein (FBP) from *Lactobacillus acidophilus*.

5.1. Introduction

The human gut typically known as gastrointestinal (GI) tract is colonised by the diverse microbial community which has a vital role in human health mainly in metabolism, nutrition and immunomodulation. These microbes popularly known as probiotics can pertain numerous health beneficial effects on the host. Adhesion to intestinal tract is primary requirement for colonization of these probiotics microbes which prolong their persistence in GI tract (Sanchez *et al.*, 2008; Sanchez *et al.*, 2010) Recent evidence suggests that adhesion to intestinal mucosa by *Lactobacillus* is mediated by surface adhesin proteins with a binding capacity to mucus and ECM components (Juge, 2012; Sengupta *et al.*, 2013). An earlier study also showed that the bacterial treatment with proteinase before binding reduced their capacity to bind mucin, which indeed suggested the involvement of surface proteins in adhesion (Jonsson *et al.*, 2001).

The epithelial cells are covered by a protective layer of mucus which consists of a complex mixture of highly glycosylated mucins which also offers an attachment site for bacterial cell surface adhesions (Dekker *et al.*, 2002). Mucins which are rich in Ser/Thr undergoes O-glycosylation primarily with N-acetyl-D-galactosamine (GalNAc). These O-glycan chains act as a base upon which long and more complex oligosaccharide chains are extended by the sequential addition of the monosaccharides like mannose, galactose, GalNAc, GlcNAc, fucose and sialic acid (Bergstrom & Xia, 2013). Microbes have evolved to express right complement of adhesins which recognises these diverse glycans/ oligosaccharides of glycoprotein mucin, proving a binding site for bacterial adherence in the gut (Johansson *et al.*, 2008).

In this chapter, we will describe two targets which are particular adhesins and are known to bind to ECM components and mucin, mucus binding protein (MubP) - LBA1018 and fibronectin binding protein (FBP) -LBA1148 from *Lactobacillus acidophilus*. Mucus binding protein is a multi-domain protein whose overall protein architecture is yet to be explored. It exhibit characteristics typical of gram-positive cell surface proteins: a C-terminal sortase recognition motif (LPxTG) which covalently anchor protein to peptidoglycan, multiple repeats of mucus binding domain (MUB), an N-terminal gram-positive secretion signal. MUB repeats domain are classified in to type 1 or type 2 depending on amino acid composition (Boekhorst *et al.*, 2006). An earlier study has revealed that the occurrence of MUB domain containing protein is most abundant in *Lactobacilli* that are found in GI tract rather than the strains which are less

frequently encounter in GI tract, e.g. *L. plantarum* (Kleerebezem *et al.*, 2003). These suggest that the MUB domain has a definite evolutionary role in GI tract adherence. Genome mining in *Lactobacillus* spp. has identified the presence of mucus-binding proteins of various size consisting of several mucus binding domain repeats (Boekhorst *et al.*, 2006). *Lactobacillus* spp. Identified the presence of various-sized putative mucus adhesins comprising one or more copies of mucus-binding repeats, which suggests that mucus binding domain containing proteins play a major role in establishing host–microbial interactions in the gut and promoting the natural selection of species as primarily GI inhabiting organisms (Boekhorst *et al.*, 2006; Kleerebezem *et al.*, 2010). Two types of domain are associated with mucus binding protein: MUB and MucBP (Pfam: PF06458) with the difference in length of the domain, MUB domain is approximately around 200 residues and later one is around 50 residues in length.

Fibronectin-binding protein (FBP), are a large protein which specifically binds to fibronectin and are found across wide variety in Gram-negative and Gram-positive bacteria. The presence of FBP in bacteria has made them effector adhesion mediating molecule both in pathogens and commensals (Henderson *et al.*, 2011; Styriak *et al.*, 2003). Fibronectin is a large dimeric glycoprotein composed of two polypeptide chains with an approximate mass of 450 kDa. Found in soluble form in body fluids and immobilised form in ECM, it has a critical role in eukaryotic cellular processes, such as adhesion, migration and differentiation and also acts as an adhesion site for the attachment of bacteria. Fibronectin binding protein is an important virulence factor for adhesion in many pathogens such as *Streptococcus pneumoniae* and *Streptococcus pyogenes* (Holmes *et al.*, 2001; Jedrzejas, 2007; Molinari *et al.*, 1997). Fibronectin binding protein (FBP) is believed to act as a bridging molecule between the bacterial cell surface and the ECM of the host, although the molecular mechanism underlying this interaction is not yet understood. Further in the chapter, we will describe the experimental part for two adhesin targets: Mucus binding protein (LBA1018) from *L. acidophilus* - LaMubP and Fibronectin binding protein (LBA1148) from *L. acidophilus* – LaFBP.

5.2. Materials

5.2.1. Chemicals and Reagents

The different chemicals and reagents used in the current study were of molecular biology grades. Enzymes used in cloning of genes like restriction endonucleases, TAQ

polymerases and DNA ligases were purchased from Fermentas and Roche. Chemicals used in protein expression and purification like Ni-NTA agarose resins, IPTG, Triton X-100, PMSF protease inhibitor, SDS, glycine, Trizma base, etc. were purchased from Sigma-Aldrich. The protein molecular marker and DNA ladders were procured from Fermentas. Molecular biology grade chemicals like sodium chloride, imidazole, sodium phosphate, potassium chloride, etc. were purchased from HiMedia. The antibiotics ampicillin, kanamycin and chloramphenicol, were purchased from HiMedia. Media for culturing of bacteria viz. Luria-Bertani (LB), yeast extract, agar, tryptone, peptone, de Man Rogosa and Sharpe (MRS) media were purchased from HiMedia, India. Paraffin oil for anaerobic culturing was purchased from HiMedia, India. LB and MRS agar plates were prepared by supplementing media with 0.75-1% agar. The glycoproteins fetuin and type III mucin was purchased from Sigma.

For autoinduction, the following media composition was prepared based on (Studier, 2005): **ZY media** (1g of Tryptone and 0.5g yeast extract in 100ml of sterile water), **50xM** (1.25M Na2HPO4, 2.5M NH4Cl, 1.25M KH2PO4, 0.25M Na2SO4), **50x5052** (25% glycerol, 2.5% glucose, 10% galactose in final volume of 100ml make up with sterile water), 1M MgSO4 in water and 1000X Trace Metals Mix (50ml of 0.1M $FeCl_3$-$6H_2O$, 1ml of 0.2M $NiCl_2$-$6H_3O$, 2ml of 1M $CaCl_2$, 1ml of 0.1M Na_2SeO_3-$5H_3O$, 1ml of 0.2M $CoCl_2$-$6H_2O$, 2ml of 0.1M $CuCl_2$-$2H_3O$, 1ml of 1M $MnCl_3$-$4H_3O$, 1ml of 0.1M $Na_2MO_4.5H_2O$, 1ml of 1M $ZnSO_4$-$7H_3O$, 1ml of 0.1M H_3BO_3 in 36ml of water and then filter sterilized). The final 100ml complex autoinduction media used for protein expression- ZYM-5052 were prepared containing 0.1ml 1M MgSO4 (final concentration 2mM), 0.01µl of 1000X trace metals, 1ml 50x 5052 and 1ml 50xM and 47.9ml autoclaved ZY media.

5.2.2. *Lactobacillus* strain and genomic DNA isolation.

The *Lactobacillus acidophilus* strain was procured from National Dairy Research Institute (NDRI), Karnal, India. The genomic DNA was isolated using a standard procedure as mentioned in section 3.3.1.

5.2.3. Cloning vectors and bacterial strains

The cloning, expression vectors and bacterial strains for cloning and expression used were similar as mentioned in section 3.2.3. Apart from that the expression strain *E. coli* BL21 (DE3) pLysS was purchased from Invitrogen Life Technologies (USA).

5.3. Methods

5.3.1. Gene amplification and cloning

5.3.1.1. PCR amplification and cloning of LaMubP and LaMubPtr: The PCR amplification of the gene LBA1018 (UniProt:Q5FKA7) encoding LaMubP and a truncated version LaMubPtr which included the domain but lacking the LPxTG motif signature (LaMubPtr) using the following forward and reverse primers (Table 5.1) was performed similarly as mentioned earlier in section 3.3.2.1.

The purified PCR amplicon was cloned in pET15b (N-terminal His-tag) expression vector. *Nde*I and *Bam*HI restriction enzymes along with 1X Tango buffer was used in the cloning procedure. The remaining steps in cloning: restriction digestion, ligation, chemically competent cells preparation and clone transformation were performed by a similar method as mentioned earlier in section 3.3.2.2 and 3.3.2.3. The purified plasmid was resolved in 0.8% agarose gel electrophoresis and a shift in plasmid band viz. control pET15b plasmid without ligation was a positive indication of a clone which was further confirmed by colony PCR and restriction digestion (*Nde*I and *Bam*HI) to visualise insert release followed by clone confirmation through plasmid sequencing. The truncated gene LaMubPtr was cloned in pET30a with their respective primers (Table 5.1) using *Eco*RI and *Not*I as restriction enzymes. All the remaining steps were performed as mentioned above.

Table 5.1: List of primers used for PCR amplification of LaMubP, LaMubPtr, LaFBP and LaFBPtr cloning.

No	Gene name	Plasmid /Vector	Primers	
1	Mucus binding protein (LaMubP)	pET15b	F	GTA **CATATG** ATGATTCAATACGTAGGTGAA
			R	GTA **GGATCC** TCATTATTTCATCTTCTTTTT
2	Mucus binding protein (LaMubPtr)	pET30a	F	GTA **GAATCC** ATGATTCAATACGTAGGTGAA
			R	AAT **GCGGCCGC** TTATATTTTTGTACTATT
3	Fibronectin binding protein (LaFBP)	pET15b pET28a	F	GTA **CATATG** ATGGCATTTGACGGATTA
			R	GTA **GGATCC** TTAAGCCATAATTTCGTCTTC
4	Fibronectin binding protein (LaFBPtr)	pET15b, pET28a	F	GTA **CATATG** ATGGCATTTGACGGATTA
			R	GGT **GGATCC** TTACAATTCGTCAGTTAT

Restriction sites are underlined and marked in bold, F- forward primer, R-reverse primer.

5.3.1.2. PCR amplification and cloning of LaFBP and LaFBPtr: The PCR amplification of the gene LBA1148 (UniProt: Q5FJY7) encoding LaFBP and a truncated version LaFBPtr which included the fibronectin binding domain (FbpA) but lacking the DUF814 domain using the following forward and reverse primers (Table 5.1) was performed similarly as mentioned earlier in section 3.3.2.1. The only change was the use of 1X Pfu reaction buffer and 2mM $MgSO_4$ with Pfu DNA polymerase because the length of LaFBP gene was more than 1000 bp. The amplification was performed using DNA thermal PCR cycler (Applied Biosystems, USA) programmed with following parameters - initial denaturation for 3 minutes at 94ºC, followed by 35 cycles of denaturation for the 30s at 94ºC, annealing at 55ºC for 30s, extension at 68ºC for 120s and final re-naturation at 68ºC for 10 minutes.

The purified PCR amplicon was initially cloned into pET15b (N-terminal His-tag) and pET28a (N-terminal His-tag) expression vector. *Nde*I and *Bam*HI restriction enzymes along with 1X Tango buffer was used in the cloning procedure. The remaining steps in cloning: restriction digestion, ligation, chemically competent cells preparation and clone transformation were performed by a similar method as mentioned in earlier section 3.3.2.2 and 3.3.2.3. The purified plasmid was resolved in 0.8% agarose gel electrophoresis and a shift in plasmid band viz. control plasmid without ligation was a positive indication of a clone which was further confirmed by colony PCR and restriction digestion (*Nde*I and *Bam*HI) to visualise insert release followed by clone confirmation through plasmid sequencing. The truncated gene LaFBPtr was cloned in pET28a and pET15b with their respective primers (Table 5.1). All the steps for cloning were performed as mentioned above.

5.3.2. Heterologous protein expression in *E.coli* strains.

5.3.2.1. Expression of r-LaMubP and r-LaMubPtr in *E.coli* strains: The confirmed positive clone with gene insert was transformed into chemically competent *E. coli* BL21 (DE3) Rosetta and BL21 (DE3) pLysS and expression trials were carried out similarly as mentioned earlier in section 3.3.2.4. Similarly for expression of recombinant proteins using autoinduction media, 500ml autoinduction media (ZYM-5052 media) was inoculated with overnight 1% pre-culture and incubated on a shaker for 5 hours at 37ºC followed by the incubation temperature of 20ºC for 20 hours. The suspension of cells in lysis buffer pH 7.8 and sonication was performed as per method mention earlier in

section 3.3.2.4 followed by protein expression detection and analysis on 15% SDS-PAGE. Similarly, like the r-LaMubP expression, the confirmed positive clone (pET30a+LaMubPtr) was transformed into chemically competent *E. coli* BL21 (DE3) Rosetta and BL21 (DE3) pLysS expression. Rest all the procedures were same as described above.

5.3.2.2. Expression of r-LaFBP and r-LaFBPtr in *E.coli* strains: The confirmed positive clone (pET15b+LaFBP and pET28a+LaFBP) with gene insert was transformed into chemically competent *E. coli* BL21 (DE3) Rosetta and BL21 (DE3) pLysS and expression trials were carried out similarly as mentioned earlier in section 3.3.2.4. Similarly for expression of recombinant proteins using autoinduction media, 500ml autoinduction media (ZYM-5052 media) was inoculated with overnight 1% pre-culture and incubated on a shaker for 5 hours at 37ºC followed by the incubation temperature of 20ºC for 20 hours. The suspension of cells in lysis buffer pH 7.6 and sonication was performed as per method mention earlier in section 3.3.2.4 followed by protein expression detection and analysis on 15% SDS-PAGE. Similarly, like r-LaFBP expression, the confirmed positive clone (pET15b+LaFBPtr and pET28a+LaFBPtr) was transformed into chemically competent *E. coli* BL21 (DE3) Rosetta and BL21 (DE3) pLysS expression. Rest all the procedures were same as described above.

5.3.3. Purification of recombinant proteins using affinity chromatography

5.3.3.1. r-LaMubP and r-LaMubPtr purification using Ni-NTA affinity chromatography: Many expression trials were conducted for soluble expression of the recombinant protein which is described in results section in detail. The purification of r-LaMubP and r-LaMubPtr using Ni-NTA affinity column was performed similarly as mentioned earlier in section 3.3.2.5 with the only difference in the pH of all buffers i.e. 7.8. Also, 0.1% mucin was added in lysis buffer along with 10% glycerol and 0.1% Triton-X 100 added in all buffers used in purification to enhance the protein stability in solution. For purification of r-LaMubPtr, the same steps mentioned above were followed.

5.3.3.2 r-LaFBP and r-LaFBPtr purification using Ni-NTA affinity chromatography: The purification of r-LaFBP and r-LaFBPtr using Ni-NTA affinity column was performed in same manner as mentioned earlier in section 3.3.2.5 with the only difference in the pH of all buffers i.e. 7.6. For purification of r-LaFBPtr, the same steps mentioned above were followed.

5.3.4. Biophysical characterization

5.3.4.1 Peptide mass fingerprinting (MALDI MS/MS) of r-LaMubP and r-LaFBP: The purified r-LaMubP and r-LaFBP was analysed by SDS-PAGE gel electrophoresis and the band corresponding to respective molecular weight size was excised from the destained gel and further in-gel digested with trypsin. Rest of the method in protein identification was performed in similar manner as mentioned in section 3.3.4.1.

5.3.4.2 Circular dichroism spectroscopy of recombinant r-LaMubP: The circular dichroism (CD) spectra to estimate the secondary structural elements of r-LaMubP protein was recorded using Jasco J-815 spectrophotometer (Japan Spectroscopic, Japan) similarly as mention earlier in section 3.3.4.2.

5.3.5. Bioinformatics analysis

5.3.5.1. Sequence homology search and phylogeny: BLAST analysis was performed to identify suitable sequence homologs using the protein sequence as a query (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) (Altschul *et al.*, 1990). Multiple sequence alignment (MSA) was performed using Clustal W/X program (Larkin *et al.*, 2007). MEGA5 package was used for the phylogenetic tree to analyse evolutionary relationships (Tamura *et al.*, 2011). Pfam and CDD tools were used to analyse the protein domain (Finn *et al.*, 2014; Marchler-Bauer & Bryant, 2004). SignalP server (http://www.cbs.dtu.dk/services/SignalP/) and TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) were used for signal peptide prediction and transmembrane region prediction respectively. ProtParam was used for the computation of various physical and chemical parameters of proteins such as molecular weight, theoretical pI, amino acid composition, atomic composition, molar extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity from the given protein sequence (Gasteiger *et al.*, 2003).

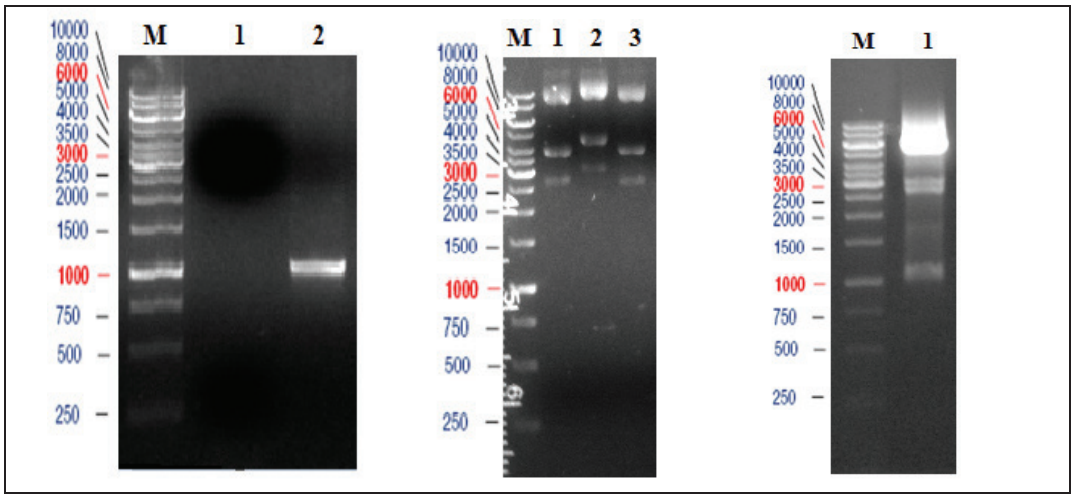
5.3.5.2. Homology modelling and structure validation: I-TASSER (Iterative-Threading / ASSembly / Refinement) server (http://zhang.bioinformatics.ku.edu/I-TASSER/) was used for protein structure prediction and the best model were selected by C-score. The model quality was improved by refinement using ModRefiner tool for model refinement (http://zhanglab.ccmb.med.umich.edu/ModRefiner/). The improved models were validated by PROCHECK (http://nihserver.mbi.ucla.edu/SAVES/) online

tool for Ramachandran statistics. The structural homology search was performed through DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server).

5.4. Results

5.4.1. Gene amplification and cloning

5.4.1.1. LaMubP amplification and cloning in pET15b: The MubP gene (LBA1018) was amplified using gene-specific primers (Table 5.1) with *L. acidophilus* genomic DNA (isolated and confirmed as mentioned in section 3.4.1) as the template. A 1041 bp long LaMubP gene was successfully amplified and resolved on 0.8% agarose gel (Figure 5.1a). The PCR amplified product was accordingly digested using *Nde*I and *Bam*HI restriction enzymes for cloning in to pet15b. After ligation, the recombinant plasmids were transformed into *E.coli* DH5α cells and the presence of LaMubP insert in the selected positive clones were confirmed through plasmid shift (Figure 5.1b), double digestion of clones with the restriction enzymes to release the insert (Figure 5.1c) and finally through nucleotide sequencing confirming the cloning of LaMubP into the desired vector (Figure 5.2). The results show the successful ligation of LaMubP gene in the pET15b vector.



(a) **(b)** **(c)**

Figure 5.1: Gene amplification and cloning of LaMubP: (a) Amplification of LaMubP using PCR, Lane M: 1kb ladder, Lane 1: blank well, Lane 2: LaMubP amplicon; (b) Plasmid shift to assess positive clones, Lane M: 1kb ladder, Lane 1: pET15b control vector, Lane 2: positive clone with pET15b and LaMubP insert resulting in a shift in plasmid size, Lane 3: negative clone without LaMubP insert; (c) Restriction digestion of positive clone, Lane M: 1kb ladder, Lane 1: digestion of positive clone with NdeI and BamHI restriction enzymes showing LaMubP insert release.

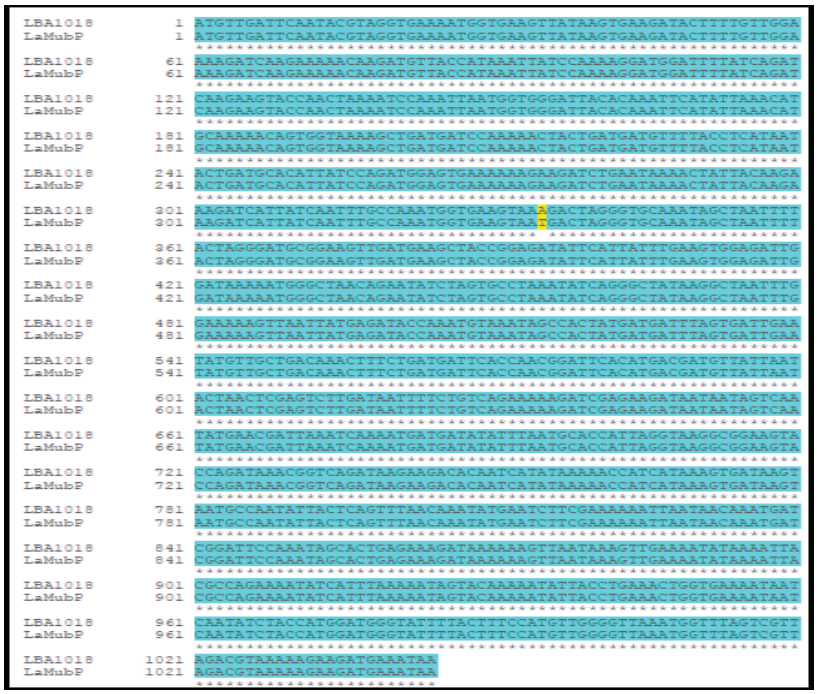
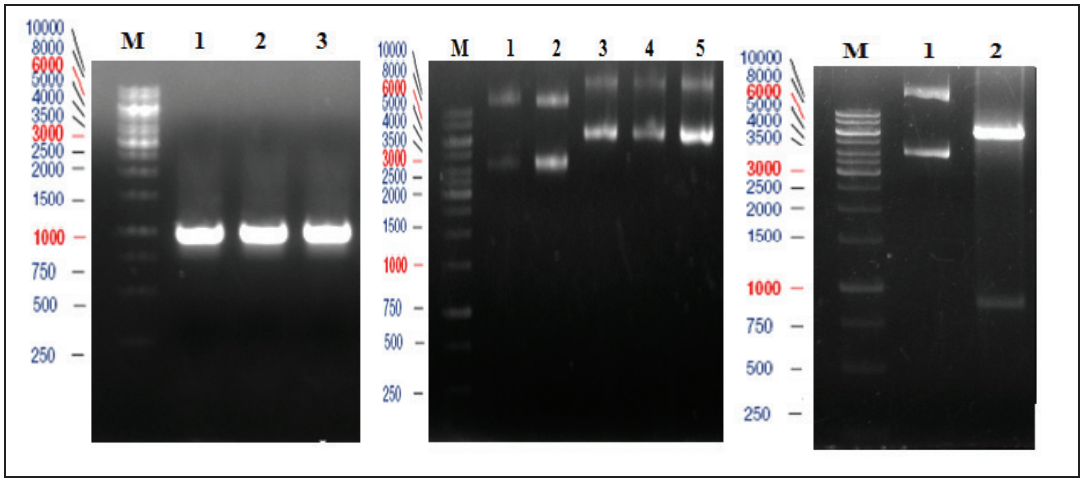


Figure 5.2: Clone confirmation of LaMubP. Sequence comparison of LaMubP nucleotide sequencing result and LBA1018 MubP entry from NCBI.

5.4.1.2. LaMubPtr amplification and cloning in pET30a: The truncated MubP gene (LBA1018) was amplified using gene-specific primers (Table 5.1) with *L. acidophilus* genomic DNA (isolated and confirmed as mentioned in section 3.4.1) as the template. Apart from the cloning of the full-length gene, a truncated gene LaMubPtr was also cloned into pET30a vector. This LaMubPtr gene was only devoid of the LPxTG motif signature (108 bp) to enhance the expression and stability of the protein. A 932 bp long LaMubP gene was successfully amplified and resolved on 0.8% agarose gel (Figure 5.3a). The PCR amplified product were accordingly digested using *EcoR*I and *Not*I restriction enzymes for cloning in to pet30a. After ligation, the recombinant plasmids were transformed into *E.coli* DH5α cells and the presence of LaMubPtr insert in the selected positive clones were confirmed through plasmid shift (Figure 5.3b), double digestion of clones with the restriction enzymes to release the insert (Figure 5.3c) and finally through nucleotide sequencing confirming the cloning of LaMubPtr into the desired vector (Figure 5.4). A few base pairs were changed when identified through sequencing, but no stop codon was incorporated in the clone.



(a) (b) (c)

Figure 5.3: Gene amplification and cloning of LaMubPtr: (a) Amplification of LaMubPtr using PCR, Lane M: 1kb ladder, Lane 1-3 LaMubPtr amplicon; (b) Plasmid shift to assess positive clones, Lane M: 1kb ladder, Lane 1: pET15b control vector, Lane 2: negative clone without LaMubP insert, Lane 3-5: positive clone with pET30a and LaMubPtr insert resulting in a shift in plasmid size; (c) Restriction digestion of positive clone, Lane M: 1kb ladder, Lane1: control pET30a vector, Lane2: digestion of positive clone with *Nde*I and *Bam*HI restriction enzymes showing LaMubPtr insert release.

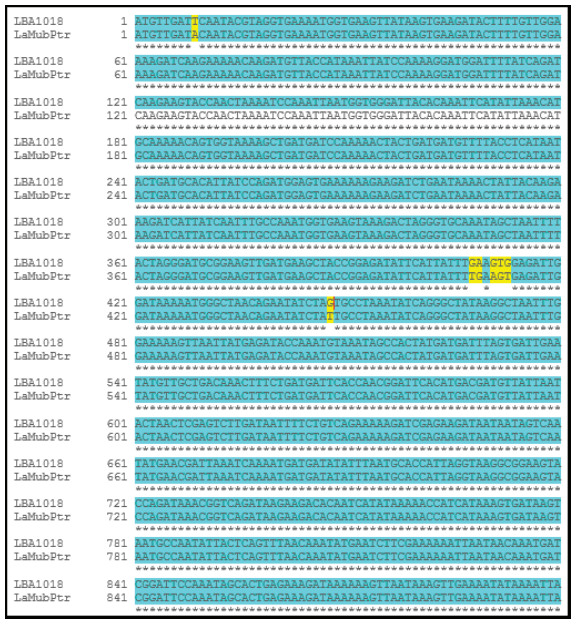


Figure 5.4: LaMubPtr clone confirmation through nucleotide sequencing: Sequence comparison of LaMubPtr protein sequence translated from nucleotide sequencing and LBA1018 MubP entry from NCBI.

5.4.1.3. LaFBP amplification and cloning in pET15b and pET28a: The FBP gene (LBA1148) was amplified using gene-specific primers (Table 5.1) with *L. acidophilus* genomic DNA (isolated and confirmed as mentioned in section 3.4.1) as the template. A 1674 bp long LaFBP gene was successfully amplified and resolved on 0.8% agarose gel (Figure 5.5a). The PCR amplified product was accordingly digested using *Nde*I and *BamH*I restriction enzymes for cloning in to pET15b and pET28a. After ligation, the recombinant plasmids were transformed into *E.coli* DH5α cells and the presence of LaFBP insert in the selected positive clones were confirmed through plasmid shift (Figure 5.5b, c), double digestion of clones with the restriction enzymes to release the insert (Figure 5.5d, e).

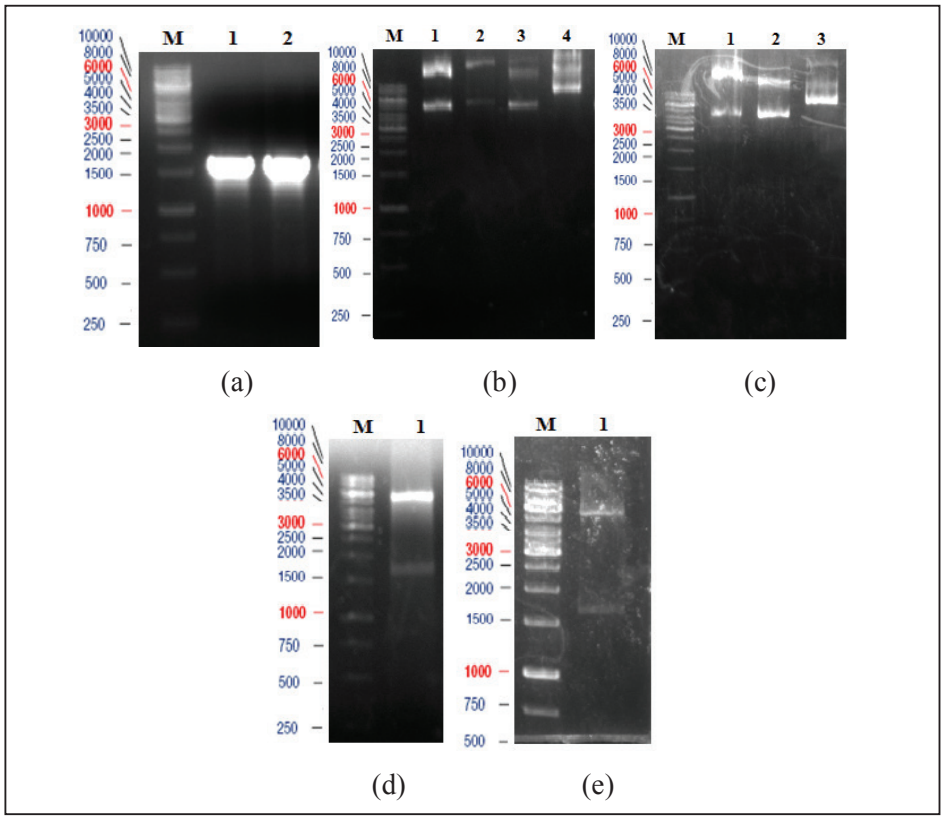


Figure 5.5: Gene amplification and cloning of LaFBP: (a) Amplification of LaFBP using PCR, Lane M: 1kb ladder, Lane 1-2 LaFBP amplicon; (b) Plasmid shift to assess positive clones, Lane M: 1kb ladder, Lane 1: pET28a control vector, Lane 2-3: negative clone without LaMubP insert, Lane 4: positive clone with pET28a and LaFBP insert resulting in a shift in plasmid size; (c) Plasmid shift to assess positive clones, Lane M: 1kb ladder, Lane 1: pET15b control vector, Lane 2: negative clone without LaFBP insert, Lane 4: positive clone with pET15b and LaFBP insert resulting in a shift in plasmid size; (d) Restriction digestion of positive clone, Lane M: 1kb ladder, Lane1: digestion of positive clone in pET28a with NdeI and BamHI restriction enzymes showing LaFBP insert release; (e) Restriction digestion of positive clone, Lane M: 1kb ladder, Lane1: digestion of positive clone in pET15b with NdeI and BamHI restriction enzymes showing LaFBP insert release.

5.4.1.4. LaFBPtr amplification and cloning in pET15b and pET28a: The truncated FBP gene (LBA1148) was amplified using gene-specific primers (Table 5.1) with *L. acidophilus* genomic DNA (isolated and confirmed as mentioned in section 3.4.1) as the template. Apart from the cloning of the full-length gene, a truncated gene LaFBPtr was also cloned into pET15b and pEt28a vector. This LaFBPtr gene was devoid of the DUF814 domain to enhance the expression and stability of the protein by restricting the size of the construct. A 1257 bp long LaFBPtr gene was successfully amplified and resolved on 0.8% agarose gel (Figure 5.6a). The PCR amplified product were accordingly digested using *Nde*I and *BamH*I restriction enzymes for cloning in to pET28a and pET15b vectors. After ligation, the recombinant plasmids were transformed into *E.coli* DH5α cells and the presence of LaFBPtr insert in the selected positive clones were confirmed through plasmid shift (Figure 5.6b), double digestion of clones with the restriction enzymes to release the insert (Figure 5.6c).

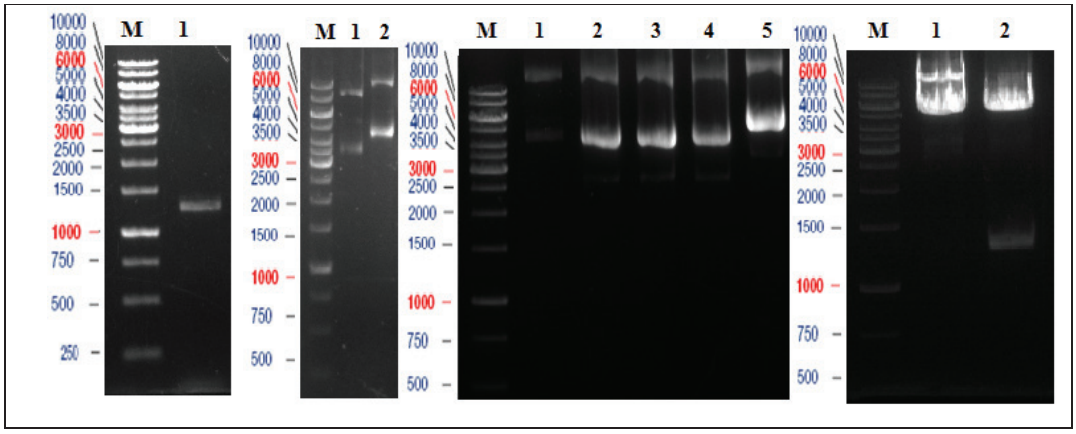


Figure 5.6: Gene amplification and cloning of LaFBPtr (a) Amplification of LaFBPtr using PCR, Lane M: 1kb ladder, Lane 1 LaFBPtr amplicon. (b) Plasmid shift to assess positive clones, Lane M: 1kb ladder, Lane 1: pET28a control vector, Lane 2: positive clone with pET28a and LaFBPtr insert resulting in a shift in plasmid size. (c) Plasmid shift to assess positive clones, Lane M: 1kb ladder, Lane 1: pET15b control vector, Lane 2-4: negative clone without LaFBPtr insert, Lane 5: positive clone with pET15b and LaFBPtr insert resulting in a shift in plasmid size. (d) Restriction digestion of positive clone, Lane M: 1kb ladder, Lane1: digestion of control pET28a, Lane 2: digestion of positive clone in pET28a with *Nde*I and *Bam*HI restriction enzymes showing LaFBPtr insert release.

5.4.2. Expression and purification of recombinant proteins in *E. coli* strains

5.4.2.1. Expression and purification of r-LaMubP: After confirming the successful cloning of LaMubP in pET15b vector, the protein expression studies were carried out using IPTG as inducers in *E. coli* Rosetta BL21 (DE3) strain. The expression of r-

LaMubP was detected in very low quantity and visualised in 15% SDS-PAGE gel electrophoresis with IPTG induction at 37°C for 4 hours. The expression was detected at around ~65 kDa which is different than its expected theoretical molecular mass of 39.6 kDa. The protein LaMubP showed an aberrant migration in SDS-PAGE. In pET15b-LaMubP construct, when expressed in *E. coli* Rosetta BL21 (DE3), the expression was detected majority in insoluble pellet fraction as well as partly in a soluble supernatant fraction at higher IPTG concentration 1.5mM and induction temperature of 37°C (Figure 5.7a). At low induction temperature, the soluble expression of r-LaMubP was slightly increased (Figure 5.7b). The expression level of r-LaMubP using autoinduction media didn't improve significantly.

The protein expression of pET15b+LaMubP construct using autoinduction was also carried out in *E. coli* BL21 (DE3) pLysS strain at 20°C induction temperature (Figure 5.8a). It was also observed that when using 0.5mM IPTG induction along with 1% glucose at 20°C induction temperature yield little more soluble protein compare to when the only IPTG is used (Figure 5.8a). The numerous expression trials were carried out to achieve more soluble protein (Table 5.2). The recombinant r-LaMubP was purified by Ni-NTA affinity column chromatography when expressed in *E. coli* BL21 (DE3) pLysS strain using autoinduction media and 20°C induction temperature (Figure 5.9).

5.4.2.2. Expression and purification of LaMubPtr: After confirming the positive clone of LaMubPtr in pET30a vector, the protein expression studies were carried out using IPTG as inducers and autoinduction in *E. coli* Rosetta BL21 (DE3) pLysS strain. The expression was detected at ~65 kDa similar to what obtained while expressing the r-LaMubP. The protein r-LaMubPtr showed an aberrant migration in SDS-PAGE. Unlike the expression of pET15b+LaMubP, pET30a+LaMubPtr expression was detected most in soluble fraction as well as partly in an insoluble fraction at 0.2mM IPTG concentration and lower induction temperature of 20°C (Figure 5.10a, b). It was also observed that the expression of pET30a+LaMubPtr at higher induction temperature of 37°C, the protein was visualised on SDS-PAGE at its right molecular size of around 39 kDa while at lower incubation temperature there was an aberrant migration on SDS-PAGE (Figure 5.11a). The expression level of r-LaMubPtr in *E. coli* Rosetta BL21 (DE3) pLysS strain using autoinduction media improve significantly. The solubility and stability of protein were increased with the addition of 0.1% mucin in lysis buffer, 10% glycerol and 0.1% Triton X-100 in all purification buffer. The recombinant r-LaMubPtr

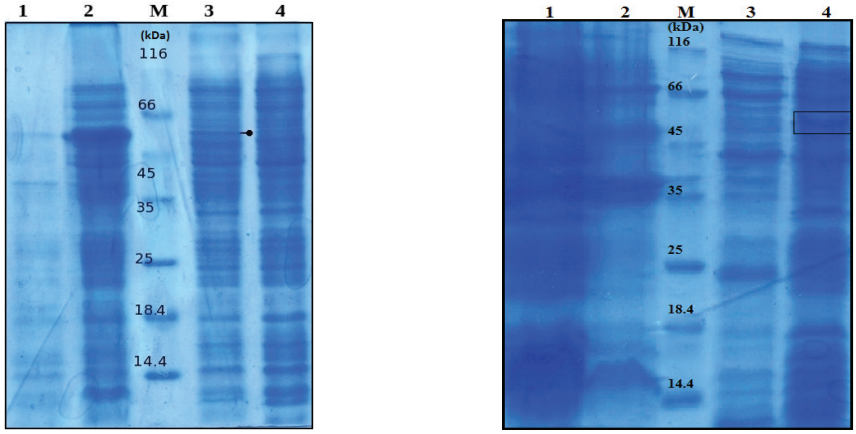


Figure 5.7: SDS-PAGE analysis of r-LaMubP (pET15b+LaMubP) in *E. coli* Rosetta BL21 (DE3): (a) Protein expression results of pET15a-LaMubP with 1.5mM IPTG and 37°C induction, Lane 1: control pET15b insoluble fraction, Lane 2: pET15b+ LaMubP insoluble fraction, Lane M: protein marker, Lane 3: pET15b+ LaMubP soluble fraction, Lane 4: control pET15b soluble fraction; (b) Protein expression results of pET15b- LaMubP with 1.5mM IPTG and 20°C induction, Lane 1: control pET15b insoluble fraction, Lane 2: pET15b+LaMubP insoluble fraction, Lane M: protein marker, Lane 3: control pET15b soluble fraction, Lane 4: pET15b+LaMubP soluble fraction.

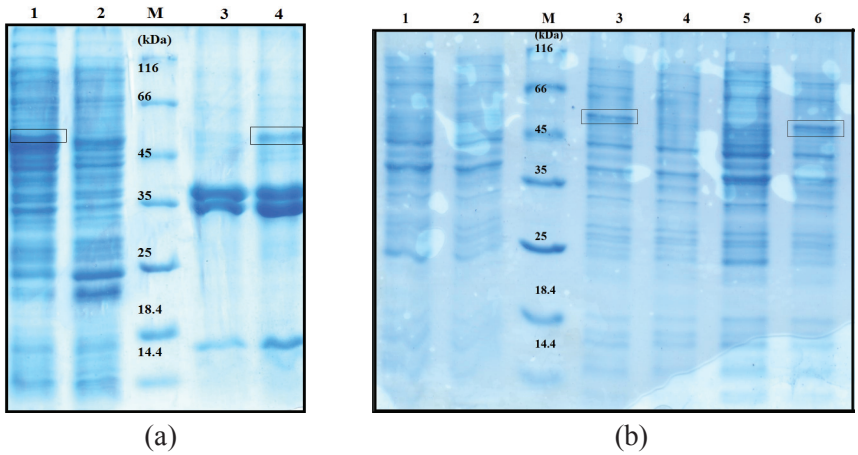


Figure 5.8: SDS-PAGE analysis of r-LaMubP (pET15b+LaMubP) in *E. coli* Rosetta BL21 (DE3) pLysS using autoinduction at 20°C induction: (a) Protein expression results of pET15b+LaMubP construct, Lane 1**:** pET15b+LaMubP soluble fraction, Lane 2: control pET15b soluble fraction, Lane M: protein marker, Lane 3: control pET15b insoluble fraction, Lane 4: pET15b+LaMubP insoluble fraction; (b) Protein expression results of pET15b+LaMubP with 0.5mM IPTG induction and 1% glucose, Lane 1: control pET15b crude expression without 1% glucose, Lane 2: pET15b+LaMubP crude expression without 1% glucose, Lane 3: pET15b+LaMubP crude expression with 1% glucose after 5 hours, Lane 4: control pET15b crude expression with 1% glucose after 5 hours, Lane 5: control pET15b crude expression with 1% glucose after 6 hours, Lane 6: pET15b+LaMubP crude expression with 1% glucose after 6 hours.

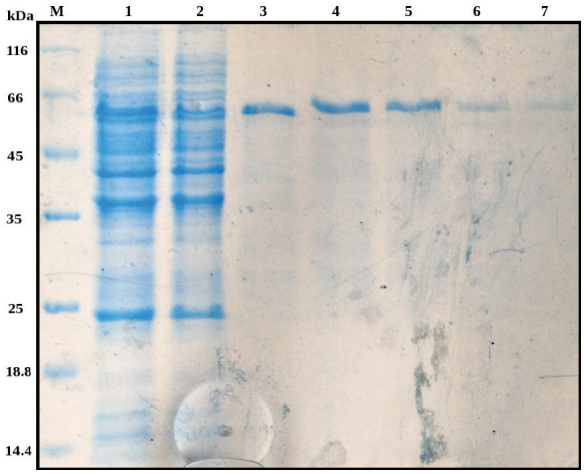


Figure 5.9: SDS-PAGE analysis of r-LaMubP protein purification: Protein purification of pET15b+LaMubP construct, Lane M: protein marker, Lane 1: crude expression of pET15b+LaMubP loaded to Ni-NTA column, Lane 2: flow through of lysate loaded in to Ni-NTA column, Lane 3-7: purified fractions of r-LaMubP protein through Ni-NTA column chromatography.

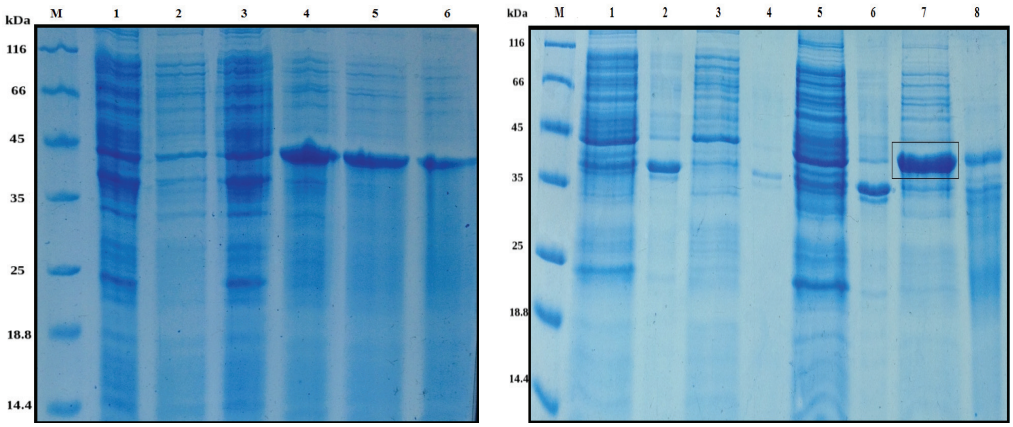


Figure 5.10: SDS-PAGE analysis of r-LaMubPtr (pET30a+LaMubPtr) in *E. coli* Rosetta BL21 (DE3) pLysS at 20°C induction (a) Protein expression crude results of pET30a+LaMubPtr construct, Lane M: protein marker, Lane 1**:** un-induced control pET30a, Lane 2: 0.5mM IPTG induced control pET30a, Lane 3: un-induced pET30a+LaMubPtr, Lane 4: 0.2mM IPTG induced pET30a+LaMubPtr, Lane 5: 0.5mM IPTG induced pET30a+LaMubPtr, Lane 6: 1mM IPTG induced pET30a+LaMubPtr (b) Protein expression results of pET30a+LaMubPtr, Lane M: protein marker, Lane 1: un-induced control pET30a soluble fraction, Lane 2: un-induced control pET30a in-soluble fraction, Lane 3: 0.5mM IPTG induced control pET30a soluble fraction, Lane 4: 0.5mM IPTG induced control pET30a in-soluble fraction, Lane 5: un-induced pET30a+LaMubPtr soluble fraction, Lane 6: un-induced pET30a+LaMubPtr in-soluble fraction, Lane 7: 0.25mM IPTG induced pET30a+LaMubPtr soluble fraction, Lane 8: 0.25mM IPTG induced pET30a+LaMubPtr in-soluble fraction.

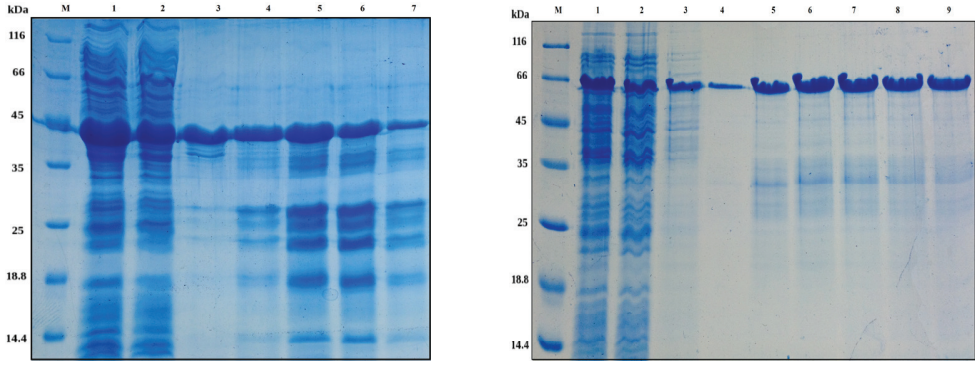


Figure 5.11: SDS-PAGE analysis of r-LaMubPtr (pET30a+LaMubPtr) purification: (a) Protein purification of pET30a+LaMubPtr construct when expressed at 37°C induction, Lane M: protein marker, Lane 1: crude expression of pET30a+LaMubPtr loaded to Ni-NTA column, Lane 2: flow through of lysate loaded in to Ni-NTA column, Lane 3-7: purified fractions of pET30a+LaMubPtr protein through Ni-NTA column chromatography; (b) Protein purification of pET30a+LaMubPtr construct when expressed at 20°C induction, Lane M: protein marker, Lane 1: crude expression of pET30a+LaMubPtr loaded to Ni-NTA column, Lane 2: flow through of lysate loaded in to Ni-NTA column, Lane 3-9: purified fractions of pET30a+LaMubPtr protein through Ni-NTA column chromatography.

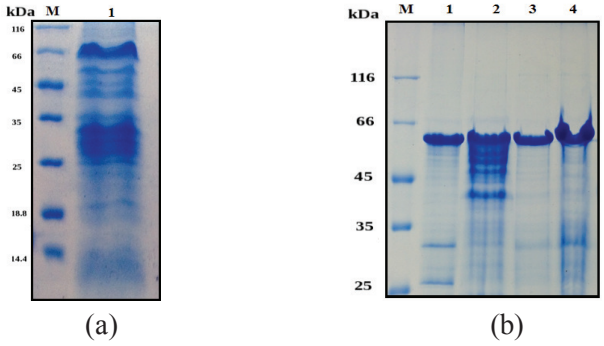


Figure 5.12: SDS-PAGE analysis of dialysed and concentrated purified r-LaMubP and r-LaMubPtr protein: (a) SDS-PAGE analysis of dialysed and concentrated purified r-LaMubP, Lane M: protein marker, Lane 1: concentrated purified r-LaMubP using centricon; (b) SDS-PAGE analysis of dialysed and concentrated purified r-LaMubPtr, Lane M: protein marker, Lane 1, 3: dialysed purified r-LaMubPtr protein, Lane 2: concentrated purified r-LaMubPtr using centricon, Lane 4: concentrated purified r-LaMubPtr using sucrose bed method.

After purification of r-LaMubP and r-LaMubPtr, the recombinant protein was further used for experiments. The purified fraction was dialysed against 20mM Tris and 100mM NaCl at pH 7.8. The dialysed fraction when further pooled to concentrate using centricon, the recombinant proteins were not stable. A sucrose bed method was used to concentrate protein which can be then further used for experiments (Figure 5.12a, b).

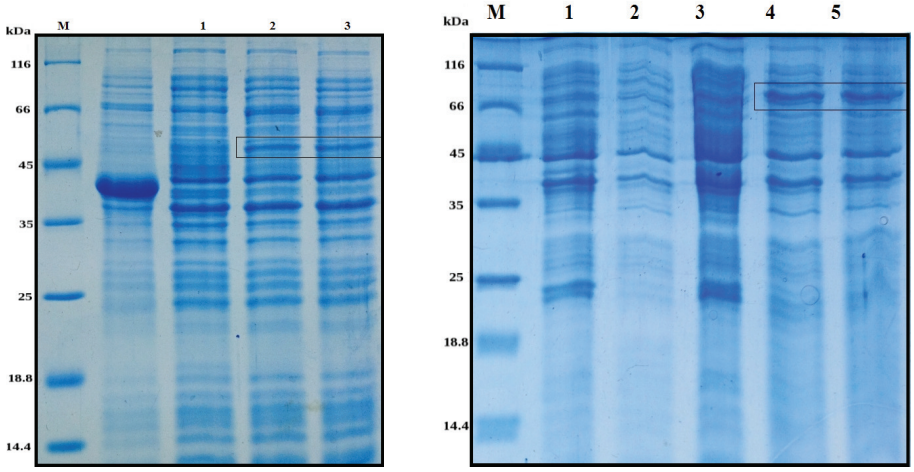
Table 5.2: Summary of expression trials of pET15b+LaMubP in *E.coli* strains.

Cons-truct	*E. coli* expression strain	Experimental conditions		Remark
		IPTG (mM)/ Auto induction	Induction temp.	
pET15b-LaMubP	Rosetta BL21 (DE3)	0.5, 1, 1.5 and 2.0	37°C	+/-
		0.5, 1, 1.5 and 2.0	30°C	+/-
		0.5, 1, 1.5 and 2.0	20°C	+/-; Little increase in solubility of expression
		Autoinduction	20°C	+/-; Little increase in solubility of expression
	Rosetta BL21 (DE3) pLysS	0.5 and 1	20°C	+/-; addition of glucose at time of induction increases the expression
		Autoinduction	20°C	+/-; LaMubP purified through Ni-NTA column in minute amount

(+) indicates soluble expression in supernatant, (-) indicates insoluble expression and (+/-) indicates expression in partially soluble and insoluble fractions.

5.4.2.3. Expression and purification of r-LaFBP: After confirming the positive clone of LaFBP in pET15b and pET28a vector, the protein expression studies were carried out using IPTG as inducers and autoinduction in *E. coli* Rosetta BL21 (DE3) pLysS strain. The expression was detected at around ~55kDa for pET28a+LaFBP and ~65kDa for pET15b+LaFBP at 0.5mM and 1mM IPTG concentration and lower induction temperature of 20°C (Figure 5.13a, b). The theoretical molecular weight of r-LaFBP is 64.7kDa. The expression level of r-LaFBP in *E. coli* Rosetta BL21 (DE3) pLysS strain was not enough to proceed with further experiments.

5.4.2.4. Expression and purification of r-LaFBPtr: After confirming the positive clone of r-LaFBPtr in pET15b and pET28a vector, the protein expression studies were carried out using autoinduction in *E. coli* Rosetta BL21 (DE3) pLysS strain. The protein expression of the pET15b-LaFBPtr construct was mostly in insoluble fraction even at lower induction temperature of 20°C. The protein expression of the pET28a-LaFBPtr construct was optimised further at lower induction temperature of 20°C using autoinduction media. The recombinant protein r-LaFBPtr was purified through Ni-NTA affinity column when expressed in *E. coli* BL21 (DE3) pLysS strain using autoinduction media at 20°C induction temperature (Figure 5.14).



(a) (b)

Figure 5.13: SDS-PAGE analysis of r-LaFBP (pET28a+LaFBP & pET15b+LaFBP) in *E. coli* Rosetta BL21 (DE3) pLysS at 20°C induction: (a) Protein expression crude results of pET28a+LaFBP construct, Lane M: protein marker, Lane 1**:** 0.5mM IPTG induced control pET28a expression, Lane 2: 0.5mM IPTG induced pET28a+LaFBP, Lane 3: 1mM IPTG induced pET28a+LaFBP: (b) Protein expression crude results of pET15b+LaFBP construct, Lane M: protein marker, Lane 1**:** un-induced control pET15b, Lane 2: 0.5mM IPTG control pET15b, Lane 3: un-induced pET15b+LaFBP, Lane 4: 0.5mM IPTG induced pET15b+LaFBP, Lane 5: 1mM IPTG induced pET15b+LaFBP.

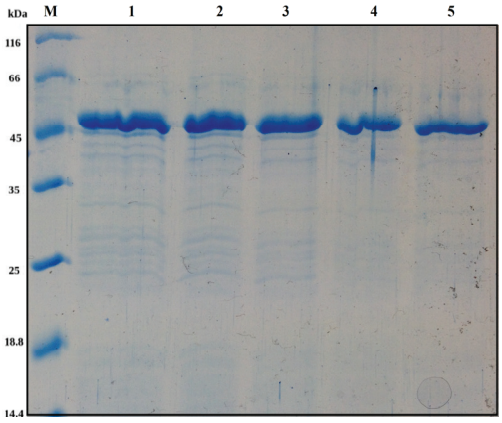
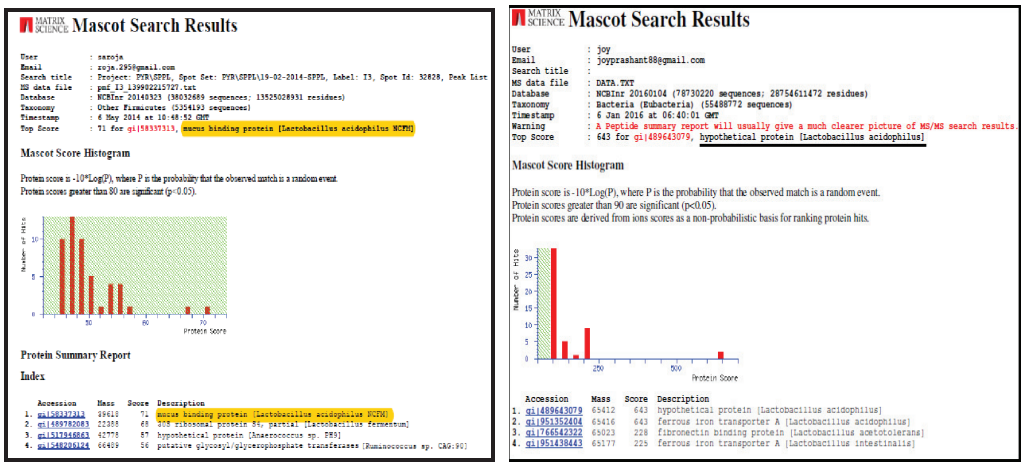


Figure 5.14: SDS-PAGE analysis of r-LaFBPtr (pET28a+LaFBPtr): Lane M: protein marker, Lane 1-5**:** purified fractions of pET28a+LaFBPtr protein through Ni-NTA column chromatography.

5.4.3. Biophysical characterization

5.4.3.1. Peptide mass fingerprinting (MALDI MS/MS) of LaMubP and LaFBPtr: The r-LaMubP protein was purified as a single band and further analysed to identify through MALDI-TOF/TOF analysis which confirmed that the expressed protein is Mucus binding protein (LaMubP) from *L. acidophilus* (Figure 5.15a). The recombinant

r-LaFBPtr protein was purified as a single band and further analysed to identify through MALDI-TOF/TOF analysis which confirmed that the expressed protein is Fibronectin binding protein (LaFBP) from *L. acidophilus* (Figure 5.15b).



(a) (b)

Figure 5.15: MALDI MS/MS analysis of purified protein: (a) MALDI MS/MS confirmation of r-LaMubP protein (b) MALDI MS/MS confirmation of r-LaFBP protein.

5.4.3.2. CD spectroscopy analysis of LaMubP

The recombinant purified r-LaMubP protein was subjected to secondary structure analysis using circular dichroism (CD) spectroscopy. The spectra obtained from far-UV (190nm- 250nm) were converted to molar ellipticity ([θ] deg.cm2.dmol-1) units. The CD spectrum obtained showed a negative band with a maximum dip at around 208 and 215nm (Figure 5.16). Though the CD spectra were not conclusive as it also indicated disorder and coil-coil regions.

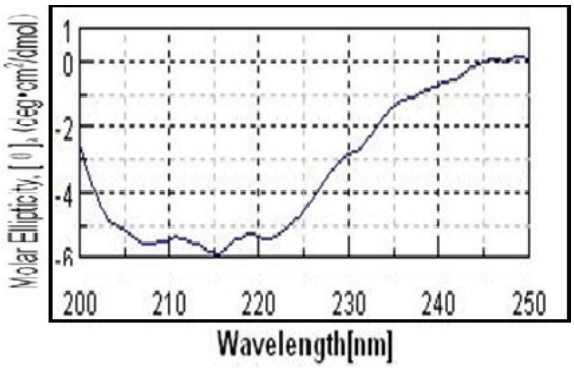


Figure 5.16: Far-UV CD spectrum of r-LaMubP: CD spectra (190-240 nm) in Far-UV range implicating disorder and coil-coil regions.

5.4.4. Bioinformatics analysis

5.4.4.1. Mucus binding protein (LaMubP): Earlier genome mining studies in *Lactobacillus* spp. has identified the presence of mucus-binding proteins of various length consisting of several mucus binding domain repeats (Boekhorst *et al.*, 2006). The genome of *L. acidophilus* was searched for the presence of Mucus binding protein, which revealed the presence of 12 proteins which are annotated as mucus binding protein or putative mucus binding protein. The search with two types of the domain, MUB (Boekhorst *et al.*, 2006) and MucBP (Pfam: PF06458) domains revealed that presence of no such protein in *L. acidophilus* genome. The target for this study (LBA1018) was also subjected to sequence alignment with other 11 mucus binding protein to assess the sequence homology on each other. The sequence homology was in between 27-41%, which showed the match between the domain regions of the proteins (Table 5.3). Blast analysis of LaMubP protein revealed that it is conserved among *Lactobacillus*. Most of the close hits are hypothetical proteins and adhesin proteins. Multiple sequence alignment (MSA) of LaMubP protein sequence with close homologues identified very few conserved residues in the alignment which were sparsely distributed. Although the positively charged tail of LPxTG motif is much conserved amongst all the sequences (Figure 5.17). TM-HMM predicted a presence of TM-helix at position 320-339 amino acids, which is an LPxTG anchoring motif region.

Table 5.3: Sequence search results of LBA1018 against all other 11 mucus binding protein from *L. acidophilus*

ID	Protein name	Length	Query coverage	E-value	Identity
LBA1652	Mucus binding protein	1174	43.00%	3.00E-04	38.00%
LBA1019	Putative mucus binding	2650	71.00%	3.00E-08	28.00%
LBA1020	Putative mucus binding	2310	67.00%	2.00E-06	30.00%
LBA1017	Putative mucus binding	294	58.00%	1.00E-08	27.00%
LBA1392	Mucus binding protein	4326	58.00%	1.00E-09	31.00%
LBA1709	Mucus binding protein	1208	73.00%	1.00E-05	29.00%
LBA1018	Putative mucus binding	346	100.00%	0	100.00%
LBA1609	Putative mucus binding	643	23.00%	2.20E-02	33.00%
LBA1377	Putative mucus binding	1017	54.00%	7.00E-03	27.00%
LBA1218	Putative mucus binding	697	39.00%	1.60E-02	41.00%
LBA1016	Putative mucus binding	185	64.00%	2.00E-16	33.00%
LBA1460	Putative mucus binding	339	26.00%	4.00E-01	27.00%

The phylogenetic analysis of LaMubP with other homologues showed that it is closely related to MucBP domain protein from *Chlamydia trachomatis* than *Lactobacillus* (Figure 5.18). The secondary structural elements of LaMubP protein were predicted by PSIPRED server, and it shows the major presence of coil and sheet region as a major secondary structural element along with a single helix in the LPxTG motif anchor which in agreement with model generated by I-tasser (Figure 5.19a, b). The modelled PDB file after resubmission to ModRefiner tool for refinement, the Ramachandran plot showed that 74.8% amino acids in the most favoured region and 21.3% amino acids in additional allowed regions while 1.9% amino acids in disallowed regions.

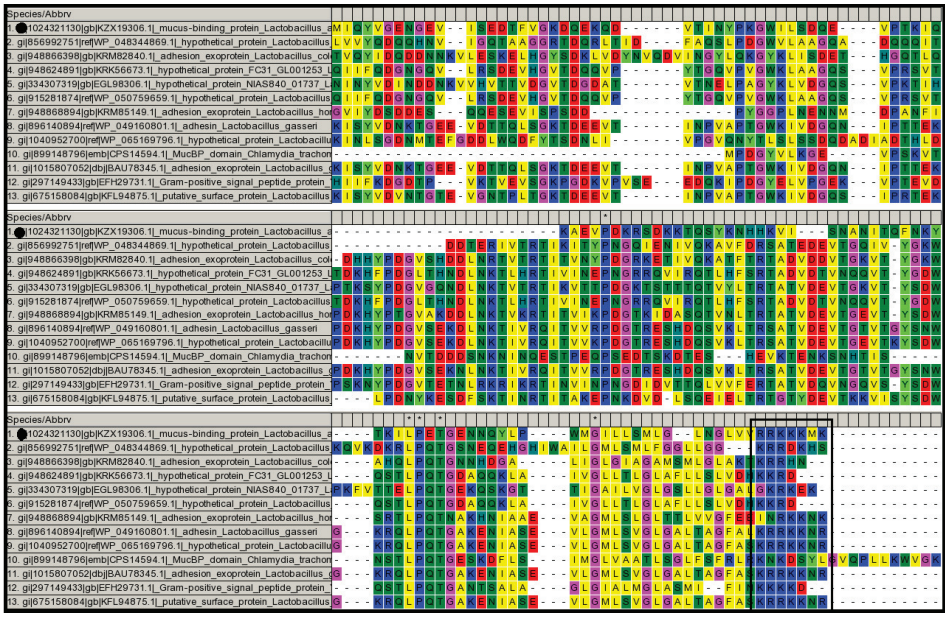


Figure 5.17: Multiple sequence alignment (MSA) analysis of LaMubP: The conserved positive charged LPxTG tail region is marked with a box. The sequence LBA1018 is marked with a black circle.

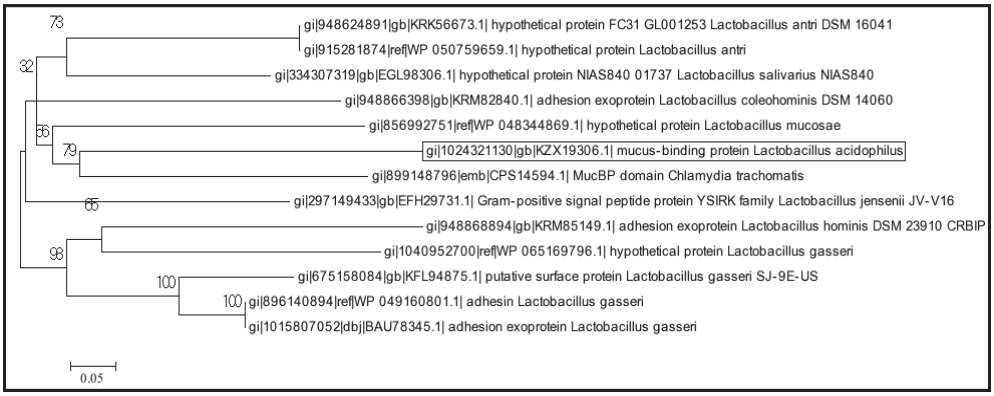
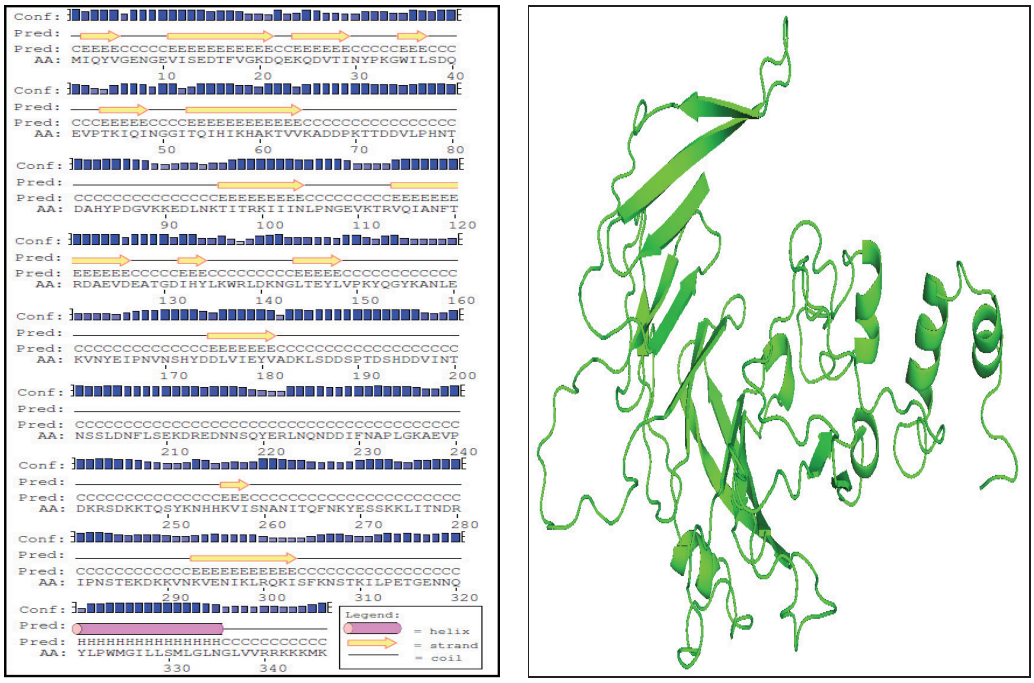


Figure 5.18: Phylogenetic analysis of LaMubP: The sequence LBA1018 is marked with a black square.



(a) (b)

Figure 5.19: Secondary structure prediction and model generation of LaMubP: (a) Secondary structure prediction of LaMubP using PSIPRED; (b) Model generation of LaMubP through I-tasser. [C-score: -2.44].

5.4.4.2. Fibronectin binding protein (LaFBP): The genome of *L. acidophilus* was searched for the presence of fibronectin binding protein, which revealed the presence of LBA1148, Fibronectin binding protein (FBP) which has Fibronectin-binding protein A, N-terminus (FbpA) domain (PF05833). Apart from the primary domain, a domain of unknown function was also present known as DUF814. The functional role of DUF814 at the C-terminal region is still unknown and the presence of which is characterised by [DE]-x-[WY]-x-H motif that may functionally important. The LaFBP sequence has this motif at 474 amino acid position (DYWFH). BLAST analysis for sequence homology search excluding *Lactobacillus* species resulted into a high matching sequence homologs. The residues in the domain of these sequences are highly conserved. Blast analysis of LaFBP protein showed that it is conserved among *Lactobacillus,* and apart from *Lactobacillus* the closest matching homologs are mostly fibronectin binding proteins, hypothetical proteins and adherence protein. Multiple sequence alignment (MSA) of LaFBP protein sequence with close homologues identified highly conserved residues in the alignment. Most of the positive charge residues in all the sequences are highly conserved (Figure 5.20). The phylogenetic analysis of LaFBP showed that it is closely related to FbpA N-terminus domain protein from *Chlamydia trachomatis* and

both of the proteins forms a separate clade (Figure 5.21).

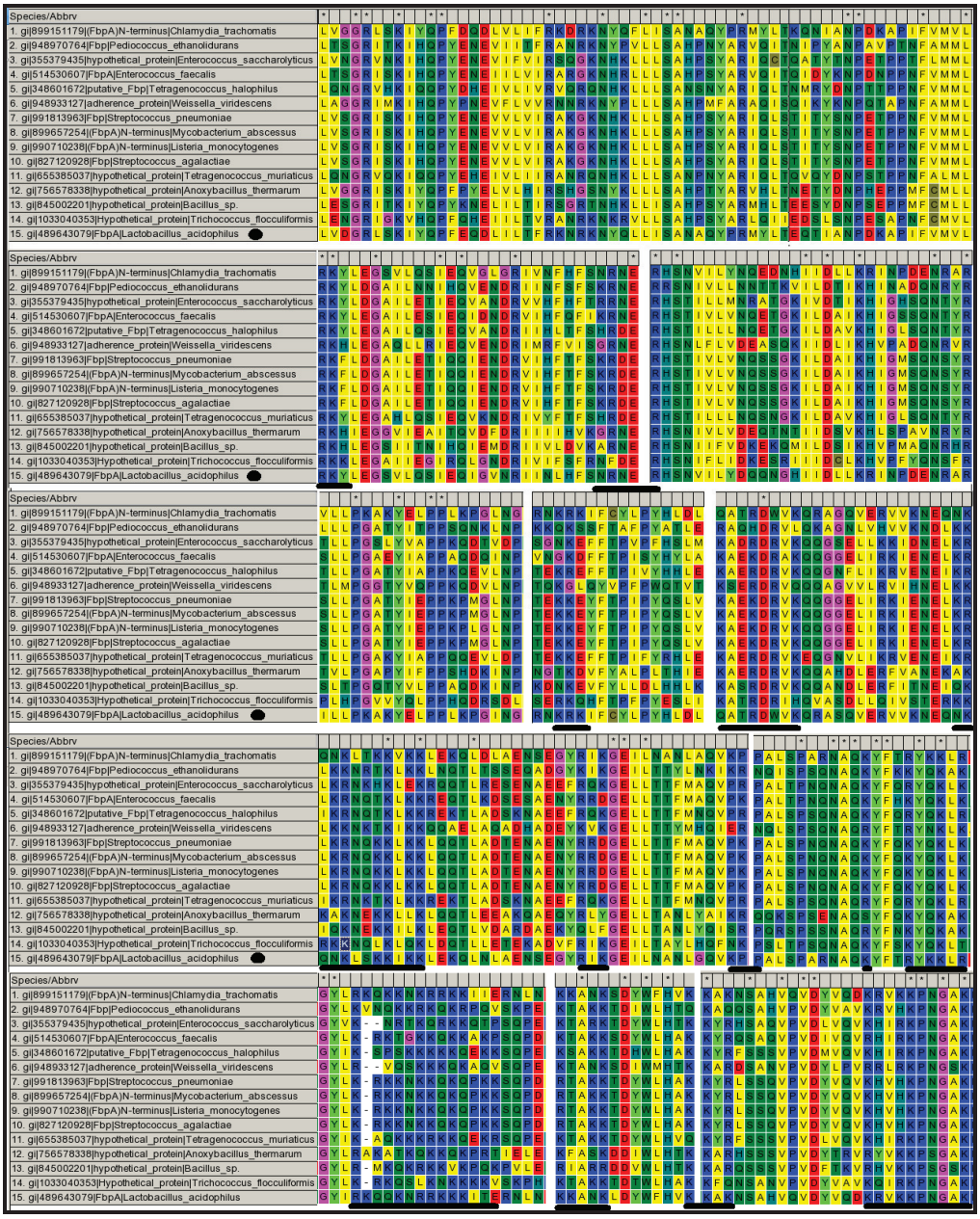


Figure 5.20: Multiple sequence alignment (MSA) analysis of LaFBP: The conserved positive charge amino acids are marked with an underline. The sequence LBA1018 is marked with a black circle. Note: The figure doesn't represent the complete sequence alignment of MSA due to the large size of the protein, rather representing only conserved regions.

The secondary structural elements of LaFBP protein were predicted by PSIPRED server, and it shows the presence of a mixture of alpha helix and beta sheet. Though alpha helix exists as a major secondary structural element which in agreement with

model generated by I-tasser (Figure 5.22a & b). The modelled PDB file after resubmission to ModRefiner tool for refinement, the Ramachandran plot showed that 77.5% amino acids in the most favoured region and 17.2% amino acids in additional allowed regions while 2.5% amino acids in disallowed regions.

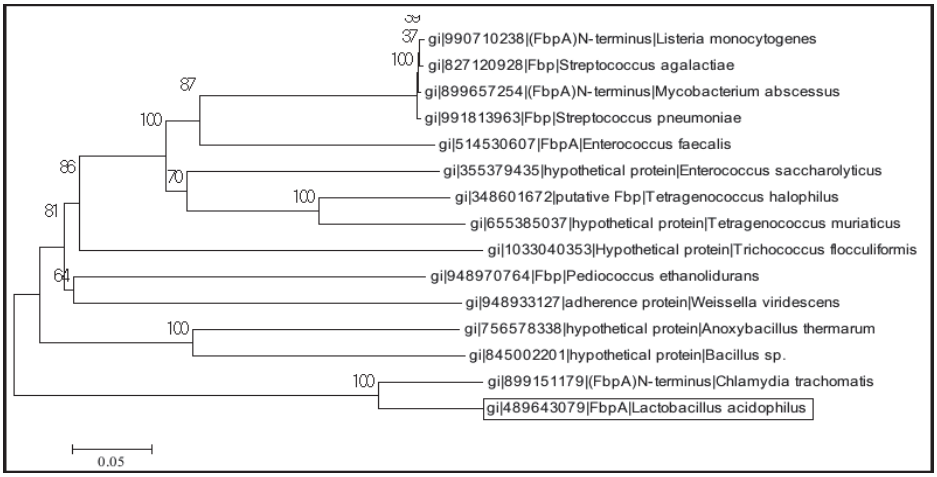
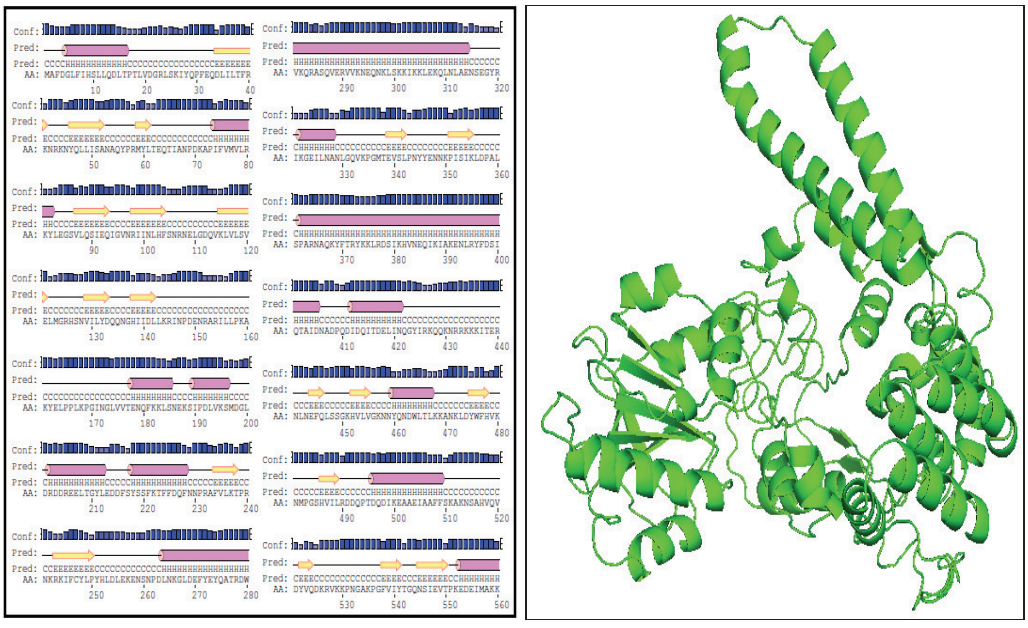


Figure 5.21: Phylogenetic analysis of LaFBP: The sequence LBA1148 (FBP) is marked with a black square.



(a) (b)

Figure 5.22: Secondary structure prediction and model generation of LaFBP: (a) Secondary structure prediction of LaFBP using PSIPRED; (b) Model generation of LaFBP through I-tasser. [C-score: -1.42].

5.5. Discussion

Adhesion of lactic acid bacteria to host tissues, epithelial and mucosa is a prerequisite step for bacterial colonisation. It is a complex process involving multiple different factors. One of the prime factors of adhesion is mediated by cell surfaces proteins also referred to as adhesins which are involved in specific recognition of host components, a central process in bacterial colonisation and adhesion. These adhesins can be broadly categorised as S-layer proteins, LPxTG anchoring proteins, transporter proteins, and N- or C-terminal membrane protein along with multifunctional house-keeping proteins which are also involved in physiological processes in the bacteria. The most studied adhesins include typical pili and fimbriae from enteropathogens (Kline *et al.*, 2009). In contrast, there is a lacuna in structural information on interactions between Gram-positive adhesins and the host extracellular matrix. These adhesins are believed to be involved in commensal or probiotics adhesion to mucus layer in the human gut. Thus for further understanding and characterization, the two adhesin molecules were selected as a candidate target for the current study are Mucus binding protein (LBA1018) - LaMubP and Fibronectin binding protein (LBA1148) - LaFBP from *L. acidophilus*.

Mucus binding protein is a multi-domain protein whose overall protein architecture is yet to be explored. It exhibits a C-terminal sortase recognition motif (LPxTG) which covalently anchor protein to peptidoglycan, multiple repeats of mucus binding domain (MUB), an N-terminal gram-positive secretion signal, all of them are typical of gram-positive cell surface proteins. MUB repeats domain are classified in to type 1 or type 2 depending on amino acid composition (Boekhorst *et al.*, 2006). Fibronectin-binding protein (FBP), are large adhesins which recognise a large dimeric glycoprotein known as fibronectin (450 kDa). Fibronectin is found in soluble form in body fluids and immobilised form in ECM by which it has critical roles in eukaryotic cellular processes, such as adhesion, migration and differentiation and also acts as an adhesion site for the attachment of bacteria. FBP are found across Gram-negative and Gram-positive bacteria and is often considered as an important virulence factor for adhesion in many pathogens such as *Streptococcus pneumoniae* and *Streptococcus pyogenes* (Holmes *et al.*, 2001; Jedrzejas, 2007; Molinari *et al.*, 1997). The presence of FBP in bacteria has made them effector adhesion mediating molecule both in pathogens and commensals (Henderson *et al.*, 2011; Styriak *et al.*, 2003).

In the current study, Mucus binding protein (LaMubP) was successfully cloned in

to pET15b vector for heterologous overexpression of the protein as recombinant HIS-tagged r-LaMubP in *E. coli* expression system. However, significant overexpression was not detected through SDS-PAGE while expressing in *E. coli* Rosetta BL21 (DE3) strain using IPTG induction. Assuming that the Gram-positive bacteria specific LPxTG motif might be toxic to heterologous host for expression, the r-LaMubP expression was also overexpressed in *E. coli* Rosetta BL21 (DE3) pLysS strain using autoinduction to overcome toxicity through leaky expression. The identity of purified r-LaMubP protein was confirmed by MALDI-TOF analysis of trypsinized digested band as MubP from *L. acidophilus* NCFM (UniProt ID: LBA1018). Though the actual calculated molecular weight of r-LaMubP should be 39.6 kDa, r-LaMubP showed expression at a molecular weight of ~65 kDa. The electrophoretic mobility of r-LaMubP proteins gave molecular weight estimates higher than predicted molecular weight by aberrant migration on SDS-PAGE, which is observed in earlier studies of mucus binding protein (Bumbaca *et al.*, 2007; MacKenzie *et al.*, 2009). This might be due to the unusual conformation of the r-LaMubP protein, aggregation of protein or membrane proteins tends to show aberrant migration on SDS-PAGE. As the expression levels were very low, a truncated gene LaMubPtr was also cloned in pET30a and expressed in *E. coli* Rosetta BL21 (DE3) pLysS. Like r-LaMubP, the r-LaMubPtr also showed aberrant migration on SDS-PAGE ruling out the involvement of LPxTG motif mainly positively charged tail causing aberrant migration. But though when expressing r-LaMubPtr at higher induction temperature (37°C), the expression analysed on SDS-PAGE showed correct molecular weight estimation. The overall stability and expression levels of r-LaMubPtr were increased when using 0.1% mucin in lysis buffer and 10% glycerol and 0.1% Triton X-100 in all purification buffer. The purified protein r-LaMubP and r-LaMubPtr when subjected to centricon, a degradation was observed on SDS-PAGE. This might be due to the removal of glycerol from the sample during centrifugation in centricon. An alternate method using sucrose bed for protein concentration was used to avoid degradation. The purified r-LaMubP and r-LaMubPtr proteins were not stable and didn't express in enough quantity, thus making it very difficult to proceed for further biochemical and biophysical studies.

Bioinformatics analyses indicated that the MUB adhesins are characterized by the type, number and domain-containing multi-Mub repeats are often associated with *Lactobacilli* that are found in GI tract, whereas strains which are less frequently encounter in GI tract or gut pathogens were more frequently limited to single or double

copies (Kleerebezem *et al.*, 2003). This arrangement is consistent with the niche adaptation of bacteria in the gut and further suggesting evolutionary role of MUB domain in GI tract adherence. Genome mining in *Lactobacillus* spp. has identified the presence of mucus-binding proteins of various size consisting of several mucus binding domain repeats (Boekhorst *et al.*, 2006). The genome of *L. acidophilus* has 12 mucus binding protein although none of them had the presence of canonical MUB (Boekhorst *et al.*, 2006) and MucBP (Pfam: PF06458) domains. LaMubP (LBA1018) was known to have 27-41% sequence homology with other 11 mucus-binding proteins due to variability in size due to the occurrence of MUB repeats. Sequence homology searches using BLAST identified a hypothetical protein, surface proteins and adhesin proteins as suitable homologues mostly from *Lactobacillus*. The LaMubP sequence has no signal peptide for secretion though it has a transmembrane helix at C-terminal which is the LPxTG anchoring motif which helps to anchor the surface protein to the cell via C-terminal and the N-terminal of the protein remains outside exposing to the cell environment. The genome analysis indicated that LBA1018 is encoded next to LBA1019 protein which does have a signal peptide, multiple MUB domain and an LPxTG motif. It indicated that they would be secreted and be functional at the cell wall through some other mechanism (Boekhorst *et al.*, 2006). MSA analysis showed that there are very few conserved amino acids sparsely throughout the mucus binding domain, but the C-terminal tail positive charge residues (K/R) of LPxTG motif are highly conserved. The closest protein in phylogenetic analysis is MucBP domain protein from *Chlamydia trachomatis* rather than proteins from *Lactobacillus.* The secondary structural elements of LaMubP protein predicted by PSIPRED server showed the presence of coil and sheet region as a major secondary structural element along with a single helix in the LPxTG motif anchor which in agreement with model generated by I-tasser and CD spectroscopy analysis.

The other target for the current study is Fibronectin binding protein, a 65kDa adhesion protein involved in binding with ECM component - Fibronectin. A Fibronectin binding protein (FbpA) from *L. acidophilus* (LaFBP) was successfully cloned in to pET15b and pET28a vectors for heterologous overexpression of the protein as recombinant HIS-tagged r-LaFBP in *E. coli* expression system. However, significant overexpression was not detected through SDS-PAGE while expressing in *E. coli* Rosetta BL21 (DE3) pLysS strain using IPTG induction as well as autoinduction method. The expression was detected at around ~55 kDa for pET28a+LaFBP and ~65 kDa for

pET15b+LaFBP at lower induction temperature of 20°C, while the theoretical molecular mass of r-LaFBP is 65kDa. A truncated gene LaFBPtr was also cloned in to pET15b and pET28a vectors for heterologous overexpression as recombinant HIS-tagged r-LaFBPtr in *E. coli* expression system. The truncated LaFBPtr clone has fibronectin binding domain A and without a DUF814 domain, a domain of unknown function. The r-LaFBPtr, a 48kDa protein was expressed and purified from *E. coli* Rosetta BL21 (DE3) pLysS strain using autoinduction and further the identity of purified r-LaFBPtr protein was confirmed by MALDI-TOF analysis of trypsinized digested band as LaFBP from *L. acidophilus* NCFM (UniProt ID: LBA1148).

The genome of *L. acidophilus* consisted of a single fibronectin binding protein (FBP), until recently an S-layer associated fibronectin binding protein was identified (LBA0191) (Hymes *et al.*, 2016). The LaFBP is a 563 amino acid protein believed to bind with fibronectin via Fibronectin-binding protein A, N-terminus (FbpA) domain (PF05833). Another domain of unknown function DUF814 (PF05670) was also present in the C-terminus of the protein. Bioinformatics analysis suggests this the occurrence of this domain is along with FbpA domain. The functional role of DUF814 is still unknown and the presence of which is characterised by [DE]-x-[WY]-x-H motif that may functionally important. The LaFBP sequence has this motif at 474 amino acid position (DYWFH) whose functional role has not been identified yet. Sequence homology search using BLAST identified that it is conserved among *Lactobacillus,* and apart from *Lactobacillus*, the closet matching homologs are mostly fibronectin binding proteins, hypothetical proteins and adherence protein. MSA analysis of LaFBP protein sequence with close homologues apart from *Lactobacillus spp.* identified highly conserved residues in the alignment. A large number of positive charge amino acids (K/R) were found to be highly conserved throughout the domain which might have a role in binding with certain negatively charged ECM components. The phylogenetic analysis of LaFBP showed that it is closely related to FbpA N-terminus domain protein from *Chlamydia trachomatis* and both of the proteins forms a separate clade. The secondary structural elements of LaFBP protein were predicted by PSIPRED server, and it shows the presence of a mixture of alpha helix and beta sheet. Though alpha helix exists as a major secondary structural element which in agreement with model generated by I-tasser. However, the FbpA lacks any known signal peptide as well as a transmembrane helix, though it is recognised as a surface molecule mediating adhesion. The underlying mechanism remains yet to be explored.

The characterization of these adhesins molecules is of great importance as both the adhesins FBP and MubP are believed to play an important role in recognition of ECM host components and thus helps *Lactobacillus* to adhere and persist in the intestinal mucus layer. FBP and MubP appear to act as a bridging protein between the bacterial surface and the ECM of the host, although the molecular mechanism underlying this interaction is not yet understood. As both the interacting ECM components mucin and fibronectin are glycoproteins, the interaction between these adhesins and ECM seems to be lectin-like recognition. Although earlier studies using competitive adhesion study showed that the binding of MubP to mucus was inhibited by fetuin, asialofetuin as well as fucose, indicating that MubP interacts with oligosaccharides, the molecular and structural mechanism is not yet understood (Roos & Jonsson, 2002). Earlier study involving isogenic mutant containing a functional inactivated FbpA (LBA1148) and MubP (LBA1392), exhibited 76% and 65% decrease in adhesion to Caco-2 cells (Buck *et al.*, 2005).

Collectively, the chapter describes the expression trials of r-LaMubP, r-LaMubPtr, r-LaFBP and r-LaFBPtr and further characterization. The expression and purification of r-LaMubP are very difficult as it is highly disorder and random coil-coil like structural protein. The r-LaFBP is too huge protein for its successful expression and purification. Although the r-LaFBPtr protein was purified successfully, further biochemical and structural characterization are needed. The r-LaMubPtr was successfully purified and stabilised using glycerol and Triton, but any further functional studies using the protein in the same condition is challenging. We hope that further structural studies on these molecules will throw some light on the underlying adhesion mechanism at molecular level.