

## **Chapter 6**

### ***Summary and conclusion***

The probiotics bacteria protect the gut against colonisation of pathogenic microbes by the direct mechanism of limiting nutrients and attachment sites or indirect mechanism of modulation of host immune system (Servin, 2004). *Lactobacilli* as well-known probiotics are naturally persistence at mucosal surfaces, principally the gastrointestinal tract, the vagina and the oral cavity. The prerequisite step in bacterial colonisation is adhesion to host tissues which promote persistence time for colonisation and in turn modulates microbe-host interaction (Beachey, 1981). The microbe-host adhesion process is mediated through the electrostatic interactions, passive forces, steric forces, hydrophobic interactions, lipoteichoic acids and specific adhesins/lectins (Bermudez-Brito *et al.*, 2012). In *Lactobacillus*, the proteins exposed/embedded in the bacterial cells membrane along with extracellular proteins plays a crucial role in interactions with the environment especially adhesion to the intestinal milieu.

In Chapter 2, a detailed bioinformatics analysis of *Lactobacillus acidophilus* genome was done to predict the secretome. The resulting secretome consists of two categories of proteins: one is the extracellular proteins which after secretion is covalently and non-covalently attached to the components of the bacterial cell wall or remain anchored to the membrane through one or more transmembrane helices or covalent coupling to lipids and another category of secreted protein released into the environment. In a total of 1,859 proteins of *L. acidophilus* genome, the experimentally reviewed proteins are 282 while the remaining 1,577 were electronically annotated proteins from which 718 proteins are putatively annotated. From our analysis, we found that 1,274 proteins are intracellular while 585 proteins are multi-transmembrane, secreted/surface associated or part of integral membrane proteins. Our final prediction resulted in a prediction of 223 extracellular proteins which are also known as secretome. The secretome comprises of 12% of the total proteins with the majority of them having motif or domain for attachment to the cell surface. The detailed bioinformatics analysis revealed that 12 proteins possess a C-terminal LPxTG anchoring motif for covalent attachment to cell wall, 42 proteins contain an N-terminal lipobox motif for covalent attachment to membrane, while 20 proteins have been found to contain a cell-wall binding domain for non-covalent attachment to the cell wall, and 16 proteins of them contain a surface layer protein domain (SLAP). The remaining 108 of extracellular proteins were found to be N-terminal anchor proteins, and the 59 proteins were predicted to be either released (i.e. secreted) or associated with cell wall with another unknown mechanism. From the total 223 proteins as secretome, 22 proteins belongs to

surface/membrane and adhesion class each, the majority of them i.e. 62 proteins belongs to enzyme class while 80 (~36%) proteins are unknown or uncharacterized. Sixteen proteins were found to contain adherence domain with 14 having mucus binding domain and 2 having fibronectin binding domain.

In Chapter 3, a detailed biochemical and biophysical characterization of *L. acidophilus* GAPDH - LBA0698 (LaGAPDH) was performed. Earlier studies on GAPDH from different organism is known to bind to host ECM components (fibronectin, mucin, plasminogen etc.) and directly to Caco-2/HT29 epithelial cell lines thus making it an essential mediator of host-microbe interactions (Kinoshita *et al.*, 2008; Pancholi & Fischetti, 1992; Ramiah *et al.*, 2008). The recombinant r-LaGAPDH protein cloned in pET28a and significant overexpression is achieved in *E. coli* Rosetta (DE3) strain at ~40kDa. Molecular weight determination revealed that r-LaGAPDH exists as a homotetramer in solution, while under a non-reducing condition in SDS-PAGE, r-LaGAPDH showed the presence of a single band at ~40kDa ruling out the possibility of oligomerization due to disulphide linkage despite the presence of two unusual cysteine residues (101 and 326) along with a conserved cysteine (156). The CD spectroscopy indicated that the r-LaGAPDH had mixed alpha helical and beta sheets as its secondary structure which complements with the secondary structure prediction and tertiary structure prediction based on homology modelling.

Finally, to evaluate and understand protein-glycan interaction, mucin binding assay and hemagglutination studies were performed. Our studies indicated that r-LaGAPDH had positive binding interaction with mucin and r-LaGAPDH agglutinated rabbit RBCs which was inhibited by complex glycoprotein fetuin. To further exploit the molecular interaction, a fluorescence-based spectroscopy study revealed r-LaGAPDH interactions with four carbohydrates mannose, galactose, N-acetyl-D-galactosamine (GalNAc) and N-acetyl-D-glucosamine (GlcNAc) primarily due to their significant presence in mucin. The study showed an interaction of r-LaGAPDH with mannose, galactose, GalNAc and GlcNAc with a  $K_d$  of  $3.6 \pm 0.7 \times 10^{-3}M$ ,  $4.34 \pm 0.09 \times 10^{-3}M$ ,  $4.0 \pm 0.87 \times 10^{-3}M$  and  $3.7 \pm 0.28 \times 10^{-3}M$  respectively. Mannose showed the highest affinity followed by GlcNAc which has little higher affinity than GalNAc for r-LaGAPDH that may be due to the difference in orientation of C4 hydroxyl group. Our enzymatic study showed that the secreted LaGAPDH from native *L. acidophilus* cells is enzymatically active which is with an agreement with the earlier study of *Streptococcus*

GAPDH (Pancholi & Fischetti, 1992). This ability indicates the maintenance of the protein structure and folds throughout the export process. Further, the crystallisation of r-LaGAPDH was carried out to evaluate the mechanistic studies of carbohydrates on adherence.

In chapter 4, a detail crystallisation setup trials, data collection and data processing of r-LaGAPDH crystals is discussed. The purified r-LaGAPDH protein crystallised in microbatch under oil method where diffraction quality r-LaGAPDH native crystals were obtained in two conditions: 25% w/v Polyethylene glycol 1,500 and 10% w/v Polyethylene glycol 1,000, 10% w/v Polyethylene glycol 8,000. Finally the data was processed in later crystallization condition in tetragonal space group P41212 with the unit cell parameters were  $a = 114.98$ ,  $b = 114.98$ ,  $c = 113.29$  Å,  $\alpha = \beta = \gamma = 90$ . A total of 661620 reflections were collected, of which 38437 were unique; the completeness of data was 99.9%. Assuming the presence of two monomeric molecules (with a calculated molecular weight of 38,933 Da each) per crystal asymmetric unit, the calculated Matthews coefficient and solvent content were  $2.41$  Å<sup>3</sup> Da<sup>-1</sup> and 48.96% respectively. The MR was performed to get the solution of r-LaGAPDH native using search molecule PDB: 4QX6, with the initial MR model with a score of RFZ=10.3, TFZ=29.1 and LLG=532. The MR solution model was further subjected to refinement using REFMAC5 in CCP4 package along with iterative rounds of model building and refinement with 20 cycles of restrained refinement. The electron density map in most of the region fitted well with model except for a few regions between 26-30 residues and 104-110 residues. The data further validated with MolProbity and PDB validation server is finally submitted to PDB with PDB ID: 5J9G.

The crystal structure of GAPDH from *L. acidophilus* (LaGAPDH) in apo-form have been determined at 2.21 Å by the MR with two monomers in the asymmetric unit, containing 5306 atoms, 676 amino acid residues and 179 water molecules with a final  $R_{\text{factor}}$  and  $R_{\text{free}}$  of 0.21 and 0.23 respectively. The condition of native data which diffracted was tried with numerous combinations of protein: precipitant ratio, protein concentration and buffer optimisation, a manual variation of PEGs and usage of cryoprotectant with the majority of them diffracted very poorly or few with no diffractions. The diffracted data were not able to solve due to errors while data processing with high  $R_{\text{merge}}$  values. A wide variety of soaking and co-crystallization trials was performed. Only four data were collected but none of them processed successfully.

In the bioinformatics analysis, sequence comparison with other GAPDH revealed that LaGAPDH contained three cysteines, one more than most of other GAPDHs which have two conserved cysteines at 156 and 160 positions. Along with LaGAPDH, only thermophilic class of bacteria was found to be lacking a cysteine at position 160. But interestingly LaGAPDH has two more unusual cysteines: one at 101 positions which are found in only seven other GAPDHs and a second at position 326 which is not conserved at all except in LaGAPDH and *Aquifex aeolicus* GAPDH. The cysteine at 101 positions is also right in the active site where NAD binds, this may be interesting to explore further. MSA analysis found a stretch of three amino acids “TAG” starting at position 305 in LaGAPDH protein which is found only in three species: *Streptococcus*, *Staphylococcus* and *Lactobacillus*. All these three GAPDH have been reported earlier to be on the cell surface or either secreted. Another stretch of amino acids “EKSK” starting at position 26 is found only in the *L. plantarum* GAPDH apart from LaGAPDH. GAPDH from *L. plantarum* is previously known to bind mucin by interacting with the human blood antigen present on mucin (Kinoshita *et al.*, 2008). The stretch at 305 positions is of more interesting as it forms an extended secondary structure on the surface exposed region of the LaGAPDH protein structure.

Chapter 5 includes the cloning, expression and purification trials for two adhesin candidate: mucus binding protein (LBA1018) -LaMubP and fibronectin binding protein (LBA1148) -LaFBP from *L. acidophilus*. A mucus binding protein (LaMubP) was cloned in to pET15b vector. However, significant overexpression was not detected in *E. coli* Rosetta BL21 (DE3) strain using IPTG induction. Also, its overexpression in *E. coli* Rosetta BL21 (DE3) pLysS strain using autoinduction was carried out to overcome toxicity if any through leaky expression. MALDI-TOF analysis confirmed the purified r-LaMubP protein as MubP from *L. acidophilus* NCFM (UniProt ID: LBA1018). The electrophoretic mobility of r-LaMubP proteins gave molecular weight estimates higher (~65kDa) than the predicted molecular weight (39.6kDa) by aberrant migration on SDS-PAGE, which is observed in earlier studies of mucus binding protein (Bumbaca *et al.*, 2007; MacKenzie *et al.*, 2009). As the expression levels were very low, a truncated gene LaMubPtr which excludes the LPxTG motif was also cloned in pET30a and expressed in *E. coli* Rosetta BL21 (DE3) pLysS. The purified r-LaMubPtr protein also showed aberrant migration on SDS-PAGE ruling out the involvement of LPxTG motif, particularly the positively charged tail. But though when expressing r-LaMubPtr at higher induction temperature (37°C), the expression analysed on SDS-PAGE showed

correct molecular weight estimation. While using 0.1% mucin in lysis buffer and 10% glycerol and 0.1% Triton X-100 in all purification buffer, the overall stability and expression levels of r-LaMubPtr increased. The protein yield along with its stability was not enough for purified r-LaMubP and r-LaMubPtr proteins, thus making it tough to proceed for further biochemical and biophysical studies.

The LaMubP (LBA1018) protein lacks the presence of canonical MUB (Boekhorst *et al.*, 2006) and MucBP (Pfam: PF06458) domains, though our bioinformatics prediction using HMM profile of known mucus binding protein predicted the presence of MUB domain profile. LaMubP was known to have 27-41% sequence homology with other mucus-binding proteins due to variability in size due to the occurrence of MUB repeats. 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 thus the N-terminal of the protein remains outside exposing to the cell environment. 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 predicted secondary structural elements of LaMubP protein shows the presence of coil and sheet region as a significant 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) is successfully cloned in to pET15b and pET28a vectors. 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. A truncated gene LaFBPptr consisting of complete fibronectin binding domain A and without a DUF814 domain was also cloned in to pET15b and pET28a vectors. The r-LaFBPptr, a 48kDa protein was expressed and purified from *E. coli* Rosetta BL21 (DE3) pLysS strain using autoinduction and was confirmed by MALDI-TOF analysis as LaFBP from *L. acidophilus* NCFM (LBA1148). The genome of *L. acidophilus* was known to encode a single fibronectin binding protein (FBP) up till the discovery of an S-layer associated fibronectin binding protein (LBA0191) (Hymes *et al.*, 2016). Through MSA analysis a significant 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 prediction secondary structural elements of LaFBP protein shows the presence of a mixture of alpha helix and beta sheet, though alpha helix exists as a significant 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 recognised as a surface molecule mediating adhesion. The underlying mechanism is not yet explored and remains worthwhile investigating.