Chapter 4

Crystallisation and structure solution of Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) from *Lactobacillus acidophilus*.

4.1. Introduction

Protein is large macromolecule which consists of long chain(s) of amino acids and plays a significant role in the maintenance of life by involvement in vast biological functions within living organisms. The dynamics and structure of these molecules provide vital information about their mode of functions, and macromolecular crystallography is a powerful tool used to determine the 3-dimensional arrangement of atoms in proteins. Thus X-ray crystallography provides vital information with valuable insights into the specific biological roles these molecules play and a crucial and foremost step towards in this technique, is the optimisation of conditions for getting suitable diffracting crystals (Blundell & Johnson, 1976).

GAPDH is a prototype moonlighting protein involved in an array of functions unrelated to its primary enzymatic functions. Reports suggest that GAPDH from gram-positive bacteria participates in host adhesions, but the detailed mechanistic and structural studies are lacking (Kinoshita, Uchida*, et al.*, 2008; Pancholi & Fischetti, 1992). Structural studies on *L. acidophilus* GAPDH combined with biochemical analysis would enable us to understand their multiple functional roles, and thus structural studies of *L. acidophilus* GAPDH (LaGAPDH) was undertaken as a part of this study and the details are described in this chapter.

4.1.1. Macromolecular crystallography and its history

Discovery of X-rays by Rontgen Conrad and the subsequent developments by Max von Laue, who observed for the first time observed the wave nature of X-rays and the diffraction of X-rays are considered as a major breakthrough for protein crystallography. These were followed by the experiments by the Brags who showed that how X-ray diffraction could be used in the determination of the atomic structure of the molecule. However, it was another 45 years later, the determination of the first protein structure of myoglobin which gave the authors, Max Perutz and John Kendrew, the Chemistry Nobel Prize in 1962. Several other crystallographic structural studies on protein have been awarded the Nobel Prize and the list includes Dorothy Hodgkin for structures of vitamin B12 (Nobel Prize for chemistry, 1964), Johann Deisenhofer, Robert Huber and Hartmut Michel for the structure of the first membrane protein (Nobel Prize for chemistry, 1988), the structure determination of ATP synthase by John Walker (Nobel Prize for chemistry, 1997), Peter Agre & Roderick MacKinnon (Nobel Prize for chemistry, 2003), Kornberg (Nobel Prize for chemistry, 2006), Venki Ramakrishnan, Ada

Yonath, Thomas A. Steitz, for the ternary structure of the ribosome (Nobel Prize for chemistry, 2009), and Brian Kobilka and Robert Lefkowitz for structural studies of GPCR proteins (Nobel Prize for chemistry, 2012).

Macromolecular Crystallography is a technique used to study biological molecules such as proteins, viruses and nucleic acids to a high resolution and under correct combinations of different conditions protein molecules tend to crystallize in a periodic regular 3-D lattice arrangement. The high-resolution crystal structure information helps us to elucidate the detailed mechanism of which structure can be linked to its functions, for rational drug design, for protein folding investigation, to study interactions of proteins with other molecules, to find out how they undergo conformational changes. The X-rays diffracted by the atoms in equivalent positions in crystal lattice give intense spots which can be used to the determined macromolecular structure by analysis of the intensities and positions of diffracted spots. The sequence of events for a typical protein structure determination by X-ray crystallography consists of following steps:

- Protein purification and crystallization.
- X-ray diffraction of the crystals and converting diffraction intensities to structure-factor amplitudes.
- Phase determination.
- Interpretation of electron density from phases and building an atomic model into electron density.
- Refining the model against the structure-factor amplitudes, validation of the refined coordinates and finally the structure-function analysis.

4.1.2. Protein Crystallization

The crystallisation of purified proteins is the first step for determination of the 3D structure of protein through X-ray diffraction methods. Often the steps of crystallization are considered as one of the major bottlenecks in the structure determination process. The proteins in solution form crystals due to supersaturation and the molecules lose rotational and translational freedom by forming many new stable noncovalent interactions, thus minimising the free energy of the system. The critical requirements for protein crystallization are: the protein should be homogenous, it should be dissolved in a suitable solvent which helps the protein to be precipitated by precipitating agents such as

PEG (polyethylene glycol), high salt, organic solvents, etc., the solution is brought to supersaturation state and formation of nuclei for actual crystal growth to begin. Methods for setting up crystallization and to bring a protein solution into a supersaturation state are there, but they have to be optimised by a trial and error process.

Main methods for protein crystallization are Vapour-diffusion method, micro batch method and dialysis method (Ducruix & Giegé., 1992; McPherson, 1999). Vapour diffusion technique is used routinely and in this method, the drop consisting of protein and the precipitant is placed on a siliconized coverslip and sealed over a reservoir containing the precipitant alone. Thus, a concentration gradient is created by causing vapour evaporation and diffusion of water between solutions of different concentrations from the drop to the reservoir which leads to the formation of crystals at a particular stage of supersaturation. The vapour diffusion method can be done in two ways: the sitting drop method and the hanging drop method. In the micro-batch method, also known as crystallization under oil, the protein solution at a particular concentration is mixed with precipitant under a layer of oil to achieve the supersaturated solution, which may lead to crystallization due to slow diffusion of water across the oil layer and it requires less protein solution. The most commonly used oil for crystallization purposes is a mixture of silicone oil and paraffin oil in equal proportions which can be varied to control the rate of diffusion. The Dialysis techniques utilise diffusion and equilibration of small precipitant molecules through a semipermeable membrane where the macromolecules are slowly concentrated to form crystals.

4.1.3. X-ray diffraction and source of X-ray

X-ray diffraction occurs due to the interaction of X-rays with electrons, and it too reveals information about the structure of material. X-rays diffracted by a set of (hkl) planes in a crystal lattice and obeying the Bragg's Law given by the following equation are useful for single molecule structural analysis.

$$n\lambda = 2d \sin\theta$$

Where n is an integer expressing the order of reflection from a set of (hkl) planes and λ is the wavelength of incident X-rays, and θ is the angle of the incident ray on this set of planes.

The most common sources for X-ray generation used in protein crystallography are rotating anode generator and synchrotron radiation. The rotating anode generator is a conventional laboratory source of X-rays, where the X-rays are generated by electrons

from a heated filament (cathode) onto a target material. A high voltage is passed through the filament producing high energy electrons, and the emerging X-rays are focused using multilayered optic systems. In synchrotrons, whenever fast moving electrons, close to the speed of light changes its path under the influence of a magnetic field, they emit radiation tangentially from their path which is called "synchrotron radiation". This radiation is an extremely bright and enables data collection at multiple wavelengths.

4.1.4. Data collection strategies and data processing

For the collection of X-ray data at room temperature, the crystal is confined between thin-walled capillary of quartz or glass. Cryo-crystallography which collects the data at a low temperature where protein crystal is kept in a stream of cold nitrogen gas at 100k-200k temperature is used now routinely, and crystal is maintained in a loop. This protects the crystals from radiation damages caused by exposure to X-rays. The screening of the crystal is assessed by its diffraction spots and pattern. The primary parameters to be considered while collecting data are exposure per frame, the oscillation angle per frame (0.5° about an axis perpendicular to the direction of the X-ray beam), the crystal to detector distance (mm), the number of frames (to ensure completeness of data), etc. The time of exposure per frame is decided by looking at the diffraction spots which indeed is a measure of the quality of crystals. The crystal to detector distance (mm) is decided such that it prevents overlaps between the spots, but this is also at the cost of resolution of data. The number of frames and the data collection strategy often depends on the crystal symmetry.

The data processing steps in a single crystal data diffraction include: preliminary analysis of the original unprocessed detector data, indexing of the diffraction patterns, pre-refinement of crystal and detector parameters along with crystal orientation, intensity integration, post-refinement and finally merging and scaling of the measurements related by space-group symmetry (Rossmann & van Beek, 1999). During indexing the diffracted spots are used to determine the cell parameters and lattice system followed by integration of intensities (reflections) which is carried out by profile fitting of the spots. At this stage, the correction factors are applied to the intensities followed by averaging the symmetry-related reflections. Finally, the post-refinement is carried out which accounts for cell parameters, crystal to detector distance, mosaicity and other data collection parameter. These steps are carried out with programs like iMOSFLM, DENZO and SCALEPACK present in CCP4 suite. It is designed to simplify data processing by

dividing the process into a series of steps such as indexing, strategy calculation, cell refinement, integration and history, which are usually carried out sequentially. The final MTZ output file of the data processing programs consists of the list of the reduced reflections (with hkl indices referred to a unique region of reciprocal space) with their corresponding intensities and the associated deviations and errors (Otwinowski & Minor, 1997).

4.1.5. X-ray data quality indicators

The quality of data collected and high-resolution limit of the scaled data is assessed based on the mean ratio of the intensity to the error $\langle I/\sigma I \rangle$, $\chi 2$ and R_{merge} as the agreement between the symmetry-related reflections as given below (Otwinowski & Minor, 1997):

$$R_{merge} = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

Where

$I_{hkl,j}$ = intensity of jth observation of (h k l) reflection

$\langle I_{hkl} \rangle$ = the mean intensity of all measured reflections.

Also, $CC_{1/2}$ must also be used as a data quality indicator which estimates intensity from half data sets and is a primary indicator for use for selecting high-resolution cutoff for data processing (Karplus & Diederichs, 2015).

4.1.6. Structure solution

4.1.6.1. Phase determination: The reflections (hkl) of X-ray diffraction data include only intensities of the diffracted waves (or equivalently structure factor amplitudes) after passing through crystals. Each reflection on the diffraction pattern or structure factor corresponds to an electromagnetic wave consisting of an amplitude and a phase where the amplitude is easily calculated by taking the square root of the intensity while the phase is lost which is vital for calculation of electron density distribution in the crystals. Four principle techniques used to obtain phases are

- Single/Multiple isomorphous replacements (SIR/MIR) method
- Multi-wavelength anomalous dispersion (MAD) method
- Molecular replacement (MR) method and
- Direct and Patterson method

Further various experimental and computational methods are used to estimate the

phases so that perfect distribution of electron density in the asymmetric unit using the Fourier transform is obtained and eventually a reasonable protein structural model is obtained. The single/multiple isomorphous replacement (MIR) methods mainly require the incorporation of heavy atoms in the native protein crystal and determination of the coordinates of these heavy atoms in the unit cell helps solve the phase problem. The heavy atom addition can be accomplished by soaking the native crystal in solutions of a heavy atom salt, co-crystallization, substituting the protein methionines by selenomethionines or by exposing the crystal to a noble gas atmosphere (Schiltz et al., 1994; Soltis et al., 1997). The isomorphous replacement method requires the following steps: preparation of a heavy-atom-containing derivatives of the protein in the crystalline state, X-ray diffraction data of the native protein crystal as well as for crystals of the derivatives is collected, application of the Patterson function for the determination of the heavy atom coordinates, refinement of the heavy atom parameters and calculation of the phases and finally calculation of the electron density map of the protein.

The most common method used for protein structure solution is the molecular replacement (MR) method in which the phases from structure factors of a known protein (about 40% identical to search protein) are used as initial estimates of phases for a new protein. Here on a map is calculated from the experimental intensities and compared to theoretical maps of the homologous structure coordinates. By rotating and then translating the maps about each other the position of the molecules can be found and subsequently the initial phases are determined. MR is computer intensive as it determines three rotational and three translational, which would describe how the search molecule is placed in the unit cell. The complexity and computational requirement were drastically reduced when Rossmann and Blow (1962) reduces this six-dimensional search to a sequence of two three-dimensional searches in which first the orientation and then the position of the search molecule are determined (Rossmann & Blow, 1962). Patterson methods are used to calculate the rotation function (Rossmann & Blow, 1962), which is used, to obtain the orientation of the model in the new unit cell, and then the translation function (Crowther & Blow, 1967), which helps in placing the correctly oriented model in the new unit cell.

Multiple-wavelength anomalous diffraction (MAD) is used to solve phase problems when no homologues are available. The method is employed in cases where the heavy metal atom is in the protein like selenomethionine or sometimes the protein is

soaked with a heavy atom. Single-wavelength anomalous diffraction (SAD) uses a single set of data from a single wavelength while in MAD (multiple anomalous dispersion) experiment data is collected at different wavelengths and phase collected from Harker diagrams (Kartha *et al.*, 1967).

4.1.6.2. Improvement and modification of electron density: The prime goal here is to minimise the noise signals in the electron density maps. If the protein is in the oligomeric state in the asymmetric unit, one can use density average method. The solvent flattening method is used to identify the solvent region in the asymmetric unit of the crystal. The major region of protein crystals is occupied by solvent, and this can be analysed by Matthews coefficient (V_M) (Matthews, 1968).

V_M can, and solvent content (V_{solv}) can be calculated from the unit cell and the molecular weight (Da) of the molecules in the unit cell as given below:

$$Vm = \frac{V}{n X M}$$

$$Vsolv = 1 - \frac{1.23}{Vm}$$

Where V_m = Matthews number; V = Volume of unit cell volume; M = molecular weight of the protein in Daltons; n = the number of asymmetric units; V_{solv} = solvent content of the protein crystals.

In the asymmetric unit, if the number of identical subunits is more than one, then non-crystallographic symmetry method is used to improve the phases. Here, the electron density for all identical molecules is forced to be equal to an average value calculated from the electron density of all the related molecules. Overall, improved phases from these modified maps are used to calculate electron density by combining with experimental amplitudes.

4.1.7. Structure refinement

The structure refinement is the crucial step in protein crystallography which aims at optimizing the agreement of an atomic model with observed diffraction data along with chemical restraints. The initial models from a structure solution often give us an

only partial information of atoms in the unit cell which contains sufficient phase information to allow the user to locate the remaining atoms. It is an iterative process of improvement of the quality of the structure model. The atomic parameters from each atom and positional coordinates x, y and z, are adjusted so as to improve the agreement between the observed structure factor amplitude |Fo| and those calculated from the structural model |Fc|. Also, further restraints are included during the process of refinement. The geometric restraints are bond angles, bond lengths, planarity, chirality, etc. (Engh & Huber, 1991). From the atom types and relative positions in the initial model, a set of structure factors can be calculated. An electron density map can then be prepared using the calculated phase angles and the observed structure factors.

The initial structural model does not represent the complete picture of the protein, and in general there is a discrepancy between the observed structure factor amplitude |Fo(hkl)| and calculated structure factor amplitude |Fc(hkl)| which is measured by a residual or R –factor, a global indicator of quality of the fit of a model. The process of making atomic parameters in more agreement with the calculated structure factor amplitude obtained from the initial model is known as refinement. Refinement is a process where the initial model parameters, positional (x, y, z) and the thermal (B) from each atom are calculated to make a good agreement between observed structure factor and calculated model. The refinement progress is monitored using the R-factor and defined as,

$$R_{factor} = \frac{\sum_{hkl} |F_o(hkl) - F_c(hkl)|}{\sum_{hkl} F_o(hkl)}$$

In macromolecular crystallography, a statistical procedure called cross-validation (R and R_{free}) is used an indicator of model quality (Brunger, 1992; Kleywegt & Brunger, 1996). This value can be gradually decreased by increasing number of adjustable parameters used in describing models and thus helps in avoiding over-interpretation of the data. The diffraction data set is randomly divided into two sets: a large working set (90-95% of the data) and s test set (5-10% of the data). The diffraction data in the working set are used in refinement process barring the use of the test set for refinement. The value of R_{free} approaches close to R-factor when the model is correct and errors are statistical. The R_{free} value should be in between the range of 3-5% over R_{work}, which is calculated without the reflections from the test set. If the value of R_{free} differs more than

5% from R_{work}, the model is probably over fitted. The refinement changes which are iteratively incorporated in the model which do not improve the model's ability to describe the test set eventually doesn't improve the fit of the model to the test set, and thus finally the R_{free} remains constant or increases.

During the refinement of crystal diffraction data, it is necessary to integrate the chemical information in to the data in the form of constraints and restraints. To improve observation to parameter ratio, some atoms can be restrained by giving limited freedom to a parameter or constrained by keeping parameter at exact values during refinement. Typical restraints include bond lengths, bond angles, chiral volumes, non-crystallographic symmetry, van der Waals' contact distances, isotropic and anisotropic thermal motions.

4.1.7.1. Maximum likelihood and simulated annealing method: The widely used methods in the refinement of a protein crystal data are maximum likelihood and simulated annealing. Both the methods uses restraints with respect to bond distances, angles and torsions and temperature factors (B-factors). In the maximum likelihood method, the phases are adjusted to minimise the R-factor while in the simulated annealing method the structure is heated to include randomness and slowly cooled before refining. The randomness reduces the chances of refinement moving to a wrong local minimum. Widely used the program in refinement are REFMAC5 (REFinement of MACromolecular structures) which uses maximum likelihood approach (Murshudov *et al.*, 1997) and the PHENIX (Python-based Hierarchical ENvironment for Integrated Xtallography) (Adams *et al.*, 2010) which includes simulated annealing along with other approaches. REFMAC tries to obtain the model data which is in agreement with the experimental data (i.e. maximises the likelihood) by refining an atomic model by adjusting the model parameters (coordinates, B-factors, TLS, etc.). The refinement progress is measured by R, R_{free}, and likelihood scores.

4.1.8. Electron density maps and model building

After every refinement cycle, the model was inspected along with manual rebuilding by visualizing $2F_O$-F_C and F_O-F_C electron density maps. The $2F_O$-F_C and F_O-F_C maps give reduced model bias by giving more weight to the observed structure factors. The $2F_O$-F_C and F_O-F_C maps were usually contoured at 1σ and 3σ respectively. The F_O-F_C map is examined for both positive and negative contours, the positive peaks

indicate the presence of unaccounted density, while a negative peak shows that the model contains features not present in the actual structure. The electron density map obtained from the phase information gives a clear picture to build the model. It is normally viewed at a contour level of around 1.0σ where σ refers to the R.M.S. deviation in the density.

4.2. Materials

For crystallization setup, crystallization buffers screen were purchased from Hampton Research, (USA) having Crystal screen 1 (Crystal Screen HR2-110) and Crystal screen 2 (Crystal Screen HR2-112). Also, an Index screen (HR144) was also used for screening crystal growth. Most of the trials were set up using microbatch under oil method using a mixture of paraffin oil (HR3-411) and silicon oil (HR3-415) also purchased from Hampton Research, (USA). A 24-well plate obtained from Laxbro (India), and 60-well Terasaki plate purchased from Grenier Bio-one (Germany) were used for hanging drop and microbatch under oil method of crystallization respectively. A 96-well plate purchased from SPL Biosciences (South Korea) was used for hanging drop setup along with microplate sealing films (Grenier Bio-one, Germany). For hanging drop, the coverslips were siliconized prior to use with Carbon disulphide sulphide (SDFine, India) and Trimethylchlorosilane (Fluka, India). The buffers for hanging drop setup were manually prepared using chemicals like TRIS (Sigma), PEGs (Sigma & HiMedia), salts (Sigma & HiMedia), etc. Double distilled autoclaved water was used for all buffers preparation.

The dataset was collected at national x-ray diffraction facility at National Institute of Immunology (NII), New Delhi, India. The Cu-Kα radiation was generated using a rotating anode X-ray generator, Rigaku FR-E+ from Rigaku (USA), at wavelength 1.5418Å. The system was equipped with a confocal mirror focusing system, and the diffraction data was collected on an R-AXIS IV++ detector. All the X-ray datasets were collected at a cryo-cooled temperature (100K) by flash cooling the mounted crystals in a nitrogen flush stream.

The X-ray intensity diffraction data collected were integrated using DENZO of HKL2000 (Otwinowski & Minor, 1997) and iMosflm (Powell *et al.*, 2013) program in CCP4 (Collaborative Computational Project No. 4). SCALEPACK of HKL2000 and SCALA program was used to scale the data during data processing. Most of the crystallization data was analysed and processed in CCP4 v 7.X software suite and Phenix suite if otherwise mentioned. The complete analysis was performed in a system installed with Linux operating system.

4.3. Methods

4.3.1. Siliconization

For hanging drop method, the glass wares and the coverslips were siliconized prior to crystallization. Siliconization makes a thin layer on the glass surface and thus makes it more hydrophobic. Briefly, the required glassware and coverslips were first washed with detergents and dried at 80ºC in a dry oven and then trimethylchlorosilane in carbon disulphide (10:1) was added. This was followed by washing with methanol and further with distilled water, and finally, the glassware and coverslips were thoroughly washed with purified water and were dried at in 80ºC dry oven prior to use.

4.3.2. Crystallization setup

The initial crystallisation trials were carried out using Hampton crystallization screens in 60-well microtiter terasaki plates using microbatch under oil method. r-LaGAPDH (2-10mg/ml) protein was dialysed against 20mM Tris, pH 7.5 buffer overnight prior to crystallization setup. For microbatch under oil method, the mixture of paraffin and silicone oil was used in (1:1) ratio. For crystallization setup the following protein to precipitant ratio were tried: (1:1, 1:2, 1:3, 2:1, 2:2, 2:3, 3:1, 3:2 and 3:3). The purified r-LaGAPDH in tris buffer was screened with different crystallization conditions and screens. The plates were set up in vibration free environment at room temperature for crystal growth. The plates were regularly monitored under a polarizing microscope to check for possible protein crystals and the promising conditions were optimised further for getting diffraction quality crystals. After initial diffraction data collection, these crystals were mounted on cryo loop for data diffraction and to improve the quality of crystals the initial hits were optimized with different combinations of protein: precipitant ratio, protein conditions and variation of PEGs. No cryoprotectant was used or optimized as the crystal condition already contained PEG and oil which would itself acts as a cryoprotectant.

4.3.3. Data collection

The crystals obtained from crystallization trials were picked with cryo-loop (Hampton Research, USA) and were flash freezed in liquid nitrogen at low temperature (100K). The initial X-ray diffraction data were collected using Cu-Kα radiations from Rigaku rotating anode X-ray generator, and the images were collected at an R-Axis IV++ image plate detector. The optics used in the system was of VARIMAX HF. The strategies

for orientation of the crystal on the X-ray beam, crystal to detector distance, exposure time and oscillation angle were considered before starting data collection. The crystal to detector distance (150-210) was adjusted to a suitable distance in mm slightly more than the longest unit cell dimension in Å. The oscillation angle per frame (0.5º) was adjusted and the choice of the exposure time(s) was selected based on the crystal quality and the oscillation range.

4.3.4. Data processing

The data processing steps of single crystal diffraction data collected on an imaging plate detector system proceeds through a series of steps involving selection of strong spots, indexing the spots, refinement of detector parameters and crystal orientation, intensity integration, post-refinement and scaling (Rossmann & van Beek, 1999). During auto indexing, after selecting some strong spots, an analysis of the position of the peaks was carried out to determine unit cell dimensions, Bravais lattice and crystal orientation using DENZO from HKL2000 package and iMosflm from CCP4 6.5.0 package followed by indexing, cell refinement and integration. The unmerged mtz file was scaled and merged further using SCALA from CCP4 package. The intensity data obtained from DENZO were scaled using SCALEPACK2MTZ in CCP4 and the multiple observations of the reflections from iMosflm were scaled using program SCALA of CCP4 suite. Before running SCALA, the unmerged mtz file from iMosflm and SCALEPACK2MTZ was run in AIMLESS to check for Laue group and preparing input for scaling and merging. During scaling in SCALA, 5% unique data was kept for R_{free} calculations. The statistics of the data quality determines the quality of the data collected and the high-resolution limit. The parameters such as $\chi 2$, R_{merge} and $<I/\sigma(I)>$ were observed for the indication of the high-quality data.

4.3.5. Structure determination

For structure determination of r-LaGAPDH diffraction data, the Molecular Replacement (MR) approach was used as a homolog structure was available for phasing - PDB: 4QX6. Before proceeding with the MR, the quality of data was checked for any possible twinning. The number of molecules in asymmetric unit was calculated using Matthew's coefficient under Cell Content Analysis program. The search model was modified to polyala model to make it more suitable for molecular replacement using CHAINSAW. Phaser-MR (McCoy *et al.*, 2007) from CCP4 was used for molecular replacement using the scaled mtz output file from SCALA. Single chain A of PDB:

4QX6 polyala model was used as an Ensemble and two molecules in asymmetric unit were assigned based on Matthew's coefficient along with setting the search parameter using two copies in Phaser-MR.

4.3.6. Structure refinement

The most common method of refinement used in protein crystallography is the 'maximum likelihood' refinement. Protein crystal structures have to be refined to improve the agreement of structure factors calculated from the model with observed structure factors and chemical nature of the model. Several rounds of alternating model building and refinement cycles were carried out till the convergence of R_{factor} and R_{free} was obtained. To have a better estimate of the quality of the fit, the data is further divided into a test set comprising 5% of randomly chosen reflections and a working set. Refinement of structures was carried out using 20 cycles of maximum likelihood restrained refinement implemented in the program REFMAC5 in CCP4 (Murshudov *et al.*, 1997). The output of initial model from Phaser was subjected to rigid body refinement and further iterative rounds of restrained body refinement using REFMAC5.

The Crystallographic Object-Oriented Toolkit (COOT) program was used to examine and interpret the electron density maps (Emsley & Cowtan, 2004). The models were fitted into the electron density, and alternating cycles of model building and refinement were carried out. The COOT program was used for the model building of the protein structure reported in this thesis.

4.3.7. Co-crystallization studies

Crystal soaking and co-crystallization studies of r-LaGAPDH were performed with mannose, galactose, N-acetyl-D-galactosamine (GalNAc) and N-acetyl-D-glucosamine (GlcNAc) at various concentrations. The purified r-LaGAPDH was dialyzed against 20mM Tris buffer pH 7.5 and as well as in sodium acetate buffer at pH 5.0 to maximize the chances to get complex crystal structure with carbohydrates. The protein was also dialyzed against with the above buffer along with 50mM sugars before crystallization setup. For crystal soaking studies, single protein crystals were grown and incubated in a solution of sugar dissolved in precipitant buffer, 20% glycerol and PEG at a range of different concentration (5-100mM). The soaking time was kept from 5 minutes to 24 hours.

4.3.8. Bioinformatics analysis

Some programs were used for sequence and structure analyses. BLAST was used for homology search from NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to find the local and global region of similarities and to identify homologues (Altschul *et al.*, 1990). CLUSTALW was used for Multiple Sequence Alignment (MSA) using the sequences downloaded from the UniProt database (http://uniprot.org) (Larkin *et al.*, 2007). Pfam was used for domain identification (Finn *et al.*, 2014). 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. The phylogenetic analysis was assessed using Muscle and Maximum-Likelihood method in Molecular Evolutionary Genetics Analysis (MEGA5) suite (Tamura *et al.*, 2011).

The homology model was built using I-tasser (Yang & Zhang, 2015) server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) and further refined using ModRefiner (http://zhanglab.ccmb.med.umich.edu/ModRefiner/) and energy minimization was done using GROMACS (Van Der Spoel *et al.*, 2005). Chimera and PyMol were used for superposing structures and molecular visualization (Pettersen *et al.*, 2004). Superpose was used to generates structure alignments of PDB coordinates and interactive images of the superimposed structures (Maiti *et al.*, 2004).

ProtParam was utilized 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). Easy Sequencing in PostScript (ESPRIPT) was used for easy visualisation of aligned sequences via a PostScript output (Gouet *et al.*, 1999). Stereo-chemical properties such as bond lengths, bond angles, planarity, main chain and side chain torsion angles for all the residues, which calculates the Ramachandran plot of the model structure were evaluated with the PROCHECK of Structural Analysis (Laskowski *et al.*, 1993) and Verification Server (http://nihserver.mbi.ucla.edu/SAVES/). CONTACT program was used for computing various types of contacts in protein structures. DSSP program was used to define the secondary structure, geometrical features and solvent exposure of proteins, given the atomic coordinates in PDB format (Kabsch & Sander, 1983). DALI server was used for comparing three-dimensional protein structures (Holm

& Rosenstrom, 2010). The database consists of all representative structures from the PDB with less than 90% sequence identity to each other. The classification and alignments are automatically maintained and continuously updated.

PDB_REDO program was used to edit and refine the model structure. PDB_REDO is a pipeline of various CCP4 programs and other programs such as WHAT_CHECK which uses its tools for the optimization of the structure model through refinement and rebuilding (Joosten *et al.*, 2014). MOLPROBITY, a web-based program was utilized for the validation of structures of proteins which adds and optimizes polar as well as non-polar hydrogen atoms. It provides detailed all-atom contact analysis of any steric problems within the molecules and can calculate the hydrogen bond and van der Waals contacts in the interfaces between rotamers (Hintze *et al.*, 2016).

4.4. Results and discussion

4.4.1. Crystallization and structure solution

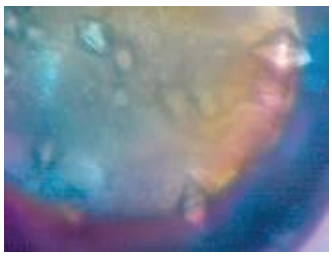
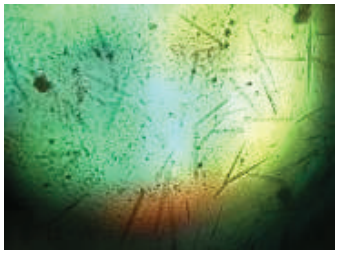
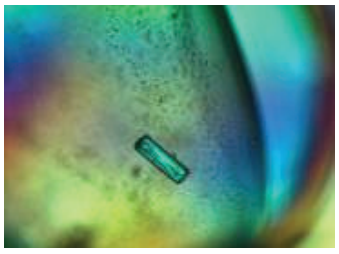
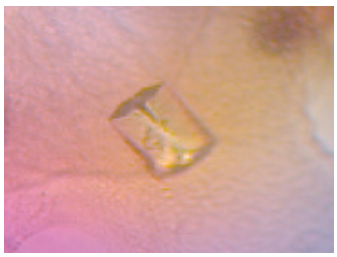
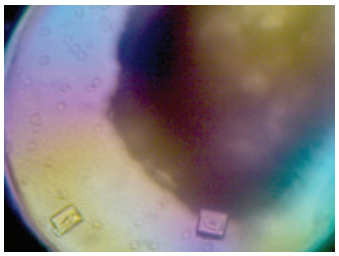
4.4.1.1. Crystallization

The putative crystals of r-LaGAPDH are shown in Table 4.1. Two particularly promising conditions (LaGAPDH_C1 and LaGAPDH_C7) gave diamond-shaped crystals and crystals obtained after two weeks of setup. The results of optimization of initial data diffraction conditions with different combinations of protein: precipitant ratio, protein conditions and variation of PEGs are mentioned in Table 4.2.

4.4.1.2. Data Collection

The data was collected for the crystals (LaGAPDH_C1, LaGAPDH_C7 and LaGAPDH_C13) under cryo conditions. The crystal (LaGAPDH_C1) in 25% w/v Polyethylene glycol 1,500 showed a diffraction with overlapping spots and twining (Figure 4.1a, b). The crystal (LaGAPDH_C7) in 10% w/v Polyethylene glycol 1,000, 10% w/v Polyethylene glycol 8,000 diffracted up to 2.52Å (Figure 4.2a, b). Data for the crystal LaGAPDH_C13 diffracted well and was further collected for processing at 2.21Å (Figure 4.3a, b).

Table 4.1: Preliminary hits obtained with different crystallization screens.

Name	Crystallization condition	Protein (protein: precipitation ratio)	Image of crystal
LaGAPDH_C1	25% PEG 1500	7.5 mg/ml 50mM TRIS, pH 7.2, (1:1)	
LaGAPDH_C2	0.1 M BIS-TRIS pH 6.5 25% PEG 3350	7.5 mg/ml 50mM TRIS, pH 7.0, (1:1)	
LaGAPDH_C3	0.1 M BIS-TRIS pH 6.5 28% PEG 2000	7.5 mg/ml 50mM TRIS, pH 7.0, (1:1)	
LaGAPDH_C4	0.1 M BIS-TRIS pH 6.5 28% PEG 2000	10 mg/ml 50mM TRIS, pH 7.8, (1:1)	
LaGAPDH_C5	0.2 M NaCl, 0.1M BIS-TRIS pH 5.5 ,25%w/v PEG 3,350	10 mg/ml 50mM TRIS, pH 7.8, (1:1)	

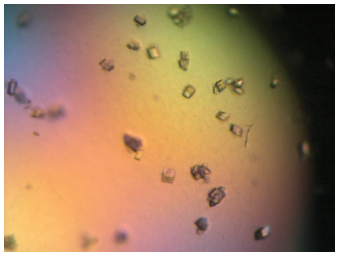
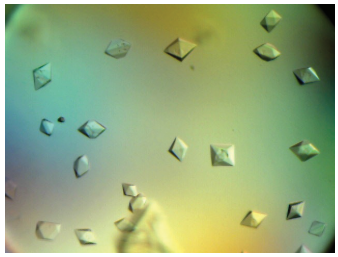
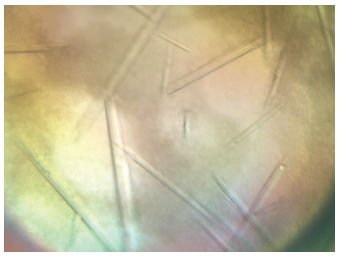
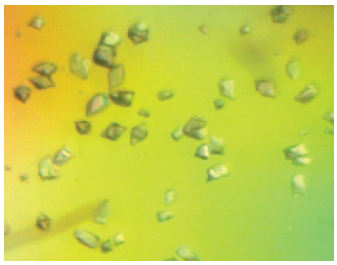
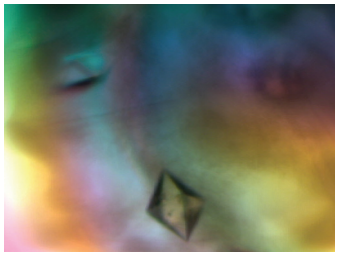
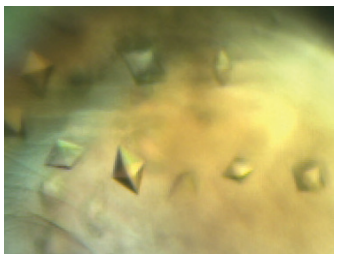
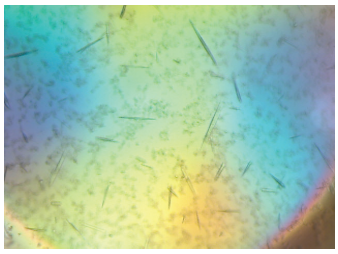
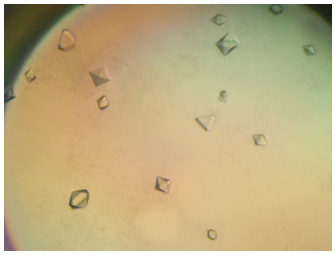
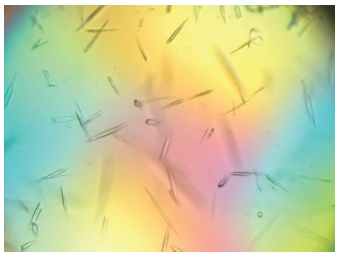
LaGAPDH_C6	0.1 M BIS-TRIS pH 6.5, 28% w/v PEG monomethyl ether 2,000	4.5 mg/ml 50mM TRIS, pH 7.5, (1:2)	
LaGAPDH_C7	10% w/v Polyethylene glycol 1,000, 10% w/v Polyethylene glycol 8,000	4.3 mg/ml 50mM TRIS, pH 8, (1:2)	
LaGAPDH_C8	0.8M Potassium sodium tartrate tetrahydrate, 0.1 M Tris pH 8.5, 0.5% w/v PEG monomethyl ether 5,000	10 mg/ml 20mM TRIS, pH 7.8, (1:1)	

Table 4.2: Optimization of initial hits from crystallization trials.

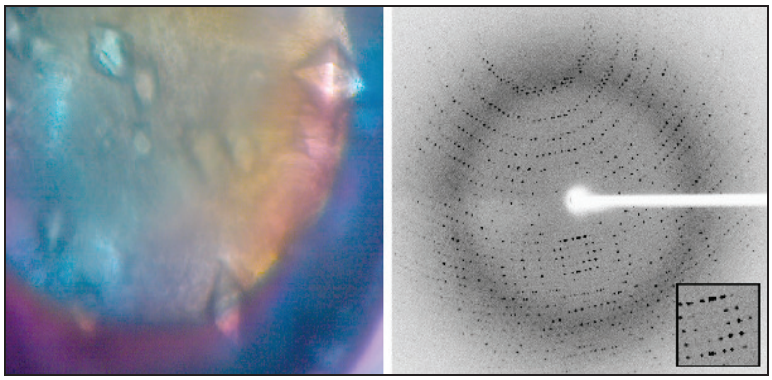
Name	Crystallization condition	Protein (protein: precipitation ratio)	Image of crystal
LaGAPDH_C9	25% PEG 1500	7.5 mg/ml LaGAPDH 50mM TRIS, pH 7.2, (1:1)	
LaGAPDH_C10	25% PEG 1500	13 mg/ml 20mM TRIS, pH 7.4, (1:1)	

LaGAPDH_C11	25% PEG 1500	13 mg/ml 20mM TRIS, pH 7.4, (3:1)	
LaGAPDH_C12	16% PEG 4000	5 mg/ml 50mM TRIS, pH 7.6, (2:2)	
LaGAPDH_C13	10% w/v PEG 1,000, 10% w/v PEG 8,000	4.5 mg/ml 50mM TRIS, pH 7.4, (1:1)	
LaGAPDH_C14	10% w/v PEG 1,000, 10% w/v PEG 8,000	5 mg/ml 50mM TRIS, pH 7.6, (2:1)	

4.4.1.3. Data processing

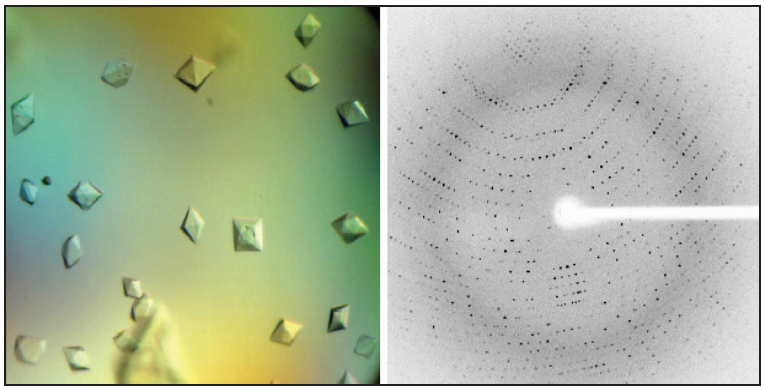
The X-ray diffraction data for LaGAPDH_C7 showed orthorhombic space group $P2_12_12_1$ and with unit cell parameters a = 110.285, b = 112.514, c = 118.361 and α = β = γ = 90° and Mosaicity of 0.45. The R_{merge} was high around 16%. The number of molecules in asymmetric unit was 4 and diffraction parameters, and statistics are shown in Table 4.3.

The diffraction data of LaGAPDH_C13 crystal was processed with iMosflm in space group $P4_12_12$ and the unit cell parameters were a = 114.98, b = 114.98, c = 113.29 Å, α = β = γ = 90. However, when it was processed with the HKL package it processed in space group $P2_12_12_1$ with unit cell dimensions a = 113.98, b = 115.09, c = 114.99 Å, α = β = γ = 90º. In tetragonal space group $P4_12_12$, a total of 661620 reflections was collected



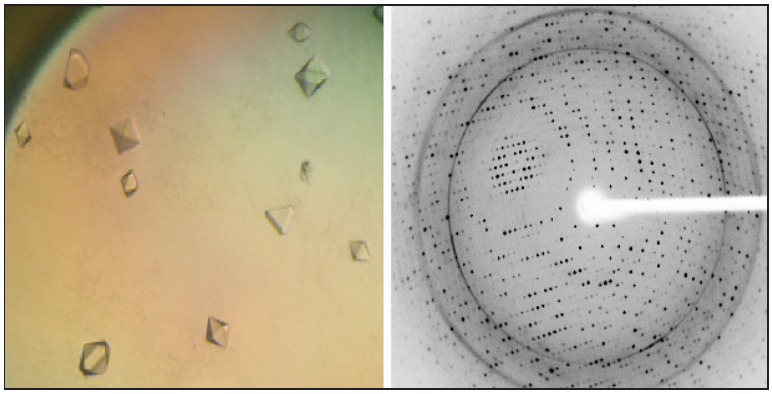
(a) **(b)**

Figure 4.1: LaGAPDH_C1 crystal and diffraction image: (a) Microscopic image of crystal; (b) Reference image file from diffraction. Inset image is magnified image of diffraction spots.



(a) **(b)**

Figure 4.2: LaGAPDH_C7 crystal and diffraction image: (a) Microscopic image of crystal; (b) Reference image file from diffraction.



(a) **(b)**

Figure 4.3: LaGAPDH_C13 crystal and diffraction image: (a) Microscopic image of crystal; (b) Reference image file from diffraction**.**

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 38,933Da each) per crystal asymmetric unit, the calculated Matthews coefficient and solvent content were 2.41Å^3 Da^{-1} and 48.96% respectively. More data diffraction parameters and statistics are shown below, and the key differences are marked in bold and underline (Table 4.4).

Table 4.3: X-ray diffraction statistics of LaGAPDH_C7 crystals

Crystal Name	LaGAPDH_C7
Diffraction source	Cu Kα
Detector	R-AXIS IV++
Wavelength (Å)	1.54178
Temperature (K)	100
Crystal-detector distance (mm)	210
Rotation range per image (°)	0.5
Total rotation range (°)	180
Space group	$P2_12_12_1$
Unit-cell parameters	a=110.285, b=112.514, c=118.361
α, β, γ (°)	α = β = γ = 90
No. of molecules in asymmetric unit	4
Mosaicity	0.45
Resolution range (Å)	50-2.52 (2.61-2.52)
Total number of reflections	1063508
Number of unique reflections	50257
Completeness (%)	99.9 (99.8)
R_{merge}	0.16 (0.45)
I/σ(I)	10.296 (4.36)
CC (1/2)	0.978 (0.892)
Multiplicity	6.3
Matthews coefficient (Å^3 Da^{-1})	2.51

Values of the outer shell are given in ().

Table 4.4: X-ray diffraction statistics of LaGAPDH_C13 crystals

Crystal Name	LaGAPDH_C13	LaGAPDH_C13
Diffraction source/	Cu Kα	Cu Kα
Detector	R-AXIS IV++	R-AXIS IV++
Wavelength (Å)	1.54178	1.54178
Temperature (K)	100	100
Crystal-detector distance (mm)	175	175
Rotation range per image (°)	0.5	0.5
Total rotation range (°)	240	240
Exposure time per image (s)	180	180
Space group	$P2_12_12_1$	**$P4_12_12$**
Unit-cell parameters	a=113.98, b = 115.09, c = 114.99	**a=114.98, b = 114.98, c = 113.29**
α, β, γ (°)	α = β = γ = 90°	α = β = γ = 90°
No. of molecules in asymmetric unit	4	**2**
Mosaicity	0.88	**0.86**
Resolution range (Å)	50 – 2.21 (2.3-2.21)	50 – 2.21 (2.3-2.21)
Total number of reflections	658295	**661620**
Number of unique reflections	74900	**38437**
Completeness (%)	99.4 (98.3)	99.9 **(99.9)**
R_{merge}	0.080 (0.542)	**0.093 (0.589)**
(I/σ(I))	13.4 (3.1)	**16.5 (4.1)**
CC (1/2)	0.998 (0.913)	**0.999 (0.945)**
Multiplicity	8.8 (7.7)	**17.2 (15.0)**
Matthews coefficient ($Å^3$ Da^{-1})	2.41	2.41

Values of the outer shell are given in (). Differences in both the data processing are mentioned in bold and underline.

4.4.1.4. Structure solution, model building and refinement

In the initial MR model using Phaser with a single chain of 4QX6: *Streptococcus agalactiae* GAPDH, a score of RFZ=10.3, TFZ=29.1 and LLG=532 was obtained indicating a good MR solution. The MR solution model was further subjected to refinement using REFMAC5 in CCP4 package. A single round of 20 cycles of rigid body refinement was performed on the initial MR model which resulted in initial R_{factor} and R_{free} of 48% and 50% percentage. The alternative model building and refinement were carried out till the convergence of R_{factor} and $R_{free,}$ and the iterative model building was performed along with density fitting in COOT. Final refinement statistics are shown in Table 4.5. The electron density map in most of the region is fitted well with the model (Figure 4.4a). Electron density could be seen for most of the part of structures except for regions between 26-30 residues and 104-110 residues in both the chains which will now be referred as chain A and chain B for further sections (Figure 4.4b, c).

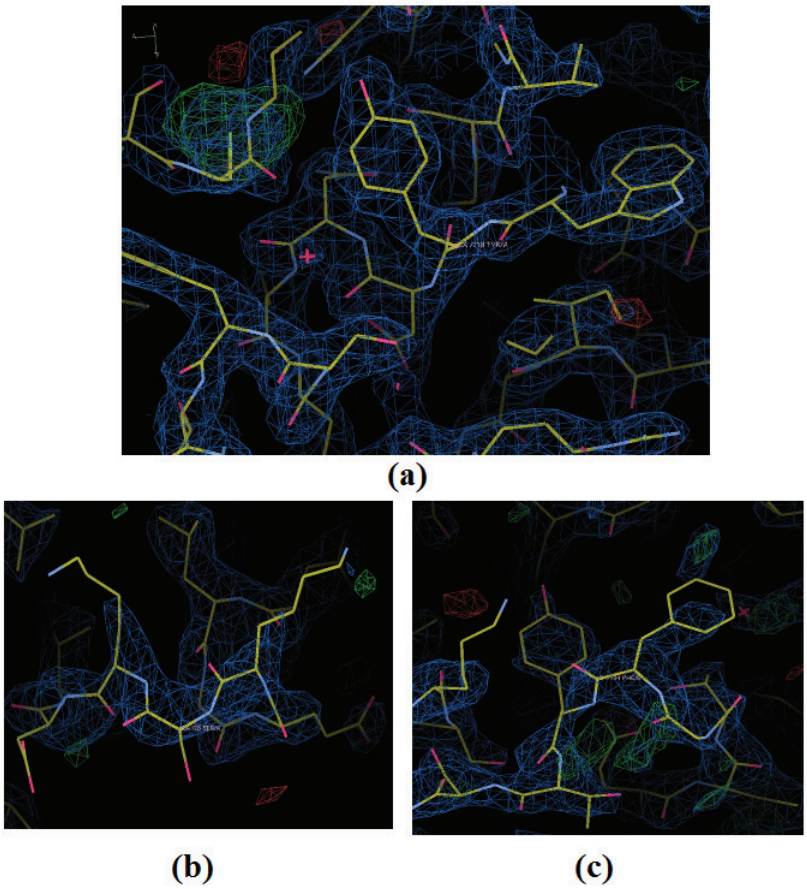


Figure 4.4: Electron density map: (a) COOT visualization of electron density map and model fitting; (b) Electron density map of 24-30 residues; (c) Electron density map of 104-110 residues.

Table 4.5: Refinement statistics of LaGAPDH_C13

Crystal Name	LaGAPDH_C13
PDB	**5J9G**
Space group	$P4_12_12$
Unit-cell parameters	a=114.98, b = 114.98, c = 113.29
No. of molecules in asymmetric unit	2
Resolution range (Å)	50 – 2.21 (2.3-2.21)
Total number of reflections	661620
Number of unique reflections	38437
Completeness (%)	99.9 (99.9)
(I/σ(I))	16.5 (4.1)
Mosaicity	0.86
R_{merge}	0.93 (0.589)
CC (1/2)	0.999 (0.945)
R_{work}/R_{free}	0.21/0.23
R_{free} **test set reflections**	1708 (4.79%)
Wilson B-factor (Å2)	40.2
No. of atoms	5306
Protein	5126
Water	179
Average B, all atoms (Å2)	44.0
Ramachandran plot	
Favoured (%)	96.7
Allowed (%)	3.3
Outliers (%)	0.0
$R_{p.i.m.}$	0.023 (0.154)
R_{meas}	0.96 (0.610)

Values of the outer shell are given in ().

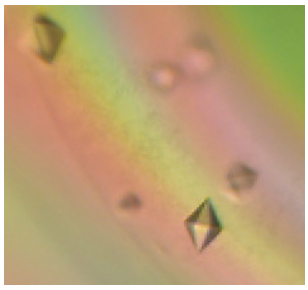
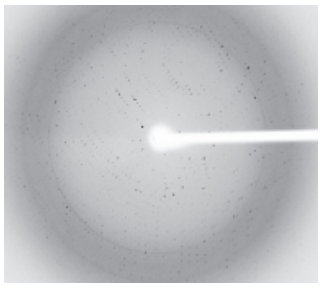
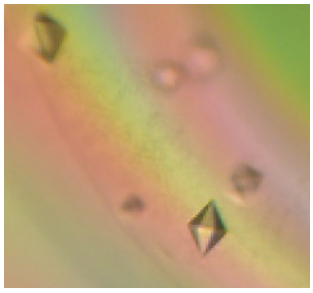
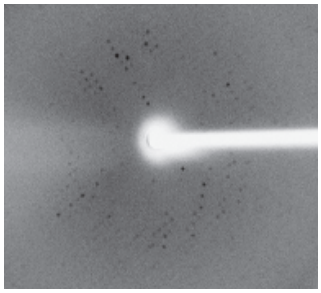
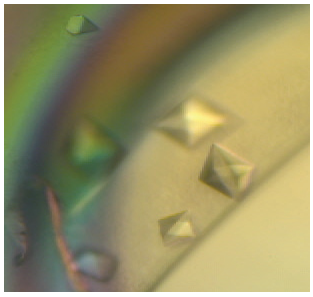
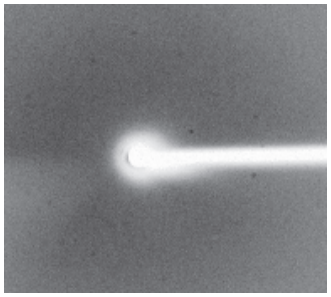
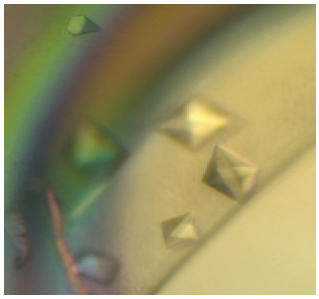
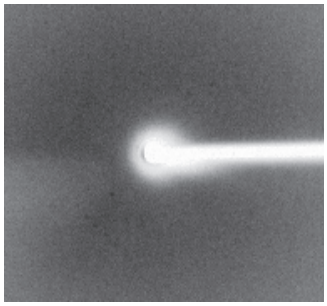
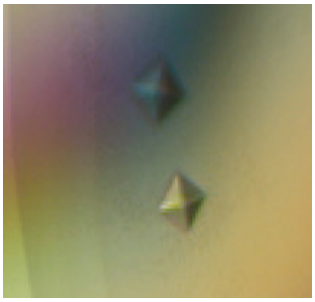
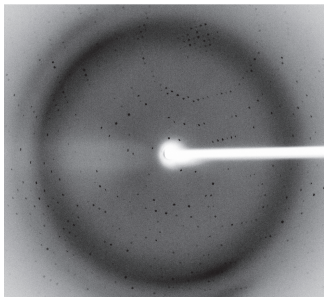
4.4.2. Crystallization, data processing and refinement of r-LaGAPDH complex

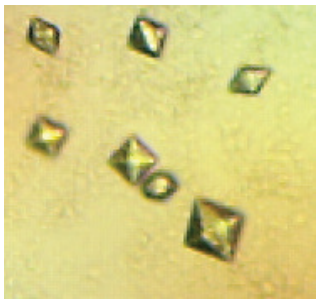
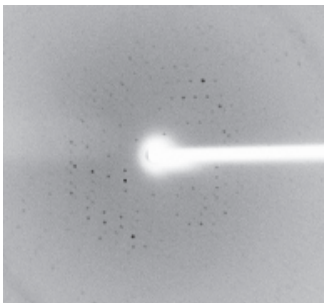
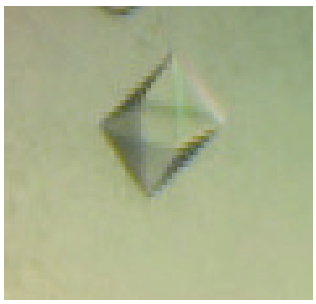
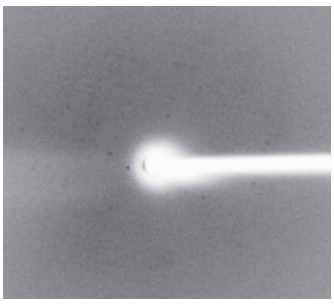
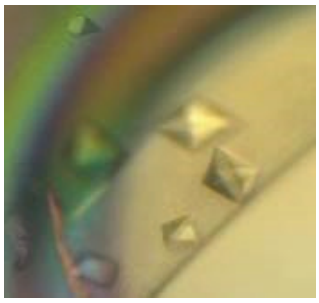
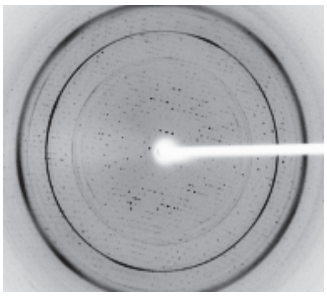
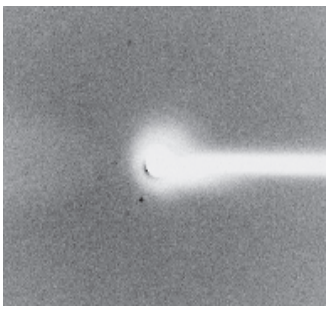
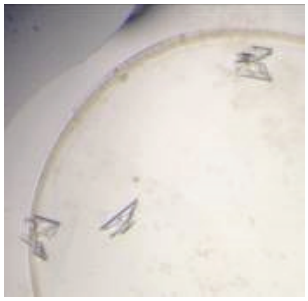
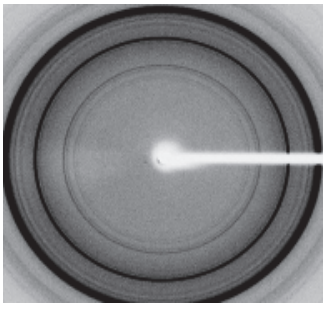
4.4.2.1. Crystallization of r-LaGAPDH complex

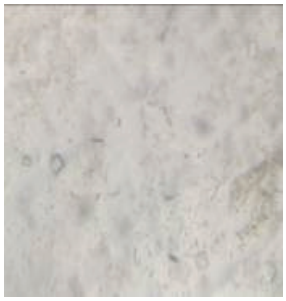
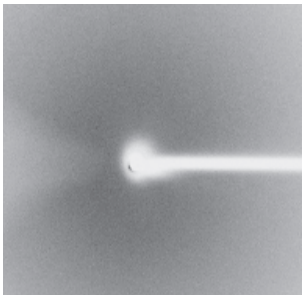
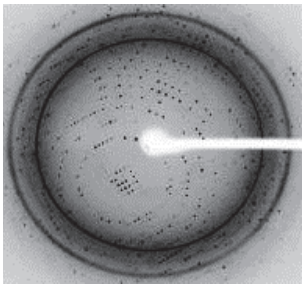
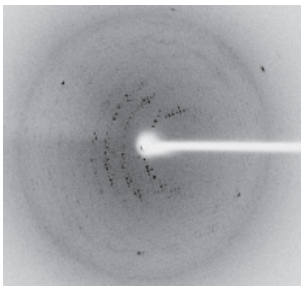
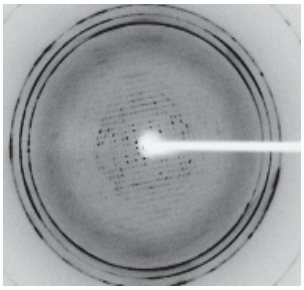
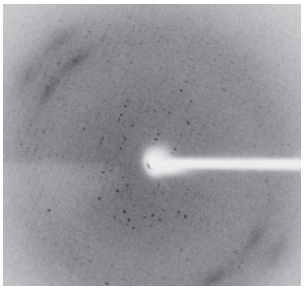
The purified r-LaGAPDH was screened for complex crystals with mannose (Man), galactose (Gal), N-acetyl-D-galactosamine (GalNAc) and N-acetyl-D-glucosamine (GlcNAc) using the above-mentioned crystallisation conditions as well as the conditions in which the native r-LaGAPDH crystals diffracted were further optimized for complex crystal generations. Numerous trials were carried out for complex crystals data including co-crystallization and soaking prior to diffraction. The one which was

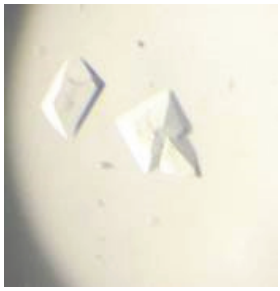
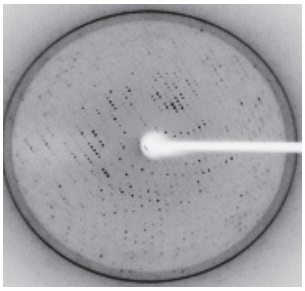
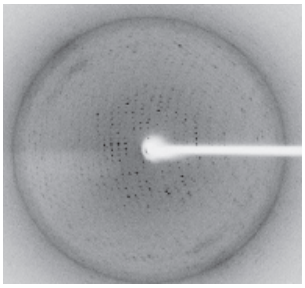
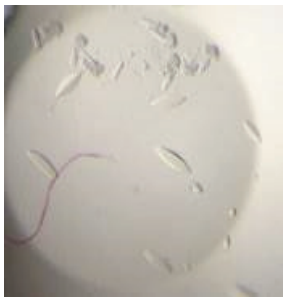
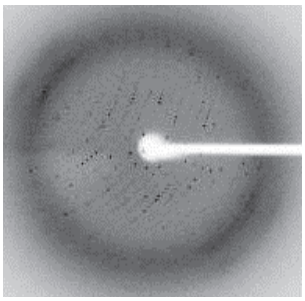
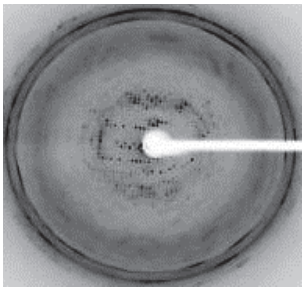
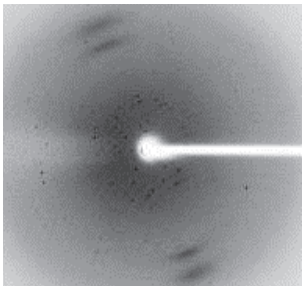
diffracted are mentioned in table 4.6.

Table 4.6. Details of crystallization trials and diffraction of r-LaGAPDH complex

Crystal	Crystal Image	Diffraction image	Remark
LaGAPDH_S1 **Condition:** 25% PEG 1500 **Protein:** 2.5 mg/ml (2:2), microbatch **Cryo**: - **Soaking**: - **Co-crystallization:** 5mM Man			Poor diffraction
LaGAPDH_S2 **Condition:** 25% PEG 1500 **Protein:** 2.5 mg/ml (2:2), microbatch **Cryo**: 30% Glucose **Soaking**: - **Co-crystallization:** 5mM Man			Poor diffraction
LaGAPDH_S3 **Condition:** 25% PEG 1500 **Protein:** 2.5 mg/ml (3:2), microbatch **Cryo**: 25% PEG 1500 **Soaking**: - **Co-crystallization:** 5mM Man			Poor diffraction
LaGAPDH_S4 **Condition:** 25% PEG 1500 **Protein:** 2.5 mg/ml (3:2), microbatch **Cryo**: 20% Glycerol **Soaking**: - **Co-crystallization:** 5mM Man			No diffraction
LaGAPDH_S5 **Condition:** 25% PEG 1500 **Protein:** 4.5 mg/ml (3:3), microbatch **Cryo**: - **Soaking**: No **Co-crystallization:** 5mM Man			Data collected but error in processing .

LaGAPDH_S6 **Condition:** 25% PEG 1500 **Protein:** 2.5 mg/ml (3:3), microbatch **Cryo**: - **Soaking**: 10mM Man **Co-crystallization:** 5mM Man			Poor diffraction
LaGAPDH_S7 **Condition:** 25% PEG 1500 **Protein:** 2.5 mg/ml (1:1), microbatch **Soaking**: 5mM Man, Gal, GalNAc, GlcNAc each. **Co-crystallization:** -			No diffraction
LaGAPDH_S8 **Condition:** 25% PEG 1500 **Protein:** 2.5 mg/ml (3:2), microbatch **Soaking**: - **Co-crystallization:** 5mM Man			Data collected but error in processing due to heavy ice-rings
LaGAPDH_S9 **Condition:** 10% PEG 8000 **Protein:** 5 mg/ml (1.5:3), hanging-drop **Cryo**: -, **Soaking**: - **Co-crystallization:** 75mM Man			No diffraction
LaGAPDH_S10 **Condition:** 25% PEG 1000, **Protein:** 5 mg/ml (1.5:3), hanging-drop, **Cryo**: -, **Soaking**: - **Co-crystallization:** 20mM Man, Gal, GalNAc, GlcNAc each.			Heavy ice-rings

LaGAPDH_S11 **Condition:** 10% PEG 8000 **Protein:** 5 mg/ml (1.5:3), hanging-drop **Cryo**: -, **Soaking**: - 25mM Man (4h) **Co-crystallization:** -			No diffraction
LaGAPDH_S12 **Condition:** 25% PEG 1000, **Protein:** 3 mg/ml (2:3), microbatch **Soaking**: 10mM Man, Gal, GalNAc, GlcNAc each. **Co-crystallization:** 25mM Man & Gal each			Data collected at 2.01Å.
LaGAPDH_S13 **Condition:** 25% PEG 1500 **Protein:** 3 mg/ml (2:2), microbatch **Soaking**: 50mM Man and 50 mM GlcNAc (overnight) **Co-crystallization:** 75mM Gal	-		Poor diffraction
LaGAPDH_S14 **Condition:** 25% PEG 1000 **Protein:** 3 mg/ml (2:2), microbatch **Co-crystallization:** - 75mM GalNAc, GlcNAc each	-		Poor diffraction
LaGAPDH_S15 **Condition:** 25% PEG 1000 **Protein:** 3 mg/ml (2:2), microbatch **Cryo**: - 25% Glycerol **Co-crystallization:** 75mM GalNAc,	-		Poor diffraction

GlcNAc each			
LaGAPDH_S16 **Condition:** 25% PEG 1000 **Protein:** 3 mg/ml (2:3), microbatch **Co-crystallization:** -25 mM Man, Gal, GalNAc, GlcNAc each.			Data collected at 2.80Å.
LaGAPDH_S17 **Condition:** 25% PEG 1500 **Protein:** 5 mg/ml (1.5:3), hanging-drop **Co-crystallization:** -25 mM GalNAc, GlcNAc each.			Poor diffraction
LaGAPDH_S18 **Condition:** 8% PEG 400 **Protein:** 5 mg/ml (3:1), microbatch, pH 5.0, 20mM sodium acetate **Co-crystallization:** -50 mM Man, Gal, GalNAc, GlcNAc each.			Data collected at 2.99Å
LaGAPDH_S19 **Condition:** 10% PEG 1000 **Protein:** 3 mg/ml (3:1), microbatch, pH 5.0, 20mM sodium acetate **Co-crystallization:** -25mM Gal & GalNAc	-		Poor diffraction
LaGAPDH_S20 **Condition:** 25% PEG 1000 **Protein:** 5 mg/ml (1.5:3), hanging-drop **Cryo**: 25% PEG 1000 **Soaking**: 50 mM Man and Gal **Co-crystallization:** -	-		Data collected at 2.5Å

4.4.2.2. Data collection and processing

The diffraction data was collected for the putative r-LaGAPDH complex crystals (LaGAPDH_S1 to LaGAPDH_S20) under cryo conditions. The respective cryoprotectant used for collecting diffraction data is mentioned in Table 4.6. The crystal (LaGAPDH_S4, LaGAPDH_S7, LaGAPDH_S9, LaGAPDH_S10, and LaGAPDH_S11) showed no diffraction (Table 4.6).

Table 4.7: X-ray diffraction statistics of LaGAPDH_S12, LaGAPDH_S16, LaGAPDH_S18 and LaGAPDH_S20 crystals.

Crystal Name	LaGAPDH_S12	LaGAPDH_S16	LaGAPDH_S18	LaGAPDH_S20
Diffraction source	Cu Kα	Cu Kα	Cu Kα	Cu Kα
Detector	R-AXIS IV++	R-AXIS IV++	R-AXIS IV++	R-AXIS IV++
Wavelength (Å)	1.54178	1.54178	1.54178	1.54178
Temp. (K)	100	100	100	100
Crystal-detector distance (mm)	200	200	200	200
Rotation range per image (°)	0.5	0.5	0.5	0.5
Space group	$P4_12_12$	P422	$P2_12_12_1$	I121
Unit-cell parameters	a=113.22, b=113.25, c=114.51	a=115.0, b=115.0, c=113.37	a=91.67, b=110.58, c=142.47	a=113.28, b=113.28, c=113.28
α, β, γ (°)	α = β = γ = 90	α = β = γ = 90	α = β = γ = 90	α = β = γ = 90
Mosaicity	0.52	0.00	0.70	-
Resolution (Å)	2.01	2.80	2.99	2.5
Total no. of reflections	611430	287531	255247	-
No. of unique reflections	46875	26763	19171	27992
Completeness (%)	93.2 (100.0)	99.4 (100.0)	99.9 (99.9)	-
R_{merge}	0.13 (4.52)	0.34 (3.169)	0.63 (5.82)	0.56 (-)
Mean (I/σ(I))	9.6 (0.7)	6.2 (1.4)	6.4 (0.9)	-
CC (1/2)	0.998 (0.302)	0.951 (0.841)	0.877 (0.049)	-
Multiplicity	13.0 (12.3)	15.0 (16.2)	8.5 (8.2)	-

Values of outer shell are given in ().

The crystal (LaGAPDH_S1, LaGAPDH_S2, LaGAPDH_S3, LaGAPDH_S6, LaGAPDH_S13, and LaGAPDH_S14, LaGAPDH_S15, LaGAPDH_S17 and LaGAPDH_S19) showed very poor diffraction spots, so the data was not further collected (Table 4.6). The crystal (LaGAPDH_S5 and LaGAPDH_S8) could not be processed (Table 4.6). The crystal (LaGAPDH_S12, LaGAPDH_S16, LaGAPDH_S18 and LaGAPDH_S20) diffracted up to 2.01 Å, 2.80 Å, 2.99 Å and 2.5 Å respectively. The details of data processing for LaGAPDH_S12, LaGAPDH_S16, LaGAPDH_S18 and LaGAPDH_S20 are mentioned below (Table 4.7.). The data for LaGAPDH_S16 and LaGAPDH_S20 were processed by XIA2-DIALS (CCP4i) program. The diffraction data of LaGAPDH complex crystals were not good.

4.4.3. Crystal structure of r-LaGAPDH

4.4.3.1 Quality of r-LaGAPDH model

The crystal structure of r-GAPDH was determined at 2.21 Å resolutions by the molecular replacement method using PDB: 4QX6 as a search model. The asymmetric unit of the crystal structure contains two monomers, A and B. The refined model consists of one dimer with subunits labelled A and B, containing 5306 atoms, 676 amino acid residues and 179 water molecules. The refinement statistics is given in Table 4.5, and the overall quality of the model was verified by MolProbity and PROCHECK using Ramachandran statistics. The electron density for almost all the amino acids in each subunit is good except in between region 24-30 and 104-110 amino acid residues. The model has 5-7% residues which are a poor fit to electron density, 5% moderate fit and 95% residues are in agreement with the electron density.

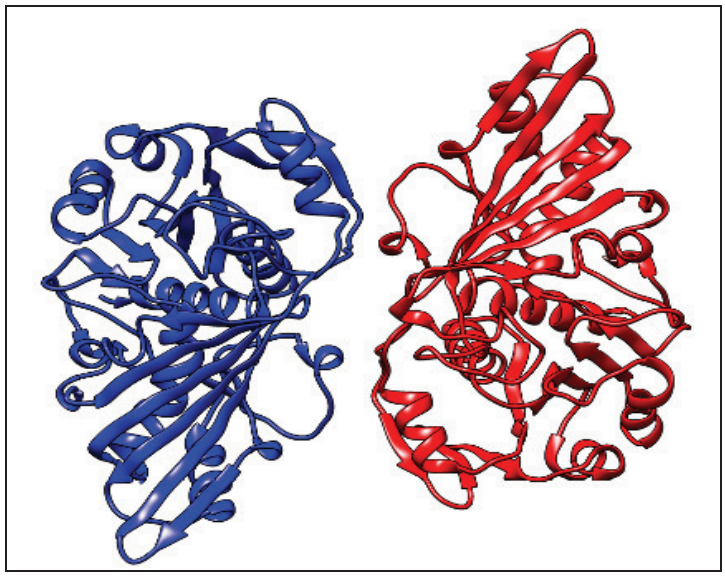
4.4.3.2. Structural features

The overall architecture of GAPDHs from a various organism is very well known to have four identical subunits (O, P, Q and R) which are arranged with a molecular 222 symmetry (Seidler, 2013). The r-LaGAPDH crystal data obtained in this study showed the presence of two monomers in the asymmetric unit (Figure 4.5.a). It has a two domains, an NAD-binding domain [1–148] residues which display the characteristic Rossmann fold and a catalytic domain [149-329] fold comprising of an eight-stranded antiparallel sheets and four helices. In between the NAD binding and catalytic domain, there is an S-loop. The r-LaGAPDH in general forms a globular structure which consists of fourteen sheets (β_1 to β_8 in the catalytic domain and β_A to β_F in NAD binding domain) and eight helices (α_1 to α_4 in the catalytic domain and α_A to α_D in NAD binding domain)

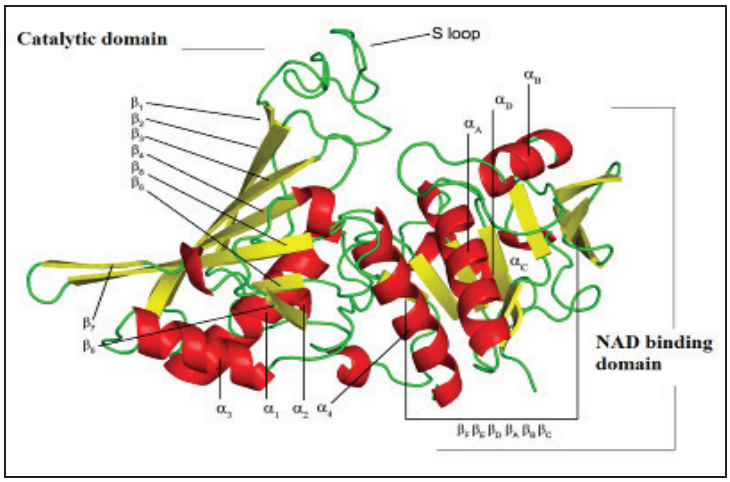
(Figure 4.5.b). As expected, the secondary and tertiary structures of r-LaGAPDH are very similar to those of other GAPDH structures (Figure 4.6).

4.4.3.3. Sequence and structural comparison with other GAPDHs.

The sequence of 41 representative GAPDHs was download from UniProt for further sequence comparison and key differences related to three cysteines at position 101, 156 and 326 are shown in the multiple sequence alignment (Figure 4.7). Also, an MSA was performed for LaGAPDH against other GAPDHs whose structure are solved.



(a)



(b)

Figure 4.5: The ribbon representation of LaGAPDH: (a) The structure of r-LaGAPDH with its two monomers. Subunits A and B are coloured in blue and red; (b) Ribbon representation of monomer LaGAPDH with the secondary structure elements indicated.

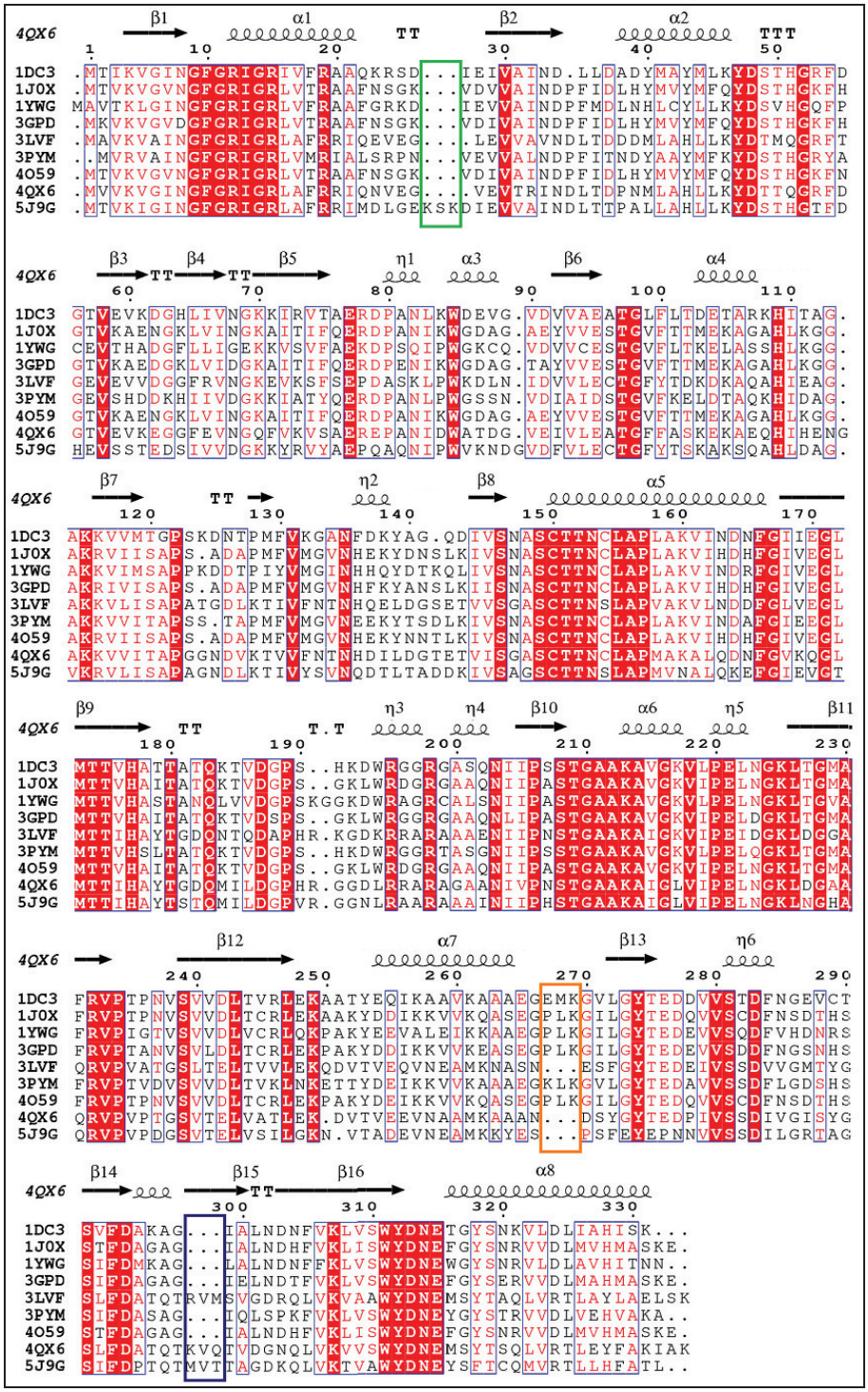


Figure 4.6: Sequence alignment of LaGAPDH with other known representative GAPDH: The secondary structural elements of MR search model, PDB 4QX6 are shown above the alignment. The alignment includes GAPDHs from E. coli (PDB 1DC3), Rabbit (PDB 1J0X), P. falciparum (PDB 1YWG), H. sapiens (PDB 3GPD), S. aureus (PDB 3LVF), S. cerevisiae (PDB 3PYM), B. taurus (PDB 4O59), S. agalatiae (PDB 4QX6) and LaGAPDH (PDB 5J9G). The green, orange and blue colour box indicated a difference in LaGAPDH with respect to other GAPDHs.

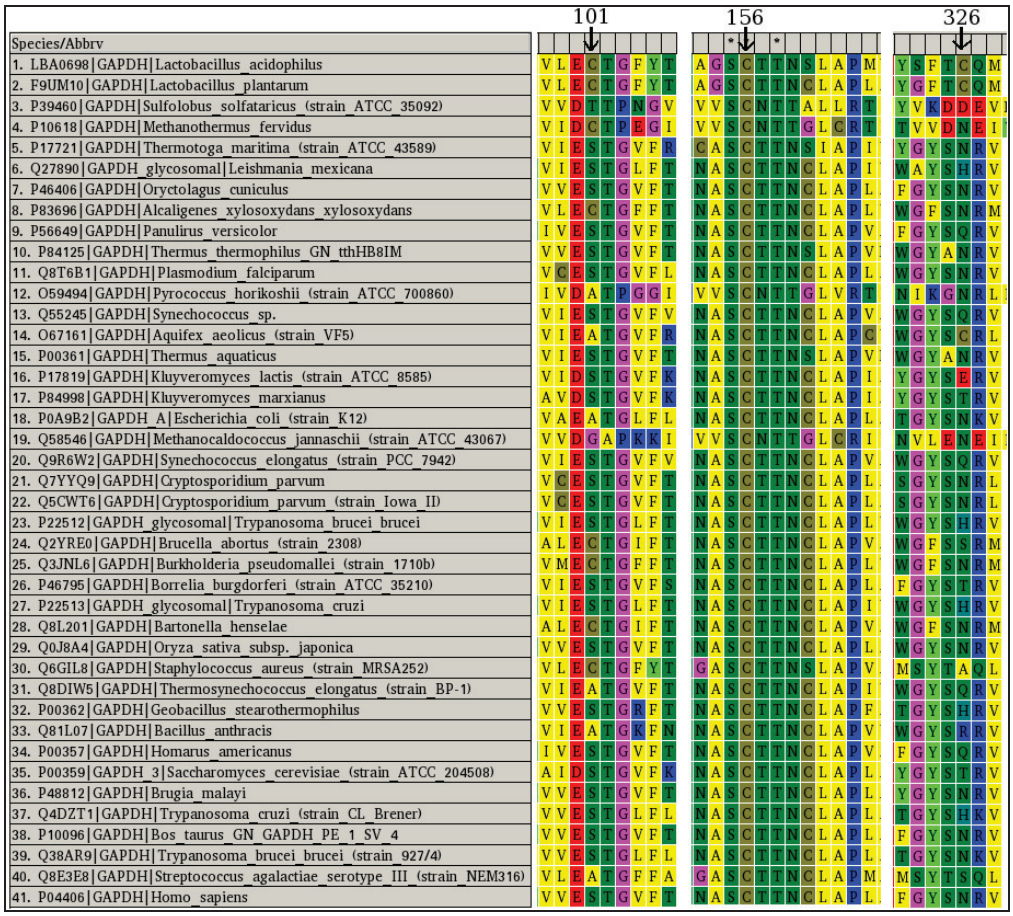
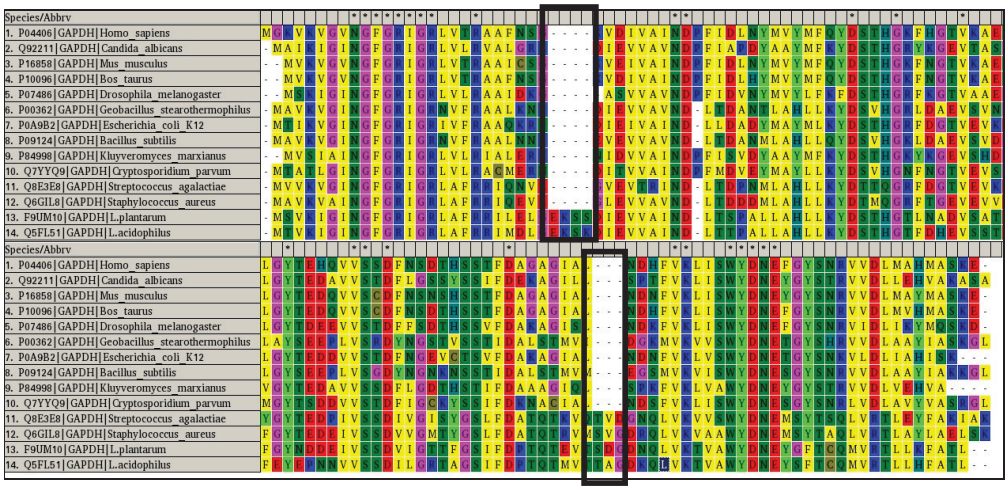
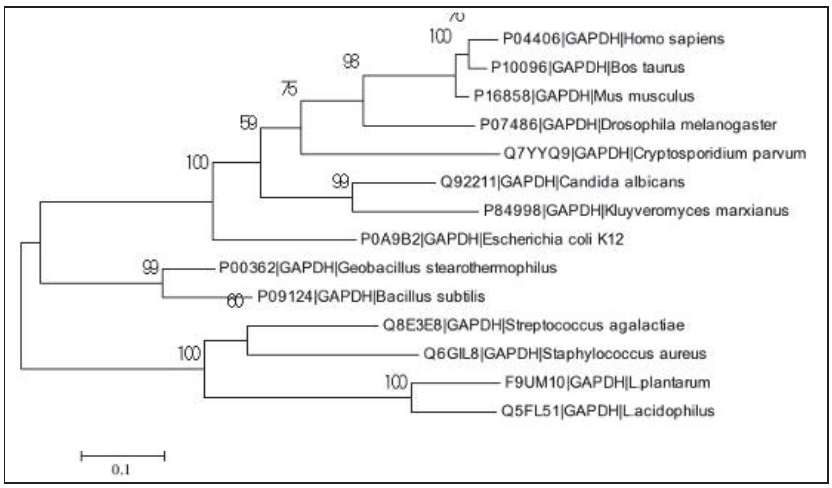


Figure 4.7: Multiple sequence alignment of GAPDH sequences and highlighting respective positions of three cysteines at 101, 156 and 326.



(a)

Figure 4.8: Multiple sequence alignment and phylogenetic analysis: (a) Multiple sequence alignments of r-LaGAPDH along with other known representative GAPDHs from different class showing the distinct region in the sequence marked with a black box. *(contd..)*



(b)

Figure 4.8: Multiple sequence alignment and phylogenetic analysis: *(contd..)* (b) Phylogenetic relationships of GAPDHs showing the distinct clade of GAPDHs from *Streptococcus*, *Staphylococcus* and *Lactobacillus* all are which known to be surface localised.

A stretch of three amino acids "TAG" starting at position 305 is found in GAPDH in three species: *Streptococcus*, *Staphylococcus* and *Lactobacillus* and an another stretch of amino acids "EKSK" starting at position 26 are found only in the *L. plantarum* GAPDH apart from LaGAPDH (Figure 4.8a). A phylogeny of these sequences also had GAPDHs from *Lactobacillus* to a different node and thus differentiating with respect to other GAPDHs (Figure 4.8b).

The structure of LaGAPDH was compared with the structure of *S. agalatiae* GAPDH (PDB 4QX6) with respect to the active site. The residues in the active sites were analyzed to check for the conservation of the crucial amino acids (Figure 4.9). The residues ARG 12, ILE 13, LEU 35, ARG 78, THR 97, THR 121, ASN 316 and an aromatic amino acid TYR 320 are involved in binding of NAD in GAPDH from *Streptococcus*. The LaGAPDH (PDB: 5J9G) active site had few of the amino acids conserved like ARG 12, ILE 13, LEU 39, ARG 82, CYS 101, SER 125, ASN 320 and an aromatic amino acid PHE 324. The 3D structure-based sequence alignment of LaGAPDH with known PDB structures reveled few stretch of amino acids in the C-terminal strands of LaGAPDH, *S. aureus* GAPDH and *S. agalatiae* GAPDH which were considerably longer and surface exposed compare to other species GAPDH (Figure 4.10). In *Streptococcus* and *Staphylococcus,* it is known that GAPDH is a surface/ secreted protein. This region might be of particular importance because of its unique presence along with surface exposure.

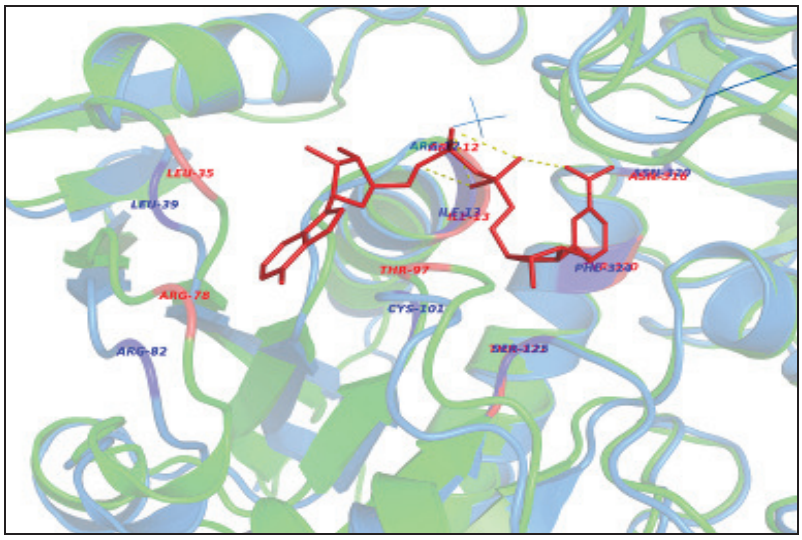


Figure 4.9: Comparison of amino acids in the active site of LaGAPDH (residues marked in red: ARG12, ILE13, LEU39, ARG82, CYS101, SER125, ASN320 & PHE324) with NAD bound active site of *Streptococcus* GAPDH (PDB 4QX6; residues marked in blue: ARG12, ILE13, LEU35, ARG78, THR97, THR121, ASN316 & TYR320).

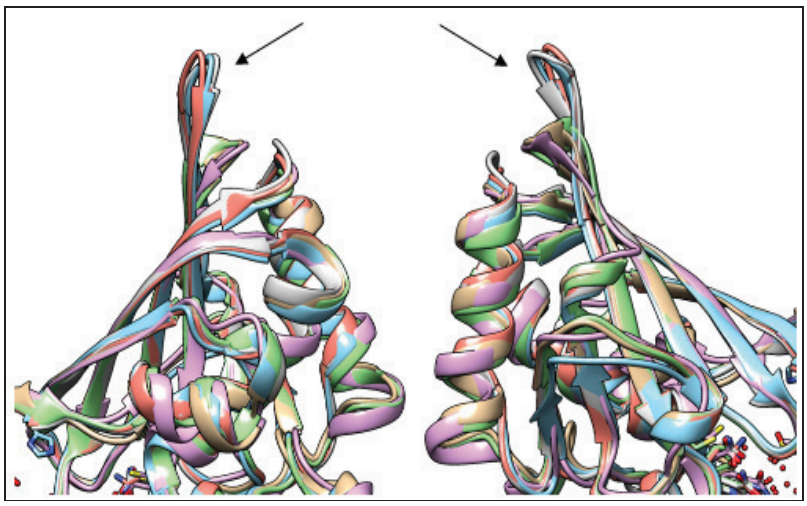


Figure 4.10: Structural difference in LaGAPDH model: The figure shows a comparison of the diverse region 303–307 as observed in LaGAPDH, *S. agalatiae* (PDB: 4qx6), *S. aureus* (PDB: 3lvf), *C. parvum* (PDB: 1vsv), *E. coli* (PDB: 1dc4) and human (PDB: 3gpd) GAPDH.

4.5. Discussion

The foremost step in bacterial colonisation is adhesion to host tissues which are primarily mediated by cell surface adhesion proteins. Most of the work on structural characterization of bacterial adhesin is focused on gram negative adhesins (Bensing *et al.*, 2016) and to a few extent gram-positive pathogenic adhesins (Liu *et al.*, 2007;

Ramboarina *et al.*, 2010). During last decade a considerable effort has been made for structural characterization of pathogenic bacteria. Currently, there is a lack of studies or very few studies focused on structural characterization of gram-positive adhesin from commensal bacteria. The first step in colonisation by gut bacteria is their adhesion to host tissue cells, ECM which is primarily based on the binding pf carbohydrate binding protein or lectin -like adhesins molecules (Ielasi *et al.*, 2016).

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a classical conserved intracellular glycolytic enzyme that catalysis the oxidative phosphorylation of G3P (glyceraldehyde 3-phosphate) to 1, 3-bisphosphoglycerate in the presence of NADH/NADPH and inorganic phosphate (Harris & Waters, 1976). GAPDH is considered the prototype of moonlighting proteins and acts as adhesin which binds to host epithelial mucin, or ECM components (fibronectin, mucin, plasminogen etc.) and thus make it an essential mediator of host-microbe interactions (Kinoshita, Uchida*, et al.*, 2008; Pancholi & Fischetti, 1992; Ramiah *et al.*, 2008).

The molecular and mechanistic aspects of moonlighting proteins not clear. There exists only a few preliminary reports of adhesins from *L. acidophilus* group which binds to ECM through lectin-like recognition (Mukai *et al.*, 1992; Yamada *et al.*, 1994) but more efforts are needed to understand the mechanistic and structural basis of these host-microbe interactions. Although the role of GAPDH in host-microbe interactions is well known, only a past known study have been undertaken to evaluate its lectin-like recognition of *Lactobacillus plantarum* cell surface GAPDH with human blood group antigens (Kinoshita, Wakahara*, et al.*, 2008). Thus structural information on GAPDH from commensal/ probiotics microbes is restricted to the crystal structure for LaGAPDH, providing the first structure from *Lactobacillus*. The cell surface GAPDH is one of the few adhesins which helps the bacteria to colonize and thus it inspired us as a principal moonlighting protein to undertake the structural characterization of LaGAPDH.

In this chapter, we discuss the detailed crystallization setup, data collection and data processing of LaGAPDH protein from *L. acidophilus*. The purified LaGAPDH protein was crystallized in microbatch under oil method. Diffraction quality 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 in a time of two weeks at room temperature. The diffraction data of LaGAPDH native crystal was

processed with iMosflm in space group P41212 and the unit cell parameters were a = 114.98, b = 114.98, c = 113.29 Å, α = β = γ = 90. However the same data when processed with the HKL package it processed in space group P212121 with unit cell dimensions a = 113.98, b = 115.09, c = 114.99 Å, α = β = γ = 90º. Finally, the data was processed in tetragonal space group P41212, a total of 661620 reflections was 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 co-efficient and solvent content were 2.41 Å3 Da-1 and 48.96% respectively. The molecular replacement was performed to get the solution of LaGAPDH native using search molecule PDB: 4QX6, the nearest homolog of LaGAPDH. The initial MR model gave a score RFZ=10.3, TFZ=29.1 and LLG=532 which indicated a good MR solution. 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 was validated with MolProbity and PDB validation server and finally submitted to PDB with PDB ID: 5J9G.

The importance of any target is very high if the structure is solved with its interacting partner. Looking forward to this and having a thorough understanding of how LaGAPDH interacts with the highly glycosylated mucin, numerous trials were carried out for LaGAPDH complex crystals with sugars as they constitute 70% of mucin glycosylated surface. Based on the experiments of fluorescence spectroscopy, a sugar which gave considerable quenching were selected for co-crystallization and soaking trials. 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, usage of cryoprotectant so that we have maximal chance of crystallizing and diffracting a complex data. In most of the case the crystal grew but diffracted very poorly or with no diffractions. Few of them diffracted, but we were not able to solve the data due to errors during data processing. The R_{merge} values were very high to have a good solution data. A wide variety of soaking and co-crystallization trials were performed. We were able to collect four data but none of them processed successfully.

The crystal structure of GAPDH from *Lactobacillus acidophilus* (LaGAPDH) in apo-form have been determined at 2.21 Å by the MR with two monomers in the asymmetric unit. The final refined model consists of one dimer with two subunits labeled A and B, containing 5306 atoms, 676 amino acid residues and 179 water molecules. The R_{factor} and R_{free} of final model were 0.21 and 0.23 respectively. The experimental results reveal that GAPDH exists as tetrameric in solution and which is in agreement with the crystallographic structures of other GAPDHs, but LaGAPDH has 2 monomer subunits in asymmetric unit as earlier observed in cell-wall anchored GAPDH from *Kluyveromyces marxianus* (Ferreira-da-Silva *et al.*, 2006). The overall structure of LaGAPDH was of Rossmann fold with a catalytic domain and NAD binding domain. Most of the GAPDHs have a conserved cysteine at 156 and 160 position. However, the latter is not conserved in few organisms. Most of the organisms have glycine or serine at conserved 160 position instead of cysteine. Sequence comparison with other GAPDH helps us to identify that LaGAPDH contained three cysteines, one more than most of other. 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 position which is 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 (UniProt: O67161). The cysteine together at 101 and 326 position are also found on *L. plantarum* GAPDH as per our analysis, which is a reported to have adherence with mucin (Kinoshita, 2008). The presence of three cysteines in LaGAPDH is not common when compared to other known GAPDHs, so the sequence was checked for any possible formation of cysteine-cysteine disulfide bonds. The distance between cysteine was large, and disulfide bonds are not possible. The cysteine at 101 position 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, 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. This region forms a loop in between α1 and β2 in the

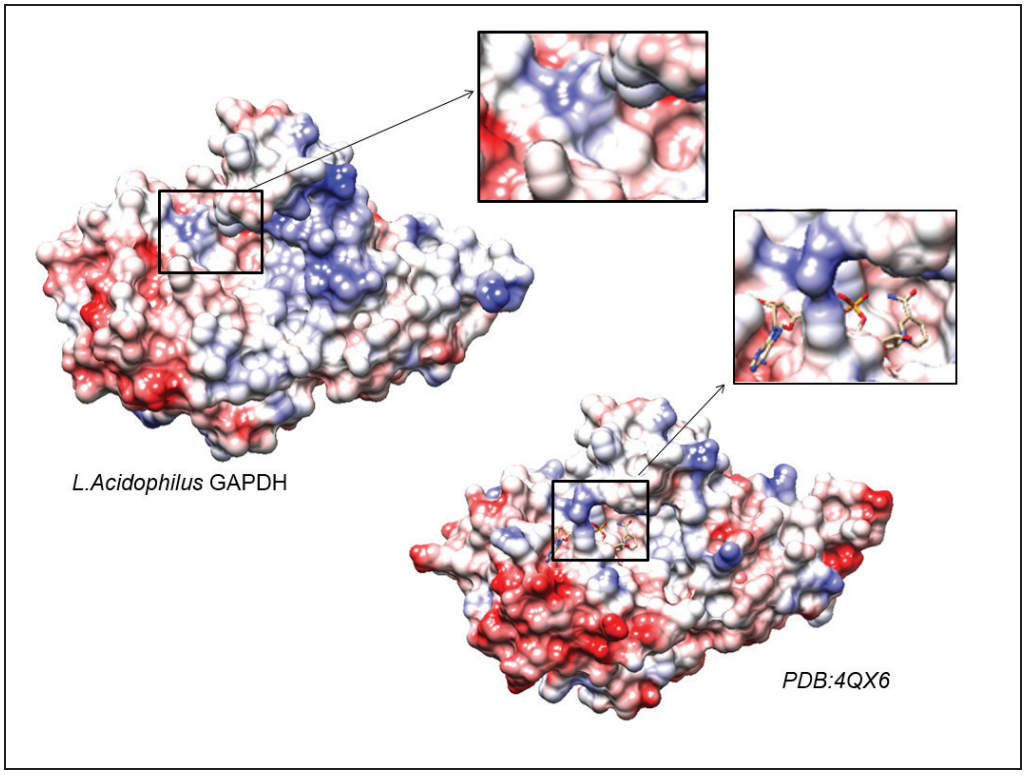


Figure 5.23: Comparison of active site between LaGAPDH and PDB: 4QX6 with respect to its surface charge. The blue color represents positive charge on surface while red represent negative charge. The active site with/ without NAD bound are shown in respective inset diagram.

N-terminal NAD binding domain. These both regions also stand out in the sequence based secondary structure alignment. Also, the electrostatic surface charges when compared with the PDB: 4QX6, indicated that there might be an open and closed form of an enzyme-substrate complex. The electrostatic surface charge when to compare to 4QX6 seems to have more positive charge residues which may help to facilitate electrostatic interaction with negatively charged mucin (Figure 5.23). The sialic acid and other chains of oligosaccharides on the mucin confer negative charge due to the carboxyl and sulphate group (Li *et al.*, 2013). Also the enzyme closed and open form with respect to NAD binding is represented above (Figure 5.23).

Though we couldn't succeed with the structure solution of LaGAPDH complex with carbohydrates, we were able to solve the structure of LaGAPDH native. Our final aim will be structure solution of LaGAPDH complex and indeed a better understanding of interaction at the molecular level.