Chapter 2

Materials and Methods

This chapter deals with the materials and methods used for the experiments done.
2. Materials and Methods

This chapter deals with the materials and methods used for the experiments that are discussed in this thesis work. Only those experimental procedures have been enlisted whose methodology and parameters/variables are necessary to be documented for repetition/verification of these experiments in future. Other experiments have been carried out using routine lab procedures and hence not mentioned below (Sambrook and Russell, 2001).

2.1 Materials

Most of the chemicals used for routine experiments such as enzymes, chemicals and reagents for molecular biology were from Bangalore Genei, Bioserve, Calbiochem, Invitrogen, Merck, NEB, Promega, Qiagen, Qualigens, Serva, Sigma, Stratagene and SRL biochemicals (listed alphabetically).

Commercial crystallization screens from Hampton Research and Jena Bioscience were used. Glass coverslips were obtained from Polar Inc. 24 well crystallization plates were obtained from Iwaki Inc. Other crystallization related chemicals and tools were obtained from Hampton Research. The resins for chromatography were bought from GE Healthcare. The primers were obtained from Bioserve and Sigma. Substrate analogs used for crystallization were obtained from RNA Tech and Jena Biosciences.
Chapter 2: Materials and Methods

2.2 Methods

2.2.1 Bioinformatics

The sequences of homologous proteins from different organisms were obtained from Basic Local Alignment Search Tool (BLAST-NCBI) server. Constraint-based Multiple Alignment Tool (COBALT-NCBI) was used for multiple protein sequence alignment using conserved domain and local sequence similarity information and generation of phylogenetic tree. Multiple sequence alignments of related sequences were also performed using ClustalW2-EBI followed by manual structure based sequence alignments using Jalview (Clamp et al., 2004). The three-dimensional protein atomic coordinates were downloaded from Protein Data Bank using the PDB accession codes (PDB id).

2.2.2 Cloning, expression and purification of Pab-NTD

The cloning, expression and purification for Pab-NTD (1-183 amino acids) were earlier standardized in our laboratory (Dwivedi et al., 2004). Change in induction temperature from 25°C to 18°C and induction period to 16 hours significantly increased the expression of Pab-NTD. Since this particular clone of Pab-NTD was cloned without tag, it was purified using three chromatographic steps as standardized before. A minor modification was made such that the gradient is much gentle in both anion-exchange and hydrophobic chromatographic steps (Figure 2.1 A, B). This resulted in a much better separation of contaminants even after the first step, i.e. anion-exchange chromatography.

The final step of purification was Gel filtration chromatography, which was done in presence of 50 mM Tris (pH 8.0) and 150 mM NaCl (Figure 2.1 C). The fractions encompassing peak-region in gel filtration chromatogram containing Pab-NTD were checked for purity on SDS-PAGE and pooled (Figure 2.1 D). This was followed by concentration of the pooled protein fraction and subsequent buffer-exchange. The protein was finally stored in 20 mM Tris (pH 7.5). Bradford method was used for protein
Figure 2.1: Pab-NTD purification

(A): Ion-exchange chromatogram
(B): Hydrophobic chromatogram
(C): Gel Filtration chromatogram
The fraction (peak) for Pab-NTD is shown in all the above chromatograms.

(D): SDS-PAGE of Pab-NTD purification steps and crystallized protein.
Lane 1: Supernatant of bacterial cell extract containing overexpressing Pab-NTD
Lane 2: Pab-NTD fraction after ion-exchange chromatography
Lane 3: Pab-NTD fraction after Hydrophobic chromatography
Lane 4: Pab-NTD fraction after Gel-Filtration chromatography
Lane 5: Pab-NTD crystals after washing in artificial mother liquor
Lane 6: Protein molecular weight marker (PMW-M) from Bangalore Genei. The molecular weight corresponding to the bands are mentioned.

(E): Pab-NTD crystals
estimation for 2-3 different concentrations in duplicates. Bovine Serum Albumin (BSA) was used as the control for protein estimation. The final yield of Pab-NTD was about 3.5 mg per litre of culture (pellet weight ~10 gm).

Generally, for archaeal proteins, a good way of initial purification is to incubate the bacterial cell lysate at 65 °C for about 20-30 mins, where most of the host proteins precipitates out. This method was checked for Pab-NTD also and it leads to no change in its crystallization ability at the defined crystallization condition mentioned below. However, for the experiments mentioned in this thesis work, the protein was purified without any heat treatment.

2.2.2.1 Cloning of Pab-NTD with His-tag and site directed mutagenesis

The structure of Pab-NTD (PDB id: 1Y2Q) revealed that the N-terminus of Pab-NTD is buried. Moreover, only 143 amino acids could be modeled in the electron density and therefore it was decided to generate another clone of Pab-NTD with only 1-143 amino acids with C-terminal 6-His tag for mutagenesis and biochemical assays. The gene comprising of amino acids 1-143 of Pab-NTD was PCR amplified and cloned with C-terminal tag into Ndel and Xhol sites of pET-21b (Novagen) expression vector. Site-directed mutagenesis was performed with the QuikChange™ XL site-directed kit (Stratagene) using recombinant plasmid of wild type (143) as template. The mutations (K121M, K121S, Y120A, E134A, and H83A) were confirmed by DNA sequencing.

2.2.2.2 Expression and purification of Pab-NTD with His-tag and its mutants

Expression of Pab-NTD with C-terminal 6-His tag was standardized by varying induction temperature, growth period and concentration of Isopropyl-D-thiogalactopyranoside (IPTG). The cells were grown at 37°C till the mid-log phase of growth (O.D. reaches 0.6) and then induced with 0.5 mM IPTG at 18°C for 16 hours. Purification of each mutant was achieved in two steps. The crude cell extract was subjected to nucleic acid precipitation and the supernatant was loaded on Ni-NTA affinity
column (Qiagen) equilibrated with lysis buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 10 mM imidazole. The elution buffer contained 100 mM imidazole. The protein was further purified using gel filtration (Superdex-75) in 50 mM Tris-HCl (pH 8.0) and 50 mM NaCl. The purified protein was exchanged with 20 mM Tris-HCl (pH 7.5) and 10 mM NaCl.

2.2.3 Fluorescence spectroscopy

Fluorescence spectroscopy is an electromagnetic spectroscopic technique that analyzes the phenomenon of fluorescence in a sample. It is primarily concerned with electronic and vibrational states of electrons in the fluorophore. In a fluorescence experiment, the fluorophore is excited by a certain ‘excitation’ wavelength of light that is absorbed by the fluorophore, thereby, promoting an electron in the ground state to an electronically excited state. The phenomenon of returning back of electron to the ground state is accompanied by emission of light of lower energy (higher wavelength) and is termed as fluorescence.

All fluorescence experiments were done using a Hitachi F-4500 fluorescence spectrophotometer. All the spectra were recorded in the correct spectrum mode with the excitation and emission band passes of 5 nm each at room temperature (298 K). Response time was 2 seconds with a scan speed of 100 nm/min. 500 µl of protein sample (0.3 mg/ml) in 20 mM Tris-HCl (pH 7.5) and 10mM NaCl was used for all measurements.

2.2.3.1 Extrinsic fluorescence binding studies of Pab-NTD using ANS

The fluorescence-binding assay of Pab-NTD (1-143 amino acids) with C-terminal tag was done by extrinsic fluorescence as it lacks tryptophan residues. ANS, an extrinsic fluorophore, has been shown to bind to pockets and hydrophobic patches in a protein. It has been used in several studies to monitor the effect of ligand binding using a displacement mechanism of the fluorophore by the ligand (Stryer, 1968).
ANS fluorescence assay was used to monitor the binding of ligand with Pab-NTD as well as the mutants. ANS (0.1 mM) was prepared in 100% methanol. The purified protein (0.3 mg/ml) and ANS (100 μM) mixture was incubated for 15 min before recording the initial spectrum. An excitation wavelength of 380 nm was used and the emission spectra were observed between 400-600 nm. The protein-ANS mixture was then titrated against increasing ligand concentrations of substrate analogs to monitor the changes in the fluorescence spectra till saturation. The protein-ANS-ligand mixture was incubated for 5-10 minutes after each titration. Blank spectra were recorded by similar titrations in buffer-ANS mixture and subtracted from individual spectrum.

2.2.4 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a physical quantitative technique used for accurate measurements of thermodynamic data of biomolecular interactions. It directly measures the heat absorbed or released during a titration and through which binding affinity \((K_a)\), binding stoichiometry \((n)\) of the interaction between two or more molecules in solution and enthalpy changes \((\Delta H)\) are calculated. Entropy changes \((\Delta S)\) and Gibbs energy changes \((\Delta G)\) can be determined using the relationship:

\[
\Delta G = -RT \ln K = \Delta H - T \Delta S
\]

(where, \(R\) is the gas constant and \(T\) is the absolute temperature).

An ITC instrument has two cells (sample & reference) kept at the same temperature \((\Delta T=0)\) in an adiabatic chamber. The protein or biomolecule is kept in the sample cell whereas the reference cell is filled with water or buffer. During the experiment, the ligand is injected into the sample cell using a syringe mounted on top of the instrument in precisely known aliquots. This causes heat to be either taken up or released (depending on the nature of the reaction) leading to temperature differences between sample and reference cell that are detected by sensitive thermocouple circuits. The time-dependent input of power required to maintain equal temperatures between the two cells is measured as a function of differential power \((dp)\) in μcal/s. Exothermic reactions manifest as inflections and endothermic reactions as peaks in the titration.
figure. Saturation of ligand binding is signaled by the reduction in peak heights to the baseline. Similar titration of ligand with identical parameters is done keeping buffer in the sample cell. This forms the blank titration that is subtracted from the actual protein-ligand titration. Heat changes per titration (kcal/mol), obtained by integrating individual peak heights, are plotted against molar ratio of [ligand]/[protein]. Non-linear curve fitting is done using different models in microcal Origin 7.0 software for the quantity of heat released.

2.2.4.1 ITC binding studies of Pab-NTD and K121M mutant

The binding affinities of Pab-NTD and K121M mutant with substrate analogs were measured using an isothermal titration calorimeter (VP-ITC, Microcal Inc.). Substrate analogs (ligand) titrations were done using 0.8-2.0 mM of ligand solution. 10 μl injection volumes were used throughout and the temperature was maintained at 30°C (303 K) with spacing of 180s. The amount of protein was also varied depending on the ligand from 50-100 μM.

2.2.5 NMR Spectroscopy

Nuclear Magnetic Resonance (NMR) Spectroscopy is a technique that utilizes the magnetic property of the nuclei called ‘spin’ that is placed in a magnetic field. This causes the nuclei to absorb energy from the applied electromagnetic pulse and radiate it back at a specific resonance frequency. Different nuclei (protons) in a sample resonate at different frequencies depending on the local chemical environment. The resonant frequency as well as the shift is directly proportional to the strength of the magnetic field. This shift when converted to a magnetic field-independent, dimensionless value is called as ‘chemical shift’ and denoted in parts per million (ppm).

15N Heteronuclear Single Quantum Correlation (15N HSQC) experiments were first suggested in 1980s (Bodenhausen and Ruben, 1980) and has emerged as a useful tool to detect the binding of ligand in an isotope-labeled protein. The output of an HSQC
experiment is a two-dimensional spectrum with $^1$H and $^{15}$N for the X and Y axis, respectively, containing peaks for each unique proton attached to the $^{15}$N. For proteins, each amino acid, except proline, gives a unique peak in $^{15}$N HSQC spectrum due to the presence of amide proton attached to the nitrogen atom in the peptide bond. The nitrogen-containing side chains give additional peaks.

In order to check for the binding of ligands, uniformly $^{15}$N labeled protein is titrated with ligands and the resulting spectra is recorded at different concentrations of ligand titrated. The shift in peaks denotes the interaction of ligand with protein. The absence of binding results in no shift in the observed peaks. NMR experiments were carried out on a Bruker 600 MHz NMR spectrometer equipped with triple resonance cryoprobe. The data were collected at 298K.

2.2.5.1 HSQC titrations with Pab-NTD

Pab-NTD was uniformly labeled with $^{15}$N by expressing it in minimal media with $^{15}$NH$_4$Cl as the sole nitrogen source. 2D $^{15}$N-$^1$H HSQC spectra for Pab-NTD was collected using 200 $\mu$M protein in 400 $\mu$l volume of 50mM phosphate buffer at pH 7.2 containing 50 mM NaCl also. The spectra were recorded upon titration of 0 $\mu$M, 100$\mu$M, 500 $\mu$M of L-Ser3AA and L-Thr3AA in different sets of experiments. Titration of D-Ser2AA was done as negative control and as expected, it showed no shifts in any peak.
2.2.6 X-ray crystallography

2.2.6.1 Crystallization

The first step towards structure determination by X-ray crystallography is crystallization. Pure solutions of proteins or other macromolecules to be crystallized are kept in specific conditions for crystallization. The purified macromolecule undergoes slow precipitation from an aqueous solution and as a result, individual molecules align themselves in a repeating series of "unit cells" by adopting a consistent orientation to form a crystal (McPherson, 1999; 2004).

2.2.6.1.1 Crystallization of Pab-NTD and K121M mutant complexes

In order to elucidate the editing mechanism, co-crystallization of Pab-NTD with different substrate analogs were attempted. The design of the substrate analogs is shown in Figure 2.2. These analogs are explained individually later along with results of respective complexes. As the crystallization condition of Pab-NTD (1-183 amino acids) was known (Dwivedi et al., 2004), initial attempts were made in the same condition. Pab-NTD was earlier crystallized in 25% PEG 8000 and 0.1 M Hepes (pH 7.0) using a concentration of 3.2 mg/ml. Crystallization attempts of Pab-NTD substrate analogs complexes in the above mentioned condition was unsuccessful.

Separate screens with expansion around the above condition were also not fruitful. Therefore, Crystal Screens 1 & 2 and Index screen from Hampton Research were set up using hanging drop vapour diffusion method in 24 well plate formats at both 4°C and at room temperature (25°C). Equal volumes (2μl) of protein (3.2 mg/ml), ligands (5 mM), and reservoir solutions were used for setting up the drop. The protein and ligands were mixed together and left for equilibration (30-60 minutes at 4°C or 15 minutes at 25°C depending upon the temperature of crystallization). The pre-mix (4μl) of Pab-NTD and ligand was used for setting up the drops. The equilibration time
Figure 2.2: Design of post- and pre-transfer editing substrate analogs

The amide-based substitutions (highlighted) prevent the analogs from spontaneous and enzyme-catalyzed hydrolysis.

(A) L-Ser3AA mimics L-ser-tRNA^{Thr}
(B) 3’-linked aa-tRNA and aa-3AA [(L/D-amino-acyl)-3-amino-3-deoxyadenosine]
(C) 2’-Linked aa-tRNA and aa-2AA [(L/D-amino-acyl)-2-amino-2-deoxyadenosine]
(D) aa-AMP (aminoacyl-adenylate) and aa-AMS (5-O-[N-(l-aminoacyl)-sulfamoyl] adenosine).

(Figures B, C and D are from Ling et al., 2009a)
and the ligand concentrations used were varied according to the expected affinity of the ligands (Appendix Tables 1 and 2).

The crystallization screens were evaluated thoroughly. Clear drops, heavy precipitates, whiskers, phase separations, crystalline precipitates, amorphous precipitates, microcrystals and needle clusters were observed. Attempts were made to figure out the constituents that are contributing positively as well as negatively towards crystallization of Pab-NTD complex. Next expansion screens were designed using the initial conditions and three different protein concentrations (2.8, 3.0, 3.2 mg/ml) were used for setting up crystallization drops. After three rounds of screening and expansion, well diffracting crystals of Pab-NTD complexes were obtained.

The crystals obtained in Tris buffer above pH 7.5 were not stable and started degrading within a couple of days of formation. Bis-Tris was found out to be the best buffer and the crystals formed in it were stable for about 20 days. Only freshly purified Pab-NTD was used for crystallization. Old proteins samples (even 3-4 days old) failed to yield well-diffracting crystals or formed needles.

All the crystal structures mentioned in this work were obtained using Pab-NTD (1-183 amino acids) (Figure 2.3). The crystals obtained, however, when ran on SDS-PAGE showed a band corresponding to lower molecular weight (Figures 2.1 D and E). Upon structure solution only electron density corresponding 1-143 or 1-147 residues were observed (mentioned below). Similar screening and crystallization attempts with Pab-NTD (1-143 amino acids) with C-terminal tags were unsuccessful. However, it is important to mention here, for future work, that Pab-NTD (1-143 amino acids) without any tag can be crystallized at 2.5 mg/ml in 25% PEG 1500 only and the crystals appear within 12 hours of setting up of crystallization drops.
Figure 2.3: Crystals of Pab-NTD and K121M mutant complexes

Pab-NTD-L-Ser3AA

Pab-NTD-L-Thr3AA

K121M mutant-L-Ser3AA

K121M mutant-L-Thr3AA

K121M mutant-Gly3AA

K121M mutant-D-Ser3AA
2.2.6.2 Cryo-protection

Since all of the x-ray crystallographic data collection reported in this work was done at 100 K, it is important to ensure that the crystal order do not break upon flash freezing in liquid nitrogen stream. The water or solvent present in crystals freezes, resulting in loss of three dimensional ‘orientation’ of repeating unit and cracking of crystals. In order to avoid this, crystals are immersed in a cryo-protectant solution before freezing that forms an amorphous glass in which the crystal experiences minimal damage (vitrification of crystals). The crystallization conditions are normally insufficient for cryo-protection and require an addition of cryo-protectant to the mother liquor. 20% glycerol in reservoir solution was used as a cryoprotectant for crystal obtained in 30% PEG 8000.

25% PEG 3350 is the precipitant concentration used in most of the crystallization experiments in this work. Ideally, this is sufficient for cryo-protection, however, 15% glycerol in mother liquor was initially used as cryo-protectant. However, with improved crystal-mounting ability towards the latter part of the thesis work, the use of glycerol as cryo-protectant was avoided and crystals were mounted directly from the crystallization drops.

2.2.6.3 Mounting of crystals

Crystals were mounted using nylon loops of appropriate size from Hampton Research. Care was taken to remove as much mother liquor as possible from the loop. This was done by tilting the loop after scooping out the crystal and gently touching the dry part of the cover slip. This was repeated for a couple of times or more, taking care that the crystal does not slip out of the loop. This removed a lot of background in the diffraction image and also the need of glycerol as cryo-protectant during most of the data collection. Hampton ‘MicroTool’ was used for manipulating clusters of crystals, or to remove crystals stuck to the cover slips.
2.2.6.4 Data collection

All x-ray diffraction data sets collected were at the home source rotating copper (Cu) anode x-ray generator (Rigaku RU-H3R). The characteristic Kα x-rays of Cu produced by the electrons falling to the innermost layer of the atom (higher binding energy) have a wavelength of 1.5418 Å. The x-rays are directed into an 'Osmic Blues' confocal mirror system consisting of 2 pairs of (horizontal and vertical) slits that produce a spectrally pure monochromatic converging x-ray beam. The x-ray generator was operated at 40kV/70mA for initial screening of crystals, whereas the x-ray diffraction data sets were collected at 50kV/100mA using Mar 345-dtb (desktop beamline) attached to the x-ray generator. Mar 345-dtb has a goniometer system with high precision phi axis and motorized translation along with a collimator consisting of beam shutter, motorized beam stop and CCD-camera for viewing and centering of crystals in the x-ray beam.

The diffracted x-rays were detected on mar345dtb Image Plate detector with 150 μm X 150 μm pixel size. The crystal-detector distance was varied for different crystals depending upon the observed diffraction. The oscillation range of Δφ = 1.0° was used for quick data collection for checking the presence of bound ligand and Δφ = 0.5° was used for final data collection. Mar345dtb, graphical user interface MAR345 Version 1.2.8, was used for collecting data. The x-ray data sets were collected at 100 K in order to reduce radiation damage of crystals. Oxford 600 Cryostream connected to a liquid nitrogen dewar with an in-line super dryer was used for cryo-cooling the crystals.

The x-rays diffraction data sets were collected in two modes for Pab-NTD: quick data collection and final data collection. Quick data collection was done to find out the presence/absence of ligand in every new ligand-protein complexes. Change of ligand concentration or change in pH in the crystallization conditions were also considered as a new ligand-protein complex. The data collection parameters were used with an objective to complete about 90% data completion in about 6-8 hours with a modest resolution of 2.5Å to 2.8Å. For final data collection, the complete diffraction potential of the crystal was exploited and the data sets were collected with few overloads per frame.
The initial x-ray diffraction image (Figure 2.4) processing for screening or setting up data collection strategy was performed using Automar version 1.3.4. Peak search, autoindexing and prediction were made to determine the tentative crystal lattice. Automar-strategy was used for devising the strategy for data collection. Two x-ray diffraction images (usually around \( \phi = 0^\circ \) and \( \phi = 90^\circ \)) from a given crystal were used for initial analysis. Most of the Pab-NTD (plate-like) complex crystals were found to be anisotropic in diffraction. In such a scenario, the strategy was devised to collect the x-ray diffraction data around the phi angle range that shows better diffraction.

2.2.6.5 Data processing

HKL program suite 1.98.2 was used for processing of the complete data sets collected. The x-ray diffraction image was displayed using ‘xdispF’ and peak search was done that picks up the diffraction spots on the image. The autoindexing of diffraction pattern was done using ‘DENZO’ (Otwinowski and Minor, 1997). The Bravis lattice was determined with the highest symmetry, which fits the data with a minimal metric tensor distortion index given that is normally below 1% for the correct lattice. The crystal and detector parameters were refined followed by reduction of diffraction data from each image into a file containing \( h, k, l \) indices, background and spot intensities and estimated error called as ‘.x’ file.

Scaling and merging of reflections from different images of the single crystal data set (.x files) was done by ‘SCALEPACK’ (Otwinowski and Minor, 1997) that was also used for the precise post refinement of crystal parameters. The output of this process is a reflection intensity file (.sca) whereas the log file contains important statistical information such as completion, chi square, Intensity (I) vs. background ratios (\( \sigma I \)), mosaicity, redundancy, reflection number, \( R_{\text{merge}} \) etc about the data set scaled. The above parameters were used to assess the quality of the data set (reviewed in Wlodawer et al., 2008).
Figure 2.4: X-ray diffraction pattern of crystals of Pab-NTD and K121M mutant complexes

Pab-NTD-L-Ser3AA complex  Pab-NTD-L-Thr3AA complex

K121M mutant-L-Ser3AA complex  K121M mutant-L-Thr3AA complex

Figure 2.4: X-ray diffraction pattern of crystals of Pab-NTD and K121M mutant with L-Ser3AA and L-Thr3AA. The insets show the diffraction limit of the given resolution.
2.2.6.6 Structure determination

X-ray crystallography is primarily used to determine the three dimensional structure (atomic arrangement in space) of the macromolecules. X-ray is used because of its shorter wavelength (in angstrom, $10^{-10}$ metre) so that information about the spatial arrangement of atoms is obtained. In macromolecular crystallography, crystal of macromolecules of interest is illuminated by x-rays and a pattern of diffracted x-rays are collected on a detector. Since diffracted x-rays cannot be focused by any optical lens, a mathematical strategy is used to obtain the structure of the macromolecule.

Structure determination of the macromolecule is all about “reversing the process of diffraction, and of getting back from the diffraction pattern to an image of the atomic arrangement” (Perutz, Noble Lecture, 1962). The diffraction spectra can be recombined to give us a representation of the x-ray scattering material in the crystal, the electron density. It is calculated by reverse Fourier transformation of the diffraction pattern.

$$p(x, y, z) = \frac{1}{V} \sum_{hk\ell} |F(h k \ell)| \exp[-2\pi i (hx + ky + \ell z) + i\alpha(h k \ell)]$$

where, $|F(h k \ell)|$ is the structure factor amplitude and $\alpha(h k \ell)$ is the phase of individual reflections.

The structure factor amplitudes of reflections were obtained from scaled and integrated intensities (.sca) in an absolute scale using Wilson’s plot by ‘Import Merged Data’ task from the CCP4 suite (CCP4, 1994) that uses the program TRUNCATE/cTRUNCATE. The output is ‘.mtz’ file in CCP4 format. Five percent of randomly selected reflections were kept aside for R-free calculation and these were not used further in structure solution or refinement.

The cell content analysis was done that provided an estimate of the number of molecules in the asymmetric unit and solvent content using Matthews’s coefficient ($V_M$) (Matthews, 1968). This information was used for searching defined number of molecules in the asymmetric unit during structure solution.
Chapter 2: Materials and Methods

The x-ray diffraction data sets measures only the intensities of reflections, however, the phase information of individual reflections is lost. This 'phase problem' in x-ray crystallography is resolved using one or more phase determination methods:

1. SIR, SIRAS, MIR, MIRAS (Single/Multiple Isomorphous Replacement with Anomalous Scattering)
2. MAD (Multiple wavelength Anomalous Dispersion)
3. SAD/SAS (Single wavelength Anomalous Dispersion/Scattering)
4. RIP, RIPAS (Radiation damage Induced Phasing with Anomalous Scattering)
5. MR (Molecular Replacement)
6. Direct methods

In this thesis work, structures have been solved using only molecular replacement as the structure of Pab-NTD was already known, which was used as search model.

2.2.6.6.1 Molecular replacement

Molecular replacement (MR) is used to determine the phases of an unknown structure when the structure of a homologous protein is known. A search model is generated from the known structure based on sequence similarity. This model is used for structure determination of the 'target' protein whose diffraction data is available. MR involves orienting and positioning of the model in the crystal lattice of the target protein. This is done by two major operations: rotation followed by translation for orienting and positioning of the model such that the equivalent points of electron density in each molecule come into coincidence.

If two position vectors $X_1$ and $X_2$ are equivalent points, then

$$X_2 = [R] X_1 + t$$

where, $[R]$ is a rotation matrix and 't' a vector defining a translation.
2.2.6.6.1.1 Patterson function

It is important to understand the Patterson function (Patterson, 1935) that is used extensively for rotation and translation search in MR using MOLREP program in CCP4 Suite. A Patterson function is a convolution function representing the vectors between the atoms in a unit cell and can be calculated directly from a set of x-ray diffraction intensities without the phase information or from the three-dimensional coordinates of a molecule. So, the Patterson function can be calculated for the model from the atomic coordinates and that for the 'target' molecule from the x-ray diffraction intensities. Patterson function has two types of vectors: self-Patterson vectors and cross-Patterson vectors. Sets of peaks representing intra-molecular vectors are called self-Patterson vectors whereas intermolecular vectors between two molecules related by crystallographic symmetry are termed as cross-Patterson vectors.

2.2.6.6.1.2 Rotation function

The self-Patterson vectors are used in rotation search to determine the orientation of a search model relative to 'target' molecules in crystal. These vectors lie within the largest intermolecular distance 'r' from the origin of the Patterson function. Therefore, the product function between the self-Patterson functions of model and target within a sphere of radius 'r' in the crystal is calculated. This gives the cross-Rotation Function that is used to determine the relative orientations of the molecules. At a certain orientation, the model self-Patterson vectors will be superimposed on the target self-Patterson vectors giving a higher value when compared to all other product functions (Rossman and Blow, 1962). The product function R is given by an equation,

\[ R = \int V P_2(X_2) P_1(X_1) \, dV \]

where, R is the product of 'target' Patterson \( P_1 \) and the rotated model Patterson \( P_2 \) within the volume V.

The magnitude of the Rotation Function is plotted in a three-dimensional space with the three angular rotations as the coordinates. The rotation solution in the case of
MOLREP (Vagin and Teplyakov, 1997) is displayed in both angle rotation systems: Euler and Polar angles. Euler angles are alpha, beta and gamma whereas the polar angles are theta, phi and chi. Statistical parameters such as Rotation Function (Rf) and Rf/σ are used for selecting the correct Rotation Function that has the highest value and differs significantly from the next best solution. If two or more molecules are present in the asymmetric unit, a self-Rotation Function, where target Patterson is compared with itself, is used to determine the orientation and the angle of rotation of non-crystallographic symmetry axis relating one molecule with another.

2.2.6.6.1.3 Translation function

Once the correct orientation of the model in the unit cell is known, Translation Function is calculated using the cross-Patterson vectors of model and target. The Translation Function T(t) (Rossmann et al., 1964; Rossmann and Blow, 1967) is calculated using the following equation,

\[ T(t) = \int_0^\infty P_{1,2}(u, t) P(u) \, du \]

where \( T(t) \) is the Translation Function, \( P_{1,2}(u, t) \) is the cross-Patterson between two symmetry related model molecules and \( P(u) \) is observed cross-Patterson vectors.

The correct Translation Function has a higher correlations or Translation function contrast (TFcnt) compared to other and is used for selecting the best translation solution. TF/σ and R-factor are also indicative of correct translation search. The translation search in three dimensions is defined in terms of fractional coordinates along three axes.

2.2.6.6.2 Molecular replacement in the case of Pab-NTD complexes

The structure was solved using the coordinates of Pab-NTD (PDB id: 1Y2Q) by molecular replacement using MOLREP in CCP4 suite. The further verification of MR solution was done by generating the symmetry related molecules and observed for clashes in Ca trace (Figure 2.5). The correct solution had no clashes with its symmetry.
**Figure 2.5: Asymmetric unit and unit cells of Pab-NTD complexes**

(A): Asymmetric unit (A.U.) [Pab-NTD dimer in yellow] and Unit Cell containing 4 such dimers in space group P2₁,2₁,2₁.

(B): A.U. [Pab-NTD monomer in yellow] and Unit Cell containing 8 such monomers in space group i222. The monomer (yellow) lies on special position such that the dimeric axis of Pab-NTD matches with that of two-fold symmetry axis to generate the second monomer of dimer (red).

1. Unit cell contains 2 dimers(yellow-red & green-blue) and 4 monomers (gray, cyan and two in magenta).

2. The two monomers in magenta can be translated by 1 cell dimension to form dimers with gray & cyan. Thus 4 dimers make up the unit cell.
related molecules. All the complexes mentioned in this thesis belonged to P2_12_1 space group except one in I222 (Figure 2.5).

2.2.6.7 Structural refinement

Once the initial phase from MR was obtained, phase improvement was carried out by refining the model by introducing some geometrical restraints and energy parameters for all the data sets. Structural refinement is the process of adjusting the model i.e. its three positional parameters (x, y, z) and an isotropic temperature factor (B) for all non-hydrogen atoms to find a closer agreement between the observed and the calculated structure factors. Refinements of all structures mentioned in this work were done using 'CNS' (Brunger et al., 1998) while model building (fitting model into electron density maps and torsion angle adjustments) was carried out using 'O' (Jones et al., 1991).

Solutions obtained from MR were used to generate the ‘molecular topology file’ (mtf) and the ‘pdb’ file using the ‘CNS’ task file ‘generate.inp’. CNS uses the ‘mtf’ file for defining restraints in subsequent refinement procedures. Rigid body refinement was performed using the ‘CNS’ task file ‘rigid.inp’ that considers the full protein as a rigid entity. It refines the position of the molecule in the unit cell. The rigid body refinement was performed in lower resolution range of 25.0 Å to 4.0 Å and 5% of randomly chosen reflections were set aside for calculating R-free. The convergence (lowering) of R-factor after rigid body refinement also indicates correct MR solutions. The rigid body refinement was followed by energy minimization using the ‘CNS’ task file “anneal.inp” that involves the use of high temperature molecular dynamics and cooling in search for stable peptide conformations. In annealing, the protein molecules were heated to 5000 K in the case of high resolution structures or 3000 K / 2000 K in the case of structure below 2.6 Å and then slowly cooled to 300 K.

After every cycle of refinement, double difference (2Fo-Fc) and difference (Fo-Fc) Fourier maps were calculated from ‘CNS’ task file “model_map.inp”. 2Fo-Fc map is generally contoured at 1.0 σ and used for complete model building, whereas
Fo-Fc map is contoured at 2.0 σ and mainly used in identification of ligands, solvent, or any other feature that is missing in the search model, for example mutated amino acids, extra residues and so on. σ refers to the standard deviation from the mean electron density and is represented as the number of electrons/Å^3.

The model was inspected and manually adjusted (model building) using (2Fo-Fc) and (Fo-Fc) Fourier maps after each round of refinement in ‘O’. Model built in ‘O’ were subsequently subjected to a few cycles of conjugate gradient minimization and restrained, individual B-factor refinement using the ‘CNS’ task files ‘minimization.inp’ and ‘bindividual.inp’, respectively, in the next rounds of refinement. Zero σ cutoff criteria were used on amplitudes for the rejection of reflections during refinement.

Clear electron densities for the ligands were observed in both 2Fo-Fc as well as Fo-Fc electron density maps in the case of all Pab-NTD complexes. The ligand was included in the refinement after two rounds of initial refinement cycles of protein alone. The parameters for refinement of ligands were obtained from PRODRG server (Schuettelkopf and van Aalten, 2004). Using Fo-Fc map at around 2.5 σ, water molecules were added to the structures using the task file ‘water_pick.inp’. The added water molecules were verified using both 2Fo-Fc and Fo-Fc electron density maps.

Simulated annealing omit maps were also computed to validate the electron density of ligand and water molecules at the editing site. The R/R_free values were used for validation of the refinement procedure. The R-factor is a measure of the agreement between the refined model and the experimental X-ray diffraction data. The R/R_free values are generally lower for high resolution data sets compared to low resolution data sets. Generally, these values are below 30 and the two R-factors should not differ by a large value. R_free value is important for cross validation of refinement procedure to avoid overfitting that takes into account 5% of reflections kept aside and not used at all during refinement.
2.2.6.8 Validation of structure

All the structures were validated using the program PROCHECK (Laskowski et al., 1993). The stereochemical parameters of the refined model were compared with those derived from well-refined structures at a given resolution. A Ramachandran Plot (Ramachandran et al., 1963; 1974) with number of residues in allowed, generously allowed and disallowed region were checked along with the analysis of all bond length and bond angles in the refined structure.

2.2.6.9 Structural analysis

The structures were analyzed manually using graphical options in ‘O’ program (Jones et al., 1991). Superposition of structures was done either using the lsq option in ‘O’ or by pair wise structural overlaps using DaliLite of EBI (Holm and Park, 2000). All structural figures were produced using SETOR (Evans, 1993), PyMol (DeLano, 2002) or GRASP (Nicholls et al., 1991).

2.2.6.9.1 Structural analysis of Pab-NTD complexes

Once the complete structure was solved and refined, it was subjected to extensive structural analysis. All the crystallographic parameters of data collection and refinement were verified again. The ‘experimental unbiased’ 2Fo-Fc and Fo-Fc maps were loaded on ‘O’ and presence/absence of ligands or water molecules at the editing site were confirmed. Next, the refined (biased) 2Fo-Fc and Fo-Fc maps were loaded on ‘O’ and any unaccounted positive density near the ligand/water molecule as well as negative density in ‘difference’ Fo-Fc maps were checked. The density for the residues lining the editing pocket was also scrutinized using both 2Fo-Fc and Fo-Fc maps. The B-factors of the ligand were compared with the residues interacting with it and found to be comparable. Simulated annealing omit maps were also calculated to validate the electron density of ligand and water molecules at the editing site.
2.2.7 Deacylation assay

Charged (aminoacylated) tRNAs are the substrates for the deacylation assays. tRNAs can be charged in vitro either by respective aaRS or can be directly purified from the bacterial cells (in vivo method). Using the in vivo method, correctly charged tRNAs are mainly purified. In vitro methods are preferred for studying mischarged tRNAs.

For in vitro method, the required tRNA is expressed and purified. Moreover, particular aaRS catalytic domain specific for the tRNA is required. The aaRS used for aminoacylation is either without the editing domain or with a mutation in the editing site so that the mischarged tRNAs are not deacylated. $^{14}$C-amino acids are used for the assay that labels the mischarged tRNA when $^{14}$C-amino acid is attached to it. The mischarged tRNAs are then purified so that there is no contamination of aaRS catalytic domain in the deacylation assay. In the deacylation assay, the reaction is started by adding mischarged tRNAs. Aliquots at different time points are obtained and precipitated by trichloroacetic acid (TCA) and the amount of remaining mischarged-tRNAs in the deacylation medium can be measured by liquid scintillation counting.

The deacylation experiments were performed in collaboration by Dr. A-C. Dock-Bregeon, IGBMC, France. For the sake of presenting a complete picture of editing mechanism, these experiments are mentioned in the thesis. The deacylation assays were performed with E. coli tRNA$^{Thr}$, produced in vivo, and $^{14}$C-L-serylated with the truncated version of E. coli ThrRS, lacking the editing domain (Sankaranarayanan et al., 1999, Dock-Bregeon et al., 2000). $^{14}$C-L-Ser-tRNA$^{Thr}$ was recovered by precipitation after phenol extraction of the aminoacylation medium. For the deacylation assay, about 3 $\mu$M of $^{14}$C-L-Ser-tRNA$^{Thr}$ was incubated with 16 $\mu$M of Pab-NTD or mutant in a reaction buffer containing 50 mM Heps-NaOH (pH 7.2), 100 mM KCl and 5 mM MgCl$_2$, at 37°C. Aliquots were TCA-precipitated, and the remaining $^{14}$C-L-Ser-tRNA$^{Thr}$ was measured by liquid scintillation counting. Each experiment was done in triplicate except for K121S mutant, which was done in duplicate. Similar procedure was followed for $^{14}$C-threonylation as well as for deacylation assay using $^{14}$C-L-Thr-tRNA$^{Thr}$. 