Chapter-2
MATERIALS AND METHODS
2.0 INTRODUCTION

In this chapter detail of the materials, methods and strategies used for purification, biochemical characterization, fluorescence studies, crystallization, X-ray data collection, data processing, structure solution and refinement of C-phycocyanins (C-PC) from cyanobacteria and xylanase from alkalophilic thermophilic bacteria (ATBXYL-C) are described. The programs used for analysis and comparison of protein structures have been detailed. Purification was carried out using different biochemical techniques such as chromatography to obtain pure samples of C-PCs and ATBXYL-C and the analysis of samples was using spectroscopic and other techniques such as electrophoresis. For crystallizing C-PCs and ATBXYL-C, mainly hanging-drop vapor-diffusion method was used as described in the following chapters. X-ray diffraction data were collected on Raxis IV++ image plate detector mounted on Rigaku rotating anode X-ray generator and processed using DENZO and SCALEPACK programs (Otwinowski, 1993; Otwinowski & Minor, 1997). Molecular Replacement (MR) method implemented in AMoRe program (Navaza, 1994; Navaza & Saludjian, 1997; Navaza, 2001) and PHASER were used to solve the structures of C-PCs and ATBXYL-C. The program REFMAC (Murshudov et al., 1997) was used for refinement and QUANTA (Accelrys) for graphics display and model fitting of the structures. Programs from CCP4 suite (Collaborative Computational Project, Number 4, 1994) were used for other calculations.

2.1 MATERIALS

Fermentation and purification of C-PCs and ATBXYL-C from the crude extract to the homogeneous protein solution involved chemicals such as Trizma, di-potassium hydrogen phosphate, potassium hydrogen phosphate, ethylenediaminetetraacetic acid (EDTA), ammonium sulfate (AS), sodium acetate, glycerol, β-mercapto-ethanol, bromophenol-blue (BPB), Acrylamide, N, N’-methylene bisacrylamide, sodium dodecyl sulfate (SDS), acetic acid, methanol, TEMED (N, N, N’, N’- Tetramethylethylenediamine), Ammoniumpersulfate (APS), CM-sephadex, DEAE–sepharose CL-6B, molecular weight marker kits for SDS-PAGE etc, were purchased from Sigma Chemicals, St. Louis, USA.
The slab gel electrophoresis unit was purchased from Tarsons, India. The following chemicals were used extensively in the crystallization trials: sodium cacodylate buffer, 3-(N-morpholino) propanesulfonic acid MOPS buffer, N-(2-hydroxyethyl)-piperazine-N’-2-ethanesulfonic acid (HEPES) buffer, tri-sodium citrate, lithium sulphate, polyethylene glycol (PEG 20K, 8K, 4K, 1K), 2-methylpentane-2, 4-diol (MPD) and isopropanal. The additives used in crystallization trials to improve diffraction quality of crystals were ethylene glycol, glycerol, sucrose and PEG-400. Multi-well trays used in crystallization experiments were purchased from Becton Dickson and Company or Falcon. All chemicals used were of analytical grade. Silicon graphics Octane workstation was used for graphics display and computations.

2.2 Fermentation of Cyanobacteria

The marine cyanobacteria *Phormidium* and *Lyngbya* spp. were isolated from the rocky surface near the sea coast of Gujarat (latitude 21° 38’ N and longitude 69°37’), in the west coast of India. These organisms were grown in batch cultures in the standard artificial seawater medium ASN III at pH 7.5 (Rippka *et al.*, 1979) and temperature 20 ± 2 °C with optimum light intensity of 60 lE m⁻² s⁻¹ provided by cool-white fluorescent tubes with a dark: light cycle of 12:12 h. The freshwater cyanobacterium *Spirulina* sp. was grown under same conditions except that the medium used was Zarouk medium (Chen *et al.*, 1996) at pH 10.

2.2.1 Extraction of phycobiliproteins: Cyanobacterial cells were harvested after 7-10 days (fresh water) and 15-20 days (marine) of incubation under laboratory-controlled conditions (temperature, pH, and light) by centrifugation at 8,000g for 30 min. The harvested cell mass was washed twice with distilled water and freeze-dried. One gram of freeze-dried cell mass was suspended in 100 ml of sodium phosphate buffer (0.1 M, pH 7.0, containing 1 mM sodium azide). Further extraction procedure of phycobiliproteins consisted of repeatedly freezing the sample at -20°C and thawing to room temperature in the dark. The mixture was subsequently centrifuged at 10,000g for 30 min at 4 °C and phycobiliprotein containing clear supernatant was collected.
2.2.2 Estimation of phycobiliproteins: The absorbance of the supernatant was measured in a VARIAN CARY 50 BIO Scan UV–Visible NIR spectrophotometer at wavelengths 620, 652, and 562 nm. Concentration of C-PCs, A-PC and PE was calculated using the following equations (Bennett et al., 1973):

\[
\text{C-PCs (mg/ml)} = \frac{[A_{620} - 0.474 \times (A_{652})]}{5.34}.
\]

\[
\text{A-PC (mg/ml)} = \frac{[A_{652} - 0.208 \times (A_{620})]}{5.09}.
\]

\[
\text{PE (mg/ml)} = \frac{[A_{562} - 2.41 \times (PC) - 0.849 \times (APC)]}{9.62}.
\]

2.2.3 Purification of C-PCs: The cyanobacteria Phormidium, Lyngbya, and Spirulina spp., produce mainly C-PCs (with absorption maxima at 620 nm) and APC (with shoulder at 652 nm). The data of the purification of C-PCs from Phormidium, Lyngbya and Spirulina spp. are summarized in Table 3.3 (Chapter 3). All purification steps were performed in the dark at 15–20°C. The clear supernatant of C-PCs was fractionated by precipitation with solid AS first at 25% and then at 50% saturation. The precipitate formed at 25% saturated AS was discarded and the supernatant was then brought to 50% AS saturation and allowed to stand for 4 h at 4°C. The precipitated protein containing mainly C-PCs was collected by centrifugation at 10,000g for 30min at 4°C, resuspended in acetate buffer (0.1 M, pH 4.5) and again centrifuged at 10,000g for 30 min at 4°C to precipitate out the basic linker polypeptides. The precipitate was discarded and the supernatant was again brought to 50% AS saturation and allowed to stand for 4 h at 4°C prior to centrifugation at 10,000g for 30 min at 4°C. The precipitated C-PC was dissolved in 5 ml of sodium phosphate buffer (0.005 M, pH 7.0) and dialyzed against the same buffer overnight at 4 °C. This dialyzed solution was chromatographed on a DEAE–sepharose CL-6B column (1.5x15 cm.) equilibrated with the same buffer. The column was eluted with a linear increasing ionic concentration gradient using NaCl solution (0–0.25 M) at a flow rate of 0.5 ml/min. The C-PC eluted between NaCl concentrations of 0.10 and 0.20 M and was collected as 2 ml fractions. The absorption spectrum from 250 to 800 nm was recorded to monitor the purity of the fractions. Those fractions having ratio of A620/A280 > 4.0 were pooled together and brought to 50% saturation of AS. The
resultant precipitate of C-PC was dissolved in minimum volume of sodium-phosphate buffer (0.005 M, pH 7.0), dialyzed against water at 4°C, and freeze-dried into powder for storage.

2.2.4 Spectroscopic measurements: Absorption spectra of C-PC samples using UV-Visible radiation were recorded on a VARIAN CARY 50 BIO Scan UV–Vis, NIR spectrophotometer. The protein concentration for these experiments was in the range 0.1–0.3 mg/ml. In the region of visible radiation, C-PC and APC show maximum absorption at wavelengths of 620 and 652 nm, respectively. The purity was evaluated based on the two absorption ratios A620/A280 and A620/A652.

2.2.5 SDS–PAGE: Electrophoresis was carried out according to Laemmli (1990) using 15% gel containing 0.1% (w/v) SDS. Samples were pre-incubated with 2% (w/v) SDS, 10% (v/v) glycerol, 4.5% (v/v) β-mercaptoethanol, 0.025% (w/v) bromophenol blue and 60mM Tris buffer (pH 6.8), for about 4-10 min at 95°C. Gels were run at room temperature and developed using Coomassie Blue R-250. Electrophore sample showed only two bands corresponding to α and β subunits of C-PCs (Fig. 3.1 Chapter 3), indicating that the linker polypeptides were effectively removed during the purification procedure. Thus, the SDS-PAGE analysis clearly indicated that the C-PC preparations were pure and free of linker peptides. The molecular weights of subunits were determined by running the low molecular weight markers (Sigma SDS) along with the sample.

2.3 Purification of xylanase C from Bacillus sp. (NCL 86-6-10)

2.3.1 Enzyme production: The alkalophilic thermophilic Bacillus (NCL 87-6-10) was isolated from soil samples of Calicut in South India rich in decomposing coconut detritus. Fermentation of Bacillus sp. was carried out in wheat bran-yeast extract medium containing wheat bran (3% w/v), tryptone (1 to 3% w/v) and tween-80 (0.5 to 1%). The xylanase activity ranged from 130 to 170 IU/ml after 48 hours of growth (Balakrishnan. et al., 1992, 2001).
2.3.2 Enzyme purification: The culture filtrate (500ml) was brought to 60% AS saturation by the addition of 215 g of the salt and the resultant solution was stirred at 4 °C for 3-4 h. The precipitate was collected by centrifugation at 10,000 x g for 30 min and dissolved in minimum volume of 20 mM potassium phosphate buffer pH 7.0. The extract was extensively dialyzed against the same buffer. The dialyzed enzyme solution was concentrated to ~5ml in Amicon ultra filtration unit using 10,000 cut off pore size membrane. The solution was loaded on to a CM-sephadex column (28x4 cm) pre-equilibrated with 20 mM potassium phosphate buffer pH 7.0. The bound enzyme was eluted with 0-0.8 M KCl gradient. The protein eluted at around 0.6 M KCl. The fractions showing ATBXYL-C activity were pooled and concentrated by Amicon ultra filtration. The purity of the enzyme was analysed by SDS-PAGE and the molecular weight was determined by mass spectrometry.

2.3.3 Enzyme Assay: ATBXYL-C activity was assayed by incubating 0.5 ml of enzyme with 0.5 ml xylan solution (substrate) and at 60 °C for 30 min. The reaction was stopped by adding 1 ml DNSA. The arrested solution was boiled for 5 min, diluted with 10 ml distilled water and the absorbance read at 540 nm. One enzyme unit is defined as the amount of enzyme required to liberate 1 µmol of xylose equivalent/ min under the assay conditions.

2.4 Chemical modification of active site residues

2.4.1 Modification of tryptophan: 1.0 ml of the enzyme solution (8-10 µM) was titrated with freshly prepared N-bromosuccinimide (NBS) (5 mM) at pH 6.0. The reaction was monitored using absorbance at 280 nm. The reagent was added till the absorbance at 280 nm started increasing. After each addition an aliquot 10 µl of the enzyme was removed and diluted to 200 µl with 50 mM tris-HCl buffer at pH 8.0 and then checked for residual enzyme activity under standard assay conditions. The number of Trp residues modified was quantified as described by Spande and Witkof (1967).
2.4.2 Modification of carboxyl groups: 1 ml of 4.0 µM ATBXYL-C solution was treated with Woodward’s reagent K (15-30 mM) at pH 7.0. Aliquots of the enzyme were removed at specific time intervals and assayed for residual enzyme activity. Number of modified carboxylate residues was determined spectrophotometrically after removing the excess reagent by passing it through sephadex G-25 as described by Sinha and Brewer (1985).

2.4.3 Modification of arginine: 5 µM enzyme was treated with phenylglyoxal (15-60 mM) in 50 mM tris-HCl buffer pH 8.0. Aliquots of the modified enzyme were removed after 5, 10, 15 and 20 min and assayed for residual enzyme activity.

2.4.4 Substrate protection studies: The effect of the presence of substrate on inactivation during chemical modification reactions was studied. The enzyme was incubated in the presence of xylan and modifier together and then the residual activity was determined. The enzyme was incubated with same amount of xylan alone as control.

2.5 Fluorescence spectroscopic studies

Intrinsic fluorescence was measured using a Perkin-Elmer LS50B fluorescence spectrophotometer connected to a Julabo F25 water bath. Background emission due to buffer solution and denaturants was measured and subtracted from the protein spectra. The protein solution (4.1 µM) was excited at 280 nm and the emission recorded in the range of wavelengths 300-400 nm at 30 °C. The slit-width on both the excitation and emission were set at 7 nm and the spectra were obtained at 100 nm/min.

2.5.1 Thermal denaturation: Effect of temperature on the fluorescence behavior of ATBXYL-C was studied using a thermostatic cuvette holder connected to an external constant temperature circulation water bath. The protein sample was incubated for 15 min at specified temperature before measuring the effect. The aliquots were removed at each specified condition for measuring the activity under standard assay conditions. For
renaturation experiments the samples were cooled back to 30 °C and left for one hour before recording the spectra. Fluorescence spectra were recorded as described above.

2.5.2 Light scattering: Fluorescence scattering at 400 nm was measured using a fluorimeter to detect and study protein aggregation under thermal denaturation conditions. The same experimental setup used for fluorescence studies was used for this purpose also. Both the excitation and the emission wavelengths were set at 400 nm. The excitation and emission slit widths were set at 5 and 2.5 nm, respectively. Scattering was observed for 120 sec.

2.5.3 Effect of pH: ATBXYL-C samples (4.1 µM) were incubated in an appropriate buffer for 4 h at 30 °C over a pH range of 1-12. The following buffers were used for these studies: 100 mM Glycine-HCl buffer for pH 1-3 and glycine-NaOH for pH 10-12. For other pH ranges, 100 mM acetate buffer (pH 4-5), 100 mM phosphate buffer (pH 6-7) and 100 mM Tris-HCl buffer (pH 8-9) were used. Fluorescence spectra were recorded as described above.

2.6.4 ANS-binding assay: The intermediate state of denatured and native ATBXYL-C at different denaturation conditions was analysed by hydrophobic dye (ANS) binding method. The final ANS concentration used was 40 µM, excitation wavelength (λ) 375 nm and total fluorescence emission was monitored between 400-550 nm. Reference spectrum with ANS at each buffer of respective pH and Gdn-HCl was subtracted from the spectrum of the sample.

2.6.5 Guanidium hydrochloride mediated unfolding: Protein samples (4.0 µM) were incubated with 1-6 M denaturant solution at pH 7 for 4 h to attain the equilibrium. After recording the scans, aliquots were removed from the sample to check for activity. Refolding experiments were conducted on ten fold diluted samples to reduce the Gdn-HCl concentration and incubated at 30 °C for 1h before recording the spectra.
2.6 Titration of xylo-oligosaccharides against xylanase

The fluorescence quenching measurements were made by titrating 2 ml of the enzyme (4.0 µM) solution with the saccharide solution (100 mM, 5 µl each time) followed by monitoring the change in fluorescence at 338 nm. The fluorescence of the buffer and saccharide solution were measured at appropriate wavelengths and these values were used to correct the observed fluorescence. Corrections were also made to compensate for the dilution upon addition of sugar and at the highest concentration of the saccharide, the volume change was less than 0.1% of the enzyme solution. The temperature of the enzyme and saccharide solutions was maintained at 30 ºC. The relative fluorescence intensity of the ATBXYL-C saturated with saccharide (F∞) was obtained from the experimental data by plotting Fo/(Fo-F) versus 1/[S] and extrapolating to y-axis, where Fo is the fluorescence intensity of the enzyme alone and F is the fluorescence intensity of the enzyme at a saccharide concentration [S]. log {(Fo-F)/(F-F∞)} was plotted against log[S] and the association constant (Ka) was determined from the plot by assuming the relation that the pKa of the complex equals the value of [S] when log {(Fo-F)/(F-F∞)} = 0.

2.7 Circular dichroism (CD) spectroscopy

CD spectra were recorded on a J-715 spectro-polarimeter with a PTC343 Peltier unit (Jasco, Tokyo, Japan) at 25 ºC in quartz cuvette. All spectra were corrected for buffer contributions and converted to mean residue weight ellipticity. Each CD spectrum was accumulated from six scans at 100 nm/min with a 1 nm slit width and a time constant of 1s for a nominal resolution of 1 nm. Far UV CD spectra were collected in the range of wavelengths 200 - 250 nm using a cell of path length 0.1 cm for monitoring secondary structure. The tertiary structure of the enzyme was monitored with near UV CD spectra in the wavelength 250 - 300 nm using path length 1 cm. Protein concentrations were 5 µM
40 µM for far-UV and near-UV CD spectra collection. The protein was equilibrated in buffers of pH 1.0 and 7.0 six hours before data collection. The CD spectrum of the native ATBXYL-C was analyzed using the algorithm in CDPro program in order to determine the percentage of secondary structures.

2.8 Amino acid composition of xylanase

Amino acid composition of ATBXYL-C was analysed with AccQ-Fluor (Waters Corporation). Salt-free lyophilized ATBXYL-C (50µg) was hydrolyzed by using 6N HCl, in vacuum-sealed tubes at 110 °C for 24 h. Excess acid was removed by evaporation in vacuum at room temperature. The sample was then derivatized with 6-aminoquinolyl-N-hydroxysuccinimiyyl carbamate (AQC) in borate buffer at pH 9. The hydrolysate (10 picomoles) was loaded on AccQ-Tag column equipped with a fluorescence detector.

2.9 CRYSTALLOGRAPHIC ANALYSIS

2.9.1 Crystallization: The first step considered crucial in protein structure determination is the growth of diffraction quality single crystals. In the absence of any single concrete theory behind the mechanism of crystallization, we have treated the protein crystallization as a trial and error procedure invoking experience and crystallization reports as guiding principles. It is accepted that the presence of impurities, ionic strength, pH, temperature, precipitating agent and several unspecified factors play role in crystallization process.

Crystallization is known to lower the free energy of proteins by ~3-6 kcal/mole relative to the solution state (Drenth & Haas, 1998). The general processes by which substances crystallize are similar for molecules of both microscopic (salts and small organics) and macroscopic (proteins, DNA, RNA) dimensions. Crystallization is one of several means by which a metastable supersaturated solution can reach a stable lower energy state by a reduction of solute concentration. There are three stages of crystallization common to all systems: nucleation, growth, and cessation of growth.
Nucleation is the process by which molecules or noncrystalline aggregates (dimers, trimers, etc.), which are free in solution, come together in such a way as to produce a thermodynamically stable aggregate with a repeating lattice. The formation of crystalline aggregates from supersaturated solutions does not however necessitate the formation of macroscopic crystals. Instead, the aggregate must first exceed a specific size (the critical size) defined by the competition of the ratio of the surface area of the aggregate to its volume (Feher & Kam, 1985; Boistelle & Astier, 1988). Once the critical size is exceeded, the aggregate becomes a supercritical nucleus capable of further growth. If the nucleus decreases in size so that it is smaller than the critical size, spontaneous dissolution will occur. The process of formation of nonspecific aggregates and noncrystalline precipitation from a supersaturated solution does not involve the competition between surface area and volume (n-mers add to the aggregate chain in a head-to-tail fashion forming a linear arrangement), and thus generally occurs on a much faster time scale than crystallization. The degree to which nucleation occurs is determined by the degree of supersaturation of the solutes in the solution. The extent of supersaturation is in turn related to the overall solubility of the potentially crystallizing molecule. Higher solubility allows for a greater number of diffusional collisions. Thus, higher degrees of supersaturation produce more stable aggregates (due to higher probability of collision of diffusing molecules) and therefore increase the likelihood of the formation of stable nuclei. In the case of a finite number of solute molecules, this condition generally results in the production of a large number of small crystals. At lower solute concentrations the formation of individual stable nuclei increases in rarity, thus favoring the formation of single crystals. Crystal growth generally starts at solute concentrations sufficient for nucleation to occur, and continues at concentrations beneath the nucleation threshold. The rate of growth is determined by a combination of the nature of the growing crystal surface and the diffusional rate.

As noted earlier, molecules crystallize from metastable supersaturated solutions as a means of lowering the overall solution free energy. Chemical precipitants are by and far the most widely used method of achieving supersaturation of macromolecules in order to induce crystallization. In general, the main influence of these compounds is on the
solvent (e.g. bulk water) rather than on the solute (the protein), with the notable exception of dye precipitants. For crystallization of proteins, the major classes of precipitants may be divided into six categories: salts, high molecular weight straight chain polymers (e.g. PEG), MPD, organic solvents, sulfonic dyes, and deionized water (Arakawa & Timasheff, 1985). Although the following discussions of the individual traits of these six categories use proteins as examples, most of these precipitants are applicable to other macromolecules including RNA, DNA, and polysaccharides.

Among various crystallization techniques known, hanging-drop vapor-diffusion method is widely adopted and has produced more crystallized proteins than all other methods combined. This method is simple, consumes less protein and it is easy to monitor the progress of crystallization. In a typical experimental set up using 24-well multiwell trays, 1-2 µl of protein solution was placed on a siliconized cover slip, mixed with 1-2 µl of the precipitant solution and allowed to slowly equilibrate against 500-1000 µl of reservoir solution of the precipitant. The concentrated solutions of pure C-PCs and ATBXYL-C were prepared in milli-Q water or buffers before setting up crystallization. The concentrations of proteins were estimated as described by Lowry et al., (1951). During initial crystallization trials it was found that in some cases the presence of certain salts (sodium formate in case monoclinic crystals of C-PCs and tri-sodium citrate in case of xylanase) was essential for growing crystals. Different precipitants such as ammonium sulfate, PEGs of different molecular weights, 2-methylpentane-2, 4-diol (MPD) and isopropanol were tried in the crystallization experiments and some were found very effective.

2. 9.2 Crystallizing proteins with PEG

Polymers: The use of high molecular weight linear polymers as precipitating agents was pioneered by Polson and coworkers who tried a variety of polymers including polyethylene glycol, dextran, polyvinyl alcohol, and polyvinyl pyrrolidone (Polson et al., 1964). Of these, polyethylene glycol (PEG) was found to be the most effective both by way of precipitating ability and cost effectiveness. PEGs are produced in a variety of
molecular weights, ranging from 200 (~3 monomers) to in excess of 1 million (~15000 monomers), and as mono- and di- methyl ethers. Like salts, PEGs compete with protein solutes for water and exert excluded volume effects (which vary according to the length of the polymer). However, unlike salts, PEGs decrease the effective dielectric constant of the solution, which increases the effective distance over which protein electrostatic effects occur. Solutions of polyethylene glycols have mean electron densities roughly equivalent to water and do not generally interact in a deleterious manner with heavy atom compounds, thus making them particularly well suited for macromolecular crystallization. PEGs with molecular weights less than 1000 are typically liquids and are generally used at concentrations above 40% v/v. PEGs with molecular weights above 1000 are generally solids and are used in the 5-50% w/v concentration range. All PEG solutions should be made with the inclusion of ~0.1% Na azide (toxic substance) to prevent bacterial growth. Also, buffering of high concentration (40%) of PEG solution with Na citrate at concentrations above 100 mM tends to cause the formation of phase transitions and color changes in the PEG solution, which suggests some form of reaction (probably cross linking by the citrate), and thus should be avoided.

2.9.3 Crystallization trials using (2-methyl-2, 4 pentane diol) (MPD) and Organic solvents (Isopropanol)

MPD is a small polyalcohol (2-methyl-2,4 pentane diol) which has properties midway between those of low molecular weight PEGs and organic solvents. MPD functions as a precipitant by a combination of activities, including competition for water, hydrophobic exclusion of protein solutes, lowering the dielectric constant of solution, and detergent-like effects. It is generally used in concentrations in excess of 40% v/v with water/buffer, and tends to cause phase transitions in the form of coacervate droplets, which are enriched in protein concentration (synonymous with those in Ray and Bracker, 1986).

Historically, organic solvents have typically been used as precipitants for protein crystallization as much by chance as by design. Crystallization by exposure to organic solvents is occasionally seen during protein purification, typically in the presence of
common solvents such as ethanol, methanol, acetone, isopropanol, DMSO, or tert-butanol (McPherson, 1990). Due to their hydrophobic nature, organic solvents cause phase transitions similar to those formed in the presence of MPD and lower the bulk dielectric of the solvent.

2.9.4 X-ray diffraction data collection

After growing protein crystals of suitable size and quality the immediate step in an X-ray diffraction experiment is to characterize the crystals and measure intensities of Bragg reflections from the crystals. Data collection is best performed as a highly interactive process. Protein crystals diffract X-rays with lesser intensity and lower resolution than the small molecule crystals. Thus, the protein data collection requires a high intensity X-ray source and a high sensitivity X-ray detector. In the last decade the combination of powerful tools such as synchrotron sources, image plate and CCD detectors and various softwares developed for collecting and processing diffraction data, increased computational and graphics power of workstations, all these have transformed protein crystallography into a powerful tool for structural study in the era of structural genomics (Beauchamp & Isaacs, 1999; Blundell et al., 2002). The X-ray storage-phosphor image plate (IP) is considered to be the most suitable detector for acquiring protein data with a home source, as well at synchrotron radiation facility. IP is a very sensitive detector, has wider dynamic range, and has high spatial resolution and high count rate capacity, which are the fundamental requirements of an X-ray area detector (Amemiya, 1997).

The IP detectors have enabled protein crystallographers to obtain very accurate data sets with laboratory X-ray sources. The plate can be erased by exposure to intense white radiation and can be used repeatedly. The size of commonly used IP is 300 mm and can be controlled by software running on a computer through an interface. Each diffraction image has a size of 18 MB, which can be stored in the computer. The crystal alignment was done through a CCD camera - TV monitor assembly. The processes of exposure, data collection, readout and storage of data are carried out automatically. All
the diffraction data of crystals described in the thesis have been collected at NCL, Pune, India. High intensity X-ray radiation usually damages protein crystals during longer exposure time for data collection. Now it is routine to collect the macromolecular data at cryogenic temperatures. This technique of flash cooling protein crystals at liquid nitrogen temperatures and collecting data offers several benefits. Some of the advantages include reducing radiation damage of the crystal on exposure to X-rays, improving the limits of resolution, decreasing thermal vibrations of the atoms, allowing storage and reuse of crystals, and helping to overcome the scaling problem by enabling collection and scaling of entire data from one crystal only (Garman & Schneider, 1997). In all the experiments described here, crystals were frozen under the cryostream of liquid nitrogen (~120 K) and different cryoprotectants like glycerol, PEG 400, MPD, 1, 2, 6-Hexanetri-ol (HXT) in the range 20 to 30% were used. Incase of C-phycocyanin PEG 400 has worked as cryoprotectant for crystals of this protein from different species in all its crystal forms (monoclinic and hexagonal).

During data acquisition the crystals were oscillated about an axis perpendicular to the X-ray beam, with a chosen, relatively small angle of oscillation, usually 0.5 to 1° per frame. Crystal to detector distance has been chosen based on the longest unit cell dimension, mosaic spread, etc., so that the diffraction spots are well resolved. The exposure time depends on the quality of crystal and oscillation range, larger the oscillation range, longer the exposure time required. However, the situation where the intensities crossing the limit of the dynamic range of image plate has been avoided. In our data collection, exposure times less than 600s per frame only was used. The range of rotation angle for data collection in order to acquire complete data set was chosen according to the symmetry of the crystal.

2.9.5 Data Processing: The analysis and reduction of single crystal diffraction data on an imaging plate detector consists of seven major steps:

1. Visualization and preliminary analysis of the original, unprocessed detector data.
2. Indexing of the diffraction pattern
3. Refinement of the crystal and detector parameters,
4. Integration of the diffraction maxima.
5. Finding the relative scale factors between frames.
6. Precise refinement of the crystal parameters using the whole data set.
7. Merging and carrying out statistical analysis of the symmetry related measurements depending upon the space group.

The first four steps of data processing were carried out by DENZO and XDISPLAYF and the steps five through seven were carried out using the program SCALEPACK (Otwinowski, 1993). The auto indexing routine deduces crystal unit cell parameters and crystal orientation parameters from a single oscillation image. The parameters that specify the orientation of the crystal relative to the X-ray beam are the vertical axis, the spindle axis and the crystal rotx, roty, and rotx values. Distortion index and unit cell parameters consistent with all possible Bravais lattices are listed by the program. The lattice with the highest symmetry that fits the data with minimal distortion is chosen. The auto indexing program also gives the crystal orientation parameters. After a successful autoindexing step, the following parameters are refined in the given order:

1. CRYSTAL rotx, roty, rotz.
2. Y BEAM, X BEAM
3. UNIT CELL
4. CROSSFIRE x, y, xy
5. CASSETTE rotx, roty
6. RADIAL OFFSET and ANGULAR OFFSET
7. DISTANCE

Some of these parameters are highly correlated. In particular, crystal to film DISTANCE and RADIAL OFFSET are correlated with the unit cell parameters. So, they can be simultaneously refined only in cases where the quality of data is high and data at high-resolution are available. The input values for the program include the value for mosaicity,
spot shape, background shape, data collection parameters like the crystal to detector distance, the kind of detector used during data collection, oscillation etc.

DENZO allows for interactive visualization and modification of parameters such as the shape, size and profile-fitting radius of the spots. The user has the choice of visually selecting reflections for the auto indexing routine. After each cycle of refinement of the parameters, DENZO updates the display and prints out the numerical summary of the refinement cycle. The output lists the new values for the refined parameters and the shift in their values during the refinement cycle. The output also gives the $\chi^2$ values for the X and Y positions of the predicted spots. The $\chi^2$ values represent the average ratio of the squared error in fitting and the expected error. A good refinement will have $\chi^2$ values near 1.0. The magnitude of the values is not a very critical indicator, as these represent only the comparison of the spatial differences between the observed and the predicted reflections to an error model that might be overly biased. $\chi^2$ values of even 2.0 or 3.0 are accepted, because the position of the predicted reflection, and hence the intensity, is still very accurate. Very large values for the $\chi^2$ parameters indicate that something is seriously wrong with indexing, refinement or the detector. DENZO finally writes out a list of hkl's and unscaled intensities for each image (Otwinowski & Minor, 1997).

The program SCALEPACK is used to scale the raw intensities output by DENZO. This program calculates single isotropic scale and B factors for each of the "films" or "batches" of the processed data that are input. The output is a scaled, merged data. The multiplicative correction factor applied to intensities (I) and $\sigma$(I) is given as

$$S = \frac{2B}{\left(\frac{\sin\theta}{\lambda}\right)^2}$$

The program SCALEPACK is used to scale the raw intensities output by DENZO. This program calculates single isotropic scale and B factors for each of the “films” or “batches” of the processed data that are input. The output is a scaled, merged data. The multiplicative correction factor applied to intensities (I) and $\sigma$(I) is given as
Merging of the symmetry related measurements is done as follows:

The multiplicative factor that is applied to each measurement is calculated from the scale and B factor of the corresponding frame and applied as follows:

\[
W = \frac{1}{(\sigma \cdot S \cdot E_1)^2 + (\{I_{corr}\} \cdot E_2)^2}
\]

Where \(E_1\) and \(E_2\) are the input variables ‘error scale factors’ and ‘estimated error’ respectively and

\[
\{I_{corr}\} = \frac{\sum I_{corr} \cdot W}{\sum W}
\]

The output \(\sigma(I)\) is

\[
\frac{1}{\text{square root } \sum W}
\]

During the refinement, the scale (and B if requested) for all frames are shifted simultaneously to minimize the difference between the \(<I_{corr}>\)’s and the \(I_{corr}\) is for individual measurements, summed over all reflections on all images. \(<I_{corr}>\) are re-determined in each cycle as described earlier. One or more ‘films’ or ‘batches’ that is
The input is designated as the reference, and its scale and B factors for the reference frame are not refined.

The assessment of the high-resolution limit of the diffraction pattern is done in two ways: The first is to calculate the mean ratio of the intensity to the error; \( I / \sigma(I) \). The second is the agreement between symmetry related reflections, i.e. \( R_{\text{merge}} \). The first parameter is an indication of the quality of data. Whereas \( R_{\text{merge}} \), although describes quality and reliability of data, depends on overcoming conditions of low redundancy of data, omission of weak or partial reflections, use of sigma cut-offs in the data set all of which can lead to artificially low \( R_{\text{merge}} \).

The quality of the scaled and merged output data is also assessed using intensity statistics.

The normalized \( \chi^2 \) given by the formula:

\[
\chi^2 = \frac{\sum (I - \langle I \rangle)^2}{\text{Error}^2 * N} \frac{1}{N - 1}
\]

\( \chi^2 \) is a good indicator of the data quality.
The $R_{\text{linear}}$ and $R_{\text{square}}$ are defined as

$$R_{\text{linear}} = \frac{\sum (I - \langle I \rangle)}{\sum I}$$

$$R_{\text{square}} = \frac{\sum (I - \langle I \rangle)^2}{\sum I^2}$$

A good data set should show $\chi^2$ values close to 1.0. The total error and the statistical error should also match closely. The $R_{\text{linear}}$ and $R_{\text{square}}$ values should also be close to each other. If $R_{\text{square}}$ is larger than $R_{\text{linear}}$, then the distribution of the deviation ($I - I_{\text{ave}}$) is skewed. This means that there are too many outliers in the data. The data collection statistics for the data set are given in a tabular form later in this chapter where the details are presented.

**Mathew's number**: Once the space group symmetry and unit cell dimensions of the crystal are known, it is possible to estimate the number of molecules in the crystallographic asymmetric unit and the solvent content of the protein crystals with the knowledge of the molecular weight of protein. The following equations are used (Matthews, 1968).

$$V_m = \frac{(\text{Unit cell volume} \times z)}{(\text{MW} \times n)}$$

$$V_{\text{solv}} = 1 - \frac{(1.23)}{V_m}$$

Where $V_m$ is the Mathew's number, $n$ is the number of molecules in the unit and $z$ is the Avogadro's number; $(V_{\text{solv}})$ is the solvent content of protein crystals. The Matthew's
number and the solvent content were calculated for all the crystal forms of C-PCs and xylanase. The solvent content of the crystal was calculated by assuming two hexamers (monoclinic system) and monomer (hexagonal system) molecule in case of C-PCs and dimer in both case of orthorhombic and monoclinic forms of the ATBXYL-C in the asymmetric unit.

2.9.6 Sequence alignment: Sequence alignments of C-PCs from Phormidium, Lyngbya and Spirulina spp. and ATBXYL-C from alkalophilic Bacillus sp. were carried using available amino acid sequence or from crystal structure determined. The pair wise alignment was carried out using the program CLAWSTAL W available at European Bioinformatics Institute (http://www.ebi.ac.uk) and EsPript (Gasteiger et al., 2003.).

2.9.7 Molecular replacement: Crystallographic technique of elucidating the molecular image as in the crystal unit cell using X-ray diffraction data involves the computation of electron density $\rho(x, y, z)$ through out the volume $V$ of the unit cell and is,

$$
\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{hkl} e^{2\pi i (hx + ky + lz)}
$$

where $F_{hkl}$ is the structure factor defined by the equation

$$
F_{hkl} = |F_{hkl}| e^{i\alpha_{hkl}}
$$

The amplitude $|F_{hkl}|$ which is proportional to the square-root of intensity $I_{hkl}$ can be obtained from intensity of diffraction spot. However, any attempt to determine the associated phases, $\alpha_{hkl}$, from experimental intensities is not straightforward. Direct methods derive the phases from amplitudes through statistical relationships satisfied by
structure invariants. It has been the traditional method for solving small molecule structures and recently being applied to determine the structures of high-resolution (~1.0 Å or better) macromolecules containing about 2000 atoms (Burla et al., 2000; Weeks & Miller, 1999). In general, macromolecular structures are solved by one or more of the three popular methods viz., Multiple Isomorphism Replacement (MIR), Multiwavelength Anomalous Diffraction (MAD) and Molecular Replacement (MR). Other hybrid methods such as Single Isomorphous Replacement with Anomalous Scattering (SIRAS), Multiple isomorphous replacement with anomalous scattering (MIRAS) and Single Wavelength Anomalous Diffraction (SAD) involve the combination of MIR and MAD. MIR, first introduced in 1954 (Green et al.,), has been used widely for macromolecular structure determination. Structure solution by anomalous dispersion, first introduced in 1956 (Ramachandran & Raman) has caught the imagination of crystallographers more recently (Hendrickson et al., 1985; Hendrickson, 1991) with the development of tunable synchrotron radiation sources and recombinant incorporation of selenomethionyl derivative into proteins (Hendrickson et al., 1990). The recent developments in preparation of halide derivatives by a quick soak and using them for phasing of macromolecule is gaining importance in the light of high-throughput structure determination efforts (Dauter et al., 2000; Dauter & Dauter, 2001). MAD method exploits the inequalities of symmetry related reflections (Friedel pairs) due to a particular scattering species to absorb X-rays of specified wavelength while MIR operates on the differences in the intensities duly generated from isomorphous derivative crystals incorporating heavy atom scatterers. The difficulty in preparing heavy atom derivatives is overcome by introducing a new method called Molecular Replacement (MR) method (Rossmann, 1972). This method requires the availability of a search molecule whose structure is similar to the structure of the target molecule. The MR method has become very popular due to availability of many accurately determined protein structures in PDB.

In this chapter, the structure determination of hexameric C–PC from *Phormidium, Lyngbya and Spirulina spp.* and a xylanase from alkalophilic thermophilic *Bacillus sp.* using the molecular replacement method, its refinement and structure validation are
discussed. A brief description of the MR method follows as it was used for the structure determination presented in this thesis.

When the structure of a similar or homologous molecule is known, MR is the simplest technique to determine the target structure. In essence, it involves generating a preliminary model of the target crystal structure by orienting and positioning the search molecule within the unit cell of the target crystal so as to account for the diffraction pattern. The problem in MR is to find in general the six parameters, three rotational and three translational. Which would describe how the search molecule is placed in the unit cell? Rossmann & Blow (1962) had shown that this six-dimensional search could be reduced to a sequence of two three-dimensional searches in which first the orientation and then the position of the search molecule is determined. The rotation function formulation developed by Rossmann & Blow (1961) involves looking for agreement between the Patterson functions of the search and target structures as a function of their relative orientation. If the two Patterson’s are calculated for the same structure the search is a self-rotation function and conveys information on the rotational transformations relating to the subunits of oligomeric proteins. If the Patterson function refers to different crystals, it is a cross-rotation function. The cross-rotation function provides information on the orientation of the known molecule in the unknown cell. First a rotational search is carried out, looking for agreement between positive peaks of the first Patterson representing mainly intra-atomic vectors of a protein subunit, with those of a rotated version of the second Patterson. A function to evaluate this agreement index is defined as

\[ R = \int P_2(x_2) P_1(x_1) \, dx_1 \]

where Patterson function \( P_2 \) is rotated w. r. t. \( P_1 \). Any point \( x_1 \) in \( P_1 \) is related to any other point \( x_2 \) in \( P_2 \) by a rotation matrix \( C \) as \( x_2 = [C]x_1 \). The integral is over a spherical volume \( U \) centered at the Patterson origin. A maxima in the rotation function indicates a potential orientation of the search molecule in the target crystal.
The two Patterson functions expressed in reciprocal space are:

\[ P_1(x_1) = l/V_1 \sum h |F_h|^2 \cos (2\pi h \cdot x_1) \]
\[ P_2(x_2) = l/V_2 \sum p |F_p|^2 \cos (2\pi p \cdot x_2) \]

where \( h \) and \( p \) are reciprocal lattice vectors and \( V_1 \) and \( V_2 \) are the volumes of the unit cells of crystals 1 and 2, respectively. The rotation function is derived by expanding the Patterson function as a Fourier series (Rossmann & Blow, 1962). The resulting expression is given as

\[ R = U/V^3 \sum h \sum p |F_h|^2 |F_p|^2 G_{h,h'} \]

where \( V \) is the volume of the crystal unit cell, \( F_h, F_p \) are structure factor amplitudes corresponding to Patterson’s \( P_1 \) and \( P_2 \), respectively, and \( G_{h,h'} \) is a spherical interference function whose magnitude depends on the reciprocal lattice vectors \( h \) and \( h' \) as well as the volume \( U \) within which the integral is evaluated. The non-integral reciprocal lattice vector \( h' \) is given by

\[ h' = p [C] \]

\[ G_{h,h'} = \frac{3 [\sin(2\pi H.r) - (2\pi H.r) \cos(2\pi H.r)]}{(2\pi H.r)^3} \]
The matrix algebra to relate these coordinate systems has been detailed by Rossmann & Blow (1962). Rotation searches are usually performed using Eulerian angles (α, β, γ) or spherical polar angles (Φ, ω, κ). Crowther (1972) developed a procedure for computing rotation functions, which is much faster than the method of Rossmann & Blow (1962). His method is based on the expansion of the Patterson function as sum of spherical harmonics. If a molecule possesses more than one non-crystallographic symmetry axis, then searching simultaneously for all the axes enhances the power of the rotation function peaks (Rossmann, 1972; Tong & Rossmann, 1990). Such a search is called locked rotation search.

Once the orientation is determined, next step is to find the translation vector (defined by translation function T) that defines the location of the molecule within the unit cell. The translation function (T) was first proposed by Crowther & Blow (1967). It is derived from the general form of the translation function given by the equation

\[
T(t) = \int \mathcal{P}_o \{ u . t \} . P(u) \, du
\]

Where \( t \) is the intermolecular vector, \( u \) is vector in Patterson space, \( V \) is primitive unit cell volume, \( P(u) \) represents the observed Patterson function of the crystal and \( \mathcal{P}_o(u, t) \) represents the cross Patterson vectors (calculated intermolecular Patterson vectors). When \( t \) becomes \( t_o \), the computed cross Patterson vectors fit correctly to the observed Patterson function \( P \), and \( T(t_o) \) will have a large positive value. There are several forms of translation function and various available packages use one or more of these formalisms. The success of translation search depends on the accurate orientation of the model during rotation, completeness and accuracy of the model, and the resolution range and selection of diffraction data.

The success of the MR method in general relies on (Navaza & Saludjian, 1997)
1. The completeness and quality of the data.
2. The extent of structural similarity between the molecular models and the actual molecules that constitute the crystal.
3. The size of each molecular model with respect to the content of the crystal cell.
4. The excellence of the criterion used as an indicator of the quality of agreement.

Several MR packages are available. AMoRe, X-PLOR/CNS, Merlot, PATSEE, GLRF are some of the packages. The principal difference between these packages is the way in which the calculations can be carried out, whether in real space or in reciprocal space (Turkenburg & Dodson, 1996). AMoRe that was designed with motto to deal with difficult problems has many novel functions incorporated in it. We have used AMoRe implemented in CCP4 suite (Collaborative Computational Project, Number 4., 1994) in the present study.

2.9.8 Structure refinement: Crystallographic refinement aims at optimizing the agreement between atomic model and observed diffraction data by making use of chemical restraints in the case of low resolution data. The structures reported in this thesis were refined using the program in CCP4. Molecular dynamics methods are exploited by CCP4i to probe conformational space of the molecule while minimizing the difference between the observed and calculated structure factors (Brunger et al., 1987, 1990, 1997). There are options for rigid body refinement, positional refinement, restrained and unrestrained individual B-factor refinement, group B-factor refinement, occupancy refinement and electron density map calculations. Features of the program pertinent to the thesis are discussed briefly in subsequent sections. Model building and map fitting were done manually using the Quanta software (Accelrys).

Refinement techniques generally fall into two categories, depending upon whether the calculations are performed either in the real space or in the reciprocal space. Refinement methods based on reciprocal space is preferred over real space because the former ones are computationally less intensive. To prevent the model from going into a
local minimum, interactive graphics was used for checking the fit of the model to the electron density. To monitor the progress of refinement the crystallographic parameter called R factor is used. This is defined as,

$$ R = \frac{\sum_{hkl} |Fo(hkl)| - |Fc(hkl)|}{\sum_{hkl} |Fo(hkl)|} $$

Where h, k, l, are the miller indices of Bragg reflections, and the summation is over all the reflections.

**2.9.9 Constraints and restraints:** During the course of refinement in macromolecules, some groups of atoms may have to be constrained or restrained to improve the ratio of observables to parameters. CCP4 suit has options to group atoms so that they move as rigid bodies, or, restrain or constrain the bond lengths, bond angles, non-crystallographic symmetry (NCS) and atomic positions to desired values by use of appropriate force constants. Restraints are used when limited freedom can be allowed for a parameter. Constraints are used to hold a parameter to an ideal value. A constraint is equivalent to a restraint having the value of force constant infinity. In NCS restraints, the molecules in the asymmetric unit are superposed by least squares superposition and the average co-ordinates \(x_{av}\) of individual atoms are computed. If x represents the co-ordinates of individual atoms then each atom can be restrained according to the mathematical term:

$$ E_{NCS} = w(x-x_{av})^2 $$

And the corresponding B-factor restraints are given by:
\[ B_{\text{NCS}} = \frac{(b-b_{av})^2}{\sigma_{\text{NCS}}^2} \]

Where \( w \) is a weight function, \( b \) and \( b_{av} \) are the respective individual and average temperature factors of NCS related atoms and \( \sigma_{\text{NCS}} \) the target deviation for B-factor restraints. Each of the hexamers in C-PC asymmetric unit has molecular point group symmetry 32 (D3). However, the application of NCS constraints imposes a strict restriction on the refinement of related subunits or parts, which is usually violated for macromolecular structure data with resolution better than 3.5 Å (Kleywegt, 1996). Hence in the initial stages, refinement was carried out with NCS restraints with different weights applied between either all 12 monomers of hexamer \((\alpha\beta)_6\) or between the two \((\alpha\beta)_6\) hexamers of the asymmetric unit. However, inspection of relative change in \( R_{\text{free}} \) factor with and without NCS restraints (tight) (Kleywegt & Brünger, 1996) led us to keep the NCS restraints (tight).

### 2.9.10 Electron density maps and Model building:

After every refinement cycle the model was inspected and manual rebuilding done by inspecting \((2Fo-Fc)\), \((Fo-Fc)\) and negative \((Fo- Fc)\) electron density maps. CCP4 has options to calculate \( \sigma_A \)-weighted maps where the structure-factor amplitudes are weighted in order to reduce the model bias of an incomplete or partially incorrect structure. The Fourier coefficients calculated are given by:

\[ F_{\text{map}} = (m \cdot Fo - D \cdot Fc) \exp (i\alpha_c) \]

Where \( m \) is the figure of merit and \( D \) is a measure of the error in the co-ordinates of the model, as defined by Luzzati (1952). Electron density maps using the Fourier coefficients derived here have been observed to be superior to the conventional 'unweighted' maps (Main, 1979; Read, 1986). The \((2Fo-Fc)\) and \((Fo-Fc)\) maps were usually contoured at 1.0 \( \sigma \) and 2.0 \( \sigma \), respectively, where \( \sigma \) refers to the r.m.s. deviation of the mean density in units of electrons/Å\(^3\) in the maps.

Model building was done using the program Quanta (Accelrys). Manual rebuilding using quanta is less subjective because of the incorporation of information
from databases of known structures. Information from the main chain database (Jones & Thirup, 1986), side chain rotamer database (Ponder & Richards, 1987), the peptide orientations database (Zou & Mowbray, 1994) are available for model rebuilding and the results have been analyzed through real space electron-density fit.

Omit maps were computed very frequently during the process of model building and refinement. After omitting residues in a region in question and map calculation after refinement, the model was built into the omit density, thus overcoming any model bias. The model bias, if any, would clearly appear as –ve density in (Fo-Fc) maps and such regions could be built successfully by omit maps. The model that emerged after each cycle of model building and refinement was checked for unusual geometry at each residue, the peptide-flip values, and high temperature factors. The other parameters checked include values of torsion angles, including (Ramachandran angles φ and ψ), side chain torsion angles such as χ1, χ2 etc. were verified between refinement cycles using the program PROCHECK. Special care was taken to resolve the side-chain flip conflicts for all Asn, Gln and His residues by analyzing their environment and interactions.

2.9.11 Refinement by maximum-likelihood method: The initial models of C-PCs and ATBXYL-C obtained from MR calculations were refined using the program REFMAC (implemented in CCP4) which makes use of maximum-likelihood equations. Bricogne (1991) have suggested a maximum-likelihood target that should be based on various probability distributions. One of the expected advantages of maximum likelihood refinement is a decrease in refinement bias, as the calculated structure-factor amplitudes will not be forced to match the observed amplitudes (Read, 1997). Use of appropriate likelihood targets through the incorporation of the effect of measurement of error and the use of cross-validation data to estimate the σ values (error) are the key ingredients in the likelihood refinement. Verification tests have shown that for protein structure refinement, maximum likelihood method is more than twice as effective compared to the least-squares method, in improving the model (Pannu & Read, 1996). REFMAC program can carry out rigid body restrained or unrestrained refinement using X-ray data (Murshudov et al., 1997).
The program minimizes the coordinate parameters to satisfy a maximum-likelihood or least squares residual. Before running the REFMAC, PROTIN program was run to invoke geometric restraints. PROTIN analyses the protein geometry and produces an output file containing restraint information. REFMAC also produces an output file with extension MTZ (named after three of its progenitors, McLaughlin, Terry and Zelinka) containing weighted coefficients for weighted mFo-DFc and 2mFo-DFc maps. About 5% of the reflections were kept aside during refinement to calculate $R_{\text{free}}$ for cross validation (Brunger, 1992). NCS averaging is effectively done for an asymmetric unit composed of N similar objects related by non-crystallographic symmetry (NCS). Rossman and Blow (1963) proposed this method, by which the current phases of reflections can be improved by averaging over the electron densities of NCS related objects. NCS averaging requires an accurate estimate of NCS operators and exact information on the position and shape of the objects whose density has to be averaged (Vellieux & Read, 1997). Since both C-PCs and ATBXYL-C structures had more than one molecule in their asymmetric units, their initial models were refined taking advantage of the presence of non-crystallographic symmetry.

2.9.12 Analysis and validation of structures: The program PROCHECK was used for checking the stereochemistry and quality of the model. This program is written by Laskowski et al., (1993) and forms part of the CCP4 suite of programs. Once the model has improved, after every cycle of model fitting and run of REFMAC it was subjected to stereo chemical quality check. The program compares and assesses the quality of the model with the available structures of similar or better resolution than the reference structure. The output contains a comprehensive residue by residue listing of the parameters and plots of their values. The program highlights the regions of the structure where the conformations deviate considerably from ideal model. These can either be due to interesting properties of the structure or possible errors in interpretation, that require further investigation during next rebuilding step or analysis.