CHAPTER 3

MATERIALS AND METHODS

3.1 Introduction

The crystallographic analysis of the anti-(4-hydroxy-3-nitrophenyl)-acetyl (anti-NP) and the anti-*p*-azophenylarsonate (anti-Ars) germline mAbs, BBE6.12H3 and 36-65 bound to various peptides derived from the screening of a random phage library would yield valuable information regarding their promiscuous binding abilities, associated with a primary immune response. The primary requirement for the crystallographic analysis was the preparation of adequate quantities of pure Fab (fragment antigen binding). Purified Fab fragment of both the mAbs was used for subsequent crystallization experiments with various peptide ligands. The X-ray intensity data for the crystals obtained were collected followed by structure determination, iterative steps of crystallographic refinement and model building. The refined models were structurally validated and then subjected to rigorous analysis. This chapter provides a brief background of the methods utilized and details of the experimental protocols followed.

3.2 Peptide synthesis

Solid phase peptide synthesis was introduced by Merrifield in 1963, and includes successive assembly of amino acid residues to build the peptide chain on an insoluble polymeric support. The C-terminal residue, with protected α-amino and side chain functional groups, is chemically attached to the insoluble resin via a flexible linker. Subsequently, in the coupling step, the α-amino group is deprotected and the next protected amino acid is reacted with the resin-bound first amino acid. This cycle of deprotection and coupling is repeated till the complete peptide chain is synthesized. After the synthesis of the desired peptide, the anchoring bond between

the peptide and the resin is cleaved using trifluoroacetic acid (TFA) to release the polypeptide.

3.2.1 Procedure for peptide synthesis

All the peptides used in this study were synthesized by solid phase method on an automated peptide synthesizer (Applied Biosystems, Model 431A), using F-moc (9-fluorenylmethyloxycarbonyl) chemistry on a *p*-hydroxymethyl phenoxymethyl polystyrene resin (Nova Biochem). For the peptide synthesis, 0.1mmol of the resin was used and deprotected using 20% piperidine in N-methyl-pyrrolidone (NMP). Subsequently 0.5nmol of the first amino acid was added and coupling was performed using DCC-HoBt (dicyclohexylcarbodiimide-hydroxybezotriazole) ester formation method. All other amino acids were coupled by DCC ester coupling. Amino acids and solutions required for peptide synthesis were procured from Nova Biochem and Applied Biosystems, respectively. After completion of synthesis, deprotection was carried out in 20% piperidine/DMF. Finally, the resin was shrunk using ether and dried under vacuum for a minimum of four hours. The cleavage was performed in dark using 94% TFA, 5% anisole, EDT and water accompanied by continuous stirring for two hours. The resin was then filtered and washed with DCM and the solution was evaporated on a rotary evaporator (Buchi, Switzerland) till only a small quantity of DCM/cleavage mixture is left. Cold anhydrous diethyl ether was added to the filtrate to aid in the separation of scavengers from the mixture. The peptides were then extracted with water using a separating funnel. Extraction was followed by evaporation of residual diethyl ether on the rotary evaporator. Total aqueous layer was then frozen as a thin film and lyophilized.

3.2.2 Peptide purification

The peptides were purified using reverse-phase HPLC. Binding occurs through hydrophobic interactions between peptide and the column support. Decreasing the ionic nature or increasing the hydrophobicity of eluant such that it competes with peptide for hydrophobic groups on the column accomplishes elution.

Crude peptides were purified on a Waters Xbridge™ BEH 130 reverse-phase C18 column (19x250mm, 10μm, spherical) on a semi-preparative HPLC system (Waters, USA) using a linear gradient of 0.1% trifluoroacetic acid (Sigma) and acetonitrile (Merck). The absorption was monitored at 214nm. After purification, the peptides were lyophilized. The purity of the peptide was checked by determination of molecular mass using single quadruple mass analyzer (Fisons Instruments, UK). Circular dichroism studies were performed on the peptides to determine secondary structural state, if any, in solution. 50 μM peptide concentrations in water were used and data was accumulated for 10 scans at a temperature of 10° C using a JASCO 710 spectropolarimeter. 1.0 nm bandwidth and 0.1nm resolutions were used, with the sample being placed in a 2mm path length cuvette.

3.3 Preparation of Fab fragment

A prerequisite for collection of X-ray diffraction data for structure determination is availability of well-ordered single crystals. Protein crystallization is considered to be an art and there are no universally applicable methods for obtaining crystals suitable for X-ray diffraction studies. However, it is unlikely that good crystals will not be obtained if attention is paid to the purity and stability of the molecule of interest. A high degree of purity is essential for successful crystallization of most proteins. Also, the molecule must be stable under the

conditions used for purification and crystallization. Therefore, in any attempt at crystallization, the homogeneity of the preparation and its stability in solution is of paramount importance.

IgG molecules possess a net negative charge, at pH 8.5. At this pH, IgG molecules, can therefore, bind to positively charged chemical moieties, an association that can be reversed using low concentrations of sodium chloride. Hence, it is possible to purify IgG using anion-exchange chromatography with a salt gradient. Since this method does not involve harsh conditions it represents an ideal way to purify IgG to be ultimately used for crystallization experiments. Ammonium sulphate fractionation prior to chromatography improves the resolution of peaks during IgG purification. The mouse IgG molecules are known to precipitate in the range of 33% to 40% (v/v) saturated ammonium sulphate.

IgG molecules have three major domains, each of which are mobile with respect to the other. Consequently, the molecule is highly flexible and not easy to crystallize. Although entire immunoglobulin molecules have also been crystallized (Harris *et al.*, 1998), the proteolytic fragments of IgG molecules have been observed to be more amenable to crystallization. Papain, a non-specific proteolytic enzyme obtained from papaya, cleaves antibodies at the hinge region to release one Fc and two Fab domains. Proteolysis by papain, if carried out carefully, followed by purification can yield a homogenous preparation of Fab. This pure preparation of Fab can be used for crystallization experiments.

3.3.1 Generation of ascites from mice

Ascites is the intra-peritoneal fluid collected from mice that have developed peritoneal tumor. Hybridoma cells, when injected into the peritoneal cavity of mice,

secrete antibody which gets collected in the peritoneal fluid. Ascites is thus a good source of monoclonal antibody. The hybridoma cells, from two different hybridoma, which were secreting IgGs, 36-65 and BBE6.12H3, were injected into the peritoneal cavity of male Balb/c mice irradiated with a dose of 400 RAD and primed with Freund's incomplete adjuvant 72 hours prior to injecting the hybridoma cells suspended in 500 µl of Dulbecco's phosphate buffered saline (DPBS). Approximately 5 x 10^5 to 5 x 10^6 hybridoma cells were injected into each mouse. Ascitic fluid could be tapped from the peritoneal cavity of mice after approximately 4-5 days.

3.3.2 Antibody purification

The collected ascitic fluid was centrifuged to remove cell debris and fat. Mouse monoclonal ascites, was filtered through glass wool to remove lipid like material left over after centrifugation. The supernatant was then subjected to $(NH_4)_2SO_4$ fractionation. Saturated $(NH_4)_2SO_4$ solution (SAS) at pH 7.0 was gradually added to the ascites in an ice bath with continuous stirring till a concentration of 40% (v/v) was achieved. The mixture thus, obtained was centrifuged to get the protein pellet and the pellet was re-suspended in buffer (0.01 M Tris-Cl, pH 8.5).

The crude antibody solution obtained from ammonium sulfate fractionation was dialyzed against the wash buffer (0.01 M Tris-Cl, pH 8.5) and then subjected to ion-exchange chromatography using 5PW-DEAE (60x150 mm) column on a Waters3000 preparative HPLC (Waters, USA), to purify IgG. All solutions used during chromatography were filtered (0.45µm) and then degassed. Following equilibration of the column with wash buffer, a 2 ml aliquot of the crude antibody

solution was injected into the HPLC. A salt gradient of 0 to 0.2 M NaCl over a period of 120 minutes was run and fractions for each peak, as detected by measurement of UV absorbance at 220nm, were collected. An aliquot of each fraction was subjected to acetone precipitation and the obtained precipitate was analyzed on SDS-PAGE to ascertain which fraction corresponds to IgG. The IgG fractions from different runs were pooled and concentrated to ~ 1mg/ml which was then dialyzed against the digestion buffer (0.15 M NaCl, 0.1M Tris-Cl, pH 7.1).

3.3.3 Generation of Fab fragment

The concentration of the purified IgG was estimated using micro bicinchoninic acid (BCA) method of protein estimation (Micro BCA Protein Assay Kit , Pierce). To ascertain the optimal ratio of IgG to papain and time of incubation for proteolytic fragmentation of the antibody, an initial analytical digestion was carried out. 1 mg/ml solution of papain (Sigma, St. Louis, USA) was first pre-activated using β-mercaptoethanol (2.0mM) for one hour. The reaction mixtures for digestion constituted 400 µg of IgG, β-mercaptoethanol (2.0mM) and EDTA (2.0mM) appropriate volumes of the activated papain solution were added so as to obtain various ratios of IgG:papain. The digestion reaction was monitored for 10 hours and 20 µl aliquots were collected every hour. The proteolysis reaction was stopped with addition of 75 mM iodoacetamide and all the aliquots were analyzed on reducing SDS-PAGE. Initially the IgG appears as two bands, one at 50 kDa and the other at 25 kDa, corresponding to the heavy and light chains, respectively. The intensity of the 50 kDa band decreases and instead, a doublet starts appearing at 25 kDa as the reaction proceeds. The Fc portion may or may not be visible as a band just above the 25 kDa doublet. The digestion is deemed complete when the 50 kDa

band has completely disappeared and only the 25 kDa doublet is clearly visible. Optimal w/w ratios of protein/enzyme and time of incubation were ascertained and preparative digestions were carried out using 20-50 mgs of IgG. The digestion mixture was then dialyzed against 10 mM Tris-Cl, pH 8.0 for Fab purification.

3.3.4 Purification of Fab fragment

Fab purification from the digestion mixture was carried out by ion-exchange chromatography using 5PW-DEAE (60x150 mm) column on a Waters3000 preparative HPLC (Waters, USA). Initially, a blank run was carried out thereafter the column was allowed to re-equilibrate with the wash buffer (10 mM Tris-Cl, pH 8.0). A salt gradient of 0 to 0.2 M NaCl over a period of 120 minutes was used to elute the Fab. An aliquot from all the collected fractions were precipitated by using chilled acetone and were analyzed on a SDS-PAGE gel to ascertain which fraction corresponds to Fab. Fab, which has low or zero net negative charge at pH 8.0, was eluted out as the first major peak early in the gradient. The Fc portion and any undigested IgG which have a higher net negative charge at pH 8.0 would elute out later in the gradient. The Fab fractions collected from various HPLC runs for both the antibodies were pooled, concentrated and dialyzed against their respective crystallization buffer (50 mM Na-cacodylate pH 6.7, 0.05% sodium azide and 50mM Tris-Cl pH 7.1, 0.05% sodium azide).

3.4 Crystallization and data collection

3.4.1 Crystallization

One of the most widely utilized methodologies of crystallization is hanging drop vapor diffusion technique (Wlodawer and Hodgson, 1975). The setup involves

a buffered protein solution in the form of a droplet in contact with the precipitant through the vapor phase. The precipitant slowly causes dehydration to occur in the protein droplet increasing the effective concentration of the protein. The hanging drop crystallization experiment is set up in 24 well tissue culture plates, with the drop of protein solution containing 50% of the precipitant in the mother liquor suspended over the precipitant solution from a siliconized cover slip. This setup is sealed with silicon grease to facilitate controlled vapor diffusion between the well and the drop.

For setting up hanging drop crystallization, a pure preparation of Fab molecules in the crystallization buffer (50 mM Na-cacodylate pH 6.7, 0.05% sodium azide or 50mM Tris-Cl pH 7.1, 0.05% sodium azide) was concentrated to a final concentration of 10 mg/ml. For the antibody-peptide complexes, 50-fold molar excess of the peptide was added to the Fab solution. Hanging drops of 8 µl volume containing 4 µl of the Fab solution and 4 µl of varying concentrations of the precipitant were set up in 24-well tissue culture plates (Nunc, Denmark). Initially, a variety of precipitants were used in the crystallization experiments. Conditions which gave indications of crystal formation were then further explored to improve the quality of the crystals. The crystallization plates were maintained at room temperature in insulated conditions so as to prevent rapid changes in temperature. For crystallization of BBE6.12H3Fab-peptide complexes, the crystallization plates were also maintained at 8°C in vibration free incubator (RUMED, Rubarth Apparate, GmbH, Germany). The plates were checked for the presence of crystals every two weeks.

3.4.2 X-ray intensity data collection

Data collection for macromolecular crystallography involves exposure of the crystal to X-rays and recording the intensities of the resultant diffraction patterns. Rapid advances in this field have made available sophisticated electronic detectors like the Image plate detector, high power X-ray generators and synchrotrons. Successful data set collection is followed by data processing to extract the hkl indices with corresponding intensities, along with an estimate of the errors involved.

At the core of the Image Plate detector is an amorphous thin film made of Barium, Europium and Bromium. This material that is coated on to a motorized plate absorbs X-rays to form F-centers. These F-centers are the regions that store photon energy as excited electrons. After the exposure is complete the plate is read by a He-Ne (2eV) red laser. Absorption of photons induces excited electrons to return to ground state with the emission of blue light (4eV) which is quantitatively read by a photomultiplier. Exposing it to intense white radiation erases the plate. While the basic technology behind the image plates remains the same, improvements in electronics and computers has led to greater automation and faster data collection cycles.

The X-ray intensity data for various Fab-peptide complexes of 36-65 were collected on the Mar345dtb, installed on a rotating anode X-ray source (RIGAKU, Japan) operating at 50kV and 100mA (CuKα radiation) with Osmic mirrors (RIGAKU, Japan). While the Mar225 image plate installed at BM14 (ESRF, Grenoble, France) was used to record three Fab-peptide complexes of BBE6.12H3. Data for antigen free BBE6.12H3 Fab and its complex with Ppy peptide was recorded on Mar345dtb image plate (Mar research, Germany), installed on the home source. For data collection at room temperature, the crystals were mounted in 0.5 mm quartz capillary tube along with some mother liquor. The capillaries were then

mounted on goniometer heads, which were in turn fixed on the oscillator dial of the image plate. However since our crystals suffered significant radiation damage at room temperature we decided to attempt cryo-crystallography and collected data at low temperature.

Radiation damage to protein crystals is greatly reduced at lower than room temperatures (D. J. Haas, 1970; Low et al., 1966). Primary radiation damage is largely caused by interactions between the molecules in the crystal and the beam. This energy is dissipated in at least two ways; it produces thermal vibrations (heat) and it provides the necessary energy to break bonds between atoms in the molecules. Secondary damage to the crystals is caused by the diffusion of reactive radicals produced due to damage to the protein. This diffusion is aided by the presence of thermal energy. At cryo-temperature of around 100K, thermal damage is limited and also the reactive products are immobilized and do not cause extensive secondary damage in areas of the crystal which are not exposed to the beam (Garman, 1999).

For low temperature data collection, the crystals were initially soaked in a cryo-protectant, which was basically the mixture of the mother liquor and antifreeze. We added 30% glycerol to our mother liquor, in which the crystals were soaked from between 1 to 5 minutes to achieve cryo-protection. The crystals were then picked up using a 20μ nylon loop, which was immediately flash frozen in a stream of nitrogen at 120k at a flow rate of 6 liters/min (Oxford cryo-systems). The crystals were centered in the beam using the two arcs and translations on the goniometer head and by viewing the crystal on the monitor of the attached CCD camera. The collimation, crystal to detector distance, oscillation angle and the exposure time per frame were optimized after a few trial frames in each case.

3.4.3 X-ray intensity data processing

The collection of macromolecular diffraction data has undergone dramatic advances during the last 20 years with the advent of two-dimensional area detectors such as image plates and CCDs, crystal cryocooling and the availability of intense, monochromatic and highly collimated X-ray beams from synchrotron sources. These technical developments have been accompanied by significant advances in the software used to process the resulting diffraction images. In particular, autoindexing procedures have improved the ease of data processing to the point that in many cases it can be carried out automatically without any user intervention. However, the procedure used to collect the diffraction images, the screenless rotation method, has remained essentially unchanged since it was first suggested for macromolecular crystals by Xuong et al. (Nguyen-huu-Xuong, 1968) and by Arndt and coworkers and popularized by the availability of the Arndt-Wonacott oscillation camera (Arndt, 1977; U. W. Arndt, 1973). In this procedure, each diffraction image is collected while rotating the crystal by a small angle (typically between 0.2 and 2°) about a fixed axis (often referred to as the φaxis). The only development of the method has been the use of very small rotation angles per image (the so-called fine φ-slicing technique) to provide improved signal to noise for weakly diffracting samples. Since, virtually all macromolecular diffraction data are collected in this way (with the exception of data collected using the Laue technique). The starting point for data integration will therefore be a series of such diffraction images and the desired outcome is a data set consisting of the Miller indices (*hkl*) of all reflections recorded on these images together with an estimate of the diffracted intensities $I(hkl)$ and their standard uncertainties $\sigma I(hkl)$. This requires the prediction of which reflections occur on each image and also the precise position of each reflection on each image (note that typically most reflections will be present on several adjacent images and

therefore only partially recorded on any individual image. For each predicted reflection, the background-subtracted diffracted intensity must be estimated. Although straightforward in principle, defects and limitations in both the sample (the crystal) and the detector can make this difficult in practice. Complicating factors include crystal splitting, anisotropic and/or very weak diffraction, high mosaicity, diffuse scattering, the presence of ice rings or spots, unresolved or overloaded spots, noise arising from cosmic rays or zingers, backstop shadows, detector blemishes, radiation damage and spatial distortion. These experimental factors will be important in determining the final quality of a data set.

The HKL2000 (Otwinowski, 1997) is GUI based suite of programs for the analysis of X-ray diffraction data collected from single crystals. The package consists of three programs: DENZO, XDISPLAYF and SCALEPACK. HKL is the program that converts the raw X-ray diffraction data, collected from an image plate and reduces it to a file containing the hkl indices, intensities of the spots on the image plate along with estimates of errors involved. DENZO initially performs peak searching. The autoindexing algorithm carries out complete search of all the possible indices of the reflections picked by peak search using a fast Fourier transformation (FFT) software module. After search for real space vectors is completed, the program finds the three best linearly independent vectors, with a minimal unit cell volume, that would index all of the observed peaks. After refining the initial cell dimensions and detector parameters, the determined values are applied to the rest of the frames and the parameters are refined for each frame. The diffraction maxima are also integrated by DENZO. The program XDISPLAYF (W., 1993) enables visualization of the peak search and processing procedures. SCALEPACK finds the relative scale factors between frames and carries out precise refinement of crystal

parameters using the whole data set. It is also used for merging different data sets and carrying out statistical analysis of the measurements related by space group symmetry. SCALEPACK also provides the detailed analysis of the merged data, and symmetry equivalent positions, space group-specific systematic absences, total percentage of data collected and the linear R_{merge} for data reduction.

MOSFLM is a package of programs with an integrated graphical user interface for processing data collected on any detectors. The programs cover all aspects of data reduction starting from the crystallographic pattern recorded on an image to the final intensities of observed reflections. In MOSFLM this entire process of integration of diffraction images is subdivided into three steps. The first is the determination of the crystal parameters, in particular the crystal lattice (unit cell) and its orientation relative to a laboratory axial system (usually based on the X-ray beam direction and the rotation axis). This is usually referred to as autoindexing. Knowledge of these parameters then allows an initial estimate of the crystal mosaicity. The second step is the determination of accurate unit-cell parameters, using a procedure known as post-refinement. This requires the integration of one or more segments of data with a few images in each segment. The final step is the integration of the entire set of diffraction images, while simultaneously refining parameters associated with both the crystal and the detector.

After integration of the data, next step is to scale and merge the data set. Scaling and merging are done with the program SCALA. This program scales together multiple observations of reflections, and merges multiple observations into an average intensity. The merging algorithm analyses the data for outliers, and gives detailed analyses. It generates a weighted mean of the observations of the same reflection, after rejecting the outliers. SCALA also provides the detailed analysis of

merging data, and symmetry equivalent positions, space group-specific systematic absences, total percentage of data collected and the linear R_{merge} for data reduction.

Finally, truncate program was used to obtain structure factor or amplitudes from averaged intensities (output from SCALA, or SCALEPACK) and write a file containing mean amplitudes and the original intensities. If anomalous data is present then F(+), F(-), with the anomalous difference, plus I(+) and I(-) are also written out. The amplitudes are put on an approximate absolute scale using the scale factor taken from a Wilson plot.

For all the Fab-peptide complexes and unliganded Fab of BBE6.12H3 antibody, the diffraction data were collected and processed using MOSFLM and subsequently merged using SCALA. For all the Fab-peptide complexes of 36-65 Fab, the diffraction data were collected and processed using DENZO and subsequently merged using SCALEPACK. The cell dimensions and space groups were unambiguously determined for each crystal. The solvent content and Matthews's constant were calculated (Matthews, 1968). The merged and scaled intensities were used for structure determination.

3.5 Structure determination

3.5.1 Structure determination using molecular replacement

The goal of diffraction analysis is reconstruction of the detailed structure of the asymmetric unit from a diffraction pattern. The diffraction pattern breaks down the structure into discrete sine waves. Any shape can be presented in three dimensions as the sum of sine waves of varying amplitudes and phases. The individual reflections of a diffraction pattern represent such waves, which have

wavelength component in three dimensions inversely proportional to their values of *h*, *k* and *l*. The image of the object can be reconstructed by recombining the individual sine waves as occur in the objective lens of the microscope. Since it is not possible to focus the X-rays, only the intensities could be recorded with the loss of phases, well known as phase problem of crystallography.

Macromolecular crystal structures are usually solved using one of the three techniques; multiple isomorphous replacement (MIR), multiple anomalous dispersion (MAD) or molecular replacement (MR). Of the three, MR is generally used in cases where a structural homolog is available. Since the structure of a number of antibodies is already known, MR is the method of choice for structure determination of antibody Fab.

The molecular replacement method, involves orienting and positioning a model molecule in the experimental unit cell through rotations and translations. The rotation function developed by Rossman and Blow (1962), involves rotation of the Patterson function of one group or molecule with respect to the other in all possible ways and the ultimate superimposition of the two Patterson functions. The translation function deals with positioning the oriented molecule in the unit cell of the unknown structure. It utilizes the cross vectors between various symmetrically related molecules for positioning the probe in the target unit cell. The translation function is carried out by moving the oriented model in small increments along all three directions and calculating the correlation between observed and calculated intensities. From the solutions obtained, the one with the highest correlation and lowest R-factor was chosen for molecular replacement.

The structure of the Fab of putative anti-NP germline mAb N1G9 was used for molecular replacement. The refined model of the native unliganded germline Fab

was subsequently used as a probe model to carry out molecular replacement for one of the Fab-peptide complex; remaining three Fab-peptide complexes were solved by using Ppy-LH as search model. The structure of antigen bound 36-65 Fab (2A6I) was used for molecular replacement of two Fab-peptide complexes of the same antibody.

AMoRe (Navaza, 1994) and Phaser packages from CCP4 suite (Elizabeth Potterton, 2003) were used for structure determination of antigen free BBE6.12H3 Fab and its complexes with peptide, respectively. The solution for 36-65 complexes was determined by using MOLREP from CCP4 suite. Both for MOLREP and AMoRe, calculations for rotation/translation functions were carried out using structure factors from 8 to 4 Å resolutions. The transformation matrices obtained from AMoRe for antigen free Fab was utilized to orient the models in the corresponding unit cell. However, both Phaser and MOLREP have a module which automatically does orientation. The packing function of Phaser also checks for possible clashes or voids between the symmetry related molecules. All the solutions were unambiguous. For outputs of AMoRe and MOLREP the crystal packing was examined using Coot (Emsley P, 2004) to ascertain the absence of steric clashes or large voids between symmetry related molecules. Calculations of the Matthews coefficient (Kantardjieff and Rupp, 2003) indicated presence of two molecules for antigen free Fab and a single Fab molecule for all Fab-peptide complexes within the asymmetric unit.

3.5.2 Automated molecular replacement package (AMoRe)

To carry out MR, the AMoRe package can be used. AMoRe constitutes a suite of programs written by Jorge Navaza (Navaza, 1993; Navaza, 1994). These

include SORTING, that sorts, packs and assesses the quality of the experimentally measured diffraction data, and is run in the first step. The program TABLING calculates the continuous Fourier coefficients from the model placed in the artificial cell. The cross-rotation function is carried out by the program ROTING, which uses Crowther's algorithm (Crowther, 1972). TRAING is used to calculate the translation function. Finally FITING is used to refine the orientational and positional parameters of the molecule corresponding to the potential solutions, as a rigid body.

3.5.3 Automated molecular replacement program (MOLREP)

MOLREP is an automated program for molecular replacement that utilizes a number of original approaches to rotational and translational search and data preparation. MOLREP can perform a variety of tasks that require rotational and/or positional search: standard MR, multi-copy search, fitting a model into electron density, heavy-atom search and model superposition.

The arsenal of rotation (RF) and translation (TF) functions includes self-RF, cross-RF, locked cross-RF, phased RF, full-symmetry TF, phased TF, spherically averaged phased TF and packing function (PF).The program is general for all space groups. The output of the program is a PDB file with the atomic model ready for refinement and a text file with details of the calculations.

The rotational search is performed using the RF of (Crowther, 1972), which utilizes the fast Fourier transform (FFT) technique. The default radius of the integration sphere is derived from the size of the search model and is usually two times larger than the radius of gyration. The RF solutions are refined prior to positional search using a rigid-body technique. The refinement is performed in space group P1 and the outcome is evaluated by the correlation coefficient. It

simultaneously uses all symmetry operators, resulting in a single peak with an improved signal-to-noise ratio which directly gives the position of the model in the unit cell. In addition, the TF is coupled with a PF to remove false maxima which correspond to interpenetrating molecules. Both the TF and PF allow the incorporation of a second model already placed in the cell. The TF solution may be subjected to rigid-body refinement incorporated in MOLREP. Non crystallographic symmetry may be imposed on the model in order to restrain the refinement. Pseudo-translation is automatically detected from analysis of the Patterson map. A significant off-origin peak gives the pseudo-translation vector, which is used to modify structure factors in the TF calculation (Navaza et al., 1998). In MOLREP multiple copies of the macromolecule in the unit cell can be searched (Vagin, 2000).

3.5.4 Automated molecular replacement program (Phaser)

Phaser is a program for phasing macromolecular crystal structures by both molecular replacement and experimental phasing methods (A. J. McCoy, 2007). The novel algorithms in Phaser are based on maximum likelihood probability theory and multivariate statistics rather than the traditional least-squares and Patterson methods. For molecular replacement, the new algorithms have proved to be significantly better than traditional methods in discriminating correct solutions from noise. One of the design concepts of Phaser was that it be capable of a high degree of automation. Phaser has novel maximum likelihood phasing algorithms for the rotation functions and translation functions in MR, but also implements other non-likelihood algorithms that are critical to success in certain cases.

3.6 Structure refinement using the Crystallography and NMR system (CNS) suite

This section describes the detailed refinement strategies used for structure determination. The structures presented in the thesis were refined using CNS (Brunger, 2007; Brunger et al., 1998).

The refinement of structures obtained after molecular replacement was done using Crystallography and NMR system (CNS) suite of programs (Brünger et al., 1998) based on the refinement of crystal structures using Cartesian (Brünger et al., 1987) or torsion-angle molecular dynamics (Rice and Brünger, 1994). This task file automatically computes a cross-validated σ_A estimate, determines a weighting scheme between the X-ray refinement target function and the geometric energy function, refines a flat bulk solvent model (Jiang and Brunger, 1994) and an overall anisotropic B-value of the model by least squares minimization, and subsequently refines the atomic positions. Available target functions include the maximum-likelihood functions MLF, MLI and MLHL (Pannu et al., 1998).

Refinement is an iterative process in which the atomic model is modified, structure factor amplitudes are calculated from the modified model, and the agreement between these calculated structure factor amplitudes ($|F_c|$) and the experimental or observed ones ($|F_{obs}|$) is determined. The goal is to find the model that produces the best agreement between $|F_{obs}|$ and $|F_c|$. Refinement is a problem of finding the minimum of a function that mathematically expresses the agreement between $|F_{obs}|$ and $|F_c|$. This function is called a target function. A commonly used target function is the crystallographic residual: SUM $\{(|F_{obs}| - k|F_c|)^2\}$, where the

sum runs over all the reflections in your data set, and k is a scale factor needed to put the F_c on the same scale as the F_{obs}.

A model consists typically of five parameters for each atom: x,y,z, B, and Q. The triplet (x,y,z) specifies the position of each atom in an orthogonal coordinate system. B is the B-factor or temperature factor of each atom, and it is related to the thermal motion of the atom. B-factor also contains information about other types of "disorder" including errors that you are being made while constructing and refining model. Q is the occupancy and it is the fraction of time that the atom spends at position (x,y,z). Typically, Q=1. If one has data better than about 1.8 Å, then occupancies between zero and one are sometimes used.

Refinement procedures for antibodies involve two basic procedures, rigid body refinement followed by positional refinement. Rigid body refinement is used to refine the results obtained from MR in terms of orientation and position of the starting model in the unit cell. Positional refinement is used to refine the positions of individual atoms in space. Both conventional R-factor (R_{cryst}) and the free R-value (R_{free}) (Brünger, 1992) were used to monitor the progress of refinement. 10% of the reflections were set aside at random to monitor the R_{free} during refinement.

Rigid body refinement was carried out to further refine the positioning of the probe molecule in the target unit cell. The probe models that gave the highest correlation coefficients were thus subjected to rigid body refinement. Refinement was initially done using data in the range of 50 Å - 4 Å; thereafter data up to the maximum available resolution were added in a step wise manner. The Fab molecule can be defined as an assembly of four domains, the V_H, V_L, C_H and C_L. Consequently, rigid body refinement where these domains were considered as discrete rigid units was carried out.

The individual atoms were then refined by several cycles of conventional positional refinement, which uses the conjugate gradient minimization method. The proper weight term called Wa was calculated which was used for subsequent positional refinement (Brunger et al., 1987). In case of CNS this value was set to -1 and the program itself calculated these weights. The refinement was started using data in the range 50 Å - 4.0 Å and higher resolution data were added in a stepwise fashion. After complete data had been added, the F_o-F_c and $2F_o$-F_c electron density maps in addition to the composite omit map were calculated and displayed on an HP xw8400workstation (Hewlett-Packard Company, U.S.A.) using Coot (Emsley P, 2004). The electron density map was examined in the context of the model and the regions of the map where the electron density was not satisfactory or the model did not fit the density were identified. Using Coot, the residues were mutated to the sequence of the molecule of interest and wherever required, moved locally to correspond to the visible density. This was followed by refinement to check if the changes made could be accepted or not. Depending on the resolution to which X-ray data was available, anisotropic or individual B-factors were also refined. This process of model building and refinement was carried out iteratively until all the differences in sequence with the probe model had been accounted for and there was no ambiguity in the fit between the model and the electron density. Water molecules were included in the model using the 'water pick' program available in CNS only after a sufficient level of refinement had been achieved. This was followed by visual examination of the waters to avoid inclusion of spurious water molecules.

The *B*-value is the measurement of the displacement of an atom from thermal motion, conformational disorder and static lattice disorder. This vibration will smear out the electron density and will also decrease the scattering power of the atom as a

function of resolution. The displacement for an isotropic *B*-value is related to the displacement *u* by the equation $B = 8\pi^2(u^2)$. The isotropic *B*-value assumes equal movements in all the directions. However, the vibration of an atom need not be the same in all the directions, and in such a case motion is described by anisotropic displacement parameter. In this formulation the motion is described by an ellipsoid that can be rotated in any direction. The entire anisotropic displacement can be described in terms of six elements: U_{11}, U_{22} and U_{33} specify the magnitude of movement in three axis and U_{12}, U_{13} and U_{23} specify the rotation off the principal axis. Anisotropic displacement parameters can be converted to the isotropic equivalent by the formula $B_{iso} = 8\pi^2(U_{11}+U_{22}+U_{33})$.

The *B*-values are restrained during refinement. Atoms that are bonded to each other influence each other's motion. *B*-values are restrained in such a manner that the average difference in the B-values of bonded atoms is kept to a target value. The *B*-values should vary smoothly along the protein chain and within the side chain. The usual target restraint for adjacent bonded main chain atom is 1.0 and for side chain the target value is 1.5 since one end of the side chain is free, ensuring the higher gradient. Similarly B-values can be graded for the one to three members of a bond angle. For main chain angles the target value is 1.5 and for the side chain angle the value is set to 2.0. Like rigid body refinement, it is done at the early stages of the refinement process. Refinement of the atomic *B*-factors is a bit tricky and is carried out in later stages of refinement.

3.7 Validation, analysis and comparison of models

Deviations from ideal geometry of the various structures were analyzed using PROCHECK (R. A. Laskowski, 1993) from the CCP4 suite. PROCHECK

checks the stereochemical quality of protein structures. The output of the program consists of comprehensive listings of the stereochemical parameters. These files were analyzed to determine if the error in the various stereochemical parameters were within acceptable limits. If any parameter of the model was found to be outside the accepted range, they were corrected by adjustment of the concerned residue followed by refinement.

The elbow angles of all the Fab structures were calculated using web based applet developed by *Standfield et al* (Stanfield et al., 2006). The buried surface areas (i.e. the area rendered inaccessible to a 1.4 Å sphere) were determined using PISA web server (Krissinel and Henrick, 2007). CONTACT program of the CCP4 package (Elizabeth Potterton, 2003) was used to determine van der Waals contacts and hydrogen bonds between peptide and Fab. Van der Waals contacts were defined to be present between atoms if they were within 4 Å of each other. Hydrogen bonds were assigned for donors and acceptor atoms when the distance between them is less than 3.5 Å. In case of hydrogen bonds where the nitrogen atom is the donor, the N-H...O angle should be greater than 120°. When an oxygen atom is the donor, a cut off value of 90° was used.

All Fab-peptide complexes of mAbs; BBE6.12H3 and 36-65 were compared in terms of various parameters such as elbow angle, peptide and CDR conformation, buried surface area, van der Waal contacts and hydrogen bonding. To compare conformations, RMSD in the position of the C_α as well as all atoms were calculated using SUPERPOSE program of CCP4 package (Krissinel and Henrick, 2004). The CDR conformations in the liganded and unliganded forms were also compared.

RESULTS