STRUCTURAL STUDIES ON ACTIN-ADP RIBOSYLATING BINARY TOXIN FROM *C. DIFFICILE*

Volume 1 of 1

AMIT SUNDRIYAL

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

February, 2010

Copyright

ॐ
असतो मा सद्गमय,
तमसो मा ज्योतिर्गमय,
मृत्योर्माऽमृतं गमय ।

O Lord, lead me from the unreal to the ultimate truth,
from the darkness to light,
and from the death of ignorance to the immortality of knowledge.

Dedicated to

MY FAMILY and MY TEACHERS

ABSTRACT

Clostridium difficile infection (CDI) is a serious problem within the healthcare environment where the bacterium causes symptoms ranging from mild diarrhoea to life-threatening colitis. In addition to its principal virulent factors, Toxin A and Toxin B, some *C. difficile* strains produce a binary toxin (CDT) composed of two subunits namely CDTa and CDTb that are produced and secreted from the cell as two separate polypeptides. Once in the gut, these fragments have the potential to combine to form a potent cytotoxin whose role in the pathogenesis of CDI is presently unclear. This thesis is a step towards understanding structural and functional aspects of the binary toxin produced by *C. difficile*.

The first half of this thesis (chapter I and II) provides a brief introduction to the method of structure determination of proteins molecules, i. e. X-ray crystallography and a detailed overview of *C. difficile* and the three known toxins from *C. difficile* namely – Toxin A, Toxin B and the binary toxin. Chapter II further focuses on *C. difficile* binary toxin and other related toxins. These toxins, known as the ADP-ribosylating toxins (ADPRTs) form a big family of potent toxins which includes Cholera, Pertussis and Diphtheria toxins and are capable of transferring the ADP-ribose part of NAD/NADPH to a varity of substrates in the target cell which ultimately results in cell death.

The second half of the thesis comprises of experimental procedures that were carried out during the course of this study and their results. Cloning and expression methods for recombinant CDTa and CDTb in bacterial system followed by their purification are described with the abnormal behaviour exhibited by CDTb (chapter III). We show for the first time that purified CDTa and CDTb can combine to form an active CDT which is cytotoxic to Vero cells (Chapter IV). The purification processes described yielded

milligram quantities of binary toxin fragments of high purity that led to the successful crystallisation of the proteins (chapter IV) for further functional and structural studies.

High resolution crystal structures of CDTa in its native form (at pH 4.0, 8.5 and 9.0) and in complex with the ADP ribose donors - NAD and NADPH (at pH 9.0) have been determined (chapter V). The crystal structures of the native protein show 'pronounced conformational flexibility' confined to the active site region of the protein and 'enhanced' disorder at low pH while the complex structures highlight significant differences in 'ligand specificity' compared with the enzymatic subunit of a close homologue, *Clostridium perfringens* Iota toxin (Ia). These structural data provide the first detailed information on protein- donor substrate complex stabilisation in CDTa which may have implications in understanding CDT recognition. Crystallisation of CDTb yielded preliminary crystals. The optimisation of these crystallisation conditions is underway. The thesis concludes with some thoughts and discussion on future directions of this research.

ACKNOWLEDGEMENTS

With profound reverence for the Supreme Ruler of the Universe, I acknowledge with grateful heart, the goodness of Almighty God for invoking His Divine guidance and blessings on all my endeavours and imploring His aid and direction, which allowed me to pursue empirical research ultimately shaping the course of my future academic and professional activities.

I am grateful to the BBSRC, University of Bath and the Health Protection Agency (HPA) of United Kingdom for their interest and confidence in me and providing me this excellent opportunity to work with all necessary facilities as well as for funding my research.

I take this privilege to express my gracious thanks and regards to my supervisor Professor K. Ravi Acharya, Department of Biology and Biochemistry, University of Bath, United Kingdom for introducing me to the secrets hidden in the reciprocal space.

It seems to be the right moment to express my thanks and regards to our collaborators Dr. Clifford C. Shone, Dr. April Roberts, Joanna McGlashan and Roger Ling at the Health Protection Agency (HPA), Porton Down, United Kingdom for providing starting material for this research in one or the other form. I extend my sincere thanks to the staff members of Diamond Light Source, Didcot, Oxfordshire, United Kingdom for their assistance and beam time at the synchrotron.

I would like to thank my colleagues at Lab 0.34, University of Bath – Dr. Haryati Jamaluddin, Dr. Umesh Singh, Dr. Talat Jabeen, Dr. Konstantina Kazakou, Dr. Hazel Corradi, Dr. Elizabeth Clark, Dr. Shalini Iyer, Dr. Nethaji Thiyagarajan, Dr. Matthew Baker, Dr. Kenneth Holbourn and Dr. Paula Darley for their support and valuable advice during the tenure of study. My acknowledgements would not be complete without mentioning my vote of thanks to my close friends Sayantan Saha, Vivek Kumar, Aishwarya Verma and Saurabh Sharma for their love, support and care.

This moment has been a great opportunity to sincerely remember and to show my heartiest thanks to all of my teachers for shaping me and making me able, and to a long list of my relatives and friends, who have

always been an integral part of my life and the powerhouse of my confidence, for their immense love, affection, care and support from behind the curtains.

My Family has played a paramount role in shaping my career. Their restless efforts, blessings, patience, moral support, love and sacrifice always motivate me and provide me the strength and courage to counter all the problems. Mere a few lines or a few pages are quite less to express my deep sense of gratitude towards my family (Pita ji and Maa, Bade Tau ji and Tai ji, Chhote Tau ji and Tai ji, Chachi ji, Deepu, Nirmala, Banti, Dimple, Pinki, Pintu, Biggu, Pappu, Neetu, Poonam, Rinki, Sumit, Ritu, Ria, Priya, Shivam, Manas, Khushi and Soham) for their complete selflessness, love, commitment and perseverance in order to fulfill my needs and ambitions without which I could never have reached to these heights. Their work ethics, personal strength, and devotion to truth always set new challenges for me. I feel fortunate and proud to have a family which understands the preoccupations that go with this type of studies.

At this moment, my heart is heavy to remember Dadaji, Dadi and two of my best teachers who are not with me today to see the light of this auspicious day of my life. My Chacha and bade Mamaji have always been a source of inspiration to me and it is their guidance and trust in me which has landed me to this stage.

- AMIT SUNDRIYAL

LIST OF ABBREVIATIONS

ADP	Adenine diphosphate
ADPRT	ADP-ribosylating
ARTT	ADP-ribosylating turn turn
Bis-Tris	Bis(2-hydroxyethyl)imino-tris(hydroxymethyl)methane
C2I	Enzymatic component of *C. botulinum* binary toxin
C2II	Transport component of *C. botulinum* binary toxin
CCD	Charged couple device
CCP4	Collaborative Computational Project No. 4
CDI	*C. difficile* infection
CDT	*C. difficile* binary toxin
CDTa	Enzymatic component of *C. difficile* binary toxin
CDTb	Transport component of *C. difficile* binary toxin
CPD	Cysteine protease domain
CRD	C-terminal repetitive domain
CST	*C. spiroform* binary toxin
DMEM	Dulbecco's Modified Eagle Medium
DTT	dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
EF2	Elongation factor 2
GDP	Guanidine diphosphate
GST	Glutathion S-transferase
GT-A	Glycosyltransferase A
GTP	Guanidine triphosphate
HDVD	Hanging drop vapour diffusion
Ia	Enzymatic component of *C. perfringens* binary toxin
Ib	Transport component of *C. perfringens* binary toxin
IPTG	Isopropyl β-D-thiogalactoside
LB	Luria Bertani media
LCT	Large clostridial toxins
MAD	Multiple wavelength anomalous scattering
MBP	Maltose binding protein
MCS	Multiple cloning site

MES	2-(*N*-morpholino)ethanesulfonic acid
MIB	Sodium malonate, Imidazole, and Boric acid
MIR	Multiple isomorphous replacement
MMT	Malic acid, MES and Tris base
MWCO	Molecular weight cut off
NAD	Nicotinamide dinucleotide
NADPH	nicotinamide dinucelotide phosphate
NMN	Nicotinamide mononucleotide moiety
PDB	Protein data bank
pI	Isoelectric point
PMC	Pseudomembranous colitis
RMSD (or r.m.s.d.)	Root mean square deviation
SDS	Sodium dodecyl (lauryl) sulphate
PAGE	Polyacrylamide gel electrophoresis
SDVD	Sitting drop vapour diffusion
S_N1	Neucleophilic substitution reaction of first order
S_N2	Neucleophilic substitution reaction of second order
SPG	Succinic acid, Na-dihydrogen phosphate, Glycine
TB	Terrific broth media
TcdA	*C. difficile* Toxin A
TcdB	*C. difficile* Toxin B
UDP	Uridine diphosphate
UDP-Glc	Uridine diphosphoglucose
VIP	Vegetative insecticidal protein

TABLE OF CONTENTS

List of Commercially Available Crystallisation Screens Used

- Molecular dimension Structure Screens 1 and 2
- Molecular dimension Clear Strategy Screen 1
- Molecular dimension Clear Strategy Screen 2
- Molecular dimension Pact Premier Screen
- Molecular dimension JCSG Plus Screen
- Hampton Research Additive Screen

Publications

- **Sundriyal A., Roberts A. K., Shone C. C. and Acharya K. R. (2009).**

Structural Basis for Substrate Recognition in the Enzymatic Component of the ADP-ribosyltransferase Toxin CDTa from *Clostridium difficile*.
***J. Biol. Chem.* 284, 28713-28719.**

- **Sundriyal A., Roberts A. K., Ling R., McGlashan J., Shone C. C. and Acharya K. R. (2010).**

Expression, purification and cell cytotoxicity of actin-modifying binary toxin from*Clostridium difficile*
***Protein Expression and Purification.* 74, 42-48.**

CHAPTER - I

INTRODUCTION TO MACROMOLECULAR CRYSTALLOGRAPHY

Introduction

Proteins are biomolecules of fundamental importance to any organism from unicellular to multicellular composition. They are one of the building blocks of the basic unit of life i. e. the cell. Proteins play a vital role in most of the cellular events such as cell growth and differentiation, signal transduction, providing mechanical strength to tissues, immune protection, storage and transport, coordinated motion of muscles and catalysis of metabolism. Structure determination of a protein molecule (and other biomolecules) at atomic resolution provides insight into its function, mechanism of recognition of substrates and the conformational changes they might undergo (Blow, 2002). The area of protein crystallography is not only of academic relevance but it is also an important gateway to structure based drug design or development of therapeutics such as engineered antibodies and enzymes to alter functional capabilities of biomolecules.

X-ray crystallography is one of the various scientific methods available to determine and study the three dimensional structures of small inorganic or organic molecules and large biological macromolecules (Nucleic acids, proteins and their complexes). However, amongst all available methods, X-ray crystallography is the most favoured method for studying biological macromolecular structures because of its unique advantage of providing details at almost atomic resolution, its accuracy and reproducibility.

Why X-rays and Why Crystals

Biological molecules are very tiny objects with their largest dimensions in Å (C-C bond is 1.54 Å, 1 Å = 10^{-8} cm). In principle, an object can be seen only if the wavelength of electromagnetic radiation used to see it is of the order of its size. Hence, the atomic details can not be resolved by using visible radiations (wavelength of 4000-7000 Å). X-rays have wavelength in the range of 100 to 0.1 Å and thus towards the lower side of their spectra, they fulfil the above requirement and can be used to visualise molecules up to a resolution that is of the order of bond lengths. Typical wavelengths used for X-ray crystallography experiments lie in the range of 1.0 to 1.54 Å.

Direct result of an X-ray crystallography experiment is the diffraction pattern of a molecule. The diffraction pattern of any molecule is its characteristic property that depends on the number of electrons present, their

relative orientation in the molecule and their location in the crystal. Diffraction from a single molecule is not strong enough to be detected above the noice level on a detector. In a crystal, identical molecules of substance are arranged in a regular repetitive fashion and thus they all diffract the incident X-ray beam in an identical manner in all directions. Diffraction from millions of identical molecules in same direction adds up and the signal can be detected easily. In other words, crystals act as an amplifier to amplify diffraction intensities of reflected X-rays.

Steps Involved in Structure Determination

In principle, the process of structure determination by X-ray crystallography is carried out by following a series of steps essentially in an order as shown in figure 1.1.

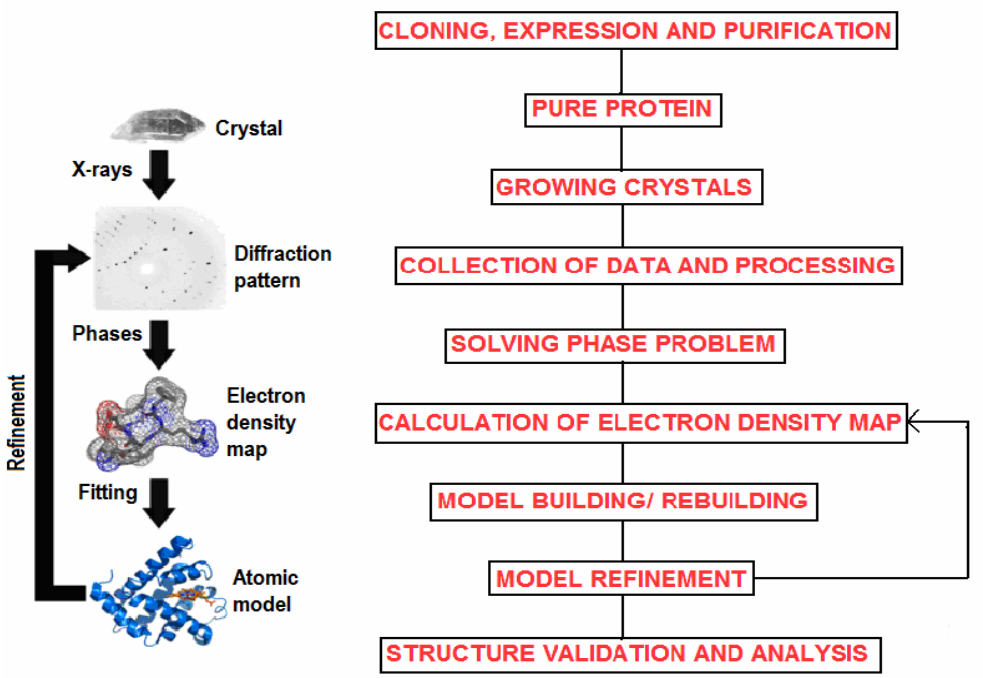

Figure 1.1: *The steps involved in the structure determination process of proteins by X-ray crystallography. (Figure partly adopted from http://en.wikipedia.org/wiki/X-ray_crystallography).*

However, the process of structure determination is not as simple and straightforward as it is illustrated above. An image of the molecule can not be drawn directly because of the unavailability of a lens to focus and recollect all scattered X-rays from the object which is a must condition (Blow, 2002). Therefore, the image of the molecule is generated by indirect methods

involving complex mathematical operations with the help of very fast and modern computers. Each step involved in the process of crystallographic structure determination is explained below in detail.

Cloning and Expression of Proteins

The first and foremost requirement in X-ray crystallography is the availability of good quality crystals. The process of structure determination by means of X-ray crystallography starts with the availability of a large quantity of extremely pure (generally > 95 % pure) homogeneous protein. As a rule of thumb, the diffraction data quality, up to a large extent, depends on the quality of crystals which in turn basically depends on the quality (i. e. purity and homogeneity) of the protein in hand.

Recombinant DNA technology provides excellent tools to produce a sufficient amount of protein in a cost and time effective manner. Discovery of several enzyme systems (and understanding of their mechanism of action) that play a vital role in the 'essential to survive' processes of central dogma (DNA, RNA and protein metabolism) have made molecular cloning almost a routine experiment in the laboratories these days.

In brief, coding DNA for the target protein can be identified. The DNA can be isolated from living cells and amplified *in vitro* using polymerase chain reaction (PCR) with the help of suitable oligonucleotide (primer) sequences and DNA polymerising enzymes. As an alternative way, coding DNA sequence for any naturally existing or hypothetical peptide sequence can be synthesised chemically. Ends of the PCR amplified or commercially synthesized DNA can be modified according to the convenience in order to construct suitable expression clones.

With the help of carefully chosen sequence specific restriction enzymes, this DNA fragment (called an insert or transgene) can be cut and inserted into a vector DNA that has compatible ends. These ends can then be sealed by using DNA ligase enzyme to produce a 'chimeric or recombinant DNA'. Positioning of the insert into the vector backbone can be regulated precisely to ensure that the inserted DNA is read in the correct reading frame. It is necessary to avoid an immature termination of transcription (and translation) and therefore, production of a mis-sense or nonsense mRNA (and thus protein), *in vivo*. The recombinant DNA is then inserted into a suitable host

organism where it replicates, transcribes and translates itself by exploiting the host cell machinery in a 'semi independent' manner.

Vector DNA is defined as a 'cloning vehicle' that has a property of self replication. Vectors, called expression vectors, are specific to carry out expression of the transgene in the host cell. These vectors generally have a promoter and other conserved sequences that are necessary for transcription of the transgene and translation of the resulting mRNA. Simpler vectors (called cloning vectors) can only replicate in the host cell but can not transcribe the gene and thus do not result in the expression of desired protein. Unlike expression vectors, cloning vectors are used only for *in vivo* amplification of the insert.

Plasmids are the most widely used cloning vectors. They are double-stranded, generally circular DNA sequences consisting of an 'origin of replication' that allows for a semi-independent replication of the plasmid in the host. Plasmids have a multiple cloning site (MCS) which consists of various restriction enzyme consensus sequences. The MCS provides freedom to choose a combination of available restriction sites for cloning purpose with a choice of reading frame to read the transgene.

In addition to plasmids, many other cloning vehicles such as cosmids, phasmids, viral vectors, bacterial artificial chromosome (BAC) and Yeast artificial chromosome (YAC) are also available and are used when convenient. Each type of vector has its own set of advantages and disadvantages over others.

Almost all vectors bear a positive selection marker usually in the form of a gene that translates for an antibiotic resistance. This property of a vector serves two elementary purposes. Firstly, it acts as a selection pressure on the host and only vector bearing cells (positive cells) can survive on a growth medium that contains that particular antibiotic. Secondly, in the presence of selection pressure (antibiotic containing medium) it becomes mandatory for the host to carry and maintain (replicate, transcribe and translate) the vector in order to survive against the applied selection pressure. Since the gene of interest is also contained by the vector, under favourable conditions, a good yield of recombinant protein is produced by the host.

Expression vectors exhibit diversity in their expression patterns. It can either be constitutive (consistent expression) or inducible (expression only

under the influence of certain growth conditions or chemicals). Expression pattern is a characteristic of the promoter that is present in the vector. Inducible expression depends on promoters that respond to specific induction conditions. Inducers are added to the growth medium and taken up by the host cell in order to start transcription of the inserted gene and hence translation of the target protein.

The next step is to select a suitable host organism for expression of the target protein. There are several different **host systems** available. They can be classified as animal cell, plant cell, yeast cells, insect cell and bacterial cell systems. It is possible to divert metabolism of the expression host towards overexpression of the target protein by providing it specific growth conditions such as substrates, aeration, effectors (inducers and enhancers) and temperature. However, the growth condition requirements of host systems differ from each other. The host system is chosen depending upon the nature of target protein. For example, if the target protein is of eukaryotic origin and requires heavy posttranslational modifications (an antibody molecule, for example), an eukaryotic expression system (animal cell, insect cells or yeast cell culture) is chosen whereas if the target protein is simple (for example a protein of bacterial origin) and does not require any post translational modification machinery, it can be expressed in prokaryotic expression systems.

Plant cell cultures suffer from the disadvantage of very slow growth rate and hence are very rarely used systems. Animal cells are commonly used to express proteins of eukaryotic origin but respond to a narrow range of growth conditions such as substrate, pH and temperature. Bacterial expression systems and specially *E. coli* bacterial cells are the most widely used prokaryotic host systems. They are easy to manipulate genetically and provide a high expression rate due to their fast metabolism and short doubling time which is beneficial to produce a large amount of the target protein in a relatively short period of time.

Generally, proteins are expressed as fusion proteins with a suitable 'tag' that is of great help in the down streaming processes. Fusion proteins are created by joining two or more genes which originally code for two separate proteins, one of which is the target protein. Translation of the fusion gene results in a single polypeptide with functional properties derived from each of

the original proteins. Production of proteins as 'fusion proteins' overcomes many of the expression-purification associated problems. Fusing the target protein to a suitable tag sometime enhances the fusion protein expression and may retain the expressed fusion protein in soluble form. Another important advantage of the tag is in purification as discussed in the next section. Both of the genes (tag and target) can be linked via a linker DNA region that codes for a peptide sequence which can be recognised by suitable protease. The fusion partner (tag) can then be cleaved off from the target protein by using these specific proteases at a carefully chosen suitable step during purification.

Protein Purification

As indicated earlier, quality of the protein to start with is one of the bottle necks in the process of crystallographic structure determination of proteins. No matter what expression system is used to overexpress the target protein, it would be expressed along with several of other proteins that are normally produced by the host. The aim of the purification process is to isolate the target protein from such a crude mixture of proteins. In principle, a protein should be more than 95 % pure for crystallisation purpose.

The overexpressed protein can be released by lysing the host cells and purified either from the crude cell lysate (in the case of soluble proteins) or from inclusion bodies (aggregated form of proteins). There are several techniques available to break open the cells such as mechanical disruption, liquid homogenisation, sonication, freeze/thaw and the enzyme mediated cell lysis. The choice of cell lysis method depends on how sensitive the protein is, the amount to be processed, how sturdy the cells are and the location of the target protein (compartmentalisation). After extraction, soluble proteins can be separated from cell membranes, DNA and insoluble proteins by centrifugation, prior to their purification whereas insoluble proteins (inclusion bodies) need to be solubilised first and then refolded prior to or during the purification procedure. Sometimes, the target protein is released into the growth media by the expression host which is mostly the case with animal cell cultures.

Ease of the purification process depends on the nature of compartmentalisation of the expressed protein as well as on the stability of the protein under the chosen physiochemical environment. Different physiochemical and biological properties of proteins can be used to develop

purification strategies ensuring high recovery of the purest form of homogeneous, stable and non-denatured protein. Listed in table 1.1 are four most basic properties that can be used.

Table 1.1: *Different properties of proteins that form the basis for different purification strategies.*

S. No	Property	Based on	Example
1	Biological activity	specific interaction	Affinity Chromatography
2	Charge	Net surface charge	Ion exchange Chromatography
3	Size	Molecular weight	Gel permeation Chromatography
4	Solubility	Hydrophobic interactions	Reverse phase Chromatography

All of the different chromatography processes described above rely on the distribution of target substance (protein) in two phases known as the stationary and the mobile phase. The mobile phase with substance (and impurities) is passed through the stationary phase where different components get separated based on their distribution coefficient between the two chosen phases.

Affinity chromatography takes advantage of the biological activity and specificity exhibited by one molecule towards the other such as antibody-antigen and enzyme–substrate systems. Generally to facilitate the purification using affinity chromatography, the target protein is expressed as a fusion protein with a tag at the N or the C terminal of the target protein (page 7). The most commonly used tags are poly-Histidine tag (His-tag), glutathione S-transferase (GST) tag and maltose binding protein (MBP). These tags (and thus fusion proteins) bind to specific molecules that have been immobilised on a stationary support matrix and thus can be trapped. Release of the bound tag (or the fusion protein) can then be achieved by altering physiochemical conditions of the mobile phase so as to alter its (tag's) affinity for the immobilised material or by using a substrate that competes with the tag to bind to the immobilised purification matrix. For example, proteins with a poly-

Histidine extension (His tag) can be trapped by a nickel chelating matrix and the bound proteins can then be released from the matrix by passing imidazole through the matrix which competes with the His-tag for binding to nickel ions immobilised on the matrix because of the imidazole ring that is present in Histidine.

Ion exchange chromatography exploits the charge property of the protein and is based on the coulombic interactions between the protein molecules and the stationary phase. Amino acids and hence proteins exhibit zwitter ion characterises. The isoelectric point (pI) of a zwitter ion is defined as the pH value at which it acquires a net zero charge. At any pH below its pI, the zwitter ion possesses a net positive charge whereas at a pH above its pI, it possesses a net negative charge. By choosing suitable buffer conditions, proteins can be forced to bind to an immobilised matrix of complementary charge (negatively charged proteins on a positively charged matrix and vice versa). The bound proteins can be released selectively by changing the pH or ionic strength of mobile phase and thus can be separated from each other.

Size exclusion or Gel filtration chromatography (GFC) is a separation technique based on the hydrodynamic volume (size in solution) of the molecules and hence the separation is achieved on the basis of differences in their molecular size. A crude protein sample is passed through a porous stationary phase. Larger molecules that can not access the pores exit the column more rapidly. Smaller molecules penetrate into the porous structure and get trapped according to their size. Retention time of a molecule in the pores is indirectly proportional to its molecular weight (size). The smaller the molecule, the longer the retention time and thus the later the molecule is released from the matrix and vice versa.

Hydrophobic interaction chromatography is another process, based on the hydrophobic interactions between the matrix and the protein molecules which can also be used effectively for purification. High pressure can be applied to drive the solute faster through a column, thereby improving the resolution. The most common form of High Pressure Liquid Chromatography (HPLC) (Regnier, 1983) is the “reverse phase” HPLC, where the column material is hydrophobic and proteins elute according to their hydrophobicity using a gradient of an organic solvent (such as acetonitrile). However, HPLC

often causes denaturation of proteins and is sometimes not appropriate for molecules that do not spontaneously refold.

In most of the cases, purification is a multistep process. More than one of the strategies listed above are chosen carefully and employed in different combinations based on characteristics of the target protein and impurities present to achieve highest purity of the target protein. During the process of purification, the quality of purified protein is monitored by Sodium Dodecyl Sulphate-Poly Acrylamide Gel Electrophoresis (SDS-PAGE) or Western blotting analysis from time to time. The quantity of proteins can be monitored by one of the several available methods such as absorbance at 280 nm or other colorimetric methods like the Bradford's method or the Lowry's method.

Growing Protein Crystals

The process of crystallisation involves controlled precipitation of the protein from its supersaturated aqueous solution such that it does not form amorphous aggregates (Rhodes, 2000). The aim of the crystallisation process is to produce large diffraction quality crystals.

Crystallisation of any substance occurs when the concentration of substance is higher than that of its saturation limit at that temperature. The state of supersaturation is a nonequilibrium state that results in precipitation of the substance from solution until the equilibrium state (saturation point) is reached. In principle, crystallisation is a two step process: nucleation and crystal growth (McPherson, 1999; McPherson, 2004). Nucleation is the step where protein molecules start aggregating in a supersaturated protein solution by overcoming an energy barrier under given experimental conditions. This is then followed by a growth step where more and more protein molecules aggregate on the formed nucleus resulting in sufficiently large crystals and the entire system attains the state of equilibrium. The process of crystal growth can be understood with the help of crystallisation phase diagram (Figure 1.2).

The phase diagram shows the solubility of a protein in a solution as a function of concentration of the protein and the precipitant present. Nucleation takes place in the nucleation zone whereas the crystal growth occurs in the metastable zone. If the concentration of the protein and/or precipitant is not enough for supersaturation to enter in the nucleation zone, no crystals would grow. However, if supersaturation is attained too quickly

and continues beyond the nucleation zone into the precipitation zone, excessive nucleation or an amorphous precipitate may result. To prevent this from happening, a careful screening of crystallisation condition variables, such as starting protein concentration, crystallising agent (precipitant) concentration, pH of solutions and incubation temperature, is needed. This process usually requires setting up hundreds of different crystallisation conditions.

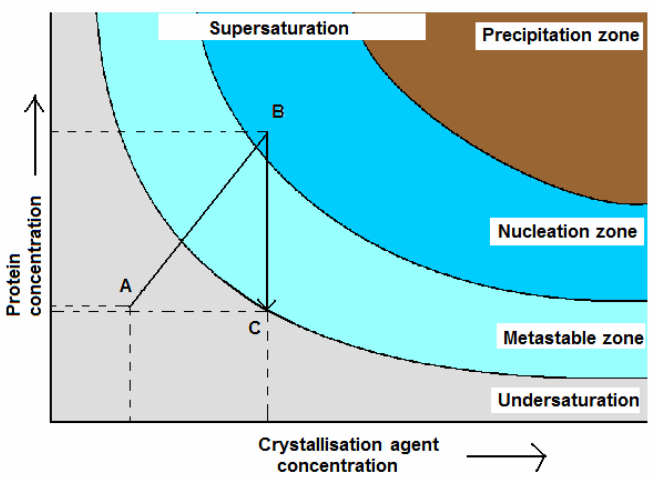

Figure 1.2: *The crystallisation phase diagram. Zones of undersaturation and oversaturations are shown in different colours.*

Though the mechanism of crystal growth is known very well, crystallisation is still the key limiting step in the success of structure determination process because of the involvement of a large number of variables. Each protein requires its own set of conditions to crystallise. Factors involved in the process can be classified into two categories.

1- Controllable parameters – pH, concentration of protein, concentration of precipitants, temperature etc.
2- Uncontrollable parameters – gravity, magnetic and electric field, vibrations, kinetics of reaction etc.

Protein molecules are big in size and have irregular shape. They never crystallise without large solvent channels between molecules. The advantage of these solvent channels is that their presence provides us an excellent way to study ligand binding. Crystals can be soaked with the ligand solution and

ligand molecules can diffuse through these channels to the active site and bind there. Also, since proteins always remain in contact with the solvent while they are in the crystal, the effect of crystal packing is negligible on their overall structure and protein structures in crystals resemble their structures in free solution. However, the presence of these solvent channels makes protein crystals extremely fragile and highly sensitive to their physiochemical environment.

Methods of Protein Crystallisation

Following are the most commonly used methods of protein crystallisation.

Batch crystallisation method is the most ancient method of crystallisation. A large volume of protein is directly mixed with the precipitating (crystallising) agent such that the state of supersaturation is reached immediately. The system is then left undisturbed for several days to achieve slow precipitation of the protein that results in the attainment of equilibrium state and thus yields crystals. A variation of this technique known as 'Microbatch' is also used where a small drop of a mixture of the protein and crystallising agent is left undisturbed under a layer of oil. Use of the oil prevents evaporation of volatile solvents from the drop.

In the method of **Dialysis,** a protein solution is separated from a large volume of crystallising agent by the use of a semi permeable membrane. Slow movement of solvent through the membrane results in an increase in the protein concentration and ultimately leads to the crystal growth. This method, however, requires a large quantity of protein in comparison to other crystallisation methods.

Vapour Diffusion technique is the most widely used technique. This technique can be used with two variations depending upon the mode of drop setting (Figure 1.3) – sitting drop vapour diffusion (SDVD) and hanging drop vapour diffusion (HDVD). More common among the two is the hanging drop method. A small volume of the protein and precipitant are mixed together and suspended over a reservoir of the precipitant in a close system. The precipitant concentration in the reservoir is maintained higher than that in the protein drop. Due to the concentration difference, the solvent molecules diffuse from the protein solution (drop) to the reservoir solution until the vapour pressure

of the drop attains equilibrium with the vapour pressure of the reservoir solution. This event leads to an increase in the precipitant and the protein concentration in the drop and thereby increasing the degree of saturation of the protein which, if the physiochemical conditions are chosen optimally, leads to the nucleation and then crystal growth.

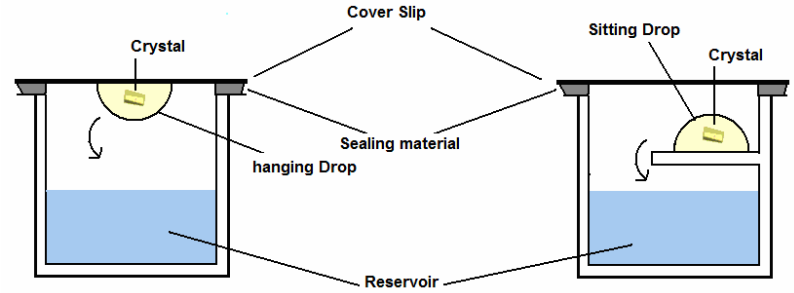

Figure 1.3: *The two variants of vapour diffusion method of crystallisation – hanging drop (left) and sitting drop (right). The arrow shows direction of diffusion of vapours in the closed system.*

The process of crystal growth can be explained with the help of a phase diagram (Figure 1.2). Crystallisation is set up at point A where the protein and the crystallising agent are mixed such that the final protein concentration in the drop remains at undersaturation state. As the drop is allowed to equilibrate against a large volume of reservoir containing a higher concentration of crystallising agent (generally twice of that in the drop), volatile solvents start diffusing in the direction from the lower (drop) to the higher concentration (reservoir solution). As a result, the concentration of the protein and the crystallising agent in the drop starts increasing and the system reaches point B which is in the nucleation zone of the supersaturation state. This state is a nonequilibrium state and hence the protein molecules start aggregating together to form the nucleus for crystal growth. The protein concentration in the drop starts decreasing. More and more molecules of protein aggregate together to attain equilibrium and soon the system enters into the metastable zone where no more nucleation can occur but the protein still keeps aggregating on already formed nucleus to attain an equilibrium state. As a consequence, the size of the growing crystals increase till the protein concentration in the drop drops down to point C where it enters into the undersaturation state again and the crystal growth ceases.

Since conditions for nucleation and crystal growth phase may differ, sometimes **seeding** becomes necessary to grow protein crystals. The technique of seeding has been used successfully when either the condition that results in an excessive nucleation does not allow further crystal growth due to protein depletion following too much nucleation or to improve crystal quality when the originally grown crystals do not diffract up to the mark. A small fraction of nucleated crystals is transferred to a new drop under suitable growth conditions which may or may not differ from the nucleation condition. Depending upon how seeding is performed and the size and the number of seeds transferred, the seeding is categorised as macroseeding, microseeding or streak seeding.

Crystals and Symmetry

Crystals are a regular repetition of objects (protein molecules in this case) in three-dimensional space. The smallest unit of a crystal that repeats itself throughout the crystal purely by its translation in three dimensions is called the **unit cell**. A unit cell can be defined by six parameters – three edges **a, b, c** and three angles **α, β** and **γ** between them. The location of an atom within a unit cell is described by a set of three cartesian coordinates (x, y, z) with respect to the origin at one of the vertices of the cell. The smallest unit of a crystal that repeats itself throughout the crystal by its rotation and translation is called the **asymmetric unit**. The unit cell may contain more than one asymmetric unit arranged in patterns that are characteristic of what symmetry the crystal possesses. The geometry of the unit cell together with the possible symmetry operations defines the **space group** of the crystal.

In addition to rotational and translational symmetry, the unit cell of a crystal can contain screw axis where the asymmetric unit is not only rotated around the axis, but also translated by a fraction of the unit cell length. Screw axis is denoted as a subscript number related to the fraction translation of the unit cell. For example, 2_1 is a two-fold rotation axis with a screw corresponding to the translation of half of the unit cell length. Furthermore, the asymmetric unit may consist of more than one molecule interrelated by the non-crystallographic symmetry (NCS). Figure 1.4 below illustrates a two dimensional lattice.

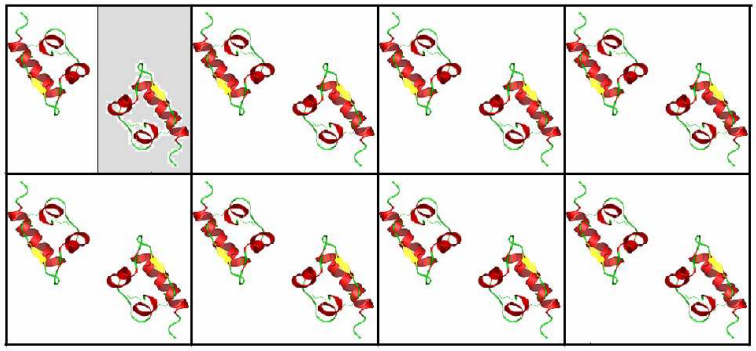

Figure 1.4: *A hypothetical protein sitting in a two dimensional lattice, with a 2 fold rotational symmetry, along the axis perpendicular to the plane of the paper. Each square block represents one unit cell and the shaded part represents the asymmetric unit.*

Symmetry poses restrictions on the shape of the unit cell. Crystals can be assigned to one of the 7 possible crystal systems which are further divided into 14 lattice types depending upon the position of lattice points within the unit cell (Figure 1.5 and Table 1.2). **Primitive lattices** are the crystal systems that contain one point at each corner of the unit cell and are designated by letter P. The **Non-primitive lattices** have additional points either at the centre of the unit cell faces (designated as face centred – C or F) or at the centre of the unit cell itself (designated as body centred – I). The seven primitive lattices along with the seven non-primitive lattices are called the **Bravais lattices** (Blundell & Johnson, 1976).

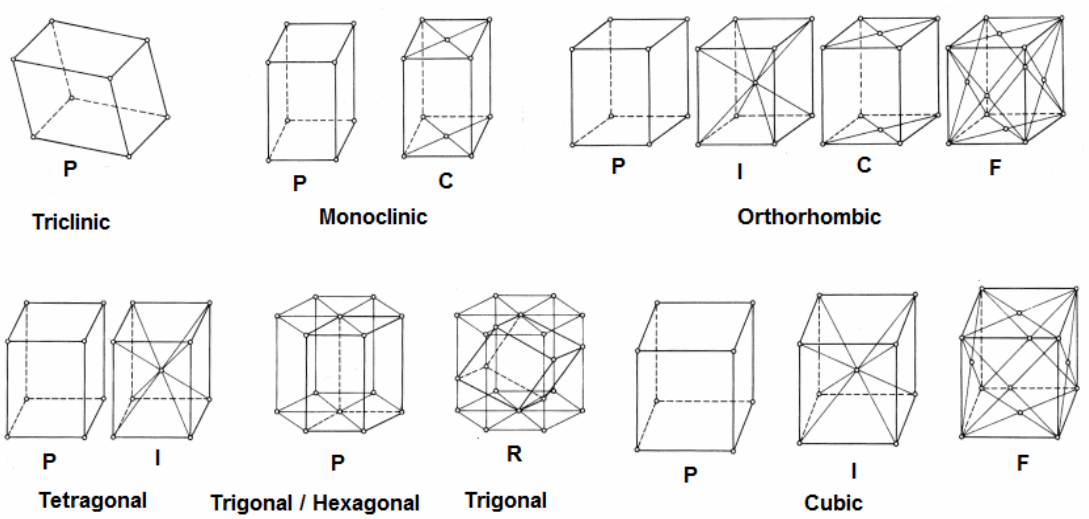

Figure 1.5: *The 7 Crystal systems and 14 Bravais lattices (P – primitive, C – centred, I – body centred, F – face centred) (adopted from http://perso.fundp.ac.be/~jwouters/DRX/diffraction.html).*

Table 1.2: *The fourteen Bravais Lattices and their associated symmetry point groups.*

Name	**Bravais lattice types**	**Restrictions on unit cell**	**Point groups**
Triclinic	P	$a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$	1
Monoclinic	P, C	$a \neq b \neq c$; $\alpha = \gamma = 90° \neq \beta$	2
Orthorhombic	P, C, I, F	$a \neq b \neq c$; $\alpha = \beta = \gamma = 90°$	222
Tetragonal	P, I	$a = b \neq c$; $\alpha = \beta = \gamma = 90°$	4, 422
Trigonal	P (or R)	$a = b \neq c$; $\alpha = \beta = 90°$, $\gamma = 120°$ $a = b = c$; $\alpha = \beta = \gamma < 120°$, $\neq 90°$	3, 322
Hexagonal	P	$a = b \neq c$; $\alpha = \beta = 90°$, $\gamma = 120°$	6, 6222
Cubic	P, I, F	$a = b = c$; $\alpha = \beta = \gamma = 90°$	23, 432

Owing to the chiral nature of biological macromolecules, not all 230 possible space groups are allowed for protein crystals as they can not possess mirror symmetry or inversion symmetry. Hence the allowed space groups for protein crystals are only 65 (Blundell & Johnson, 1976).

Diffraction and Bragg's Law

When a crystal is exposed to a beam of X-rays, the incident beam is scattered in all possible directions. This scattering can be of two types- coherent scattering and non-coherent scattering. Diffraction results from the coherent scattering whereas the non-coherent scattering leads to the absorption of energy by atoms in the crystal. Coherently scattered (diffracted) X-rays interfere constructively ('in phase' with each other) in certain directions and give rise to the observed diffraction pattern that is recorded on a detector.

In 1913, the phenomenon of diffraction was explained by W. H. Bragg and W. L. Bragg. They considered diffraction as a result of simple reflection taking place from a plane mirror. Crystals are made up of several families of planes (called lattice planes or Bragg's planes) passing through the lattice points (Figure 1.6). Any family of planes is identified by its Miller indices h, k,

and l which are integers representing how many times that particular plane repeats itself in a unit cell in all three x, y and z directions respectively.

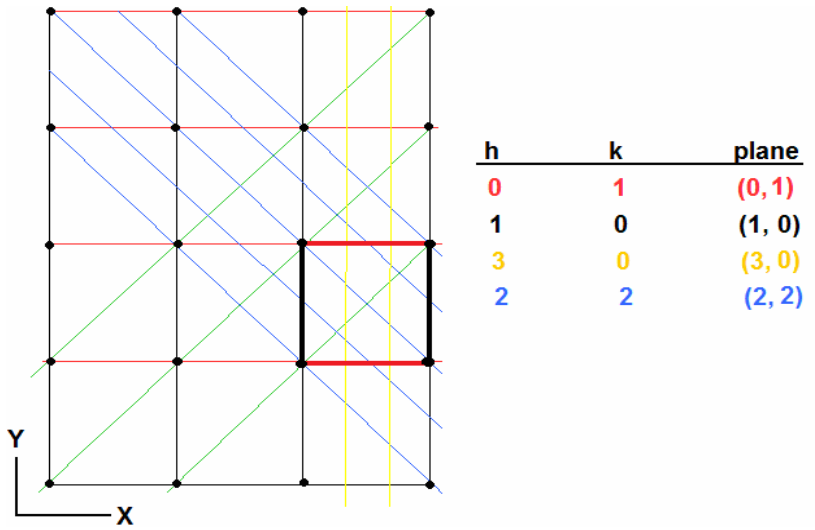

Figure 1.6: *A representation of different families of Bragg's planes through a two dimensional crystal lattice.*

For example, in figure 1.6, a family of planes shown in blue lines, repeats itself twice in the X direction and twice in the Y direction, in the unit cell (which is shown in thick lines). Hence, its miller indices will be h = 2, k = 2 and this particular family of planes will be denoted as (2, 2).

Braggs proved that for a given angle of incidence (θ) of X-rays on a plane, any family of planes would diffract (scatter coherently) the incident X-rays only and only if the interplanar distance (d) between two consecutive planes in the family and the wavelength of incident X-rays obeys the following relation.

$$2d \sin\theta = n\lambda \quad \text{---------- (1)} \qquad \text{(Where n is an integer)}$$

Figure 1.7 below is a schematic representation of Bragg's law. This law implies that for a given wavelength of X rays, reflections resulting from the diffraction of X-rays from closely spaced families of planes will be at a larger angle of reflection and thus will be recorded away from the centre of the detector and vice versa.

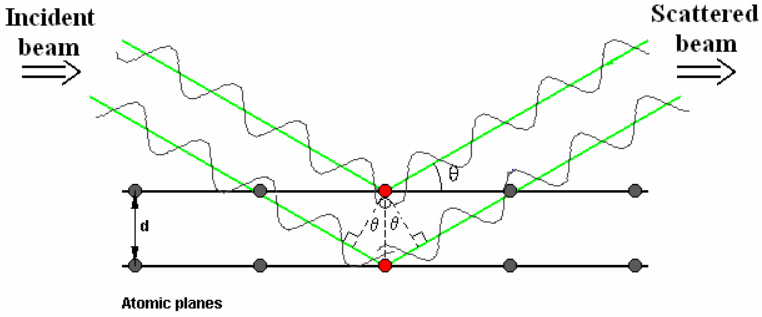

***Figure 1.7**: A schematic representation of Bragg's law. D is the interplanar distance between two consecutive planes in the family and θ is the angle of incidence of X-rays on the plane.*

Reciprocal Lattice and Ewald's Sphere

Bragg's law can give an exact estimation of the angle of diffraction but does not provide any information about the position of reflection with respect to the origin in 3-dimensional spaces. This piece of information is obtained from the reciprocal space concept and the Ewald's sphere.

The reciprocal space is an imaginary 3-dimensional space where the reflected spots (as a result of diffraction) are assumed to be situated. It is clear from the Bragg's law (equation 1) that for a given wavelength of incident X-ray, a family of planes with a narrower interplanar distance would lead to a reflection observed at a wider angle on the detector and vice versa. An arbitrary origin is chosen and a perpendicular is drawn on any family of parallel planes (Bragg's planes) with an interplanar distance d in the real space and a spot at a distance 1/d from the chosen origin is identified on this perpendicular. This spot is called reciprocal lattice point corresponding to that set of planes. This essentially means that one family of planes in the real space lattice produces only one spot in the reciprocal space. The position of any spot in the reciprocal space can be given by three indices h, k, l, known as Miller indices, which are none other than the indices of the set of planes that gives rise to that reciprocal point and hence that particular reflection on the detector. All such spots together constitute a reciprocal lattice corresponding to the real space lattice.

The Ewald's construction (Figure 1.8) is a geometrical representation of the reciprocal lattice. It is a sphere of radius $1/\lambda$ (where λ is the wavelength of incident X-ray) and the crystal is assumed to be situated at the centre of the sphere (point A). The origin of the reciprocal space is assumed at the point where the direct beam leaves the sphere (point O).

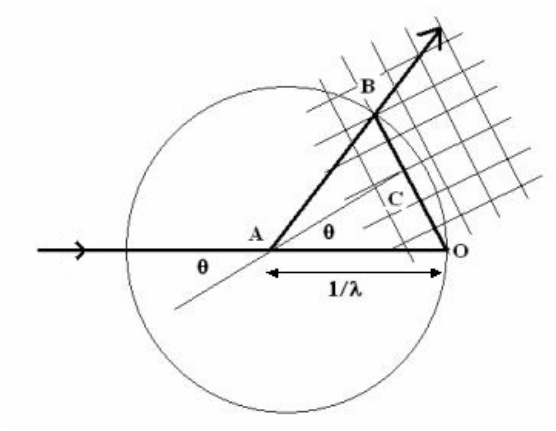

Figure 1.8: *The Ewald's sphere and the reciprocal lattice construction (adapted from Rhodes, 2000). Direct beam leaves the sphere at O (origin of the reciprocal lattice) and the crystal is situated at the centre A. Reciprocal lattice point B is in the diffracting condition and line AB shows the direction of the diffracted ray whereas the point C can be brought in the diffracting condition by rotating the sphere around the origin O.*

It can be shown with the simple laws of geometry and trigonometry that when a reciprocal point (point B) falls on the surface of the Ewald's sphere, it fulfils the condition given by the Bragg's law and thus gives rise to a reflection in the direction along the line joining the centre of the sphere to that reciprocal point on the surface of the sphere (along the line AB). In figure 1.8 the reciprocal space point B is in the diffracting condition.

However, in any particular orientation of the crystal (and thus of the reciprocal lattice) not all reciprocal points can be brought on the surface of the Ewald's sphere to give rise to a diffracted reflection. This also means that in any particular orientation of the crystal in the beam (which, from Bragg's law, essentially implies that at a particular angle of incidence of X-rays, *θ)* not all families of Bragg's planes can be brought into the diffracting positions. To do so, the Ewald's sphere has to be rotated with respect to the reciprocal lattice keeping the origin fixed in order to make all the reciprocal points fall on the surface of the Ewald's sphere (in diffracting position) in one or the other orientation of the reciprocal lattice. This forms the basis of the most commonly used method of data collection – 'the rotation method'.

The higher the intensity of the incident ray, the more intense will be the reflections in the reciprocal space for a given time of exposure. Also, for a given intensity of the beam, higher the electron density corresponding to a set of planes in the real space, the more will be the number of waves that will be

coherently scattered from that particular set of planes and more intense the corresponding spot will be. Two pieces of information for any reflection (spot on detector) that can be drawn directly from the collected diffraction data are the position (h, k, l,) and the intensity (I_{hkl}) of the reflection.

X-ray Generators and Detectors

X-rays of wavelengths in the range of 1.0 – 1.54 Å are usually used for crystallography purposes and are obtained from one of the two types of x-ray generators.

In **laboratory (or home) sources,** a beam of electrons originated at a cathode is focused onto a metal anode target through a strong electric potential. These high energy electrons cause transitions of the metal atoms at anode which result in the production of electromagnetic radiation of a wide range of energies and hence of varying wavelength, known as 'white radiation'. This includes some strong characteristic radiations corresponding to the excitation of inner cell electrons of the metal. Copper is the most commonly used target metal and produces characteristic radiation of CuKα, (1.54 Å wavelength) and CuKβ (1.39 Å wavelength). Molybdenum may be used as an anode if X-rays of shorter wavelength are required (MoKα and CuKβ, 0.71 and 0.63 Å wavelengths respectively) (Blundell & Johnson, 1976). Using appropriate filters, the white radiation can be converted to a monochromatic X-ray beam by removing the weaker (Kβ and other much weaker) radiation. A filter made of an element with atomic number Z-1 effectively blocks the Kβ radiation produced by a metal of atomic number Z (Ni is an effective CuKβ blocker) (Rhodes, 2000). Home sources can again be classified as 'sealed tube' (stationary anode) and 'rotating anode' generators. Although, in-house X-ray sources are very convenient and reliable, their use is limited by their low intensity beam and inability of tuning the wavelength.

Tremendous advancements in technology in the past 25 years have made data collection much quicker. At **synchrotron radiation sources**, electrons are generated in an electron gun (Figure 1.9) and are accelerated with the energy of several giga-volts (Helliwell, 1997). These electrons, moving almost at the speed of light, under the influence of an electric field, are fed into an outer storage ring (Blow, 2002). In the storage ring, these fast moving electrons are forced to revolve in a circular path via a magnetic field and hence

to emit electromagnetic radiations in the line tangential to their path (Figure 1.9). The emitted electromagnetic beam is then carried from the storage ring to the experimental area through a high vacuum beamline (Helliwell, 1992) collimated by the mirrors and filtered to make it a monochromatic beam. Synchrotrons have the advantage of fast data collection using highly intense beam. Another major advantage of synchrotron sources is the ability to tune the wavelength of the X-ray beam.

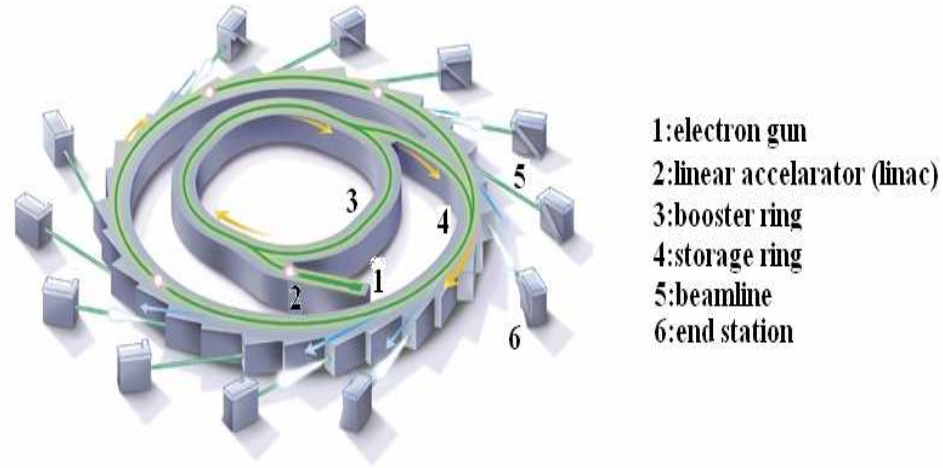

Figure 1.9: *A schematic representation of a Synchrotron source and its parts (adopted from http://www.warren.usyd.edu.au/bulletin/NO51/ed51art8.htm).*

Detection of the diffracted X-rays is a crucial part of an X-ray crystallography experiment. There are several types of detectors available to record the diffraction data. Recording the data on X-ray films is now an obsolete method. **Charged couple device (CCD) detectors** are the most advanced types of detectors**.** These detectors are characterised by their fast read out time and high noise reduction capability.

Crystal Mounting and Data Collection

For mounting in the beam, crystals can be loaded into a glass capillary with the crystallisation solution (the mother liquor), and sealed at both the ends. Another approach is to loop mount the crystals. Crystals are scooped into a tiny loop, made of nylon or plastic, supported by a solid rod and then held in the beam.

The capillary or the loop containing crystals is then mounted on a goniometer, which allows it to be positioned accurately within the X-ray beam and rotated. Since both, the crystal and the beam are often very small, the

crystal must be centred within the beam. The most common type of goniometer is the "kappa goniometer", which offers three angles of rotation: the ω angle, which rotates about an axis perpendicular to the beam; the κ angle, about an axis at ~50° to the ω axis; and, the φ angle about the loop/capillary axis. The oscillations (rotation) carried out using the rotation method of data collection involve the ω axis only.

The primary data quality plays an important role since data collection (Figure 1.10) is the last experimental step in X-ray crystallography (Dauter, 1999). While collecting the data one needs to ensure that the collected data is complete as much as possible,

1) – quantitatively, and
2) – qualitatively

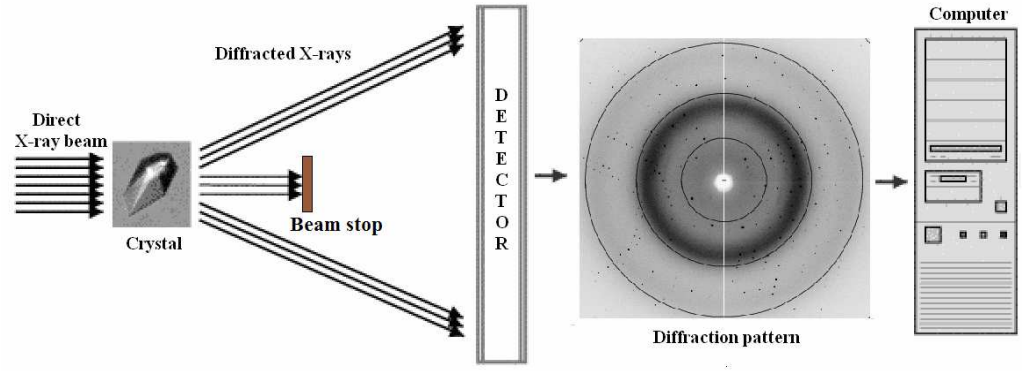

Figure 1.10: *Arrangement of a typical X-ray crystallographic data collection experiment.*

Factors that influence data collection can be classified into two classes. **Quantitative factors** (such as total rotation angle for which the data has to be collected and the wavelength of incident beam) ensure that we record as many reflections as possible. Quantitative factors basically depend on crystal geometry and the experimental set up.

The **wavelength** of x-rays can be chosen based upon the nature of the experiment. Any wavelength is suitable for native data collection. Usually, a higher resolution can be obtained using a shorter wavelength X-rays. Shorter wavelength also reduces damage to the crystals due to the absorption of radiation, termed as **radiation damage**.

Exposure time affects the intensity of each spot (reflection) in the diffraction pattern. A shorter exposure time leads to the loss of high resolution weak reflections whereas a longer exposure may result in the saturation of low-resolution spots (termed as overloads). Hence, the exposure time should be chosen carefully to compensate the both. In addition, a longer exposure time increases radiation damage to the crystal.

One image of the spots is insufficient to reconstruct the diffraction pattern of the whole crystal. Hence the crystal is rotated in the beam and many images are collected. The total **angle of oscillation** required to collect a complete data set depends on the symmetry of the unit cell. For a crystal possessing no symmetry (Triclinic, page 16), at least 180^0 rotation data is needed to ensure completeness of the collected data. For higher symmetry space groups the total angle of rotation required is less, as there are more symmetry related reflections. Usually, data over a larger range of oscillation is collected to reduce the signal to noise ratio and to improve the redundancy of the data (Blundell & Johnson, 1976). The total range of oscillation is achieved in several steps of small angle of rotation per image ($\Delta\Phi$, usually of 1^0 per image for protein crystals). $\Delta\Phi$ is chosen depending upon the unit cell parameters and the arrangement of spots on the image to avoid overlapping of the spots or collecting too many of partially recorded spots.

Crystal-to-detector distance determines the resolution of the collected data. The further the detector from the crystal, the lower will be the resolution (Figure 1.11) (Evans, 1999).

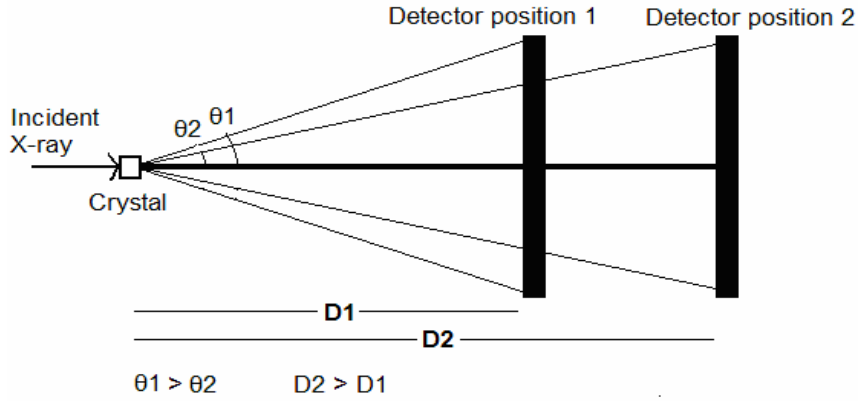

Figure 1.11: *Effect of the crystal to detector distance on resolution range (the larger the θ, the higher will be the resolution).*

Mosaicity of the crystal refers to the internal disorder of the crystal. Ideal crystals are like a brick wall where bricks are arranged regularly. However, real crystal lattices can deviate from the ideal and are not the perfect lattices. High mosaicity can result in overlapping spots and data loss. High mosaicity can easily be detected on a diffraction pattern as broadened spots (more like a smear) than circular.

Another factor that influences the data collection is the movement of incident radiation beam. X-rays are never ideally monochromic (single wavelength X-rays). This phenomenon is known as the **beam divergence**. Combined effects of the crystal mosaicity and the beam divergence can cause a particular reflection to be spread over a range of crystal rotations (same reflection appearing over more than 1 image, partially) (Dauter, 1999).

On the other hand, **qualitative factors** (such as R factor and signal to noise ratio) indicate that the collected data is of the best possible quality under the given experimental conditions. These factors depend on the method employed in the data collection (Dauter, 1999) and are discussed in the data processing section on page 26.

Cryogenic Data Collection

The crystallographic data can be collected either at room temperature or at lower temperatures. Low temperature (at 100 K) data collection is more common. In theory, lowering the temperature increases molecular order in the crystal and thus improves the diffraction pattern. However, the crystal is soaked in a cryoprotectant before freezing it to avoid the formation of ice crystals during the data collection.

Another advantage of maintaining the crystal at a cryogenic temperature is that it prevents diffusion of free radicals from the site of primary radiation damage in the crystal and thus saves the crystal from further damage called secondary radiation damage. It provides the crystal with a longer life span and allows the experimenter to collect more and more data without damaging the crystal in the beam.

Cryogenic data collection, however, has some disadvantages as well. Selection of cryoprotectant is a trial and error method. The wrong choice of cryoprotectant may lead to cracking or even shattering of the crystal.

Sometimes, transferring the crystal to a low temperature may also result in an increased mosaicity.

Concept of Resolution

The amount of structural information that can be extracted from a crystal depends on the resolution to which the crystal diffracts the incident beam (Table 1.3). Being able to bring families of planes with narrower interplanar distances to the diffracting positions essentially means that being able to acquire higher resolution data for the crystal.

Table 1.3: *The structural information obtained from a crystal based on the resolution.*

Resolution (Å)	Structural information that can be obtained
6.0	Outline of the molecule and secondary structure features (e. g. helices, strands) can be identified.
3.0	Course of the polypeptide chain can be traced and topology of the folding can be established. With the aid of the amino acid sequence, it is possible to place the side chains within the electron density map.
2.0	Main chain conformations can be established with great accuracy. Details of the side chain conformations, bound water molecules, metal ions and cofactors can be identified.
1.5	Individual atoms are almost resolved. It is possible to figure out almost all solvent molecules.
1.0	Hydrogen atoms may become visible.

A family of closely spaced planes diffracts at a higher angle of diffraction (Bragg's law, $\theta \propto 1/\lambda$). Hence, higher resolution spots are always collected far from the centre of the detector. Also, Bragg's law clearly indicates that for a given angle of incidence, with a shorter wavelength of incident X-rays, families of planes with smaller interplanar distances can also be brought in the diffracting positions ($d \propto \lambda$) which means that a higher resolution can be obtained. A high resolution data gives information about the finer details of the structure. However, low resolution data is equally important for structure

determination as it contains information about the overall structure. For example, a 6 Å (low resolution) data set can provide information about outlines of the molecule and its secondary structure features while individual atoms can be easily fitted into higher resolution data (Table 1.3).

Data Processing

The crystallographic diffraction data is collected as two dimensional images full of diffracted reflections. To determine the structure of the molecule, this data needs to be processed. Data processing is a complex multi step process which includes –

(1) - indexing of the data and measurement of cell parameters,
(2) - refinement of cell and detector parameters,
(3) - integration of the data, and
(4) - scaling of the data.

The first step in data processing is the determination of the unit cell dimensions and the crystal system. At this stage, based on the diffraction pattern, peaks are picked and indexing of the diffraction pattern is performed depending upon the position of peaks (Rossmann and van Beek, 1999). A complete search of all possible indices is performed. Finding values (integers) for one index (for example, h) for all reflections is equivalent to having found one real-space direction of the crystal axis (for example, **a**). After the search for the real space vectors is completed, the program finds three linearly independent vectors with minimal determinant (unit cell volume) that would index all the observed peaks to determine unit cell dimensions, Bravais lattice and the crystal orientation.

This procedure usually provides with more than one choice for space group with their respective distortion coefficient which is an indication of to what extent the unit cell parameters for that particular space group have to be distorted in order to make it a perfect cell. The selected space group would be the one that has highest order of symmetry with lowest distortion coefficient. Further processing of the data proceeds using the initial estimates of cell parameters for selected space group as reference. Crystal to detector distance,

wavelength and oscillation range (phi values), are the input values needed in order to complete the process of autoindexing.

Autoindexing is followed by refinement of cell and detector parameters and integration of whole data. Usually, autoindexing is done with only one or a few of the recorded images. Integration of data refers to the conversion of hundreds of collected images to one file consisting of the Miller indices and corresponding intensities for each reflection.

Scaling is the final step in data processing. A scale factor is applied so that the intensities from all images of the data set can be related. The scaling of intensities is needed because the diffraction quality of the crystal degrades with time as it depends on mosaicity, air and crystal absorption, radiation damage etc. The first image usually has a scale factor of 1 and all the subsequent images will be scaled up to this (Smyth and Martin, 2000). The step of scaling averages the processed data while accounting for errors that occur during the data collection. The output of the scaling process is a list of reflections with systematic absences that is characteristic of the space group that had been chosen during indexing.

The whole process of 'processing the data' produces a list of indices and their corresponding scaled intensities for all the recorded reflections and provides important statistical information about the quality of data such as completeness of data, signal to noise ratio and reliability factor. Each of the above steps involves many complex calculations. Therefore the entire process is carried out with the help of computer programs using sophisticated algorithms. The most frequently used programs are MOSFLM (Leslie, 1992) and HKL 2000 / HKL Package (Otwinowski and Minor, 1997). The quality of processed and scaled data can be assessed by following statistics:

Completeness of data is the ratio of the number of unique reflections recorded to the total number of unique reflections possible. The higher the value, the more information can be obtained from the processed data.

$$\text{Completeness} = \frac{\text{Number of recorded unique reflections}}{\text{Total number of possible unique reflections}}$$

$\mathbf{R}_{sym}$ is an estimate of disagreement between the measured intensities of symmetry related reflections. A low R_{sym} value indicates less errors in the data

collection and hence more precision. If two or more data sets are scaled together, the R value is termed as $\mathbf{R}_{merge}$. For a typical data set of 2.0 Å, an R_{sym} of 10-12 % is within the acceptable limit (Blow, 2002).

Signal to noise ratio (I / σI) is the ratio of intensity (I) to the error in recording that intensity (σI). This value is indicative of the accepted resolution of the data set as reflections with error ratio of (I/σI) < 2.0 can not be distinguished from the background noise and may contain errors.

Redundancy, or multiplicity of the data refers to how many times all symmetry related reflections have been recorded. High redundancy is an indicative of accuracy in intensity measurement.

$$\text{Redundancy} = \frac{\text{Total number of reflections recorded}}{\text{Total number of unique reflections recorded}}$$

Interpretation of Data – Diffraction to Structure

Fourier proposed a method called Fourier Transformation (FT) to analyse complicated mathematical functions which are repetitive in nature. These complicated functions can be represented as a series of functions that are an integral multiple of a fundamental function. Since, crystals are also a repetitive function of a fundamental function i. e. the unit cell, they also can be analysed by applying the Fourier transform on them. More accurately, crystals are built from repetitive blocks of electron density. This electron density varies from point to point inside the unit cell but if we look at the crystal as a whole, this electron density repeats itself again and again in a regular fashion. Hence, by applying the Fourier transformation, the electron density at any point in a unit cell can be used to determine all its Fourier components (in case of waves; the amplitude, frequency and phase). Therefore, by working in the opposite direction (known as the inverse Fourier transformation) if the amplitude, frequency and phase components of the function are known, they can be used to calculate electron density at any point in the unit cell. This situation, however, is more complicated because a unit cell is a three dimensional object. Furthermore, each spot observed in the diffraction pattern appears not as a result of scattering from one electron in the unit cell but as a result of the constructive interference between waves scattered from all of the electrons present in the unit cell. Therefore, we need

to apply the inverse Fourier transform in all 3 directions within the volume of the unit cell to calculate the electron density at each and every point in that volume.

The recorded reflections on the diffraction pattern, represent a sum of waves, diffracted from atoms on planes in the real space and are known as 'structure factors'. A three dimensional wave can be expressed in the following form:

$$f_{(xyz)} = f_{hkl}\, e^{2\pi i\alpha} \quad \text{---------- (2)}$$

Where f_{hkl} is the amplitude component of the wave and α is the phase component. The h, k, and l are the frequency terms of the wave in all three directions respectively i. e. by definition, how many times the wave repeats itself per unit cell in all three directions. Hence, the sum of all of the waves coherently interfering and producing a reflection in the reciprocal space can be represented as:

$$F_{(hkl)} = \Sigma_h\Sigma_k\Sigma_l f_{hkl}\, e^{2\pi i\alpha} \quad \text{---------- (3)}$$

This equation is known as the 'structure factor equation' corresponding to the reflection hkl (the Miller indices of that reflection or of the family of planes from which that particular reflection is originated). This way, structure factors for each and every reflection recorded on the detector can be calculated. Since the structure factor is the Fourier transform of electron density (ρ), another form of equation (3) can be written as

$$F_{(hkl)} = \int_v \rho_{xyz}\, e^{2\pi i (hx+ky+lz)}\, dx.dy.dz \quad \text{---------- (4)}$$

Where, v is the volume of the unit cell. An inverse Fourier transform of equation (4) results in equation (5) -

$$\rho_{xyz} = 1/v\, \Sigma_h\Sigma_k\Sigma_l F_{(hkl)}\, e^{-2\pi i (hx+ky+lz)}$$

$$= 1/v\, \Sigma_h\Sigma_k\Sigma_l f_{(hkl)}\, e^{2\pi i\alpha}\, e^{-2\pi i (hx+ky+lz)} \quad \text{---(5)}$$

Equation (5) gives us the value of electron density at any point x, y, z in the volume of the unit cell provided that we have estimated all the structure factor amplitudes, frequencies and phases.

Obtaining Phases

The recorded diffraction pattern of a crystal is the Fourier transform of the electron density of its unit cell content. In principle, the Fourier transform is reversible and therefore it is possible to reconstruct the electron density in a unit cell from its diffraction pattern. However, from the electron density equation (equation-5) it is clear that to determine the electron density at any particular point in the unit cell (x, y, z) we need to know three parameters – (i)- the amplitude factor (f_{hkl}), (ii)- the frequency factor (h, k, l) and (iii)- the phase (α) components for all the Fourier terms i. e. for all the diffracted waves.

The amplitude and the intensity of a wave are interrelated (the intensity is directly proportional to the square of the amplitude). In the process of data collection we only record intensities corresponding to reflections and hence amplitudes for all reflections can be easily determined. The frequency terms are nothing else but the Miller indices of the reflections. These values for all observed reflections have already been determined during the process of indexing the data.

However, the third vital piece of information for each reflection – 'The Phase', is lost during the process of data collection and needs to be determined indirectly. This problem of losing phases in the data collection is termed as the **Phase Problem**. There are three main methods of solving the phase problem which can be used depending upon the type of the problem encountered. Any of these methods, however, does not provide with the actual and accurate phase information. An initial estimate of phases is calculated which is refined and improved subsequently (Taylor, 2003).

Isomorphous replacement is a classical method of solving the phase problem. The principle of this method is that the contribution of any atom to the structure factor arising from the plane that intersects its position is proportional to the number of electrons present in the atom. Proteins are formed of C, N, O and S atoms which share almost same number of electrons in them. If a heavy atom, with an exceptionally large number of electrons is introduced uniformly in the crystal, the intensities of reflections corresponding

to the families of planes containing the heavy atom increase because of the additional scattering of waves by the heavy metal atom. This ultimately increases the amplitude factor of the corresponding structure factor equation for that reflection. In this method two different data sets are collected - one for the native crystal and the other for the heavy metal derivative crystal (Green *et al.*, 1954).

The condition that applies in this method is that both of the crystals should essentially be isomorphous i. e. both crystals should belong to the same space group with not more than 5% change in their cell parameters. The resulting intensities for both data sets are compared to retrieve phase information of heavy metal atom substructure that is present in the crystal. Positions of heavy metal atoms in the unit cell can be identified and can further be used to build the protein model. In normal practice, more than one heavy metal derivative is used and the method is called multiple isomorphous replacements.

The method of **anomalous scattering** exploits the property of **Friedel's law**. According to Friedel's law, each set of planes produces two reflections given by hkl and –h-k-l which are equal in their intensities ($I_{hkl} = I_{-h-k-l}$) but differ in their phases exactly by 180^0. This makes all diffraction patterns centrosymmetric. Heavy metal atoms are incorporated into the protein crystal and the diffraction data is collected at the absorption edge of the incorporated heavy atom. This results in the absorption of radiation and Friedel's law brakes down ($I_{hkl} \neq I_{-h-k-l}$). The absorption edge of an atom is defined as the wavelength at which the atom absorbs X-rays.

By comparing the intensities of Friedel's pairs of native and anomalous data (collected at the absorption edge), positions of heavy atoms in the unit cell can be determined. The anomalous scattering technique overcomes the problem of isomorphism as both, the native and the anomalous data sets can be collected from one single crystal by changing the wavelength to the absorption edge of the incorporated heavy atom. Usually, data sets are collected at several wavelengths in order to maximise the absorption (Taylor, 2003) and the method of phase extraction is called Multiwavelength Anomalous Dispersion (MAD) method (Hendrickson and Ogate, 1997).

The use of a combination of above two methods is also becoming common. This technique is known as Single Isomorphous Replacement with Anomalous Scattering (SIRAS).

Molecular replacement (Rosmann and Blow, 1962) is a method for phase estimation where a similar structure is known (Figure 1.12). Popularity of molecular replacement is increasing as more and more structures are being deposited in the Protein Data Bank (PDB). The success of molecular replacement method depends on the availability of sufficiently homologous structure. The higher the primary sequence identity, the higher are the chances, that the proteins will assume similar kind of three dimensional fold. As a rule of thumb, if the structure to be solved shares more than 30 % sequence identity with another protein whose structure is available, the molecular replacement method of phase estimation can be applied.

In principle, this method exploits the property of reversibility of the Fourier transform and Patterson synthesis. A suitable protein with known structure (and hence known phases) is selected as a model and **Patterson maps** of the model molecule and the target unit cell content are calculated. A Patterson map is a Fourier transform of the structure factor amplitudes only and does not require phases. It represents all possible atom to atom vectors and thus relative positions of atoms with respect to each other. The Patterson map of the model is rotated first and then translated (Figure 1.12) within the unit cell to obtain the correct orientation of the target in the unit cell relative to the origin (Taylor, 2003).

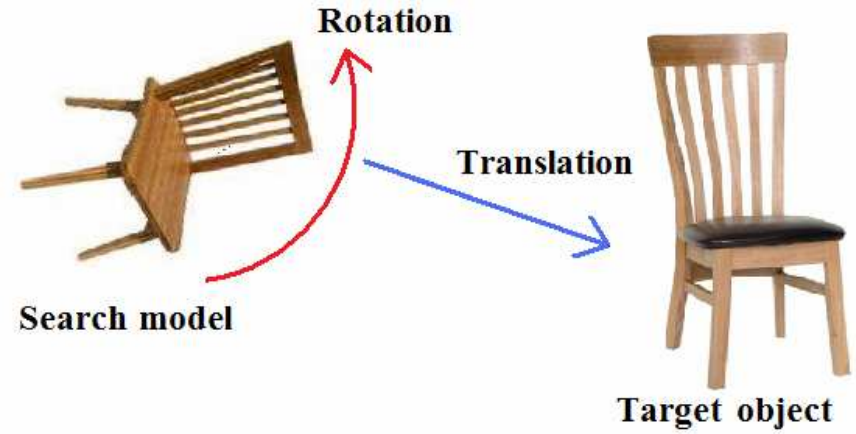

Figure 1.12: *A schematic illustration of the process of molecular replacement. The target is similar but not identical to the model.*

This operation of finding a rotation matrix '[**R**]' and a translation vector '**t**' relates the Patterson map of the model (**M**) to the Patterson map of the target structure (**X**) according to the flowing equation -

$$\mathbf{X = [R].M + t}$$

Phases from the model (called calculated phases) are then associated with the observed structure factor amplitudes of the target molecule from diffraction data to calculate an initial electron density map according to equation – (5). Several computer programmes such as AMoRe (Navaza, 1994), MolRep (Vagin and Teplyakov, 1997) and PHASER (McCoy *et al.*, 2007) are available to assist the whole operation.

Model Building and Refinement

The calculated phases are combined with the observed structure factor amplitudes from the diffraction pattern and a starting set of structure factor equations is calculated. These structure factor equations are used to calculate an initial electron density map of the molecule by using equation-(5). The quality of map at this stage depends on the quality of collected data and errors in phase estimation. Electron density maps can become biased towards the model if phases have been estimated by molecular replacement. This is termed as **model bias**.

Many rounds of crystallographic model building and refinement are then carried out in a cyclic process (Figure 1.1, page 03) aiming to improve the agreement between the observed data and the atomic model that has been calculated by using phases from the search model. The cyclic process of model building and refinement is usually repeated until, ultimately, a model is generated which represents the observed data as closely as possible.

In order to reduce the model biasing of phases, usually a $2F_o$-F_c Fourier map is calculated. In this map electron density at any point is calculated using the structure factor amplitudes equal to a sum of twice the observed structure factor amplitudes ($|F_{obs}|$), minus the calculated amplitudes ($|F_{calc}|$) $\{\mathbf{(2|F_{obs}| - |F_{calc}|)}\}$ in equation-5. This map represents a positive continuous density of the model. The structure factor amplitudes of this map are = $|F_{obs}|$ if the model is perfect. This map has a larger contribution of $|F_{obs}|$ and hence

if the model misses parts or is not perfect, this map shows the missing parts up with less intensity.

Another map, known as the F_o-F_c maps is also generated by using a sum of the observed structure factor amplitudes, minus the calculated amplitudes $\{(|\mathbf{F_{obs}}| - |\mathbf{F_{calc}}|)\}$ in equation 5. The electron density corresponding to this map is zero if the model is perfect, positive if some parts are missing in the model but present in the structure and negative if parts are absent in the structure but present in the model.

In addition to protein molecules, crystals contain water molecules which are bonded to the protein by hydrogen bonding and ligands that were incorporated during the crystallisation or by soaking the crystal in ligand solutions. These molecules also need to be modelled. The electron density corresponding to these molecules is visible in the Fo-Fc map. Water molecules can be added manually or automatically with the help of computer programs used in refinement such as ARP/wARP (Lamzin and Wilson, 1993). Atomic coordinates for ligand molecules can be obtained from the database such as the HIC-UP server (Kleywegt and Jones, 1998). Alternatively, coordinates for ligand can be obtained from the PRODRUG server (Schuettelkopf and van Aalten, 2004) or from the Sketcher application of CCP4 (CCP4, 1994).

Interpretation of electron density maps and model building, however is a laborious exercise which has been made easier by the development of several softwares such as O (Jones *et al.*, 1991) and COOT (Emsley and Cowtan, 2004).

Model building is followed by refinement where adjustments are made to bring the calculated structure factors close to the observed structure factors. Preliminary progress of refinement is assessed by the reliability factor (or the R factor) and improved model geometry. More appropriately, refinement is a process that produces the most biologically meaningful structure from the experimental data. The model parameters that are refined in each cycle of refinement include the position (x, y, z), occupancies and thermal factors (B-factors) of atoms. Generally used programmes for refinement are REFMAC (Murshudov *et al.*, 1997) and CNS (Brunger *et al.,* 1998).

Refinement can be started at a low resolution in order to reduce the model bias and to avoid the entrapment in local minima. The resolution can subsequently be increased in one or more steps. This strategy proceeds with

the correction of the gross features of the model first and ensures that the wrongly assigned details do not bias the model. Initially, the model is refined by rigid body refinement in which the protein molecules are refined as if they are rigid bodies and no relative movement of different domains is allowed.

Another basic type of refinement is restrained refinement where some freedom of movement within a narrow range of limits is allowed to the parameters to be refined. Adding restraints increases the observations to parameter ratio and therefore a good technique is to restrain the geometry of the protein tightly so that the phases could become more accurate as distortions in local geometry cannot be assigned without good phases.

B-factors are the atomic displacement factors. B factors represent the distribution of positions occupied by an atom over a period of time (dynamic disorder) as well as variations in the position of an atom between different unit cells (static disorder) (McRee, 1993). B factor refinement is an example of restrained refinement. Large B-factor values are usually indicative of errors in the model coordinates.

Structure Validation

Validation is used to access the quality of the refined structure. It is a process of checking quality of the structure against basic laws and known knowledge of science. The measure of success of refinement process can be assessed by several means.

The **R-factor** is the primary quality parameter of a structure. At the end of every cycle of model building and refinement, the difference between the calculated structure factors amplitudes and the observed structure factor amplitudes begin to converge and the value of R-factor drops. The R factor is calculated as below

$$R_{cryst} = \frac{\Sigma \, ||F_o| - |F_c||}{\Sigma |F_o|} \quad \text{---------- (6)}$$

Another important quality accessing parameter is R_{free}. This concept was first coined by Brunger (Brunger, 1992). A set of randomly selected reflections (known as the test set) is taken out from all the available data and not used in the refinement. The rest of the data with which the refinement is

carried out is termed as working set. The R_{cryst} indicates agreement (or disagreement) between the observed data (working set) and the calculated data. On the other hand, the R_{free} is calculated in a similar way but for the test set. Since, the R_{free} is calculated against the experimental data and not the model, there is no model bias in the refinement of the test set. The advantage of R_{free} is that it indicates about wrongly or over fitted data.

Root mean square deviation (r.m.s.d) is another statistical parameter that helps in assessing the quality of a structure by indicating the deviation of covalent bond lengths and bond angles from their ideal values. A low r.m.s.d. value indicates that the geometry of the molecule is good and that the refinement was carried out properly. Usually an r.m.s.d. value < 0.2 for bond angles and <0.02 for bond lengths is considered acceptable.

Ramachandran plots (Ramachandran *et al.*, 1963) are good indicators of accuracy of protein models. A Ramachandran plot indicates whether the main chain backbone dihedral angles (Φ – Ψ angles) fall into the allowed range to form protein secondary structure elements. PROCHECK (Laskowski *et al.*, 1993) and MOLPROBITY (Davis *et al.* 2007) are two useful programs that can assist in the assessment of the quality of the structure at various stages of refinement.

Deposition of Atomic Coordinates with the Protein Data Bank

There is no definitive point when refinement of a structure is completed. As a rule of thumb, when the R_{cryst} and the R_{free} stabilise, structure refinement is considered to be completed. Refined and validated structures are then made available to the public. The protein data bank (at either European Bioinformatics Institute, EBI; http://www.ebi.ac.uk/ or with the Research Collaboratory for Structural Bioinformatics, RCSB; http://www.rcsb.org/) is a global repository for structural information for X-ray crystallographic data. Not only the refined atomic coordinates for the protein but the experimental data, protein sequence and other parts of information are also deposited through a web interface such as AutoDep. The PDB facilitates an open access to all structures deposited world wide.

CHAPTER - II

CLOSTRIDIUM DIFFICILE AND ITS KNOWN TOXINS

Introduction to *Clostridium difficile*

Clostridia are Gram positive, spore forming, anaerobic, rod shaped bacteria. They are motile bacteria that are widely distributed in nature with their special prevalence in soil. They are commonly found in the gastrointestinal track of many animals including humans (Barth *et al.*, 2004). Clostridia are closely related to *Bacillus* genera (Shimizu *et al.*, 2002; Read *et al.*, 2003). Along with *Bacillus*; they are thought to constitute the first bacterial population on the earth (Fox *et al.*, 1980). Beside their genetic similarities, the two genera are well known for their ability to produce a variety of toxins which makes them potent pathogens of eukaryotic cells.

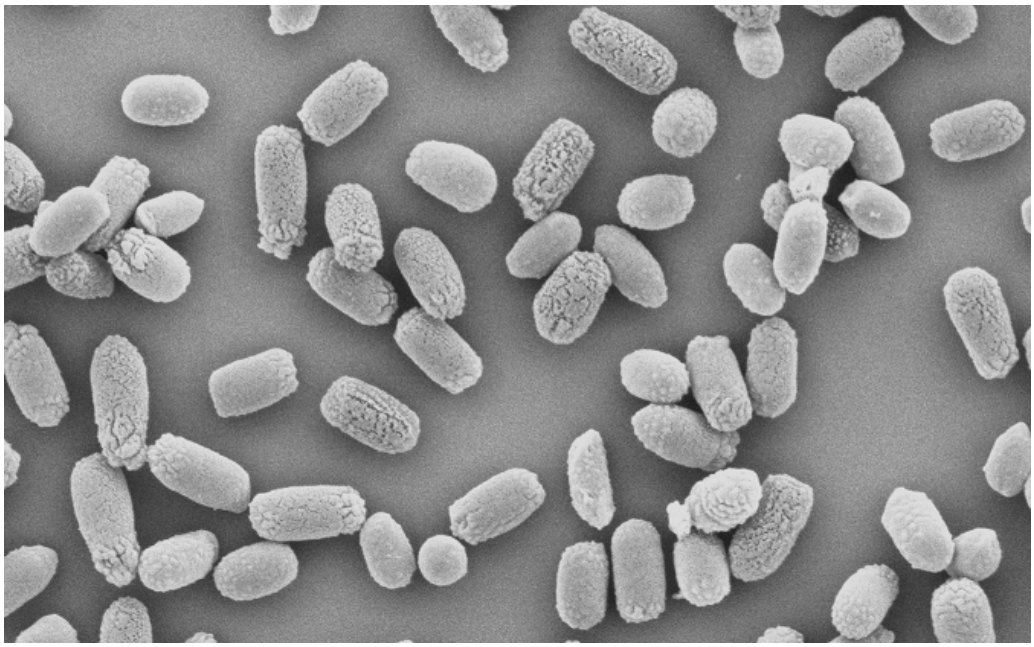

Figure 2.1: *An electron microscopic photograph of Clostridium difficile spores (figure obtained from Health Protection agency, U.K.)*

Clostridium difficile (Figure 2.1), originally known as *Bacillus difficile* was described in 1935 for the first time (Hall and Toole, 1935). In 1978, *C. difficile* was isolated from patients undergoing antibiotic treatment. The bacterium was soon identified as the primary cause of pseudomembranous colitis (Voth and Ballard, 2005). It was found that *C. difficile* causes disease almost exclusively in the presence of exposure to antibiotics. *C. difficile* is the only known anaerobic bacterium that produces toxins in the colon (Bartlett and Perl, 2005).

***Clostridium difficile* Infection**

Clostridium difficile is an important nosocomial pathogen. *C. difficile* infection (CDI) is known to be responsible for almost all cases of pseudomembranous colitis (PMC) and hospital acquired diarrhoea worldwide (Elliott *et al.*, 2007). Elderly people are more at risk. CDI is recognised by a wide variety of symptoms ranging from mild self limiting diarrhoea to more severe life threatening pseudomembranous colitis. The more serious issue is that the infection occurs in hospitalised individuals who have undergone antibiotic treatment (Hurley and Nguyen, 2002) and hence the disease is called Hospital Acquired Diarrhoea. *C. difficile* is resistant to several antibiotics and antimicrobial agents which give the bacteria a selective advantage over other microbes that results in *C. difficile* associated outbreaks in healthcare facilities.

Reports suggest that almost 3% of healthy and up to 40% of hospitalised individuals are colonised with *C. difficile* (McFarland *et al.*, 1989). In healthy individuals, the bacteria remain in spore form under normal conditions and only go back to their active – vegetative form, when the normal intestinal flora gets disturbed upon exposure to antibiotics (Bartlett and Perl, 2005). In general, any therapeutic agent, procedure or illness that disturbs the normal intestinal flora may give rise to CDI (Riley, 1998). Clindamycin and cephalosporins have been considered as significant cause of PMC (Tedesco *et al.*, 1974; Gerding, 2004).

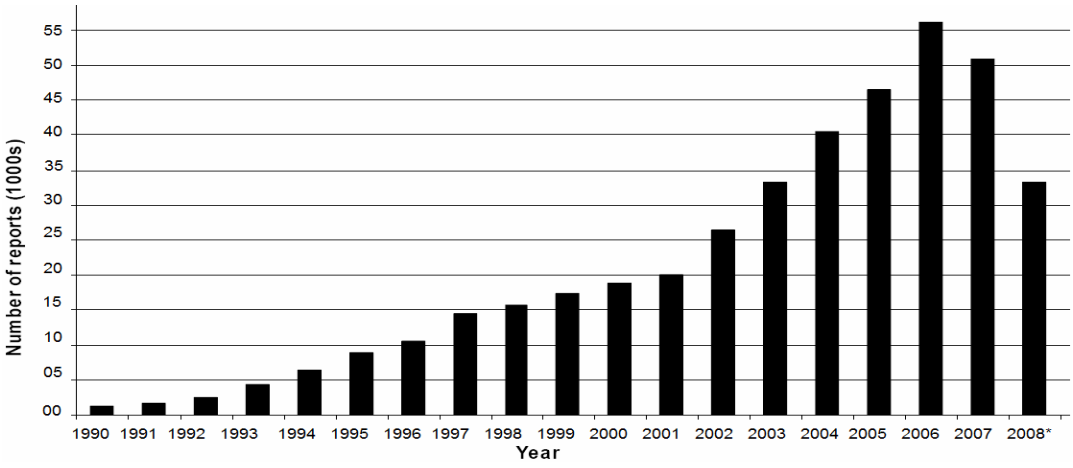

Figure 2.2: *The number of reported cases of Clostridium difficile-infection in the United Kingdom (Source - Health Protection Agency and Office for National Statistics). * - the 2008 data is for 3 quarters (Jan. 2008 – Sep. 2008).*

A survey report by the Health Protection Agency (HPA), in 2006 revealed that there was a 30 fold increment in the reported case of CDI in 15 years between 1990 to 2005 (Figure 2.2). In the year 2005, more than 3500 individuals lost their lives to CDI in the United Kingdom (Figure 2.3). This number was more than 8 % of the reported cases of CDI in that year. Although the number of reported cases of CDI decreased from 2006 to 2007 (Figure 2.2), the severity of infection kept on increasing with a death toll of 6500 in 2006 and more than 8000 in 2007 (Figure 2.3).

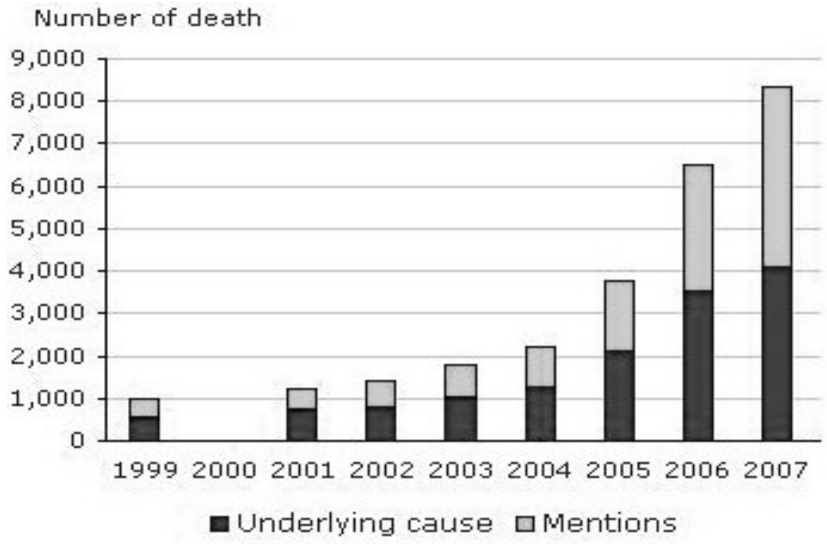

Figure 2.3: *The number of deaths associated with Clostridium difficile infection in England and Wales (Source- Office for National Statistics).*

The pathogenesis of *C. difficile* has been attributed to its three well known toxins – Toxin-A (TcdA), Toxin-B (TcdB) and a binary toxin (CDT). All three toxins are discussed below in detail.

Clostridium difficile **Virulence Factors**

Toxin-A and Toxin-B (TcdA, 308 kDa and TcdB, 270 kDa) are two proteins that have been considered to be the main virulence factors of *C. difficile* (Thelestam and Chaves-Olarte, 2000; Elliott *et al.*, 2007). They along with several other closely related Clostridial toxins such as *C. sordellii* lethal toxin (TcsL) and haemorrhagic toxin (TcsH) and *C. novyi* alpha toxin (Tcn-α), constitute a group known as Large Clostridial Cytotoxins (LCT) (Just *et al.*, 2000). All members of this family are single chain proteins of high molecular

weight ranging from 200 to 300 kDa (Rupnik *et al.*, 2003) and are among the largest known bacterial toxins.

TcdA and TcdB from *C. difficile* have been studied in great detail. Both proteins are expressed efficiently by the host during the late log phase or stationary phase of growth (Voth and Ballard, 2005). However, the precise environmental signal that modulates toxin expression is still unclear. Studies suggest that the production of both proteins by *C. difficile* can be enhanced under stress conditions such as in the presence of antibiotics vancomycin and penicillin (Dupuy and Sonenshein, 1998). A recent study (Lyras *et al.*, 2009) emphasises the essentiality of Toxin-B in *C. difficile* infection.

TcdA and TcdB toxins are encoded by two separate genes namely *tcdA* and *tcdB*. Along with three other genes *tcdC, tcdD* and *tcdE*, these toxins form a pathogenicity locus (Figure 2.4) which spans over a 19 kb region on the genome of the bacterium (Hammond and Johnson, 1995). Translation products of these genes (*tcdC, tcdD* and *tcdE)* are suspected to be involved in the pathogenicity of the organism by regulating the expression of TcdA and TcdB and their release from the cell (Hammond and Johnson, 1995). A high sequence similarity and functional homology between TcdA and TcdB indicates that the two genes may have arisen as a result of gene duplication (von Eichel-Streiber *et al.*, 1992).

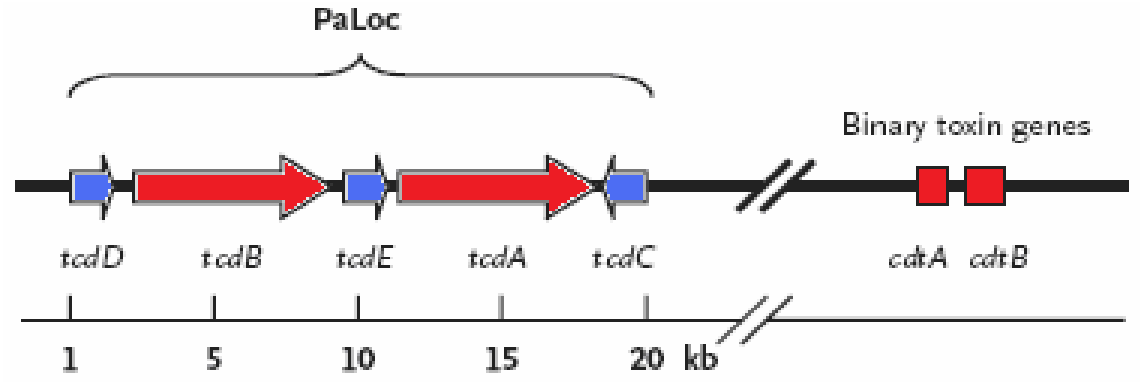

Figure 2.4: *The arrangement of tcdA and tcdB toxin genes along with their regulators (tcdC tcdD and tcdE) in the C. difficile pathogenicity locus and the relative position of binary toxin genes (Adopted from McDonald et al., 2005).*

Less prominent is the *C. difficile binary* toxin (CDT). The pathogenic role of CDT in *C. difficile* infection is still a question of debate. About 6 to 12.5 % strains of *C. difficile* that have been isolated from patients suffering from CDI

are found to contain CDT genes (Stubbs *et al.*, 2000; Popoff, 2000; Geric *et al.*, 2003). CDT is a genome encoded toxin. CDT coding genes are located at an unknown position outside the pathogenicity locus (Figure 2.4) (McDonald *et al.*, 2005). A recent study has highlighted on an 18 base pair deletion in *tcdC* gene (one of the negative regulators of TcdA and TcdB, Figure 2.4). The deletion is found closely associated with the prevalence of *C. difficile* strains carrying CDT encoding genes (McDonald *et al.*, 2005). The importance of this deletion and its correlation with the presence of CDT genes is not understood yet. It is suggested that the presence of CDT contributes towards the severity of infection (Perelle *et al.*, 1997). However, to date, there is no report available to evaluate the cytotoxicity effect of complete CDT in isolation.

***Clostridium difficile* Binary Toxin (Actin-ADPRT)**

Similar to many other Clostridial binary toxins such as *C. perfringins* iota toxin, *C. botulinum* C2 toxin, and *C. spiroforme* toxin, CDT is composed of two components – an enzymatically active component (CDTa) and a catalytically inert transport component (CDTb) (Barth *et al.*, 2004). Domain organisation of CDTa and CDTb is shown in Figure 2.5.

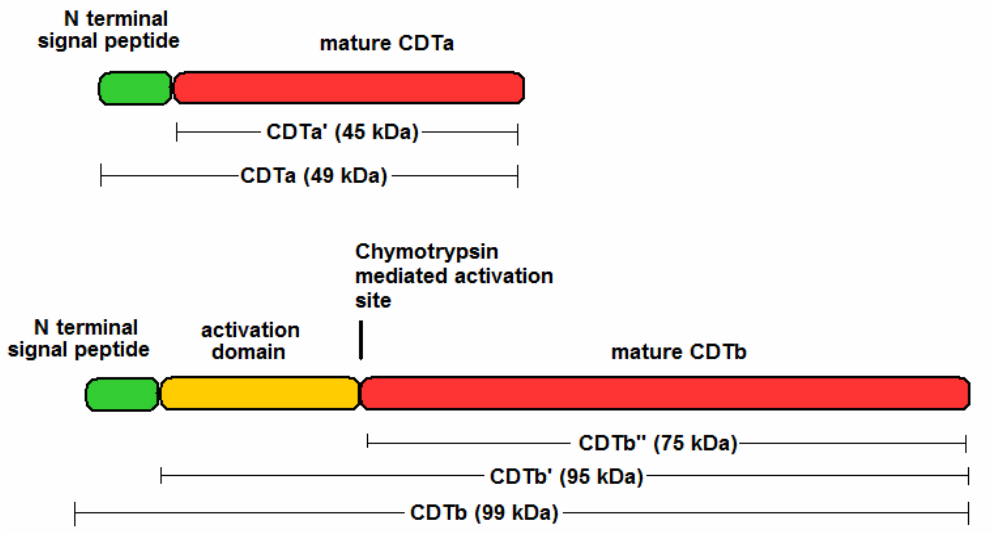

Figure 2.5: *The domain organisation of CDTa and CDTb. CDTa'- mature CDTa fragment (without signal peptide), CDTb'- CDTb fragment without signal peptide, CDTb'' – fully mature CDTb fragment.*

The enzymatic component of *C. difficile* binary toxin (CDTa) is a 462 amino acid protein with a total molecular weight of 49 kDa. The first 42-N

terminal residues of CDTa have been predicted to form a transmembrane peptide segment that acts as a signal peptide. CDTa gets activated by proteolytic cleavage of the signal peptide and the cleavage site has been identified at Lys 42-Val 43 (Perelle *et al.*, 1997). Amino acid residues (Arg 295, Glu 378 and Glu 380) that are essential for catalytic (ADP ribosylation) activity of the enzymatic component of *C. perfringens* Iota toxin, Ia, the closest homologue of CDTa (Perelle *et al.*, 1996; van Damme *et al.*, 1996), are well conserved in CDTa. Precursor and mature CDTa share 81% and 84% sequence identity with the corresponding lengths of the enzymatic component of Iota toxin (Perelle *et al.*, 1997; Voth and Ballard 2005).

The transport component of *C. difficile* binary toxin (CDTb) consists of 876 amino acid residues with a molecular weight of 98.9 kDa (Figure 2.5). The protein itself is catalytically inert but plays an important role in transporting the enzymatic component (CDTa) into the target cells. The first 42 N terminal residues of CDTb have also been predicted to be a signal peptide that displays features of a transmembrane segment (Perelle *et al.*, 1993). Precursor CDTb shares 81.2 % and 38 % sequence identity with the transport component of *C. perfringens* Iota toxin (Ib) and *C. botulinum* C2 toxin (C2II), respectively (Barth *et al.*, 2004). CDTb undergoes a proteolytic cleavage by chymotrypsin and the cleavage site has been proposed to be at Lys 209-Leu 210 (Perelle *et al.*, 1997). As a result of cleavage, a 25 kDa N terminal fragment of CDTb falls apart and the remaining larger C terminal fragment functions as an active (mature) CDTb. The mature CDTb is 82 % identical to Ib and 40 % identical to C2II (Barth *et al.*, 2004).

Clostridial Actin-ADPRTs

Several species of *Clostridium* and *Bacillus* produce binary toxins that belong to the ADP ribosylating toxin (ADPRT) superfamily. They all target actin molecules in the target cell. These toxins are composed of two subunits (components) which are transcribed, translated and secreted out of the cell as two separate proteins (Barth *et al.*, 2004) encoded by two distinct genes. The G+C content of these genes vary between 27 to 31 % among different Clostridial species (Popoff, 2000). A significant difference at the genetic level between different Clostridial binary toxins is that *C. difficile* CDT, *C. botulinum* C2 and *C. spiroforme* CST are chromosome encoded toxins whereas *C. perfringens* iota toxin is a plasmid encoded toxin (Barth *et al.*, 2004).

The smaller component (known as A or I) of Clostridial binary toxins possess the enzymatic activity of the toxin (Figure 2.5) and is responsible for the covalent ADP ribosylation of monomeric actin molecules in the target cell (Aktories and Wegner, 1992). These toxins utilise NAD or NADPH as the ADP-ribose donor. The larger component (known as B or II) of these toxins is enzymatically inactive (Figure 2.5). The B component is responsible for the translocation of the A component into the target cell (Ohishi *et al.*, 1980).

Clostridial binary toxins are further classified into two main classes based on their substrate specificity (Schering *et al.*, 1988; Rupnik *et al.*, 2003; Barth *et al.*, 2004). Toxins belonging to the Iota family can ADP-ribosylate all three isoforms of actin whereas toxins from C2 family are specific for only smooth muscle actins (β and γ isoforms of actin) (Vandekerckhove *et al.*, 1987; Popoff *et al.*, 1988; Aktories *et al.*, 1986; Mauss *et al.*, 1990). Binary toxins produced by different *Clostridium* species are listed in table 2.1.

Table 2.1: *Classification of costridial binary toxins based on their substrate specificity.*

Family	Toxin and Components	Specificity
Iota family	*C. perfringens* toxin (iota)	α / β / γ Actins
	C. spiroforme toxin (CST)	
	C. difficile toxin (CDT)	
C2 family	*C. botulinum* toxin (C2)	β / γ Actins

Another basis of their classification is the sequence identity between different binary toxins. Members of the Iota family share more than 80% sequence identity in the family, while when aligned against the C2 family members, the sequence identity is much less – around 30 to 40% (Barth *et al.*, 2004).Toxins from one family also show immunological cross reactivity. In addition, the transport component of one toxin can transport the enzymatic components of other toxins into the target cell and thus can be exchanged among different toxins within the family (Rupnik *et al.*, 2003).

Common Mechanism of Action of Clostridial Actin-ADPRTs

The B (or the transport) component of these binary toxins is produced as an inactive precursor molecule which gets activated on proteolysis by

various serine proteases such as furin, trypsin and chymotrypsin (Fernie *et al.*, 1984; Klimpel *et al.*, 1992; Perelle *et al.*, 1997; Stiles, 1987). This activation results in the loss of about 20 to 25 kDa N terminal fragment from the precursor molecule (Figure 2.5). The large C terminal fragment of the B component undergoes a conformational change that facilitates the formation of a homo-heptameric transport component complex (Barth *et al.*, 2004).

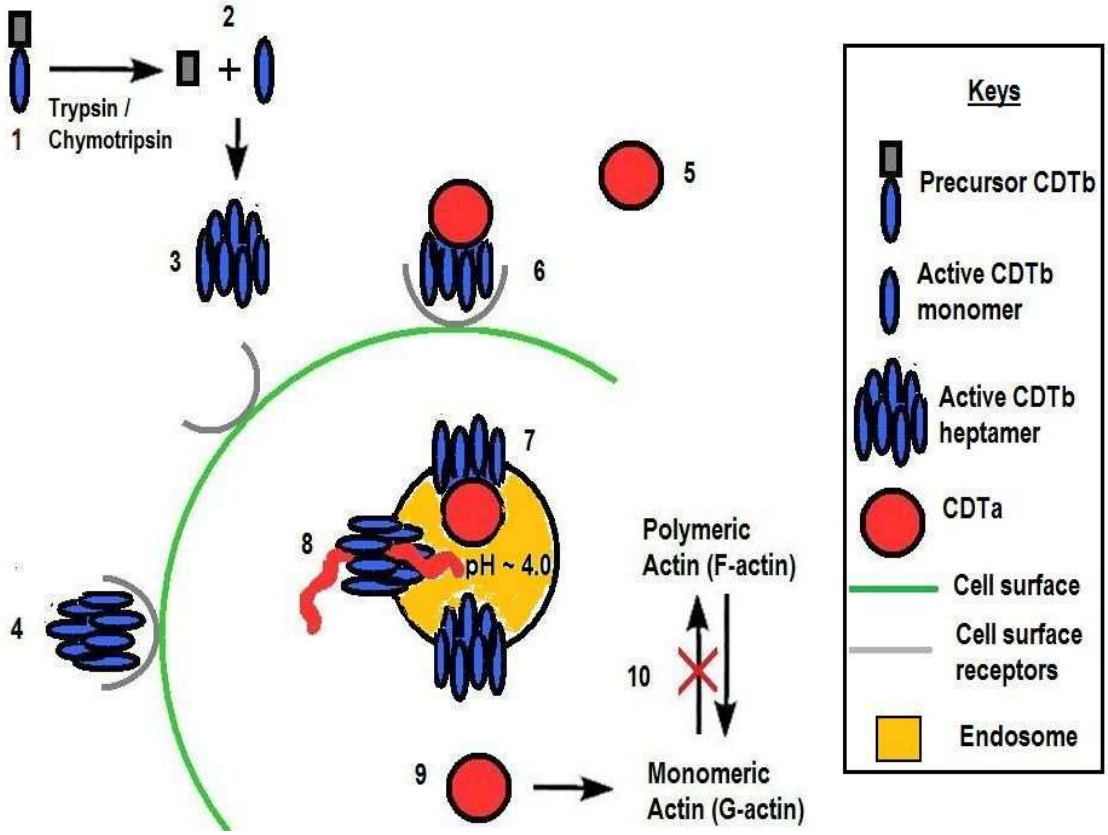

Figure 2.6: *The process of cell intoxication by Clostridial binary toxins. The B subunit is activated by chymotrypsin (1, 2) and forms a heptameric pore like structure (3) that binds to the unknown cell surface receptors (4). The A component (5) then docks on the assembly (6) which then gets endocytosed via early endosomal pathway (7). The A component translocates through the pore into the cytosol (8) and irreversibly modifies monomeric actin which blocks its polymerisation (10).*

The enzymatic component of the toxin then docks on the cell surface receptor bound heptameric transport component complex. The N terminal domains of both components are believed to be involved in docking on each other (Barth *et al.*, 2004). The entire assembly of cell surface bound toxin is then translocated into the cytosol via acidified early endosomal pathway similar to the single chain diphtheria toxin or multi chain *B. anthrcis* lethal

and edema toxin (Madshus *et al.*, 1991; Friedlander, 1986). Late endosomes are not involved in the transport as inhibitors of late endosomes are not found to affect the biological activity of C2 or iota toxins on the cells. However, the biological activity of C2 or iota toxin can be blocked by bafilomycin-A, which is known to inhibit the acidification of early endosome (Barth *et al.*, 2000; Blocker *et al.*, 2001; Werner *et al.*, 1984).

Highly acidic environment of the endosomal compartment (pH < 5.0) has been suggested to facilitate membrane insertion of the transport component heptamer generating a tunnel through the endosomal membrane. The low pH also induces a drastic conformational change in the enzymatic component. The enzymatic component is then translocated from the ensdosomal compartment into the cytosol via the tunnel. In the cytosol, the enzymatic component regains its three dimensional structure and becomes catalytically functional again. It is not clear whether the transport component heptamer also enters cytosol with the enzymatic component or remains attached to the endosomal membrane (Ohishi and Yanagimoto, 1992; Richard *et al.*, 2002). A heat shock protein (Hsp90), a well conserved ATPase in eukaryotic cells has been thought to be involved in the transportation of enzymatic components of iota, CDT and C2 toxins across the endosomal membrane but the mechanism is yet to be understood (Haug *et al.*, 2003a; Haug *et al.*, 2003b).

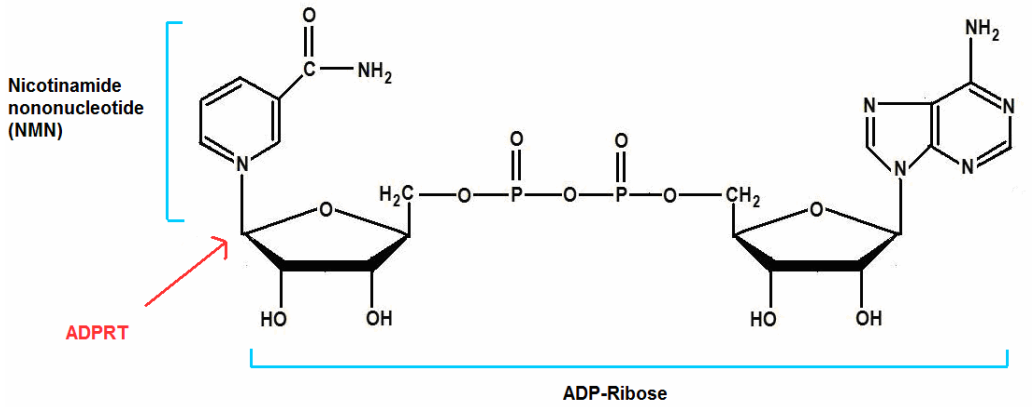

Figure 2.7: *Site of cleavage on the NAD molecule by ADPRTs.*

The enzymatic component of these toxins transfers the ADP ribose moiety (Figure 2.7) of NADH or NADPH to monomeric actin (G-actin) molecules (Aktories and Wegner, 1989; Considine and Simpson, 1991). Monomeric actin

(G-actin) is a single peptide chain of 375 amino acid residues. 14 G-actin molecules interact together to produce a long thread like structure. Two of these strands then produce a right handed double stranded helix known as polymeric actin (F-actin). The polymeric form of actin is a polar molecule. Polymerisation of actin molecules takes place mainly at one end of the polymer known as the barber end (Figure 2.8), whereas depolymerisation occurs at the other end of the molecule known as the pointed end (Figure 2.8) at a faster rate (Aktories and Wegner, 1992).

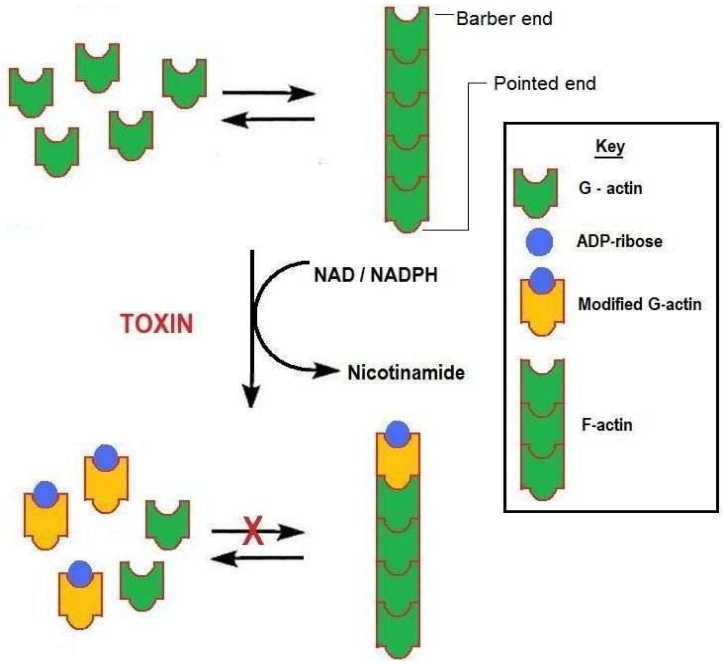

Figure 2.8: *A schematic representation of mechanism of actin cytoskeleton disruption by Clostridial binary toxins. An irreversible modification of monomeric actin at Arg-177 prevents stacking of newly coming momonomeric actin on the growing polymeric actin chain.*

All ADP-ribosylating toxins transfer the ADP-ribose of NADH to Arg-177 residue of monomeric actin (Vandekerckhove *et al.*, 1988). Arg-177 of actin is located in the domain of newly entered G-actin molecule which interacts with the next coming G-actin (Figure 2.8). In the process of polymerization, Arg-177 gets buried in the polymer and remains unaccessible to the toxin (Figure 2.18). Hence, the polymeric form of actin is not a substrate for the ADP ribosylation by these toxins (Aktories *et al.*, 1986).

The irreversible modification of G actin results in disruption of the F-actin - G-actin equilibrium in the cell as the polymerisation of actin molecules ceases (Aktories and Wegner, 1992; Barth *et al.*, 2002). Eventually the cell

cytoskeleton, which is totally dependent on this equilibrium, collapses. These events result in excessive fluid loss from the cell (Simpson, 1982), increased intestinal fluid accumulation (Ohishi, 1983), rounding of the cell (Reuner *et al.*, 1987) and finally cell death.

Research has been carried out to identify cell surface receptors of these binary toxins but only a limited amount of knowledge is present in literature. Cell surface receptor/s of *C. difficile* CDT have not been identified. Cell surface receptors for *C. botulinum* C2 toxin have been identified as asparagine linked complex/hybrid carbohydrates (Eckhardt *et al.*, 2000; Sugii and Kozaki, 1990) whereas the receptors for *C. perfringens* iota toxin have been found to be proteins which are resistant to proteases (Liu and Lappa, 2003; Stiles *et al.*, 2000; Stiles *et al.*, 2002).

Bacterial ADPRTs and Their Classification

Bacterial pathogens utilise a whole range of toxins to modify or kill the target cell. ADP ribosylation (Collier and cole, 1969), glucosylation (Sehr *et al.*, 1998), acetylation (Mukherjee *et al.*, 2006), deamidation (Schmidt *et al.*, 1997) and proteolysis (Schiavo *et al.*, 1992) of host proteins are some of the favoured methods of cell intoxification. ADP-ribosylation of elongation factor-2 (EF-2) was the first covalent modification shown to be performed by any toxin (diphtheria toxin) (Collier, 1975).

ADP ribosylating toxins (ADPRTs) are a large family of potentially lethal toxins that transfers the ADP-ribose portion of NAD, covalently, to their targets (Deng and Barbieri, 2008). Producers of this family of toxins belong to a vast range of bacterial pathogens including Clostridia and *Baccilus.* These organisms are the principal causative agents of several serious diseases (Holbourn *et al.*, 2006) such as cholera, diphtheria and hospital acquired diarrhoea. Targets of these ADPRTs are the key regulators of cellular functions such as small GTPases or Actin. Covalent modification of these proteins by toxins results in the serious collapse of key cellular processes and eventually cell death (Holbourn *et al.*, 2006). The ADPRTs have been classified in 4 major classes based on their domain organisation and target specificity (Table 2.2).

The **AB_5 class** consists of some of the most well known toxins such as cholera, pertussis and *E. coli* enterotoxin (Figure 2.9). The catalytically active subunit (A subunit) of the toxin docks on a doughnut shaped pentamer of B

subunit that comprises the cell binding and translocation domains (Stein *et al.*, 1994; Zhang *et al.*, 1995; Gill *et al.*, 1981; Finkelstein *et al.*, 1987; Sixma *et al.*, 1991). The hetero-hexamer assembles in the bacterial cell itself prior to its secretion (Sandkvist *et al.*, 2000). The A subunit undergoes a proteolytic cleavage to release a disulphide linked A1 domain from the rest of the complex. The A1 domain is then transported into the target cell where it undergoes another activation process in order to become fully functional (Holbourn *et al.*, 2006). Targets for this family of toxins are small regulatory G proteins (Table 2.2).

Table 2.2: *Different classes of the ADPRTs and their substrates.*

ADPRT class	**Toxin (PDB ID)**	**Bacterium**	**Target**
AB_5	Cholera (1XTC)	*Vibrio cholerae*	Gs
	Pertussis (1PRT)	*Bordetella pertussis*	Gi, Gt and Ga
	E. coli Enterotoxin (1LTS)	*Escherichia coli*	Gs
AB	Diphtheria (1TOX)	*Corynebacterium Diphtheriae*	eEF2
	Pseudomonas exotoxinA (1AER)	*Pseudomonas Aeruginosa*	
A-B binary	VIP (1QS1)	*Bacillus cereus*	G-Actin
	Iota (1GIQ)	*Clostridium perfringens*	
	CDT (2WN4)	*Clostridium difficile*	
	C2	*Clostridium botulinum*	
Single polypep -tide	C3bot (1G24)	*Clostridium botulinum*	RhoA, B, C
	C3stau (1OJZ)	*Staphylococcus aureus*	RhoA, B, C, E and Rnd3

Diphtheria toxin belongs to the **AB class** of ADPRTs (Figure 2.9). Members of this family are highly potent toxins. Lethal dose of diphtheria toxin for humans is as low as 0.1 µg of toxin per kilogram (Deng and Barbieri, 2008). These toxins are multidomain proteins with their receptor binding, translocation and catalytic domain residing on one single polypeptide chain (Hwang *et al.*, 1987; Allured *et al.*, 1986; Morris *et al.*, 1985; Sandvig and Olsnes, 1980; Collier, 1975; Wilson and Collier, 1992). The substrate for AB class of ADPRTs is a diphthamide residue (a His residue that has been modified by addition of diphthamide side group) (Van Ness *et al.*, 1980) on elongation factor-2 (Table 2.1) (Wilson and Collier, 1992). Interruption of

elongation factor-2 (EF2) function disrupts protein synthesis which leads to cell death (Collier, 1975).

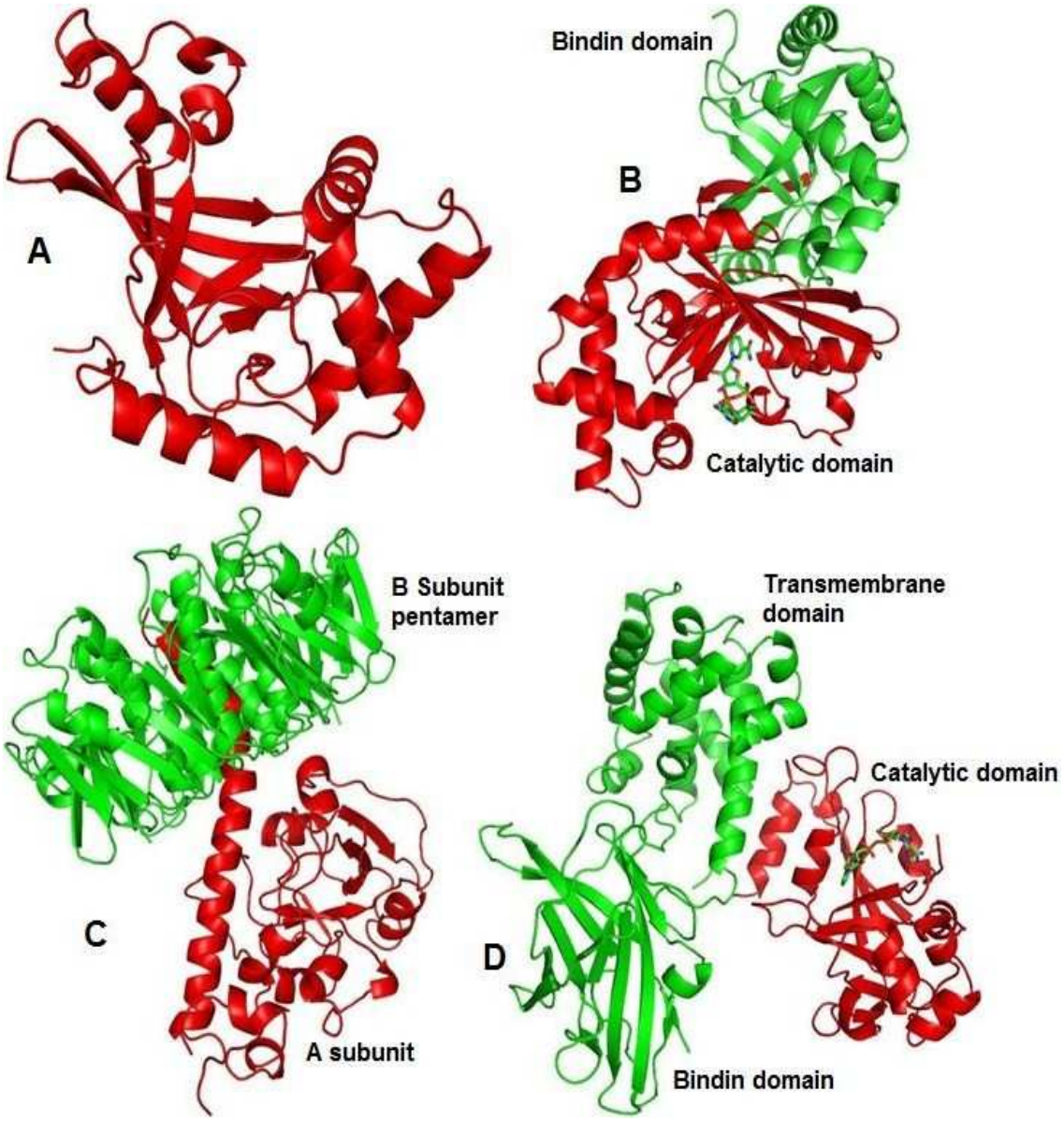

Figure 2.9: The s*tructural comparison of all 4 classes of the ADPRTS with representative members from each class: A- C3Bot (PDB ID - 1G24) (Han et al., 2001), B- Iota Toxin (PDB ID - 1GIQ) (Tsuge et al., 2003), C- Cholera toxin (PDB ID – 1XTC) (Zhang et al., 1995), D- Diphtheria toxin (PDB ID - 1TOX) (Bell and Eisenberg, 1996). The catalytic domains of each protein are shown in red (figure adopted from Holbourn et al., 2006).*

The third class of ADPRTs comprises **small single domain C3 coenzymes**. An example of this class is *C. botulinum* C3bot toxin (Figure 2.9) (Aktories *et al.*, 1987). This family of ADPRTs targets small GTPases such as RhoA, B and C (Table 2.1) at an exposed Arg-41 (Chardin *et al.*, 1989; Sekine

et al., 1989). Covalent modification of Rho proteins as a result of ribosylation prevents its switching to active GTP bound state and leads to the loss of control over the cytoskeleton and eventually to cell death (Wilde and Aktories, 2001).

The **A-B binary ADPRTs** comprise the fourth class of the superfamily. This family includes toxins from a wide range of *Clostridium* species such as *C. perfringens* iota toxin, *C. botulinum* C2 toxin, *C. difficile binary* toxin (CDT), and vegetative insecticidal protein (VIP2) from *Bacillus cereus* (Han *et al.*, 1999; Aktories *et al.*, 1986; Stiles and Wilkins, 1986; Simpson *et al.*, 1989; Popoff and Boquet, 1988). As the name suggests, these toxins are binary in nature. These toxins are composed of two independently transcribed and translated gene products. A larger subunit (B subunit), that is known to form a heptameric pore like structure upon proteolytic activation translocates the catalytically active A subunit into the cytosol of the target cell. These toxins ADP-ribosylate monomeric actin in the target cell and thus are responsible for the collapse of the cell cytoskeleton (Aktories and Wegner, 1989).

Mechanism of Action of *C. difficile* Toxin A and Toxin B

TcdA and TcdB toxins from *C. difficile* utilise a well defined mechanism of action. Both the toxins possess glucosyltranferase activity and are capable of transferring the glucose moiety of UDP-glucose to small GTPases of the Rho superfamily in the target cell (Just and Gerhard, 2004; Just *et al.*, 1995a; Just *et al.*, 1995b; Lyras *et al.*, 2009). Rho proteins are the primary regulators of actin cytoskeleton (Hall, 1990). Irreversible glucosylation by TcdA and TcdB results in the inactivation of these small GTPases and thus disruption of vital cell signalling pathways (Just *et al.*, 1995a; Just *et al.*, 1995b) which ultimately leads to cell death.

Internalization of TcdA and TcdB in to the target cell takes place through nonproteinaceous cell surface receptor mediated endocytosis (Florin and Thelestam, 1983; Mitchell *et al.*, 1987) via acidified endosomal pathway. The low pH of the endosome induces conformational changes in the toxin structure and exposes a hydrophobic domain (discussed in the next section) of the protein that is then inserted into the endosomal membrane (Qa'Dan *et al.*, 2000). The formation of such channels in the lipid bilayer by TcdB in a pH-dependent process has indeed been reported (Barth *et al.*, 2001).

Structural Organisation of TcdA and TcdB

Structurally, these proteins are described as ABCD type protein (figure 2.10) (Jank and Aktories, 2008). The full length protein can be divided into 4 domains according to their function (Giesemann *et al.*, 2008).

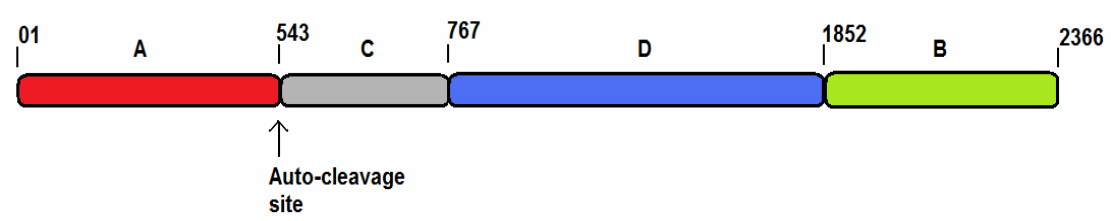

Figure 2.10: *The domain organisation of toxin-A and Toxin-B (ABCD model). A – Activity domain, B – Binding domain, C – Cutting domain and D – Delivery domain. The amino acid residue numbering is based on toxin-B (figure adopted from Jank and Aktories, 2008*).

The N terminal catalytic domain (activity or A domain) possesses full biological activity of the molecule (Hofmann *et al.*, 1997; Faust *et al.*, 1998). A repetitive oligopeptide sequence (binding or B domain) at the C terminal end of the protein has been suggested to be involved in receptor binding (Tucker and Wilkins, 1991; Wren, 1991; Frisch *et al.*, 2003; Ho *et al.*, 2005). The cell surface receptors of TcdA are carbohydrates in nature including Gal-α1, 3-Gal-β1, 4-GlcNAc (Krivan *et al.*, 1986; Pothoulakis *et al.*, 1996).

The central part of the protein constitutes the other two domains. Very little is known about its exact function (Giesemann *et al.*, 2008). However, a small hydrophobic stretch (delivery or D domain) is suggested to mediate membrane insertion during the translocation process (Qa'Dan *et al.*, 2000). The fourth functional domain of the protein (cutting or C domain), is characterised by its resemblance to a putative catalytic triad of a cysteine protease and is thought to be responsible for autoproteolytic cleavage of the protein (Pruitt *et al.*, 2009) to facilitate transport of the A domain into the cytosol.

In spite of availability of adequate information about their mode of internalization into the target cell as well as their mechanism of action, the structural information about *C. difficile* Toxin-A and Toxin-B is limited. The three dimensional structure of full length TcdA or TcdB are yet to be determined. The crystal structure of the catalytic domain of Toxin-B at 2.2 Å

(Figure 2.11) with its donor substrate UDP-glucose (UDP-Glc) and co factor (Mn^{2+}) ion has recently been reported (Reinert *et al.*, 2005).

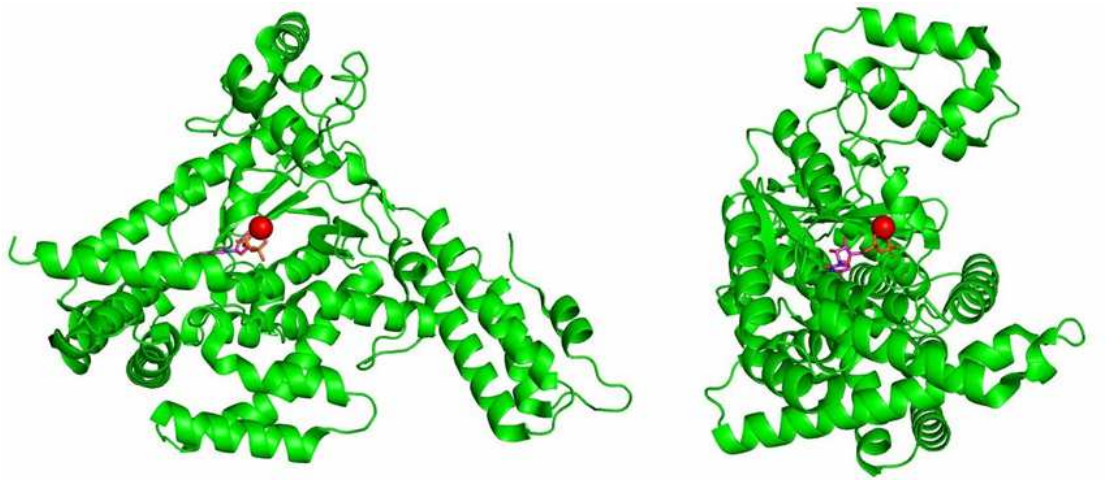

Figure 2.11: *The crystal structure of the catalytic domain (domain A) of TcdB (Reinert et al., 2005) with bound manganese (shown as shphere) and UDP-glucose (shown in sticks). The two orientations are at 90^0 to each other (PDB ID - 2BVL).*

The N terminal catalytic domain of TcdB consists of the first 543 amino acid residues of the protein. The overall fold of the catalytic domain resembles that of the members of glycosyltransferase-A (GT-A) family proteins (Reinert *et al.*, 2005). Like other GT-A family proteins, a common D-X-D motif exists in TcdA and TcdB which is involved in the binding of Mn^{2+} ion and glucosyl group. As a result of intoxification, only the A domain of the protein is translocated into the cytosol of the target cell (Pfeifer *et al.*, 2003; Rupnik *et al.*, 2005; Reineke *et al.*, 2007). The Large Clostridial toxins (LCTs) undergo an autoproteolysis that has been attributed to a cysteine protease activity located in the C domain (also known as cysteine protease domain or CPD) of the protein (Figure 2.12) (Egerer *et al.*, 2007). Inositolhexaphosphate (IP6) has been suggested to mediate this autoproteolytic process (Reineke *et al.*, 2007; Egerer *et al.*, 2007). The crystal structure of C domain of TcdA in complex with bound IP-6 at 1.6 Å resolution has been reported (Pruitt *et al.*, 2009). The C-domain of TcdA spans from residue 543 to 809. The CPD of TcdA is composed of 9 stranded β sheet flanked by 5 α- helices (Figure 2.12). A trio of Asp, His and Cys have been shown important for autoproteolytic activity of TcdA (Pruitt *et al.*, 2009) and TcdB (Egerer *et al.*, 2007).

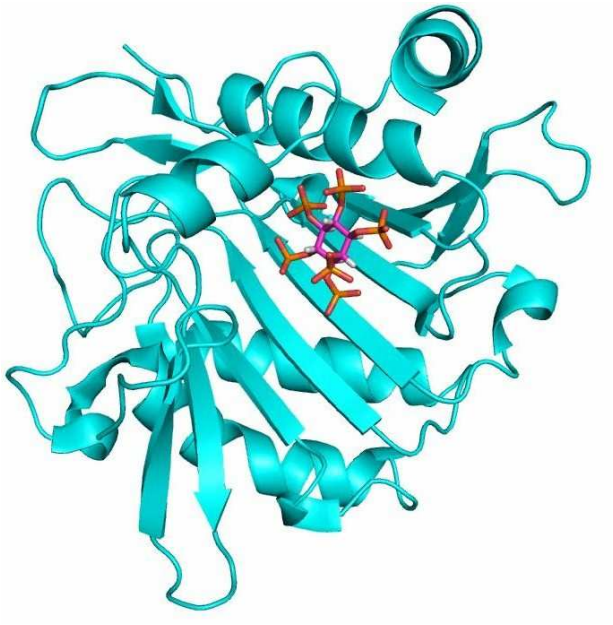

Figure 2.12: *The C domain (or Cysteine protease domain or CPD) of TcdA with bound IP-6 (shown in sticks) (PDB ID - 3HO6) (Pruitt et al., 2009).*

At least two independent high resolution crystal structures of different lengths of the receptor binding C terminal repetitive domain (CRD) of TcdA (Figure 2.13) have been determined (Ho *et al.*, 2005; Greco *et al.*, 2006).

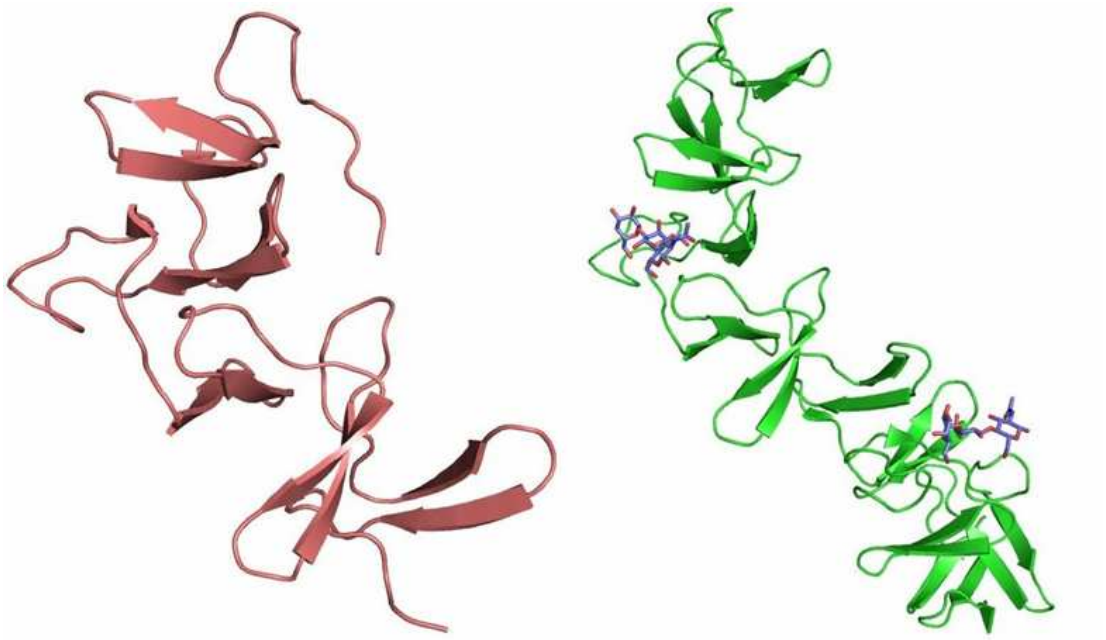

Figure 2.13: *LHS – the crystal structure of C terminal repetitive domain (127 residues) of TcdA (PDB ID - 2F6E) (Ho et al., 2005). RHS – the crystal structure of C terminal repetitive domain (255 residues) of TcdA in complex with a synthetic derivative of its natural carbohydrate receptor (shown in sticks)(PDB ID - 2G7C) (Greco et al., 2006).*

The presence of repetitive units of 21, 30 or 50 amino acid residues is the most striking feature of the C terminal repetitive domain of TcdA and TcdB (Dove *et al.*, 1990; von Eichel-Streiber *et al.,* 1990; von Eichel-Streiber *et al.,* 1992; von Eichel-Streiber *et al.,* 1996). In TcdA, there are 30-38 repeats present whereas in TcdB the number of repeats are 19 to 24 (Ho *et al.* 2005). The CRD of TcdA is composed of 32 short repeats (SR) and 7 long repeats (LR) with each repeat consisting of a β hairpin followed by a loop (Ho *et al.*, 2005). The carbohydrate binding site (Figure 2.13, RHS) in the CRD is a shallow trough between a LR and the hairpin turn of the following SR (Greco *et al.*, 2006). It is suggested that the CRD of these toxins adopts an elongated serpentine shape in which all carbohydrate binding sites are presented on the same face of the structure. This arrangement allows for a multivalent interaction of the toxin on the cell surface (Greco *et al.*, 2006).

Main Experimental Aims of This Thesis

Limited information is available for both, the structures and mechanistic details of Clostridial binary toxins. The available structures to date include a high resolution (1.8Å) structure of the enzymatic component (Ia) of Iota toxin (Tsuge *et al.*, 2003) in ligand bound form and a 2.1 Å resolution structure of the enzymatic component (C2I) of C2 toxin in native state (Schleberger *et al.*, 2006). Both of these toxins belong to two different classes of Clostridial binary toxins (Table 2.1) on the basis of their substrate specificity which makes comparison of the two available structures difficult at the molecular level.

A partially incomplete structure of the transport component (C2II) of C2 toxin in monomeric form has been determined (Schleberger *et al.*, 2006). The structure provides limited amount of information due to its poor resolution (3.1Å). In addition to that, the C terminal receptor binding domain of the protein could not be modelled in this structure. The mature transport components of Iota family toxins are about 120 amino acid residues longer than the C2 family transport component (Barth *et al.*, 2004). Since it is the large C-terminal fragment of the transport component that heptamerises and is functionally active; this difference in the length of mature proteins may provide some crucial information. It would be interesting to establish the functional implications of this extra length of the protein.

C. difficile is resistant to commonly used antibiotics and is capable of causing infection in their presence. An alternative approach to control *C. difficile* infection can be designed based on targeting its toxins. To do so, it is necessary to know the 3-dimensional structure of these toxins. Structural details of both components of binary toxin can provide important clues about their interaction, mechanism of cell intoxication, about their domains that should be targeted to make the toxin ineffective and what kind of molecules can efficiently inhibit the function of the toxin.

In addition, the first 42 N terminal residues in CDTa and CDTb have been reported to function as signal peptides (Perelle *et al.*, 1997; Rupnik *et al.*, 2003). Cleavage of the signal peptide is essential for both components to become fully mature and functionally active. It would be interesting to see what, if any conformational changes, the absence of the signal peptide and further proteolytic cleavage induces in mature CDTa and CDTb. Hence, the significance of determining the 3D structure is apparent and this leads to the aims of the study expected in this thesis.

In order to provide a structural basis of the understanding of CDT function and to determine its role in pathogenesis, a full scale structure function study on CDT was initiated with following specific aims:

- To establish methods of cloning, expression and purification of both the components of *C. difficile* binary toxin.
- To assess the cell cytotoxicity potential of complete *C. difficile* binary toxin.
- To determine and analyze the structure of enzymatic as well as transport components of *C. difficile* binary toxin, and
- To understand the mechanistic details of binary toxins using protein engineering approach.

CHAPTER - III

CLONING, EXPRESSION AND PURIFICATION OF *C. DIFFICILE* BINARY TOXIN

A - CLONING EXPRESSION AND PURIFICATION OF ENZYMATIC COMPONENT OF *C. DIFFICILE* BINARY TOXIN: CDTa

MATERIALS AND METHODS

Primer Design, PCR Amplification and Subcloning

A set of primers (Table 3.1) was designed to PCR amplify the coding sequences of CDTa without its N terminal signal peptide sequence. This protein fragment was named CDTa′ and the corresponding coding DNA was named *cdtA*′.

Table 3.1: *The primer sequences for amplification of cdtA′.*

Fragment	Primer Sequence
cdtA′	F= *AGCA* GGATCC GAA ATC GTG AAC GAA GAT ATT C
	R= *AGCA* GTCGAC TTA* ATC CAC GCT CAG AAC C

*F – forward primer, R - reverse primer, In italics - random 5′ overhang, underlined - restriction sites, * - stop codon. The amplified DNA product was named as cdtA′.*

Table 3.2: *The PCR composition and reaction conditions for cdtA′ amplification.*

Fragment	Reaction mixture (50 µl)	Reaction conditions
cdtA′	Templet DNA= 2 µl, Forward primer=2.5 µl, Reverse Primer=2.5 µl, 10X KOD buffer =5 µl, 25mM MgSO4= 2 µl, 8 mM DNTP mix=5 µl, 5M Betaine=10 µl, DMSO=2 µl, KOD polymerase=1 µl, Water= 18 µl.	 95^0C– 300 secs, [95^0C – 60 secs, 48^0C – 60 secs, -(40 cycles) 72^0C – 60 secs]

A recombinant DNA construct (*pPCRscript-cdtA)* containing the coding region of full length CDTa was kindly provided by our collaborator (Dr. Clifford C. Shone, HPA, Porton Down) and was used as template DNA for the

amplification reaction. The PCR composition and reaction conditions are given in table 3.2. The amplified product was run on a 0.8% agarose gel in Tris Acetate EDTA (TAE) buffer at 100 volts for 45 minutes and the product was eluted from the gel by using a Promega Wizard SV Gel and PCR Clean-up system.

Three different clones of *cdtA'* were prepared with the vector backbones of *pMAL-HT, pMAL-p2x* and *pGEX-6p1*. The PCR amplified product (*cdtA'*) and each of the vector backbones were double digested (Table 3.3) in separate reactions in a total volume of 50µL each, with BamHI and SalI restriction enzymes to produce compatible sticky ends. The reaction mixtures were incubated at 37^0C overnight to allow the complete digestion of DNA.

Table 3.3: *Composition of the restriction digestion reactions.*

Ingredient	**Reaction volume**	
	50 µl	**10 µl**
Substrate DNA	30.0 µl	2.0 µl
10 X buffer D	5.0 µl	1.0 µl
BamHI and SalI	2.0 µl and 1.0 µl	0.4 µl and 0.2 µl
100 X BSA	0.5 µl	0.1 µl
Nuclease free water	11.5 µl	6.5 µl

The digested products were run on a 0.8% agarose gel in 1X TAE at 100 volts for 45 minutes and the desired DNA fragments were eluted from the gel using Promega Wizard SV Gel and PCR Clean-up system.

Table 3.4: *Composition of the ligation reaction.*

Ingredient	**Reaction volume = 10 µl**
Vector DNA	5.0 µl
Insert	3.0 µl
T4 DNA ligase	1.0 µl
10 X ligase buffer	1.0 µl

In the next step, the double digested insert (*cdtA'*) was ligated with the double digested vector backbones (*pMAL-HT, pMAL-p2x and pGEX-6p-1*) to

produce the desired recombinant constructs (Table 3.4). The reaction mixtures were incubated at 4^0C overnight to allow the ligation reaction to complete.

Screening of Positive Recombinant Clones

E.coli DH5α competent cells were transformed separately with the ligated products. Transformed cells were then plated on LB agar media (Table 3.5) supplemented with 100 µg/ml ampicillin and incubated at 37^0C overnight.

Table 3.5: *Composition of different growth media used for protein expression.*

Media	**Ingredients**
LB broth media	Tryptone = 10 gm, Yeast extract = 5 gm, NaCl = 10 gm, dissolve and make up volume to 1000 ml.
LB Agar media	LB media + 1 to1.5% Agar
TB broth media	Tryptone = 12 gm, Yeast extract =24 gm, Glycerol = 4 ml, dissolve and make up volume with water to 900 ml. sterilised and allow to cool. Add 100 ml of separately sterilised 10 X TB salts solutions.
TB salts (10X)	K_2HPO_4 =125.4 gm and KH_2PO_4 =22.70 gm, dissolve and make up volume to 1000 ml.

On the next day, overnight grown single colonies for each construct were selected randomly and inoculated in 10 ml of LB media supplemented with 100 µg/ml ampicillin, in separate tubes. The cultures were allowed to grow at 37^0C with continuous shaking at 200 rpm overnight. Plasmids were isolated from these cultures using a Promega Wiazrd Plus SV Minipreps DNA purification system.

The isolated plasmids were subjected to double digestion reactions for preliminary analysis. The digestion reactions at this step were carried out in a total volume of 10 µl each (Table 3.3). The reaction mixtures were incubated at 37^0C for 4 hours and analysed on a 0.8% agarose gel in 1X TAE at 100 volts for 45 minutes. The isolated plasmid DNA that was cleaved into two fragments (vector back bone and insert) of the expected size as a result of digestion were selected and the presence of the correct DNA fragments (both, inserts and vector back bone) was confirmed in those by sequencing (Eurofins, MWG).

Preparation of Expression Host

The new clones that were prepared for CDTa′ expression were *pMAL-HT-cdtA′, pMAL-p2x-cdtA′ and pGEX-6p1-cdtA′*. 20 µl of competent *E. coli* BL21-CodonPlus (DE3)-RIPL cells were transformed with all three recombinant DNA in separate reactions. The transformed cells were plated on the Luria-Bertani agar (LB-agar) media (Table 3.5) containing 100 µg/ml ampicillin. All plates were incubated at 37^0C and cells were allowed to grow overnight.

Next day, 20 ml of Luria-Bertani (LB) broth (Table 3.5) supplemented with 100 µg/ml ampicillin was inoculated with a single overnight grown colony from each plate in sepearte flasks and these cultures were allowed to grow at 37^0C with shaking at 200 rpm overnight. 500 µl of each of the overnight grown cultures were mixed with equal volumes of 30% glycerol and stored at -80^0C. These glycerol stocks were used as seed cultures for the subsequent expression trials.

Expression Trials for New Clones

Table 3.6 provides the details of three sets of expression trials carried out for all three newly constructed clones. The frozen glycerol stocks were used to inoculate 20 ml of LB broth supplemented with 100 µg/ml of ampicillin in separate flasks. These cultures were allowed to grow at 37^0C with shaking at 200 rpm overnight. The expression experiments for all three clones were conducted simultaneously under identical conditions so that the results could be compared.

500 ml of fresh sterile media (Table 3.5) was inoculated with the above grown seed cultures in separate flasks giving it a 1% final inoculum. Appropriate amount of ampicillin (100 µg/ml) was added to the media prior to inoculation. These cultures were incubated at 37^0C with shaking at 200 rpm. For low temperature trials, the incubation temperature was shifted to the desired value when the culture OD_{600} was 0.60 – 0.80. Cultures were induced with the Isopropyl β-D-thiogalactoside (IPTG) giving a final concentration of 1 mM when the culture OD_{600} was 0.90 to 1.0. Incubation at the set temperature was continued up to 20 hours post induction.

Expression samples were collected at different post induction time points to analyse on tris-glycine SDS-PAGE during each expression run. These

samples were centrifuged at 10,000 rpm for 10 minutes and cell pellets were resuspended in 75 µl of water. 25 µl of 4X SDS-PAGE loading dye was added to them. These samples were then heated at 100^0C for 5 to 10 minutes and stored at 4^0C till they were analysed on a Tris-glycine SDS-PAGE.

In addition to that, expression samples of 1 ml volume were collected separately and centrifuged at 10,000 rpm for 10 minutes at 4^0C. Cell pellets were resuspended in 500 µl of 1X Bug Buster solution (Novagen) and the suspensions were incubated at room temperature for 30 minutes. The suspensions were centrifuged at 10,000 rpm for 10 minutes at 4^0C and supernatants were collected. 25 µl of 4X SDS-PAGE loading dye was added to 75 µl of each collected supernatant and these samples were also run on a Tris-glycine SDS-PAGE along with the harvested whole cell samples.

Table 3.6: *Different expression trials carried out for CDTa' expression in the shake flask method using three different DNA constructs.*

Parameters	**Trial 1**	**Trial 2**	**Trial 3**
Media	LB	LB	TB +1X TB salts + 0.5% glucose
Host	*E. coli* BL21-CodonPlus (DE3)-RIPL		
Incubation Temperature	Started at 37^0C, 200 rpm		
	Continued at 37^0C	Shifted to 20^0C at OD_{600} = 0.6 to 0.8	
Induction	1 mM IPTG at OD_{600} = 0.9 to 1.0		
Harvest	4 hours/ 8 hours / 20 hours post incubation		
Results	Visible expression, insoluble protein for all the 3 clones		Visible expression, soluble protein for all the 3 clones

10% resolving, 5% stacking Tris-glycine SDS-PAGE gels were run for all samples. To ensure loading the same amount of total protein on the gel, volume equivalent to the $9/OD_{600}$ of each sample was loaded on the gels. All gels were run at 200 volts at room temperature till the dye front migrated out of the lower end of the gel. Gels were stained with brilliant blue R stain for an hour and washed with destaining solution until the protein bands were clearly visible. Sample preparation and gel run were carried out essentially in an identical manner at all times unless otherwise stated.

Large Scale Expression of CDTa′

Based on the results of expression trials, clone *pMAL-p2x-cdtA′* was chosen for the large scale production of the protein which was carried out in shake flask method. Seed culture was grown from the glycerol stock in the ampicillin containing LB media at 37^{0}C with continuous shaking at 200 rpm overnight. Sterilised TB media (supplemented with 1X TB salts, 0.5% glucose and 100 µg/ml ampicillin) was inoculated with the seed culture to give 1 to 2% inoculum. The culture was allowed to grow at 37^{0}C at 200 rpm. The incubation temperature was decreased to 20^{0}C when the culture OD_{600} reached to 0.6-0.8. The culture was induced with the IPTG to a final concentration of 1 mM at OD_{600} = 0.9-1.0 and continued to grow at 20^{0}C. The culture was finally harvested at 4 hours post induction time and centrifuged at 10,000 rpm at 4^{0}C for 10 minutes. The cell pellet was stored at -80^{0}C until further processing.

Purification of CDTa′

The cell pellet was resuspended in buffer A (Table 3.7) (10 ml/gram of cell pellet) and the cells were lysed using a French press in two cycles at 2000 bar pressure. The cell lysate was centrifuged at 25,000 rpm at 4^{0}C for 30 minutes and the supernatant was collected. A Q sepharose ion exchange resin was equilibrated with buffer A and the clear supernatant was loaded onto it. The column was washed with plenty of buffer A until the base line was reached. The bound protein was then eluted in steps of 10%, 20%, 30%, 40%, 50% and 100% of buffer B in buffer A (Table 3.7). All elution fractions were collected separately and run on a 10% separating SDS-PAGE. The MBP-CDTa′ fusion protein was identified on the gel and fractions containing the desired protein were pooled for tag cleavage reaction.

Sufficient amount of factor Xa (1 Unit / 50 µg of fusion protein) was added to the protein and incubated at 20^{0}C for 24 hours with gentle shaking. On the next day, completion of the tag cleavage reaction was confirmed by running the reaction product on a 10% Tris-glycine SDS-PAGE. The protein was dialysed overnight against a 50 volumes of buffer C (Table 3.7) using a 12-14 kDa cutoff dialysis tubes.

Table 3.7: *Composition of buffers used in the CDTa' purification.*

Buffer	**Composition**
Buffer A	20 mM NaCl, and 5 mM $CaCl_2$ in 50 mM Tris-HCl, pH 8.0
Buffer B	1 M NaCl and 5 mM $CaCl_2$ in 50 mM Tris-HCl, pH 8.0
Buffer C	20 mM NaCl in 50 mM Tris-HCl, pH 8.0
Buffer D	1 M NaCl in 50 mM Tris-HCl, pH 8.0
Factor Xa cleavage buffer	100 mM NaCl and 5 mM $CaCl_2$ in 50 mM Tris-HCl, pH 8.0

The dialysed protein solution was collected in a fresh tube and centrifuged at 10,000 rpm for 10 minutes at 4^0C to remove the insoluble debris and precipitated protein. Clear supernatant was passed through a Q sepharose ion exchange resin that was pre-equilibrated with buffer C. Purified CDTa′ was collected in the column flow-through.

The bound uncleaved fusion protein and the cleaved MBP tag were eluted from the column in two steps of 10% and 50% of buffer D (Table 3.7) in buffer C and were collected separately. The column flow-through and eluted fractions were analysed on a 10% Tris-glycine SDS-PAGE to assess the purity of the protein. The purified protein was concentrated to 0.5 mg/ml using a 10 kDa MWCO Millipore concentrator at 4000 rpm at 4^0C and was stored at -80^0C in 1 ml aliquots.

RESULTS AND DISCUSSION

Primer Design, PCR Amplification and Subcloning

The DNA construct (*pPCRScript-cdtA*) was kindly provided by our collaborator. It was commercially synthesised by GENEART (Germany). The construct was to facilitate *in vivo* amplification of the 'insert' (CDTa coding sequence) for further development of new clones. The coding sequence of full length CDTa was inserted in the *pPCRScript* vector backbone between BamHI and SalI restriction sites. The insert was provided with a stop codon immediately after the CDTa coding sequence.

The Primer set was designed to amplify the desired DNA fragments and to clone it into the first reading frame between BamHI and SalI sites. Both, the forward and the reverse primers were designed with a 5′ overhang of 4 random nucleotides followed by the restriction enzyme recognition sequence. Figure 3.1 shows the PCR amplified DNA fragment on a 0.8% agarose gel.

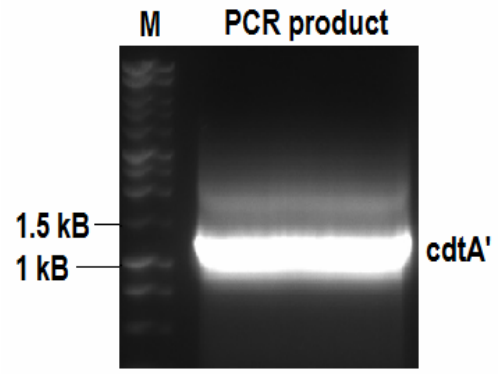

Figure 3.1: *The PCR amplified cdtA′ on a 0.8% agarose gel.*

While going for the sticky end cloning with a PCR amplified product containing restriction sites at the ends, it is always advantageous to have such random 5′ overhangs. It is a well established fact that the restriction enzyme recognition sites at the end of the sequence are cut with a poorer efficiency than the recognition sites in the middle of the sequence. These 5′ random overhang sequences bring the recognition sequences in the middle of the DNA fragment and thus provide a better place for the restriction enzyme to latch on the DNA and to have a better grip on the DNA to cut it.

A schematic arrangement of the domains of CDTa is presented below (Figure 3.2). The primer set was designed to amplify the DNA fragments (*cdtA*′), without the coding sequence of its signal peptide. CDTa (49 kDa) is produced by the bacterium as an inactive precursor which is activated by a proteolytic cleavage of the signal peptide (Perelle *et al.*, 1997). The signal peptide of CDTa comprises of the first 42 N terminal residues of the full length protein and has a molecular weight of about 4 kDa. The expressed protein (CDTa′, ~45 kDa), thus lacks the N-terminal signal peptide.

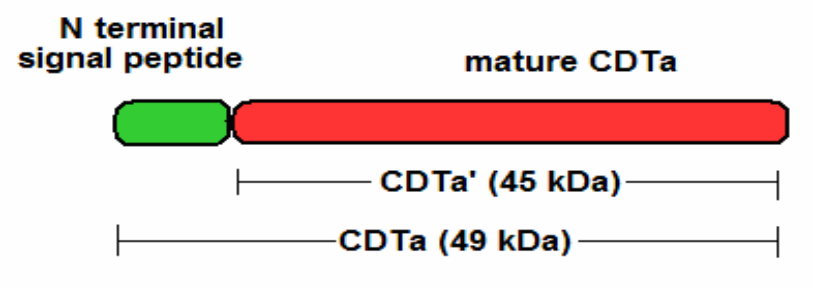

Figure 3.2: *The domain organisation of CDTa.*

Three different clones – *pMAL-HT-cdtA', pMAL-p2x-cdtA' and pGEX-6p1-cdtA'*, were prepared at this stage. Preliminary analysis of new recombinant DNA clones was performed by double digestion reaction (Figure 3.3). All three used expression vectors accept the insert in the first reading frame between BamHI and SalI restriction sites. The PCR amplified fragment had its first codon immediately after the BamHI recognition sequence. Therefore, we had the possibilities of inserting the amplified *cdtA'* into all three used expression vector backbones to produce the correct fusion protein in the first reading frame with the chosen set of restriction enzymes. The Sequence of all recombinant constructs producing DNA fragments of the expected size on an agarose gel as a result of digestion were subsequently confirmed by sequencing (Eurofins, MWG).

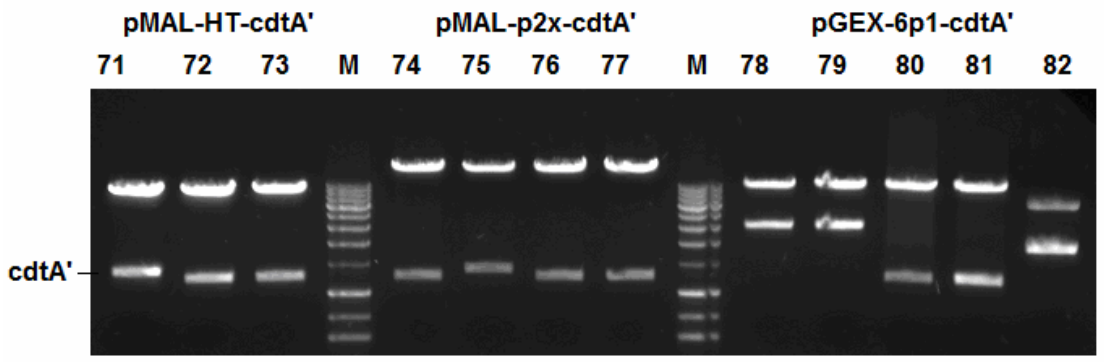

Figure 3.3: *Preliminary analysis of isolated plasmids from the selected colonies by enzymatic double digestion. The numbers were randomly assigned to colonies. DNA No. 73, 77 and 81 were sent for sequencing and were found correct. DNA No. 78, 79 and 82 did not have correct insert.*

The *pMAL-p2x* and *pGEX-6p1* vectors were obtained commercially from New England Biolabs and GE Healthcare respectively. The *pMAL-HT* vector is a modified version of *pMAL-c2x* expression vector (New England Biolabs). The

original *pMAL-c2x* vector is designed to produce a fusion protein with a cleavable N terminal MBP (Maltose Binding Protein) tag. The pMAL-HT was produced by deleting the MBP coding sequence from the commercially available *pMAL-c2x* vector and replacing it with the coding sequence for 6-His tag. This modification allowed the construct to produce an N terminal 6-His tagged fusion proteins which could be cleaved from the fusion protein with the help of factor Xa.

Expression Trials for New Clones

The new clones are capable of producing the desired protein fused with three different cleavable tags at the N terminal (Table 3.8). Based on the tag and its properties, a suitable purification strategy can be developed for purification of the fusion protein and finally the target protein i. e. CDTa′.

All different expression trials (table 3.6) produced the desired fusion proteins (Figure 3.4). The expression trial set 1 (Table 3.6) produced insoluble fusion proteins in the form of inclusion bodies for all three clones.

Table 3.8: *Details of fusion proteins produced by using different recombinant clones for CDTa′* expression.

Clone	**Fusion Protein**	**Molecular Weight (Tag + Protein)**	**Nature of Tag**
pMAL-HT-cdtA′	6His-CDTa′	48 kDa (1 kDa + 47kDa)	N terminal, cleavable by Factor Xa
pMAL-p2x-cdtA′	MBP-CDTa′	92 kDa (45 kDa + 47 kDa)	N terminal, cleavable by Factor Xa
pGEX-6p1-cdtA′	GST-CDTa′	74 kDa (27 kDa + 47 kDa)	N terminal, cleavable by PreScission protease

The TB media is richer than the LB media and supports a much faster growth of the organisms. The higher the growth rate of the organism, the higher would be the rate of fusion protein expression and thus higher the chances of protein forming inclusion bodies. Therefore, no expression trial in TB media was carried out at 37^0C.

Few more optimisations of the expression conditions (Table 3.6, trial set 3) resulted in all three fusion proteins expressed in soluble form. The level of

expression of the three fusion proteins under identical conditions was in the following order (Figure 3.4).

MBP-CDTa′ > GST-CDTa′ > HT-CDTa′

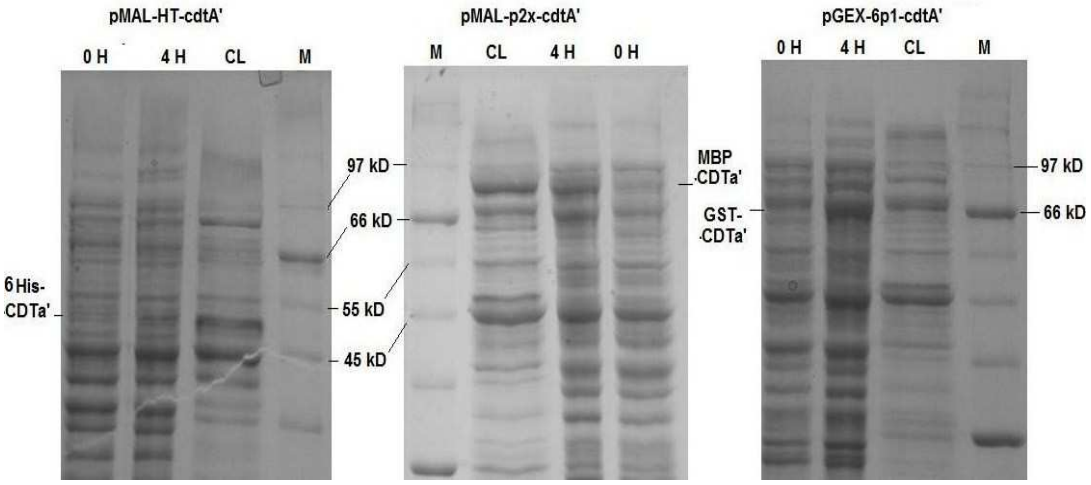

Figure 3.4: *The expression of all 3 fusion proteins in expression trial 3 (Table 3.6). 0H – pre induction sample, 4H – 4 hours post induction whole cell sample, CL – bug buster treated supernatant of 4 hours harvested sample.*

Prolonged post induction incubation for 8 hours or longer did not improve the expression of fusion proteins any further. Based on their level of expression, clone *pMAL-p2x-cdtA′* was selected for the large scale production of the protein that was carried out in shake flask method.

Purification of CDTa′

The purification of CDTa′ was completely based on its net surface charge distribution. At a pH below its isoelectric point (pI), any given protein has a net positive charge on it whereas at a pH above its pI, it has a net negative charge. Based on its net charge, proteins bind to an anion or a cation exchanger resins from which they can be selectively eluted and thus separated from each other. In this particular case, the fusion protein (MBP-CDTa′) and the tag (MBP) have their pIs in the range of 5.0 to 5.5, and therefore bear a net negative charge at the pH of buffers (i. e. pH 8.0) that are used in purification process. The theoretical pI of CDTa′ is 8.9 and hence, it is expected to have a net positive charge at pH 8.0.

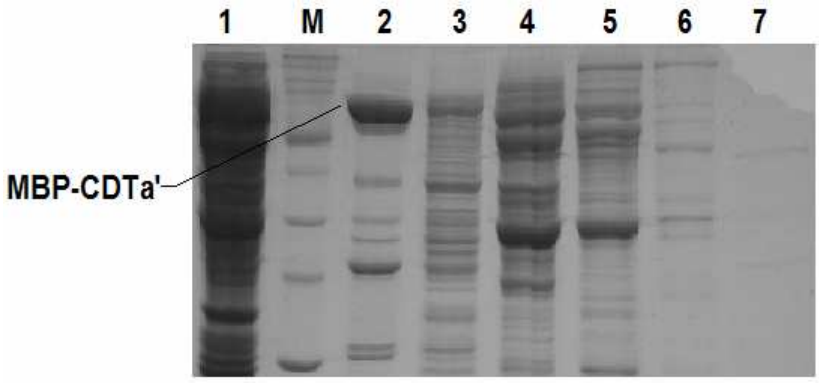

Figure 3.5: *Step elution from first anion exchange column. MBP-CDTa' is present in the first elution fraction with 10% elution buffer. 1- crude cell lysate, M – marker protein ladder, 2, 3, 4, 5, 6 and 7 – elution steps with10%, 20%, 30%, 40%, 50% and 00% buffer B in buffer A respectively.*

Out of these three species, only the fusion protein was present in the crude cell lysate. Because of its net negative charge at pH 8.0, it bound to the Q sepharose anion exchange resin and could be eluted from the column with an increased salt concentration in the mobile phase i. e. the elution buffer. Elution in steps was carried out to provide an idea of suitable range of the concentration of salt needed to elute the fusion protein from the column. Most of the desired fusion protein elutes with 10% buffer B which corresponds to 100 mM of NaCl concentration (Figure 3.5).

Results of the first anion exchange elution pattern were useful in the sense that factor Xa is most efficient in the presence of 100 mM NaCl salt. The pH of all buffers (pH 8.0) was also optimum for factor Xa mediated tag cleavage reaction to take place (Table 3.7). Factor Xa cleaves the MBP- CDTa′ fusion protein (in 24 hours at 20^0C) into two species (Figure 3.6, lanes 2 and 3) namely, MBP (~42kDa) and CDTa′ (~45 kDa).

The dialysed, tag cleaved protein sample (Figure 3.6, lane 4) contains all three species – CDTa′, MBP and the uncleaved MBP-CDTa′. The CDTa′ did not bind to the second Q sepharose column under the conditions that were identical to the loading conditions for the first anion exchange column run because of its net positive charge at pH 8.0 and could be collected in the column flow-through (Figure 3.6, lane 5) in second Q sepharose run. Whereas, the MBP and MBP-CDTa′ along with other impurities bound to the column again and thus could efficiently be separated from CDTa′ (Figure 3.6, lanes 6 and 7).

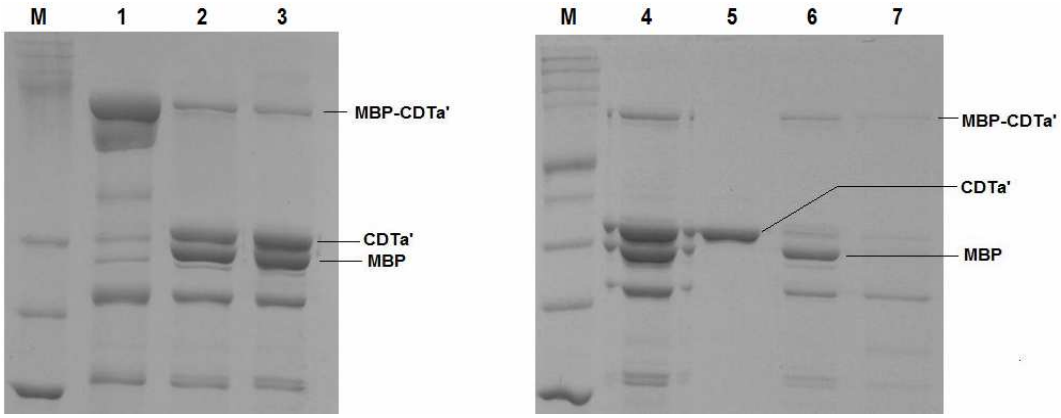

Figure 3.6: *Step wise progress of CDTa' purification process. Lane 1- 10% elution fraction, lane 2 and 3 – cleaved protein. Lane 4 - dialised cleave protein, lane 5- flow-through from second anion exchange, lane 6 and 7- elution fractions (20% and 50% of buffer D in buffer C respectively) from second anion exchange column.*

SUMMARY

Three different recombinant constructs namely – *pMAL-HT-cdtA'*, *pMAL-p2x-cdtA'* and *pGEX-6p1-cdtA'* were prepared for the expression of CDTa fragment without its signal peptide. Based on the expression pattern, CDTa' was expressed in soluble form as MBP-CDTa' fusion protein. The purification process yielded a protein (CDTa') of high purity that was stored at -80^{0}C.

B - CLONING, EXPRESSION AND PURIFICATION OF TWO DIFFERENT CONSTRUCTS OF TRANSPORT COMPONENT OF *C. DIFFICILE* BINARY TOXIN: CDTb′ and CDTb″

MATERIALS AND METHODS

Recombinant DNA Construction

Two different versions of CDTb were studied in this research. Sets of specific primers (Table 3.9) were designed to PCR amplify the DNA fragments for coding sequences of both the versions. The first version, *cdtB′*, was the DNA fragment that lacks the coding region for the N terminal signal peptide of the protein. We named this protein fragment as CDTb′. The second DNA fragment (named *cdtB″*) was the coding sequence for mature CDTb fragment and the expressed protein was named CDTb″. A recombinant construct (*pPCRscript-cdtB)* containing the coding region of full length CDTb was provided by our collaborator and was used as template DNA for both of the amplification reactions.

Table 3.9: *The primer sequences for the amplification of cdtB′ and cdtB″.*

Fragments	Primer Sequence
cdtB′	F= *GTAT* GGATCC GTG TGC AAC AC R= *AGCA* GTCGAC TTA* ATC CAC GCT CAG AAC C
cdtB″	F= *AGTA* GGATCC CTG ATG AGC GAT TGG R= *AGCA* GTCGAC TTA* ATC CAC GCT CAG AAC

F – forward primer, R - reverse primer, In italics - random 5′ overhang, underlined - restriction enzyme recognition sequence, * - stop codon.

The reaction composition and reaction conditions for both the amplification reactions are provided in table 3.10. The amplified products were analysed on a 0.8% agarose gel in 1X TAE, at 100 V for 45 minutes and the products were eluted from the gel using Promega Wizard SV Gel and PCR Clean-up system.

Both the amplified PCR fragments were then cloned into the *pGEX-6p-1* expression vector (GE Healthcare). The PCR amplified products (*cdtB′* and *cdtB″*) and the vector backbone were double digested in a 50µL reaction each,

with BamHI and SalI restriction enzymes to produce compatible sticky ends as described in the section 3A previously. Further steps of ligation, transformation, positive clone selection, plasmid isolation, DNA sequencing, preparation of *E. coli* BL21-CodonPlus (DE3)-RIPL expression host and glycerol stocks were also carried out as described in the section 3A.

Table 3.10: *The PCR reaction composition and reaction conditions for cdtB′ and cdtB″ amplification.*

Fragment	**Reaction mixture (50 µl)**	**Reaction conditions**
cdtB′	Templet DNA= 2 µl, Forward primer=2.5 µl, Reverse Primer=2.5 µl, 10X KOD buffer =5 µl, 25mM MgSO4= 2 µl,	95^0C– 300 secs, [95^0C – 60 secs, 55^0C – 60 secs, -(40 cycles) 72^0C – 90 secs]
cdtB″	DNTP mix (2mM each)= 5 µl, 5M Betaine=10 µl, DMSO=2 µl, KOD polymerase=1 µl, Water= 18 µl.	95^0C– 300 secs, [95^0C – 60 secs, 51^0C – 60 secs, -(40 cycles) 72^0C – 90 secs]

Preliminary Expression Trials of GST-CDTb′ and GST-CDTb″

The expression trials for both the fusion proteins were conducted separately but in a similar way. 200 ml of sterilised TB media (with 1 X TB salts, 100 µg/ml ampicillin and 0.5% glucose) was inoculated with an overnight grown seed culture to give 1% inoculum and incubated at 37^0C with shaking at 200 rpm. The culture was induced with the IPTG to a 1mM final concentration when the culture OD_{600} was in the range of 0.8 to 1.0. Incubation at 37^0C was continued for the next 4 hours. Samples at different post induction time were collected to run on a Tris-glycine SDS-PAGE. Sample preparation and gel run was carried out as described previously (section 3A).

More expression trials were carried out in order to express the desired protein in soluble form (Table 3.11). The parameter varied during different expression trials was temperature. All other parameters were kept identical. The culture was started at 37^0C in shake flask method at 200 rpm and was allowed to grow until the culture OD_{600} reached to 0.6 -0.8. The temperature

was then decreased to the desired value and the culture was induced with 1 mM of IPTG at OD_{600} = 0.9 - 1.0. The culture was harvested at 16-20 hours post induction, centrifuged at 10,000 rpm for 10 minutes at 4^0C and the cell pellet was stored at -80^0C.

Table 3.11: *Different expression trials for CDTb′ and CDTb″ using pGEX-6p1-cdtB′ and pGEX-6p1-cdtB″ clones in E. coli host.*

Parameters	**Trial set number**				
	1	**2**	**3**	**4**	**5**
Media	TB +1X TB salts + 0.5% glucose + ampicillin				
Host	*E. coli* BL21-CodonPlus(DE3)-RIPL				
Method	Shake flask				
Incubation conditions	Started at 37^0C, 200 rpm				
Temperatures	37^0C	30^0C	24^0C	20^0C	16^0C
Induction	1 mM / 0.5 mM IPTG at OD_{600} = 0.9 to 1.0				
Harvest	4 hours/ 8 hours / 20 hours post incubation				
Results	overexpreesed protein seen at expected position on SDS-PAGE				
Location of protein	Inclusion bodies			Soluble form	

A small sample (1ml) of harvested culture was collected separately and centrifuged as above. The pellet was resuspended in buffer F (150 mM NaCl in 50 mM Tris-HCl, pH 7.5) and was sonicated in 3 cycles of 10 sec on 20 seconds off. The sonicated sample was centrifuged at 10,000 rpm for 10 minutes at 4^0C and the supernatant was run on a 10% separating Tris-glycine gel along with the uninduced and induced (hourly) whole cell samples to confirm the presence of overexpressed protein in soluble form.

Large scale Expression of GST-CDTb′ and GST-CDTb″

The large scale expression of both of the fusion proteins (GST-CDTb′ and GST-CDTb″) in soluble form was carried out in identical manner using a BIOFLO 3000 bioreactor (NewBrunswik). TB media supplemented with 1X TB salts, 100 µg/ml ampicillin and 0.5% glucose was inoculated with 1% (v/v) of the overnight grown seed culture at 37^0C. The incubation temperature was lowered down to 16^0C when the culture OD_{600} reached to 0.6 - 0.8, and the

culture was induced with the IPTG to a final concentration of 1 mM at OD_{600} = 0.9 to 1.0.

The temperature of the bioreactor was maintained at a set value by running hot/chilled water into the vessel jacket. A mixture of air and oxygen was sparged continuously in a ratio of 40:60 (air to oxygen) at 0.5 bar pressure each into the bioreactor to maintain a minimum of 60% dissolved oxygen (DO) at all times. The pH of the culture was maintained at 7.0 throughout the process with intermittent addition of 10% orthophosphoric acid and 10% ammonium hydroxide, as and when required. An agitation rate of 150 rpm was also maintained during the run. The pH, DO and the temperature of the bioreactor were controlled in proportional-integral-derivative (PID) manner. Incubation at 16^0C was continued and the culture was harvested at 20 hours post induction time. The harvested culture was centrifuged at 10,000 rpm for 10 minutes at 4^0C and the cell mass was stored at -20^0C.

Afinity Purification and Tag Cleavage of CDTb′

The cell pellet was resuspended in lysis buffer (Table 3.12) and the cell suspension was sonicated in 5 cycles of 20 second on, 40 second off. The cell lysate was centrifuged at 25,000 rpm for 30 minutes at 4^0C and the clear supernatant was loaded onto a GST affinity column pre-equilibrated with lysis buffer.

Table 3.12: *Composition of the lysis buffer.*

Buffer	**Composition**
Lysis buffer	150 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5

The column was washed with plenty of lysis buffer until the base line was reached and the bound protein was eluted with 20 mM reduced glutathione in lysis buffer. The eluted protein was run on a 10% Tris-glycine SDS-PAGE to analyse the quantity and quality of the eluted protein.

To cleave the GST tag, sufficient amount (1 Unit / 100 µg of fusion protein) of PreScission protease (GE Healthcare) was added to the fusion protein solution and the reaction mixture was incubated at 4^0C, overnight with

continuous stirring. The cleavage reaction product was analysed on a 10% resolving Tris-glycine SDS-PAGE in parallel with the uncleaved protein.

Gel Filtration

The affinity purified fusion protein (GST-CDTb′) was loaded onto a Superdex-200 gel filtration column pre-equilibrated with lysis buffer (Table 3.12). Elution fractions of 1 ml each were collected and fractions corresponding to the eluted peak were analysed on a 10% resolving Tris-glycine SDS-PAGE. Three different gel filtration runs at a flow rate of 1.0 ml/minute, 0.5 ml/minute and 0.2 ml/minute were carried out in order to achieve best separation.

Effect of the Cell Lysis Method on Fusion Protein

The affinity and gel filtration procedures described above did not produce satisfactory results and were associated with a severe protein degradation problem. Hence, an extensive search for an appropriate purification strategy was carried out starting from the cell lysis method. The cell pellet was resuspended in lysis buffer (Table 3.12) as described before and cells were lysed using homogenizer at 400 bar pressure in 2 cycles. The cell lysate was centrifuged at 25,000 rpm for 30 minutes at 4^0C and the clear supernatant was collected. An affinity purification step was carried out in an identical way as it has been described before. At a different occasion, cell lysis using a French press at 2000 bar pressure was also tried.

Search for Suitable Purification Strategy

The affinity purified, tag cleavage reaction product was loaded again onto a GST affinity resin pre-equilibrated with lysis buffer (Table 3.13). The column flow-through was collected and was dialysed against 50 volume of dialysis buffer (Table 3.13) at 4^0C overnight. The dialysed protein was centrifuged at 10,000 rpm for 15 minutes at 4^0C to remove any particulate material and precipitate present. The clear supernatant was collected and loaded onto a Q sepharose column pre-equilibrated with dialysis buffer (Table 3.13). The loaded column was washed with plenty of dialysis buffer and the bound protein was eluted with 100 ml of 0 to 100% gradient starting with dialysis buffer and ending with elution buffer (Table 3.13). Fractions

corresponding to the eluted peak were analysed on a 4-12% Bis-Tris SDS-PAGE. Fractions containing the desired protein (CDTb') were pooled.

A need for further purification of the pooled protein was felt. Several different strategies from this point onwards were employed to serve the purpose. A flow chart shown in figure 3.7 explains some of the applied strategies. Composition of all the buffers that were used at different steps is provided in table 3.13.

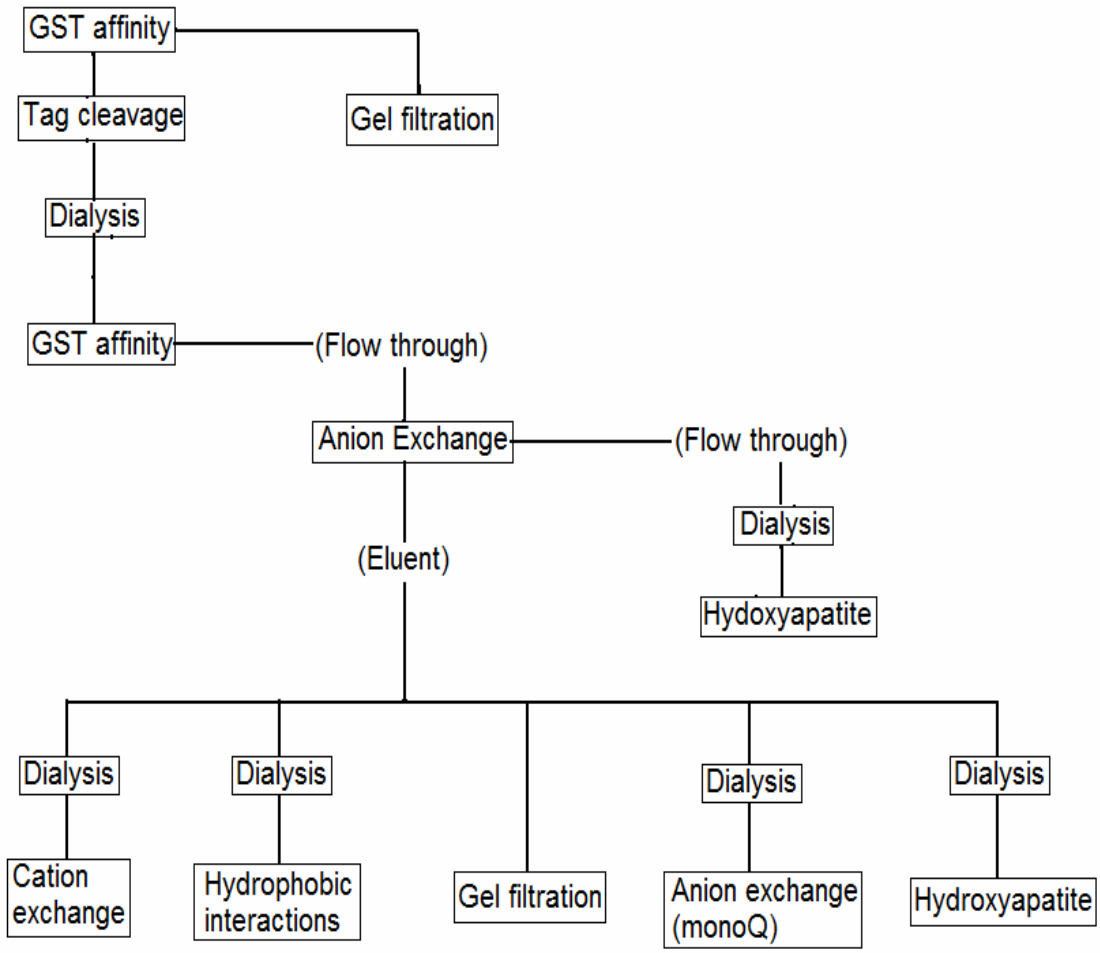

Figure 3.7: *A flow chart of different purification strategies employed.*

It became clear that several steps of dialysis were needed to be incorporated at the different stages of purification to keep the protein in an appropriate buffer to match the physiochemical conditions suitable for loading the protein onto a given particular purification resin. All the dialysis steps were carried out using 12-14 kDa cut off dialysis tubes at 4^0C overnight with continuous stirring.

Table 3.13: *Composition of buffers used at different steps of purification.*

Step	**Buffer**	**Composition**
Lysis / affinity purification and tag cleavage	Lysis buffer	150 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5
	Elution buffer	20 mM reduced Glutathion in lysis buffer
Q sepharose Anion exchange / MonoQ anion exchange	Dialysis / Equilibration buffer	50 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5
	Elution buffer	1 M NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5
SP sepharose cation exchange	Dialysis / Equilibration buffer	20 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Succinate buffer , pH 5.5
	Elution buffer	1 M NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Succinate buffer, pH 5.5
P sepharose hydrophobic exchange	Dialysis / Equilibration buffer	1M NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5
	Elution buffer	2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5
Hydroxyappatite column	Dialysis / Equilibration buffer	10 mM Sodium di hydrogen phosphate, pH 7.0
	Elution buffer	250 mM Sodium di hydrogen phosphate, pH 7.0
Gel filtration	Equilibration buffer	150 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5

A More Efficient Purification Strategy for CDTb′

Different purification strategies employed to solve the problem of degradation did not bring any significant improvement in the final purity of CDTb′. Therefore another method was employed. The cell lysis and affinity

purification steps were carried out in lysis buffer (Table 3.14) as it has been stated before.

Table 3.14: *Composition of buffers used in the alternative strategy for CDTb′ purification.*

Step	**Buffer**	**Composition**
Lysis/affinity purification/tag cleavage	Lysis buffer	150 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5
	Elution buffer	20 mM reduced Glutathione in lysis buffer
MonoQ anion exchange	Dialysis buffer	50 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5
	Elution buffer	500 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5
Concentration	Concentration buffer	200 mM NaCl, 2 mM DTT and 1 mM EDTA in 50 mM Tris-HCl, pH 7.5

The affinity eluted fraction (fusion protein) was then directly loaded onto a MonoQ anion exchange column pre-equilibrated with lysis buffer. The loaded column was washed with plenty of lysis buffer and the bound protein was eluted with a 0 to 100% gradient starting with lysis buffer and ending with elution buffer (Table 3.14). All the fractions were analysed on a 4-12% Bis-Tris SDS-PAGE and fractions containing the desired fusion protein were collected and pooled. Sufficient amount of PreScission protease (1 Unit / 100 µg of fusion protein) was added to the pooled fusion protein solution and left overnight at 4^0C with gentle shaking to cleave the GST tag.

The cleavage reaction product was once again passed through a GST affinity column pre-equilibrated with lysis buffer following a brief spin at 10,000 rpm for 15 minutes at 4^0C to remove any particulate material and precipitate present. The column flow-through was collected and dialysed against 50 volume of dialysis buffer (Table 3.14) overnight at 4^0C. The dialysed protein was loaded onto a MonoQ column pre-equilibrated with dialysis buffer and the bound protein was eluted in 0 to 100% gradient starting with dialysis buffer and ending with elution buffer (Table 3.14). Fractions containing the free CDTb′ protein were identified by Bis-Tris SDS-PAGE and pooled.

Routine Quality Check and Mass Spectrometric Analysis of CDTb′

The protein was concentrated to 7 mg/ml using 10 kDa MWCO concentrators. The protein quantity was estimated by recording absorbance at 280 nm and the concentrated protein was stored at -80^{0}C. The stored protein sample was run on a 10% Tris-glycine SDS-PAGE after a week's time for routine analysis.

For mass spectroscopy analysis, the protein was buffer exchanged into water using a concentrator of 30 kDa MWCO, at 4^{0}C and 2000 rpm. The buffer exchanged protein was run on a 10% Tris-glycine SDS-PAGE along with the original concentrated sample under reducing as well as non-reducing conditions. The protein was analysed by mass spectroscopy facility at the Department of Chemistry, University of Bath.

Final Purification of CDTb′ and CDTb″

The cells were lysed by homogenisation in buffer F (Table 3.15) and the cell lysate was centrifuged as explained before. The clear supernatant was loaded onto a GST affinity column and the column was washed with plenty of buffer F until the base line was reached. The bound protein was eluted with 20 mM of reduced glutathione in buffer F.

Table 3.15: *Composition of buffers used for final purification of CDTb'.*

Buffer	**Composition**
Buffer F	150 mM NaCl and 2 mM DTT in 50 mM Tris-HCl, pH 7.5
Buffer G	1000 mM NaCl, 2 mM DTT in 50 mM Tris-HCl, pH 7.5
Buffer H	50 mM NaCl, 2 mM DTT and 0.2% tween-20 in 50 mM Tris-HCl, pH 7.5
Buffer I	1000 mM NaCl, 2 mM DTT and 0.2% tween-20 in 50 mM Tris-HCl, pH 7.5

The eluted fusion protein was loaded onto a MonoQ anion exchange column pre-equilibrated with buffer F followed by extensive washing of the column with plenty of buffer F. The bound protein was eluted in a 0 to 60% gradient of buffer G (Table 3.15) in buffer F. All of the fractions containing the desired fusion protein were collected and pooled based on a 4-12% Bis-Tris

SDS-PAGE result. Sufficient amount of PreScission protease (1 Unit / 100 µg of fusion protein) was added to the pooled fusion protein solution and left at 4^0C overnight with gentle shaking for the tag cleavage reaction to take place.

The cleavage reaction product was dialysed against 50 volume of buffer H (Table 3.15) overnight at 4^0C. The dialysed protein was centrifuged at 10,000 rpm for 15 minutes at 4^0C and the supernatant was passed through a GST affinity column pre-equilibrated with buffer H and the column flow-through was collected. The collected flow-through was then loaded onto a monoQ column pre-equilibrated with buffer H and the bound protein was eluted in a 0 to 60% gradient of buffer I (Table 3.15) in buffer H. Fractions containing the free CDTb′ (or CDTb″) protein were identified on a 4-12% Bis-Tris SDS-PAGE and pooled. The pooled protein was stored at -80^0C.

Quality Analysis and Quantification of Protein

The quality of proteins (CDTb′ and CDTb″) was assessed by running protein samples on a freshly casted 10% resolving Tris-Glycin-SDS-PAGE in Tris-Glycin buffer using a standard protocol. In addition to that, two different commercially available SDS-PAGE systems – a 4-12% NuPAGE Novex Bis Tris Gels (BT gels) in MES buffer and a 4-20% Novex Tris Glycine gels (TG gels) in Tris Glycine buffer were also employed for the analysis of both the proteins. Both the gel systems and running buffers were purchased from Invitrogen and were used as per the instructions provided by manufacturers.

The protein quantity in all the samples except the crude cell lysates was estimated by recording the absorbance at 280 nm wavelength. Theoretical absorbance (for 1 mg protein per ml sample) was calculated by submitting protein sequence to the ProtParam application of Expasy proteomic server (http://www.expasy.org).

RESULTS AND DISCUSSION

Recombinant DNA Construction

The primers were designed to amplify the desired DNA fragments and to clone it into the first reading frame between BamHI and SalI sites. The amplified DNA fragments (*cdtB*′ and *cdtB*″) code for two different versions of

the transport component of *C. difficile* binary toxin, named CDTb′ and CDTb″ respectively. All sets of primers were designed with a random 4 nucleotides overhang followed by the restriction enzyme recognition site at the 5′ end of the coding sequence for the reasons that have been explained before (section 3A). The PCR amplified cdtB′ fragment is shown in figure below (Figure 3.8)

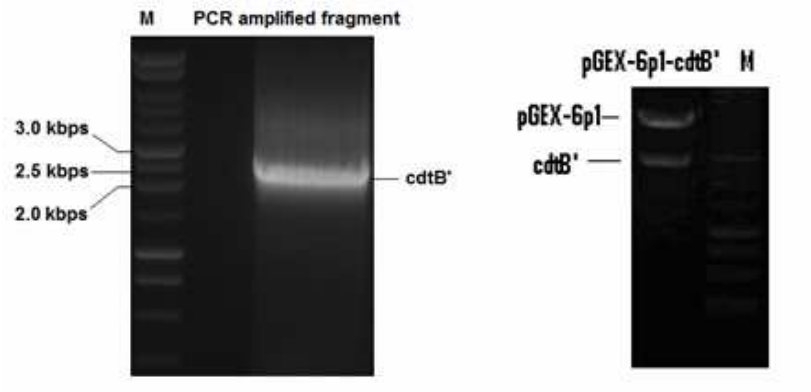

Figure 3.8: *The PCR amplified cdtB′ (left) and double digestion of positive pGEX-6p1-cdtB′ clone (right).*

Positive colonies were successfully grown on LB-agar plate for both of the desired clones namely *pGEX-6p1-cdtB′* and *pGEX-6p1-cdtB″*. A preliminary analysis of these clones by double digestion method yielded DNA fragments of the expected size on an agarose gel (Figure 3.8). Sequencing of both recombinant constructs confirmed the presence of the correct vector backbone and insert in the correct orientation and position.

Expression of Proteins

The transport component of *C. difficile* binary toxin, CDTb (99 kDa) is produced as an inactive precursor molecule with an N-terminal signal peptide of 42 residues. Figure 3.9 shows the domain organisation of CDTb. CDTb′ is the name given to the fragment of CDTb that lacks the N-terminal signal peptide.

However, to transport the enzymatic component (CDTa) into the target cell, CDTb has to be activated by a proteolytic cleavage (Perelle *et al.*, 1997) mediated by chymotrypsin. As a result of chymotrypsin mediated activation, a 25 kDa N-terminal fragment of precursor CDTb is cleaved from the protein and the remaining C-terminal large fragment (75 kDa) has been suggested to

oligomerise to form a heptameric pore like structure (Barth *et al.*, 2000; Blocker *et al.*, 2001).

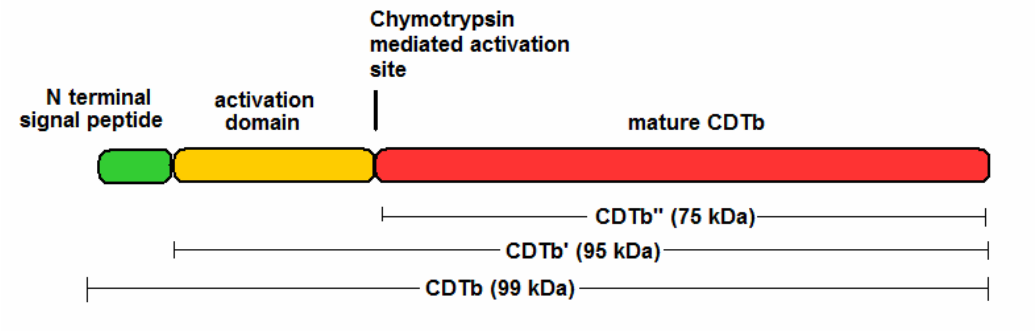

Figure 3.9: *The domain organisation of CDTb.*

This essentially means that the expressed CDTb′ requires a proteolytic activation by chymotrypsin to become fully functional protein whereas the expressed CDTb′′ should be a fully active (mature) fragment of CDTb (Figure 3.9). With an N-terminal GST fusion partner (27 kDa), expected molecular weight of both of the expressed fusion proteins (GST-CDTb′ and GST-CDTb′′) are 122 kDa (Figure 3.10) and 102 kDa respectively.

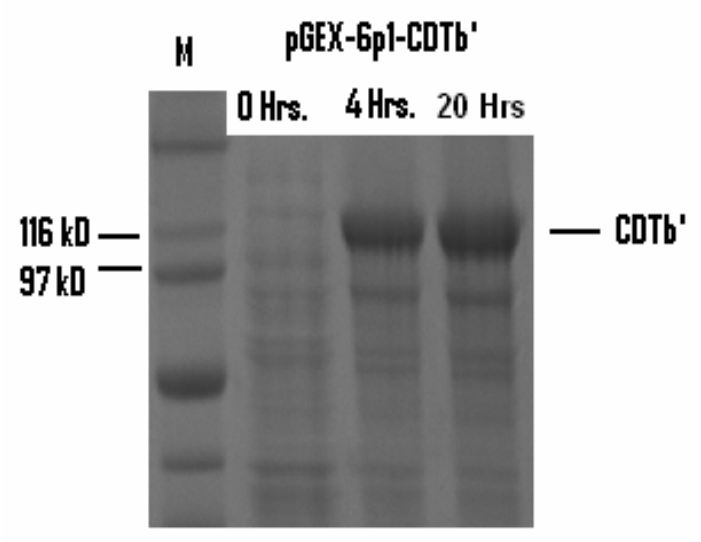

Figure 3.10: *The expression samples of GST-CDTb′ fusion protein on a 10% Tris-glycine SDS-PAGE.*

Affinity Purification and Tag Cleavage of CDTb′

The gel analysis revealed that the affinity eluted fraction consisted of two major proteins bands (Figure 3.11, lane 1). The upper observed band

present at about 120 kDa corresponded to the expected molecular weight of GST-CDTb′ fusion protein. The lower major band on the gel was observed at about 97 kDa marker protein. The expected molecular weight of GST tag is 27 kDa and it was initially thought that the tag had been cleaved off the fusion protein in solution to give free CDTb′. It could be confirmed by any of the following two methods.

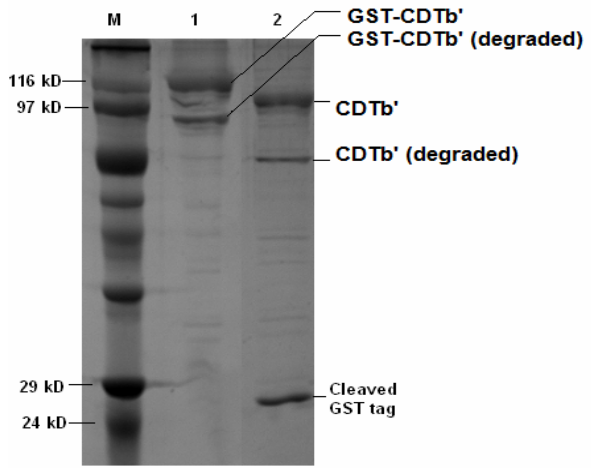

Figure 3.11: *The affinity purified GST-CDTb′ protein and the tag cleavage reaction results.*

The first method exploits the biological specificity by western blot analysis of eluted fraction against anti-GST antibody. If the lower band present on gel (Figure 3.11, lane 1) was tag cleaved free protein, it should not be detected by the anti-GST antibody on the blot. The second method was to perform a tag cleavage reaction. This method takes advantage of the size of the tag. If the lower band was the tag cleaved protein, no observable shift in the position of the band on SDS-PAGE should be detected. However, there should be a position shift for the upper 120 kDa band yielding two bands on SDS-PAGE – one corresponding to the free CDTb′ at 95 kDa and the other corresponding to GST tag at 27 kDa. In the absence of anti GST-antibody, the method of tag cleavage was employed.

A shift in positions of both the protein bands was observed. Three bands were clearly visible on SDS-PAGE (Figure 3.11, lane 2). The top most protein band at 95 kDa (CDTb′) and the bottom most band at 27 kDa (GST tag) of lane 2 of Figure 3.11, were generated from the upper GST-CDTb′ fusion protein band (120 kDa) of lane 1 (Figure 3.11). However, the middle band at

about 66 kDa (Figure 3.11, lane 2) is the tag cleaved product of the lower band in the affinity purified fraction (Figure 3.11, lane 1). The result clearly indicated that the lower band in the affinity purified fraction was not the tag free CDTb′ as it was thought resulting from auto-tag cleavage in solution.

What does the presence of the lower band in affinity purified fraction (Figure 3.11, lane 1) imply? This protein band results from the degradation of the fusion protein. The GST tag is at the N-terminal of the fusion protein and is intact in both the major components of the elution fraction (Figure 3.11, lane 1) and we could still see a position shift for both the protein bands (Figure 3.11, lane 1) as a result of tag cleavage reaction (Figure 3.11, lane 2). Hence, this degradation of the protein must be taking place at the C terminal end. Since it is the large C terminal fragment that becomes functionally active on chymotrypsin mediated activation; this degradation somewhere at the C terminal end defeats the whole purpose of purification process completely. Several other pilot purification trials in the presence of different concentrations of protease inhibitor solutions / PMSF / DTT and/or EDTA did not yield any improvement.

Gel Filtration

The gel image (Figure 3.11) provides a fair idea of the molecular weight difference between the fusion protein and its degraded companion which is about 25 to 30 kDa.

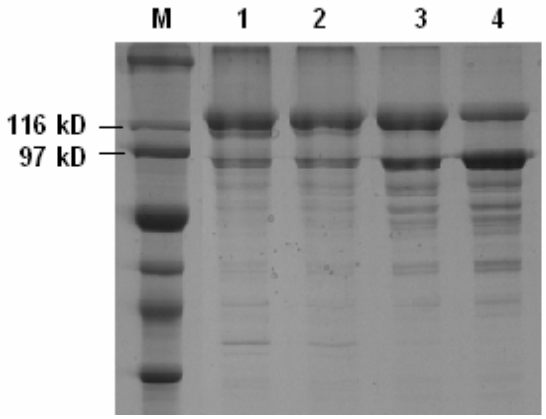

Figure 3.12: *Results of the Gel filtration run. Lanes 1 to 4 are the early to late fractions of elution at a flow rate of 0.2 ml / minute.*

The gel filtration chromatography separates proteins on the basis of their molecular weight (size) and could have been a method of choice to remove the contaminating protein from the protein of interest. The separation pattern of all the three gel filtration runs at different flow rates was similar to each other (Figure 3.10). There is a variation in the ratio of higher to lower band intensity as we proceed from lane 1 to 4 (early to late fractions of elution). This pattern was expected as the high molecular weight protein elutes first from the column. None of the run, however, could separate the two proteins efficiently.

Effect of Cell Lysis Method on Fusion Protein

The first change that was incorporated in the purification protocol was the use of homogeniser replacing sonicator for the cell lysis. A drastic decrease in the amount of the lower molecular weight protein in the affinity elution fraction was observed on SDS-PAGE (Figure 3.13, lanes 1 and 3).

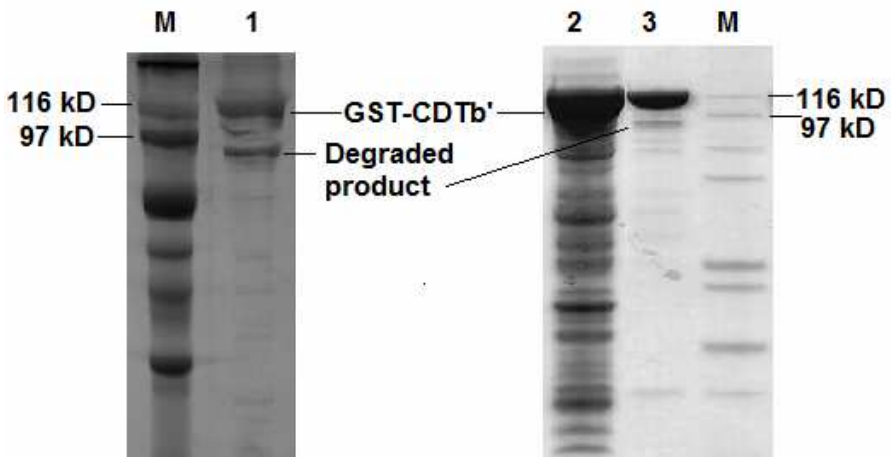

Figure 3.13: *Comparison of the affinity eluted protein using sonication (lane 1) and homogenisation (lane 3) as the method of cell disruption. Lane 2 - Crude cell lysate.*

Homogenisation is a milder method of cell disruption. It is based on mechanical shearing. In homogenization, the cell suspension is passed through a small orifice and made to strike against a metallic O ring at a very high speed causing rapture of the cell wall. On the other hand, in sonication, the cell disruption is carried out by using the energy of ultrasonic waves which produces a significant amount of heat during the process that may result in

denaturation and degradation of the protein. Cell lysis using another method (French press), also did not prove effective (data not shown).

Search for Suitable Purification Strategy

Having found a suitable method for cell lysis, further attempts to purify the protein were continued. None of the various strategies employed (Figure 3.7) produced crystallisation quality protein. The best quality (purity) of protein was produced by following the strategy shown by shaded path in the following flow chart (Figure 3.14). Protein from the first Q sepharose anion exchange looked pretty much clean and promising when it was run on a 4-12% Bis-Tris SDS-PAGE (Figure 3.14).

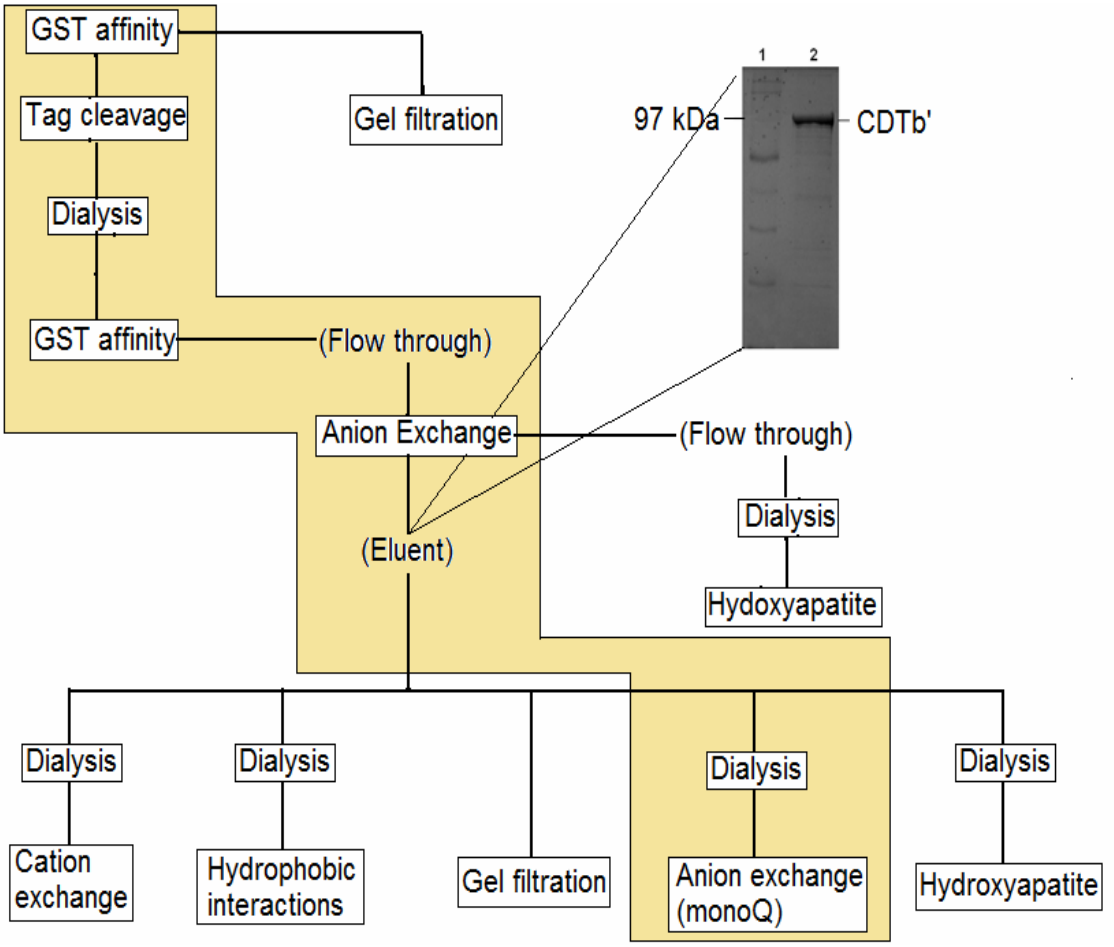

Figure 3.14: *The best protein producing strategy (shaded) and the quality of the protein after first Q sepharose anion exchange.*

Further purification of the protein shown (Figure 3.14), however, was not successful. The second anion exchange step on a MonoQ column removed many more impurities and concentrated the desired protein to a small volume.

The MonoQ sepharose is an anion exchanger with much finer particle size and hence provides an improved resolution and separation of proteins as compared to other Q sepharose resins. It was this stage, where impurities still present in the sample were also concentrated and became visible on the gel (Figure 3.15, lanes 4 to 26).

The elution fractions highlighted in the rectangle (Figure 3.15), consisted of highly concentrated protein. However, this protein was not considered suitable for crystallisation trials, and hence, further attempts were made to purify the protein.

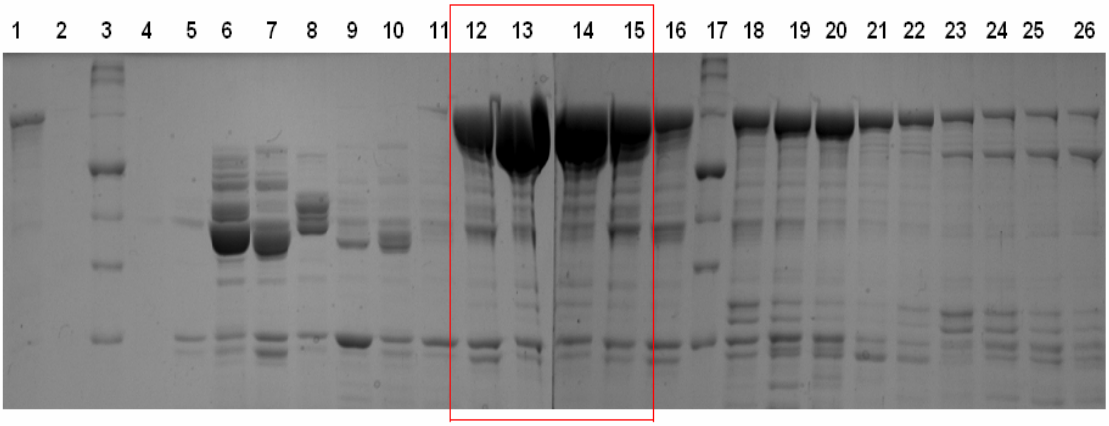

Figure 3.15: *The quality (purity) of protein at the end of the shaded strategy in figure 3.14. 1- monoQ load, 2- monoQ flow-through , 3- protein marker, 4 to 26 – eluted protein fractions.*

Purification, Concentration and Storage of CDTb′

A different purification strategy based on the difference in pI of the two proteins i.e. fusion protein (GST-CDTb′) and tag cleaved free CDTb′, was tested. Theoretical pI values for these two proteins are in the range of 4.8 to 5.0 and differ from each other only by 0.12 pH unit. It indicated that it would be almost impossible to separate them from each other on an ion exchange column and both of them would elute over the same range of salt concentration from the column. However, it was found that the fusion GST-CDTb′ protein elutes at about 200-250 mM of salt while the free CDTb′ elutes at about 150-200 mM of salt concentration in the elution buffer from the MonoQ anion exchange column under identical conditions (Figure 3.16).

This difference in elution pattern was not observed using ordinary Q sepharose resin due to the lack of resolution. However, this difference could be used in purification to improve the purity of the protein. Hence, the purification method used at this stage was as follows (Figure 3.17).

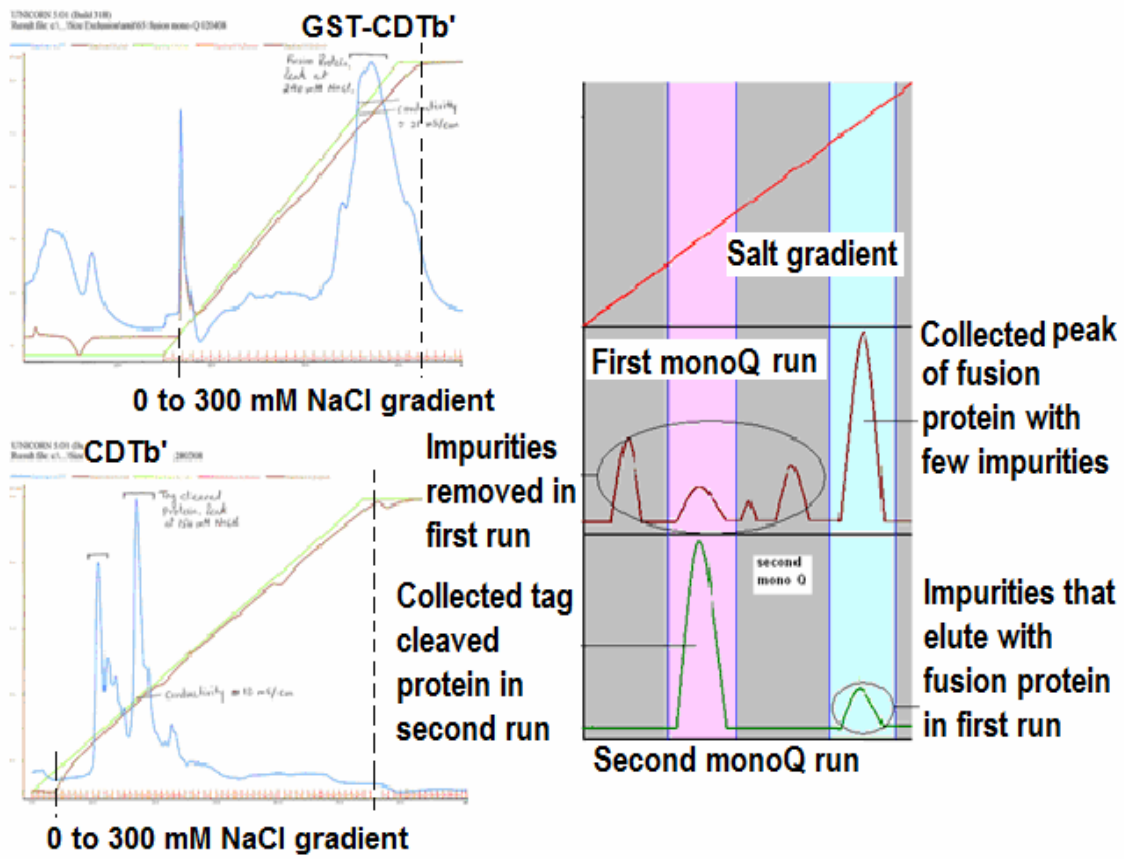

Figure 3.16: *The final purification strategy used for CDTb' purification.*

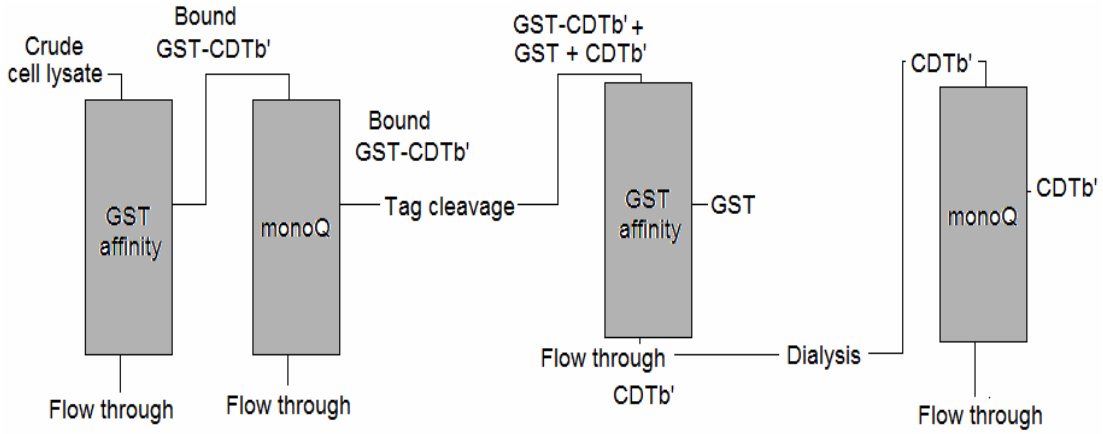

Figure 3.17: *The CDTb' purification strategy based on the pI difference.*

In the first anion exchange, the elution fractions corresponding to 200-250 mM salt (containing GST-CDTb′) were collected to get rid of the impurities that were eluted at 150-200 mM salt. In the second mono Q run, fractions containing the desired free protein (CDTb′) were collected corresponding to 150-200 mM salt concentration. The impurities that were eluted with the fusion protein in the first anion exchange step and remained unmodified

would still elute at the same 200-250 mM salt concentration under identical conditions of protein loading and elution and thus can be separated from the target protein i. e. CDTb′, in the second anion exchange step.

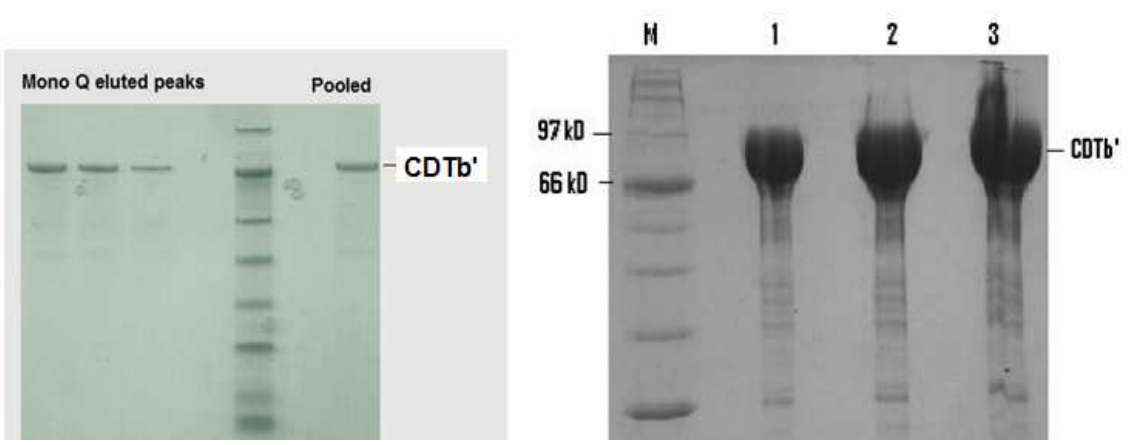

Figure 3.18: *The purified CDTb′. A – purified (pooled) CDTb′ from 3 different purification batches. B – concentrated CDTb′ protein.*

The eluted fractions were analysed on a gel and fractions containing CDTb′ were collected and pooled. Figure 3.18 A shows purified CDTb′ from 3 different batches on a 4-12% Bis-Tris SDS-PAGE. Protein quality from all three batches does not seem to differ from each other. Though the protein is in diluted form, it was certainly much better than the protein purity that was achieved by means of any other purification strategy used until then (Figure 3.15). Figure 3.18 B shows concentrated CDTb′ protein on an SDS-PAGE. A direct comparison of Figure 3.18 B with Figure 3.15 clearly shows the improvement in the quality of purified protein.

Routine Quality Check and Mass Spectroscopic Analysis of CDTb′

The concentrated protein, stored at -80^{0}C was analysed on 10% Tris-glycine SDS-PAGE after one week of purification. On the gel, CDTb′ looks degraded resulting in a major protein band at around 65 kDa molecular weight (figure 3.19). Looking at the molecular weight difference of CDTb′ and the degraded product on the gel, it was suspected that it was the same degradation at the C terminal of the protein that occurred during the cell lysis step by sonicating the cell suspension.

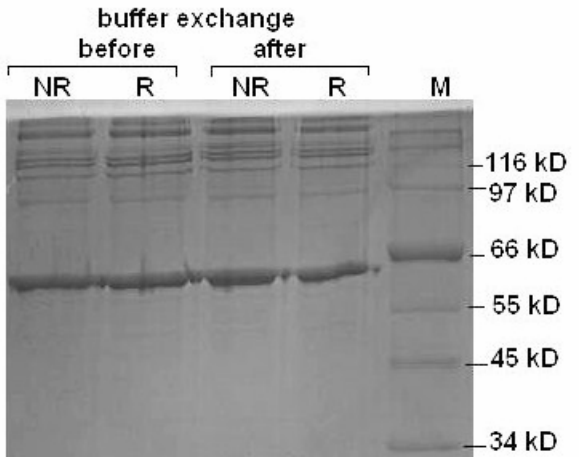

Figure 3.19: *Assessment of the protein (CDTb') quality on buffer exchange into water. NR- non reducing condition, R- reducing condition.*

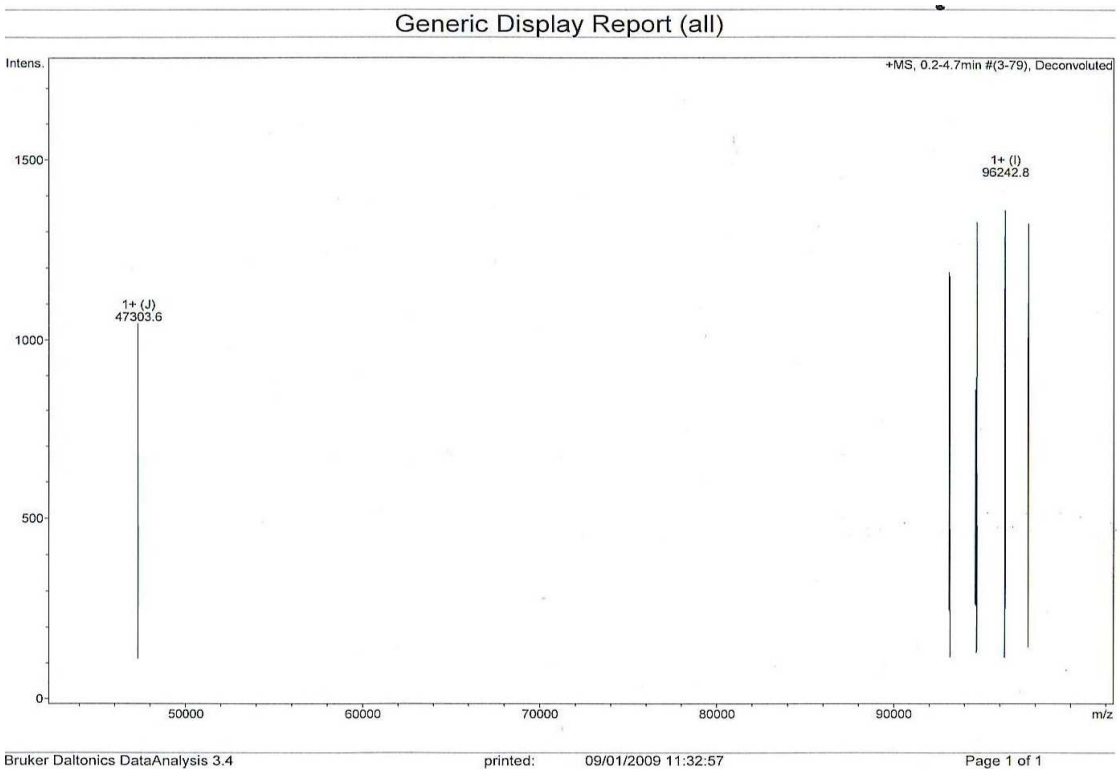

Figure 3.20: *The mass spectroscopy results for CDTb'.*

To find out the molecular weight of the degraded product, mass spectroscopic analysis of the protein sample was carried out. The protein was stored in 50 mM Tris-HCl containing 200 mM NaCl. Tris is not a recommended buffer for mass spectroscopy and hence, the protein was buffer exchanged into water prior to its mass spectroscopic analysis. The buffer exchanged protein was run on a 10% Tris-glycine SDS-PAGE along with the original concentrated sample. Buffer exchange into water did not cause any

observable change in the quality of the protein under reducing as well as non-reducing conditions (Figure 3.19).

The mass spectrometric analysis did not show any peak at or around 65 kDa molecular weight. A peak at 47 kDa was detected (Figure 3.20). However, there was no protein band detected on the gel at 47 kDa (Figure 3.20). The analysis was conducted within few hours of buffer exchange to avoid any ambiguity in the results.

There are several examples of proteins which do not appear at their theoretical molecular weight on SDS-PAGE. The TcdC protein from *C. difficile* is one such protein. It has a molecular weight of 27 kDa but appears at about 34 kDa on Tris-glycine SDS-PAGE (Govind *et al.*, 2006; and unpublished data from our laboratory). No other reason could be thought for the absence of a peak at 65 kDa when we have most intense band on SDS-PAGE at that position Figure 3.19). However, whether it is the 47 kDa molecular weight protein that appears at 65 kDa on Tris-glycine gel was not clear.

Final Purification of CDTb′ and CDTb″

Both of the target proteins (CDTb′ and CDTb″) were purified successfully (Figure 3.21). Addition of 0.2% Tween 20 enhanced the purity of both the proteins.

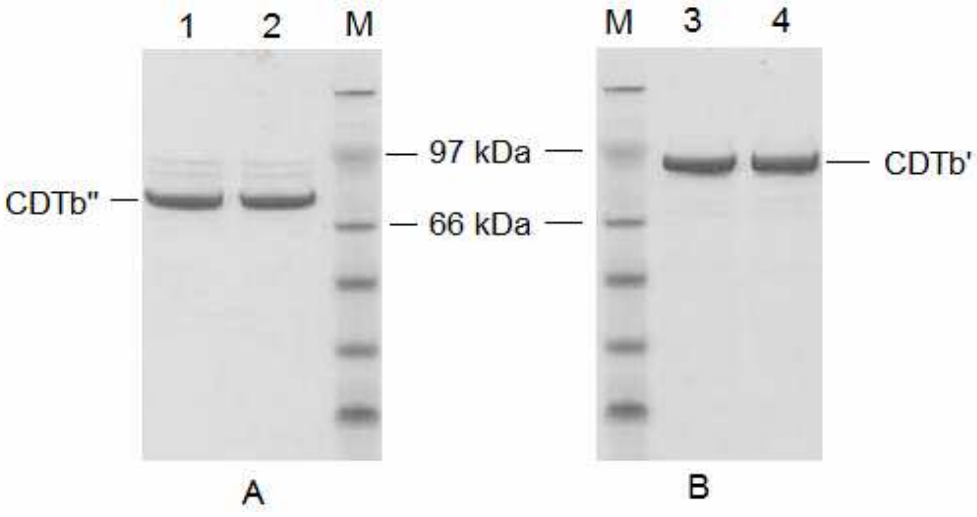

Figure 3.21: *The purified CDTb′ and CDTb″ on the Bis-Tris SDS-PAGE system. Lane 1- 3rd day CDTb″ stored at -20^{0}C, lane 2- 3rd day CDTb″ at -80^{0}C, lane 3 - 11th day CDTb′ stored at -20^{0}C, lane 4 - 11th day CDTb′ stored at -80^{0}C.*

The storage concentrations of CDTb′ and CDTb″ were about 0.20 mg/ml and 0.15 mg/ml respectively. The purified protein was tested on Bis-Tris SDS-PAGE over a period of several days to analyse the degradation of protein over long term storage and to check the effect of freeze-thaw process on protein quality. No observable difference in the protein quality was detected on the gel (Figure 3.21).

Abnormal Behaviour of CDTb′

The purity of proteins was regularly checked on SDS-PAGE during the process of purification and storage. To understand the ambiguous results obtained for SDS-PAGE analysis and mass spectrometry, for CDTb′, two different types of SDS-PAGE systems (Bis-Tris system and Tris-glycine system) were tested. Figures 3.22 A and 3.22 B compare identical protein samples of CDTb′ and CDTb″ along with the purified CDTa′ on both types of the gel systems following multiple cycles of freeze-thaw. Both of the gels were run simultaneously to avoid any ambiguity in comparing results.

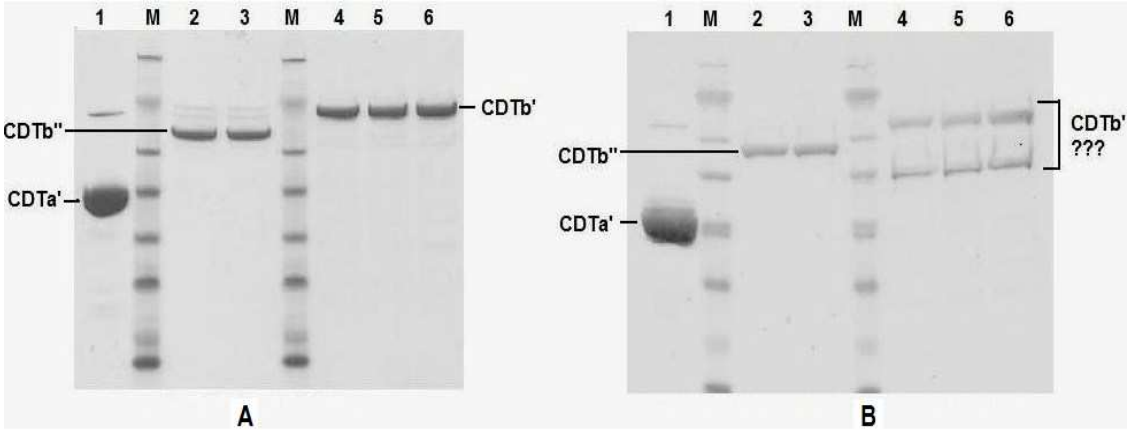

Figure 3.22: *The purified CDTb' and CDTb" (A)- on the Bis-Tris SDS-PAGE system, (B) - on the Tris-Glycine SDS-PAGE system. Lane 1- CDTa', 2- 3rd day CDTb" stored at -20^{0}C, 3- 3rd day CDTb" at -80^{0}C, lane 4 - 11th day CDTb' stored at -20^{0}C, lane 5 and 6- 11th day CDTb' stored at -80^{0}C.*

The degradation of free CDTb′ was observed only on Tris-Glycine SDS-PAGE (Figure 3.22B, lanes 4, 5 and 6) whereas the protein appeared to be perfectly fine on a Bis-Tris SDS-PAGE system (Figure 3.22 A, lanes 4, 5 and 6) even after multiple cycles of freeze-thaw. The other two proteins, CDTa′ (Figure 3.22 A and B, lane 1) and CDTb″ (Figure 3.22 A and B, lanes 2 and 3) appear as a single band at the correct position on both types of the gel systems and

hence were used as controls. The protein bands in the protein standard used (SeeBlue plus2, from Invitrogen) appear to have different mobility on both types of gel systems. Therefore, all three tested proteins (CDTa′, CDTb′ and CDTb″) appear at different positions with respect to the standard used on the two gel systems.

An excellent study by Hachmann and Amshey (Hachmann and Amshey, 2005) helps in understanding the abnormal behaviour of proteins on SDS-PAGE systems. The authors have suggested that in general, proteins are more prone to degradation on a Tris-Glycine gel than on a Bis-Tris gel for the reasons such as pH-dependent modifications in proteins, modification of sulphydryl groups and formation of acrylamide adducts to sulphydryl groups or amino groups on the protein. These effects are expected to be high at a higher pH which is a condition for Tris-Glycine gel. Hydrolysis of aspartate-proline (DP) bonds has been reported to occur when traditional Laemmli method of sample preparation is employed (Tang, 1997; Kubo, 1994).

However, it is also worth noting that not all DP bonds are liable to hydrolysis under these conditions and this is not the only mode of peptide bond cleavage under these particular conditions. Thus, it is possible that the local environment of a DP bond is responsible for its hydrolysis. As a result of these modifications, the peptide band may not be recognized and the protein could appear as multiple bands on the gel (Hachmann and Amshey, 2005) (Figure 3.22 B, lanes 4, 5 and 6).

A closer look of the protein sequences reveals that CDTa′ (appendix I) lacks DP bonds and no abnormality was observed for CDTa′ on the two tested gel systems. However, presence of 7 such potential sites in CDTb′ (appendix I) makes this protein highly prone to the above stated modifications. Five of these DP bonds are present in CDTb″ too (appendix I). The CDTb″ did not show any abnormal behaviour (Figures 3.22 A and B, lanes 2 and 3). It is possible that these DP bonds present in CDTb″ are the less labile sites for modification. We can also not rule out the possibility that this abnormal behaviour of CDTb′ (Figures 3.22 B, lanes 4, 5 and 6) could be a result of a more complex modification process.

Another important point to make here is that, this abnormal behaviour of CDTb′ was observed only when the GST tag was removed from the protein. The GST-CDTb′ fusion protein was prone to degradation by sonication (Figure

3.11) but did not show such abnormal behaviour. The fusion protein appeared as a single thick band on Tris-Glycine gel (Figure 3.10). How the fusion partner (GST) protected the protein (CDTb′) from showing abnormal behaviour is also not clear.

The observed difference in the gel pattern of CDTb′ on different SDS-PAGE systems was in agreement with mass spectrometry results (Figure 3.20) for the protein where no peak was detected in the range where degraded product (lower protein band) was present (i.e., at 65 kDa (Figures 3.19 and 3.22B, lanes 4, 5 and 6) as observed on the Tris-Glycine gel.

CDTb is the transport component of CDT which corresponds to Ib of C. perfringens iota toxin and C2II of C. botulinum C2 toxin. The crystal structure of C2II has been published along with its purification method (Schleberger, 2006). In terms of sequence, closest members to CDTb, is the transport component of iota family binary toxins. Such abnormal behaviour has not been reported for C2II and Ib despite the presence of such potential DP sites for modification.

SUMMARY

Two different constructs of CDTb were cloned into *pGEX-6p1* system. Both the constructs were overexpressed in soluble form as GST fusion proteins. CDTb′ protein was found to be sensitive to the method of cell lysis. Furthermore, the protein (CDTb′) exhibits abnormal behaviour on a Tris-glycine SDS-PAGE which made its purification a time consuming task. Purification of CDTb″ was relatively easy once a method for CDTb′ purification was established.

CHAPTER – IV

CHARACTERISATION AND CRYSTALLISATION OF *C. DIFFICILE* BINARY TOXIN

MATERIALS AND METHODS

Chymotrypsin Mediated Activation of CDTb′

Chymotrypsin and trypsin inhibitor from hen egg white (both from Sigma) were dissolved in buffer F (Table 3.15) at 1 mg/ml concentration. Chymotrypsin was added to the protein (CDTb′) to yield 1:10 (chymotrypsin to protein) ratio. The mixture was incubated at room temperature (25^0C) and samples were taken at 10, 15, 20, 25, 30, 40, 50 and 60 minutes time points.

Trypsin inhibitor was added to the samples to give a 1:2 ratio (chymotrypsin to trypsin inhibitor). 4X NuPage gel loading dye (Invitrogen) was added to the samples followed by heating of the samples for 5 minutes at 95°C. The samples were analysed on a 4-12% Bis-Tris SDS-PAGE system in MES buffer. Suitable heated and non heated control protein samples (CDTb′) were also run in parallel. The resulting activated fragment of CDTb′ was named CDTb′#.

Vero Cell Culture

Vero cells (kidney epithelial cells extracted from African green monkey) were grown in complete Dulbecco's Modified Eagle Medium (DMEM, supplemented with 10% heat inactivated fetal calf serum (FCS) and 2 mM glutamate) at 37^0C in the presence of 5% CO_2 in air. Cells were routinely trypsynised and passaged twice a week. For the cytotoxicity assay, cells were trypsinised and used to coat 96 well plates in a total volume of 200 µl of complete DMEM (medium supplemented with FCS and glutamate). The plates were incubated as above for 16-24 hours to allow the formation of a confluent monolayer. To perform the cytotoxicity assay, the medium was removed gently from the wells without disturbing the cell layer and the cells were washed twice with Dulbecco's phosphate buffered saline (DPBS). 100 µl of serum free DMEM was added to the wells and the cells were incubated at 37°C in the presence of 5 % CO_2 in air.

Cytotoxicity Effects of Complete CDT

To assess the cytotoxic potential of CDT, both components of CDT were added to Vero cell monolayers. In the first set of cytotoxicity assays, the cells were incubated with CDTa′+CDTb′ and CDTa′+CDTb″ (250ng + 250ng) at

37^0C. Suitable buffer and protein controls were also set up by incubating the cells with buffer, CDTa′, CDTb′ and CDTb″ (250 ng each) alone, under identical conditions. A positive control with 50 ng/ml of *C. difficile* Toxin A was also set up. All experiments were set up in duplicate in a total volume of 200 µl each and the cells were examined at 4 hours post incubation time using an Olympus CK2 inverted microscope.

In the second set of experiments, Tween-20 was completely removed from CDTb′ and CDTb″ protein stocks before testing the proteins on the cells. Proteins were dialysed against 50 volume of buffer P (50 mM NaCl in 50 mM Tris-HCl, pH 7.5) overnight. Each dialysed protein was loaded onto a Q sepharose column and the column was washed with buffer P until the base line was reached. Bound protein was eluted in one step with buffer Q (600 mM NaCl in 50 mM Tris-HCl, pH 7.5). An equal volume of 50 mM Tris-HCl, pH 7.5 was added to the protein to bring the final salt concentration to 300 mM and the protein was stored at -80^0C.

Varying amounts of CDTa′ (50, 100, 150, 200 and 250 ng) were mixed with equal amounts of CDTb′ or CDTb″ in different combinations in a total volume of 100 µl of serum free DMEM and added to the monolayers in separate wells. A set of tests with identical amounts of CDTa′, but with chymotrypsin activated CDTb′ and CDTb″ (named as CDTb′$^{\#}$ and CDTb″$^{\#}$) were also prepared. Suitable negative controls with individual proteins CDTa′, CDTb′, CDTb″, CDTb′$^{\#}$ and CDTb″$^{\#}$ (250 ng each) and positive controls with *C. difficile* Toxin A (50 ng/ml) and Toxin B (0.5 ng/ml) were also set up under identical conditions. All the experiments were performed in duplicate and the cells were incubated at 37°C in the presence of 5 % CO_2 in air. Cells were examined for evidence of cytotoxic effect after 24 hours incubation using an Olympus CK2 inverted microscope.

CDTb Oligomerisation in Solution

Eight different buffer systems (MIB pH 4.0, MMT pH 4.0, SPG pH 4.0, Na-Acetate pH 4.0, Bis-Tris pH 5.5, Bis-Tris pH 6.5, Tris-HCl pH 7.5 and Tris-HCl pH 8.5) were screened for CDTb oligomerisation experiment. CDTb′ was treated with chymotrypsin to produce an activated CDTb fragment. Chymotrypsin was deactivated by adding trypsin inhibitor to the reaction mixture as described before. 20 µl of chymotrypsin activated protein was

mixed with 2 µl of 1 M of each buffer in separate reaction tubes and incubated at 4°C overnight. On the next day, 5 µl of NuPAGE gel loading dye was added to all samples. The samples were analysed on a 4-12% Bis-Tris SDS-PAGE in MES buffer.

Concentration and Crystallization of CDTa′

The protein (CDTa′) was concentrated further to 10 mg/ml using a 10 kDa molecular weight cutoff (MWCO) concentrator (Millipore) at 4000 rpm at 4^{0}C. Primary crystallisation of CDTa′ was set up with the help of Phoenix crystallisation robot using five different commercially available crystallisation screens from Molecular Dimensions Limited, namely: (i)- Structure Screens I and II, (ii)- Clear Strategy Screen I, (iii)- Clear Strategy Screen II, (iv)- Pact Premier Screen and (v)- JCSG plus Screen. Detailed composition of each screen is provided in appendix II.

Table 4.1: *Some of the crystallisation conditions from commercial screens that produced preliminary CDTa′ crystals.*

Screen	**Condition number (appendix II)**
Structure Screen 1 & 2	C7, D1
Clear Strategy Screen 1	E4, G1, G5,
Clear Strategy Screen 2	F3,
Pact Premier Screen	A3, A4, A5, A6, B2, B3, B4, B5, B6, C3, C4, C5, C6, C7, C8, C9, D1, D3, D4, D5, D6, D7, D8, D9, E1, E7, E10, F1, F10, G1, G6, G7, G10, H1, H6, H7, H10
JCSG plus Screen	A10, C6, D1, D12

Each screen comprises 96 conditions and hence in total 480 different conditions were set up using the sitting drop vapour diffusion (SDVD) method. 150 nl of crystallisation solution was added to an equal volume of protein and allowed to equilibrate against a reservoir of 50 µl at 16^{0}C.

Eight, out of the many conditions (Table 4.1) that produced primary hits for the crystallisation were selected for further optimisation. Crystals were reproduced in a 2 µl drop containing the protein and the reservoir solution in

a 1:1 ratio using the hanging drop vapour diffusion (HDVD) method in 24 well plates under identical incubation conditions.

Final crystals for native CDTa′ were grown in three different conditions (Table 4.2) using the HDVD method by streak seeding the drops. Reservoir solution was added to an equal volume of the protein at 4 mg/ml concentration and allowed to equilibrate against a reservoir of 500 ul for 60 to 90 minutes at 16^0C. The equilibrated drops were then streak seeded with thin plate crystals that were grown previously under identical conditions.

To grow CDTa′ crystals in complex with NAD and NADPH, ligand solution at 100 mM was added to the protein at 5 mg/ml and diluted with CDTa concentration buffer (20 mM NaCl in 50 mM Tris-HCl pH 8.0) in such a way that the final concentration of the ligand was 10 mM and that of the protein was 4 mg/ml. Crystallisation was set up using the HDVD method under the condition containing 20% PEG 1500 in 0.1 M MIB buffer pH 9.0 (Table 4.2) by streak seeding the drops as described for the crystallisation of native protein above.

Table 4.2: *The CDTa′ final crystal growth conditions.*

Crystal name	**Composition of crystallisation condition**
CDTa-8.5	0.2 M Potassium Thiocyanate, 0.1 M Tris pH 8.5, 20% PEG 2K MME
CDTa-9, CDTa+NAD and CDTa+NADPH	0.1M MIB buffer pH 9.0, 20% PEG 1500 (MIB = sodum malonate, imidazole, boric acid buffer)
CDTa-4,	0.1M MIB buffer pH 4.0, 20% PEG 1500

Concentration and Crystallisation of CDTb′ and CDTb″

CDTb′ and CDTb″ were concentrated to 7 mg/ml with the help of Millipore 10 kDa MWCO concentrators at 4000 rpm at 4^0C. As mentioned before, both of the proteins were stored in a buffer containing 0.2 % Tween-20. Initial set of crystallisation trials for each protein was set up in the presence of Tween-20 using all five commercially available crystallisation screens from Molecular Dimensions Limited in identical manner as described for CDTa′.

An additional set of crystallisation trials for CDTb′ was also set up in the absence of tween-20. Tween-20 was removed from the protein as descried

before. Table 4.3 below provides a list of crystallisation conditions that produced primary hits for all different crystallisation trials.

Table 4.3: *the primary crystallisation hits obtained for CDTb′ (with and without tween-20) and CDTb″ (with tween-20) using commercial screens.*

Screen	**Condition number (appendix II)**		
	CDTb′ (with tween-20)	**CDTb″ (with tween-20)**	**CDTb′ (without tween-20)**
Structure Screen 1 & 2	B10, B12, C3, C9, C11, D9, E2,	B10, C9, H6	G6
Clear Strategy Screen 1	C8, D9, E3, E9, F3, F9, G2, G3, G9, H3, H8, H9,	D8, E3, E9, F3, F9, G2, G8, G9, H3, H8, H9,	F9, F10, G10, H10,
Clear Strategy Screen 2	E2,	E2, G2,	--
Pact Premier Screen	C9, C10, D10,	C10, D10,	E10,
JCSG plus Screen	B10, D2, D3, D6, D7, F5, G5,	D6, F5,	C4,

The primary hits obtained for CDTb′ and CDTb″ crystallisation in the presence of tween-20 were then optimised using the HDVD and SDVD methods in 24 well plates. 1 µl of the protein was mixed with an equal volume of reservoir solution and allowed to equilibrate against a 500 µl volume of reservoir solution at 16^0C. Optimisation of CDTb′ crystallisation primary hits in the absence of tween-20 were also performed in 24 well plates using the HDVD as well as SDVD method.

The primary hit that produced the best looking crystals (Pact premier screen – E10) for CDTb′ crystallisation in the absence of Tween-20 was also optimised further using the additive screen from Hampton Research Limited in a 96 well plate with the help of crystallisation robot. Table 4.4 lists all crystal producing conditions from the additive screen (for detailed composition please see appendix II). The obtained hits were then optimised manually using HDVD method.

Table 4.4: *CDTb' crystallisation hits using the additive screen with Pact Premier E10 condition as the basic condition.*

CDTb′ (without Tween-20 – Additive Screen	
Basic condition	**Additive Screen condition number**
20% PEG 3350	A8, B7, B8, C1, E11, F2, G1

RESULTS AND DISCUSSION

Chymotrypsin Mediated Activation of CDTb

Available literature and experimental evidence suggest that transport components of Clostridial binary toxins have to be activated by tripsin or chymotrypsin to become fully functional (Perelle *et al.*, 1997; Fernie *et al.*, 1984). Activation of the transport component of *C. botulinum* C2 toxin (C2II) by trypsin has been reported (Ohishi, 1987). At least two different studies (Blocker *et al.*, 2001; Gluke *et al.*, 2001) describe activation of the transport component of *C. perfringens* iota toxin (Ib) by chymotrypsin under the conditions of temperature, pH and incubation time similar to that reported for C2II.

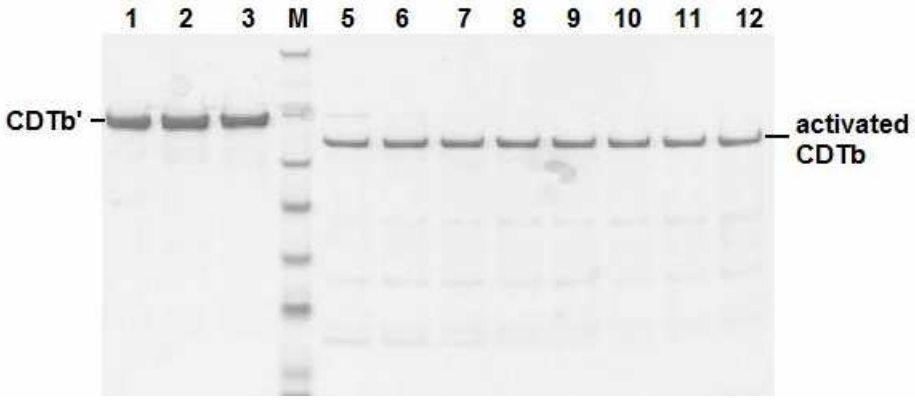

Figure 4.1: *The chymotrypsin mediated activation of CDTb' on Bis-tris SDS-PAGE. Lane 1 and 2- non-heated CDTb', lane 3- 5 minute heated CDTb', lane 5, 6, 7, 8, 9, 10, 11 and 12- 10, 15, 20, 25, 30, 40, 50 and 60 minute chymotrypsin treated, heated samples.*

Figure 4.1 shows the control protein, CDTb′ (lanes 1, 2 and 3) and the chymptrypsin activated CDTb fragment (lanes 5 to 12) on a Bis-Tris-SDS-PAGE. A gradual decrease in the amount of precursor protein (CDTb′) can be

seen which completely disappears between 20 and 30 minutes time under the tested conditions. The reaction was continued for 60 minutes and no further cleavage or degradation of the activated fragment was observed in spite of the presence of chymotrypsin in the mixture. Chymotrypsin is a non specific protease and is known to cleave its substrate at random sites. However, successful production of the protein fragment of the correct size is strong evidence that activation of the transport component of *C. difficile* binary toxin by chymotrypsin is a highly specific process. These results also indicate that the expressed and purified protein (CDTb′) is correctly folded.

Cell Cytotoxicity Effects of Complete CDT

The first set of cytotoxicity experiments was conducted with CDTb′ and CDTb″ protein samples that were stored in a buffer containing 0.2% Tween-20. Previous data produced in our laboratory showed that Tween-20 has a lethal effect on growing Vero cells.

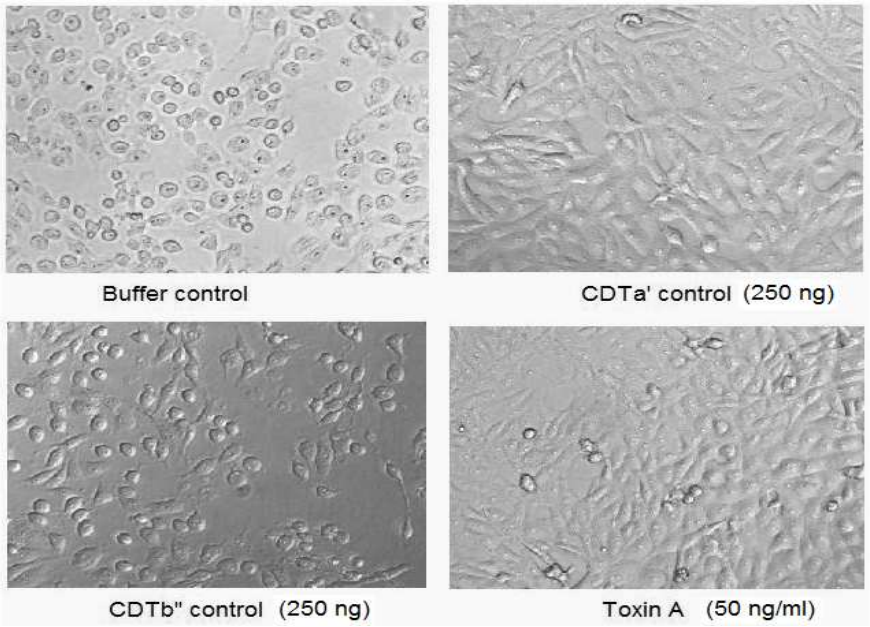

Figure 4.2: *The effect of Tween-20 on growing Vero cells in 4 hours time.*

Figure 4.2 demonstrates the effect of Tween-20 on growing Vero cells at 4 hours post incubation time. CDTb is the transport component of the binary toxin and possesses no catalytic activity. The observed cell death in CDTb′ and CDTb″ alone controls was due to the presence of Tween-20 in the used protein samples (Figure 4.2). CDTa is the catalytic component of the binary toxin. It

requires CDTb as its translocation partner to access entry into the target cell. CDTa′ alone controls did not show any cell death (Figure 4.2).

Due to the lethal effect of Tween-20 on growing cells, it was necessary to remove it from the stored protein samples. Figures 4.3, 4.4 and 4.5, below summarise the results of the second set of cytotoxicity experiments showing the effect of toxins on the cells at 24 hours time point. Observations were not made post 24 hours incubation as the cells were maintained in serum depleted medium.

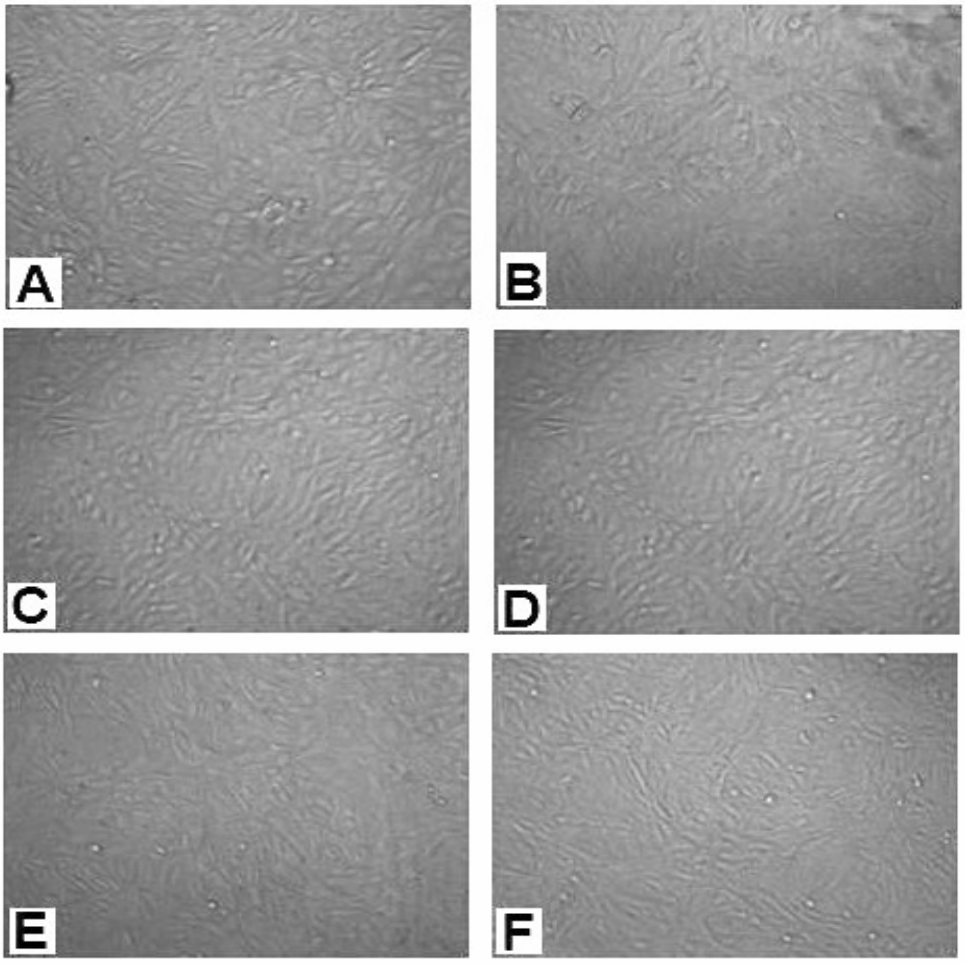

Figure 4.3: *Controls for the cell cytotoxicity test. (A) – Blank (buffer control), (B) – CDTa' alone control (250 ng), (C) – CDTb' alone control (250 ng), (D) – CDTb'' alone control (250 ng), (E) – CDTb′# alone control (250 ng), (F) – CDTb''# alone control (250 ng),in a total volume of 200 µL.*

No cell death was observed in the buffer control (Figure 4.3 A) and in the presence of individual toxin components CDTa, CDTb′, CDTb′′, CDTb′# and CDTb′′# (Figures 4.3 B, C, D, E and F). These observations proved that the individual components of CDT are not cytotoxic. Healthy cells in CDTb′# and

CDTb′′# controls indicate that the chymotrypsin was inactivated completely by trypsin inhibitor. Toxin A and Toxin B are the best characterized toxins from *C. difficile* and were used as positive controls for Vero cell cytotoxicity in this study (Figures 4.3 G and H).

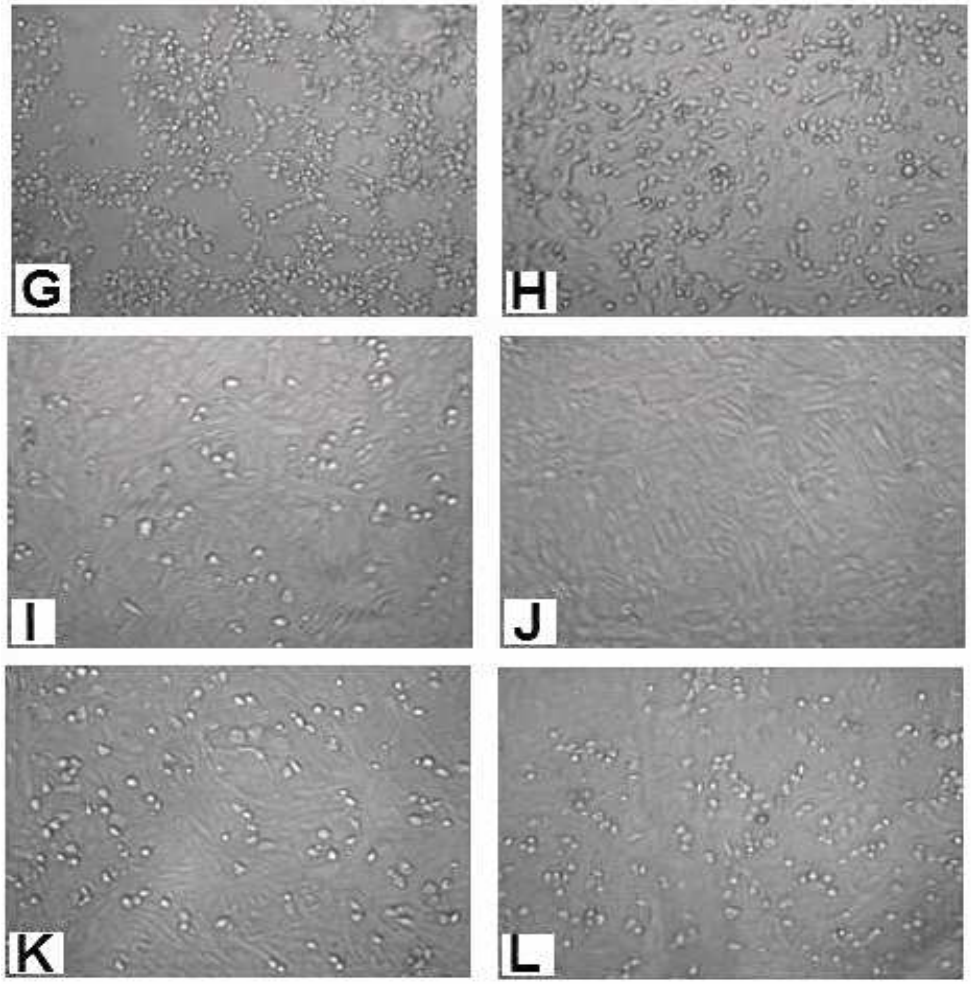

Figure 4.4: *The effect of binary toxin on growing Vero cells. (G) – Toxin A control (5ng), (H) – Toxin B (0.5 ng), (I) – CDTa′ + CDTb′# (10 ng+50 ng), (J) – CDTa′ + CDTb′ (250ng+250 ng), (K) – CDTa′ + CDTb′# (50ng+250 ng), (L) – CDTa′ + CDTb′# (250ng+250 ng), in a total volume of 200 µL.*

No cell death was observed in any of CDTa′+CDTb′ mixture test cases (Figure 4.4 J). CDTb′ is the recombinant inactive precursor fragment of CDTb. It requires chymotrypsin mediated activation to become fully functional. Vero cells are kidney epithelial cells and are not known to produce chymotrypsin which is necessary to activate CDTb′. In all previously reported studies on different Clostridial binary toxins (Gluke *et al.*, 2001; Blocker *et al.*, 2001; Ohishi *et al.*, 1980; Kaiser *et al.*, 2006), the corresponding transport

components activated by trypsin or chymotrypsin have been used. Cell death was observed for CDTa′ in the presence of CDTb′# test cases (Figures 4.5 P, Q and R). Cell death was recorded for all CDTa′+CDTb′′ cases but to a considerably lower extent (Figures 4M and 4N) when compared with its chymotrypsin activated product, CDTb′′# (Figures 4.5 M and O).

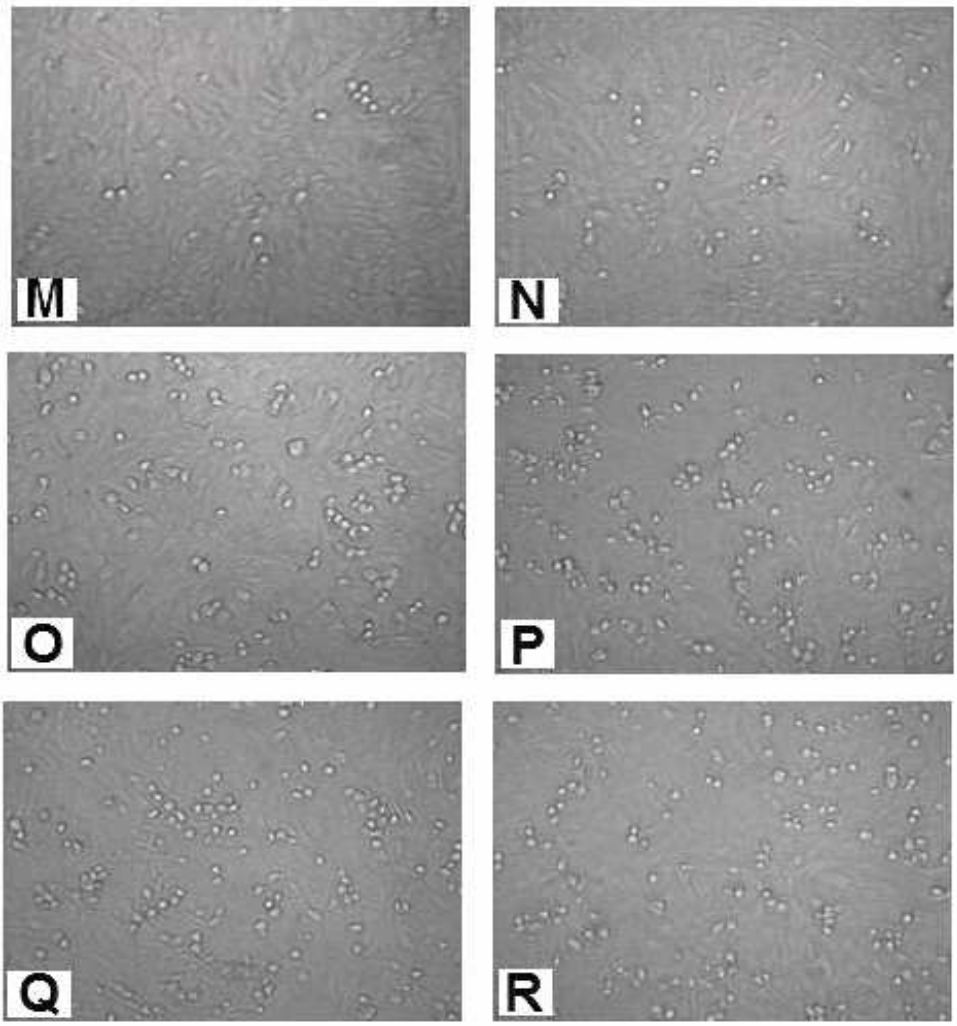

Figure 4.5: *The effect of various concentrations of binary toxin components on growing Vero cells. (M) – CDTa′ + CDTb′′(50ng+250 ng), (N) – CDTa′ + CDTb′′(250ng+250 ng), (O) – CDTa′ + CDTb′′# (50ng+250 ng), (P) – CDTa′ + CDTb′# (50 ng+50 ng), (Q) – CDTa′ + CDTb′# (50ng+150 ng), (R) – CDTa′ + CDTb′# (50ng+250 ng), in a total volume of 200 µL.*

A five-fold variation in the concentration of CDTa and CDTb was screened in different combinations. However, it was observed that variation in CDTa′ amount from 50 to 250 ng, keeping CDTb′# concentration constant, did not increase cell death significantly (Figures 4K and 4L). No significant difference in cell death was observed when CDTb′# concentration was varied

keeping that of CDTa′ fixed (Figures 4P, 4Q and 4R). However, our experimental results are insufficient to show conclusively whether the amount of activated CDTb or the concentration of CDTa present is the rate determining step in binary toxin mediated cell death.

Previously, the cytotoxicity effect of CDTa has been studied by Gluke and co-workers using the transport component of iota toxin (Ib) to mediate the cell entry of CDTa because CDTb could not be well expressed (Gluke *et al.*, 2001). Hence, their study could not provide a definitive answer as to whether full length CDT (CDTa+CDTb) is cytotoxic. Our study presented here, provides a clear picture as we have used both components of CDT and have conclusively shown that the complete *C. difficile* binary toxin is capable of killing cells at as low as 50 ng/ml of CDTa′ and 250 ng/ml of CDTb′# (Figure 4.4 I) (at the tested amount of CDTa (Gluke *et al.*, 2001) and iota toxin (Blocker *et al.*, 2001) reported in previous studies).

Perelle and co-workers have demonstrated the cell cytotoxicity of complete CDT on Vero cells. In their experiments, Vero cells were incubated with the binary toxin producing C. difficile bacterial cell culture supernatant (Perelle *et al.*, 1997). The culture supernatant contained Toxin A and Toxin B in addition to CDT. Authors suggest that incubation of the culture supernatant at -20^{0}C deactivated both *C. difficile main toxins*. However, the purified Toxin A and Toxin B that have been used in our study as positive controls, were always stored at -20^{0}C and no deactivation of either toxin was observed as both toxins were still capable of killing cells (Figures 4.4 G and H). In our study, both components of CDT were expressed recombinantly and purified. Hence, our results leave no ambiguity about cell death mediated by the complete CDT. Our report is the first report on CDT mediated cell cytotoxicity in isolation.

Formation of CDTb Oligomer in Solution

Binary toxins from various Clostridial and bacillus species follow a similar mechanism of cell entry. Transport components of these toxins form a heptameric pore like structure upon activation by trypsin/chymotrypsin/furin. Transport components from two different Clostridial actin modifying binary toxins have been studied in this regard. However, the two studied toxins belong to two different classes of Clostridial

actin-ADPRTs (Table 2.1) and the physiochemical conditions of heptamer formation vary among them. C2II from *C. botulinum* has been reported to form an SDS resistant heptamer whereas the Ib heptamer (from *C. perfringens*) was found susceptible to SDS (Blocker *et al.*, 2001).

The protein (CDTb′) was activated by chymotrypsin in a buffer containing 50 mM Tris-HCl pH 7.5 and 300 mM NaCl, as described previously. 8 different buffer systems of different pH were screened for the oligomerisation of CDTb. Figure 4.6 below shows all samples on a Bis-Tris SDS-PAGE under SDS conditions.

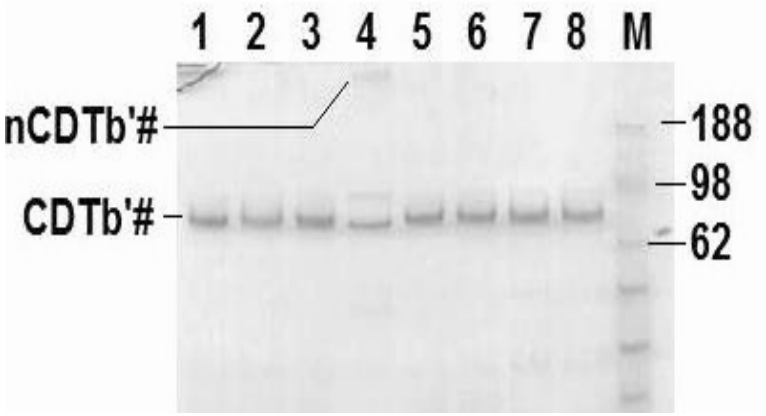

Figure 4.6: *The formation of the CDTb oligomer in solution**.** Lane 1 – MIB pH 4.0, 2 – MMT pH 4.0, 3 – SPG pH 4.0, 4 – Na-acetate pH 4.0, 5 – Bis-Tris pH 5.5, 6 – Bis-tris pH 6.5, 7- Tris-HCl pH 7.5, 8 – Tris-HCl pH 8.5*

A faint protein band was observed well above the 188 kDa marker protein band for activated CDTb oligomer. Intensity of this protein band varied in different lanes. However, it was clearly visible in Na-Acetate buffer, pH 4.0 test condition (Figure 4.6, lane 4). This observation agrees with results reported by Blocker and co-workers (Blocker *et al.*, 2001) for the oligomerisation of iota toxin transport component (Ib). The formation of the CDTb oligomer could not be tested under non-SDS conditions.

Concentration and Crystallisation of CDTa′

The protein was concentrated to 10 mg/ml without any difficulty. Figure 4.7 below, shows the concentrated CDTa′ protein on a 10% Tris-glycine SDS-PAGE.

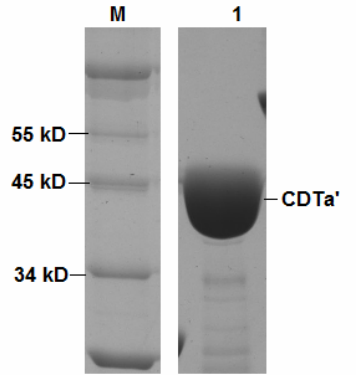

Figure 4.7: *The concentrated CDTa' protein on Tris-glycine SDS-PAGE (lane 1)*

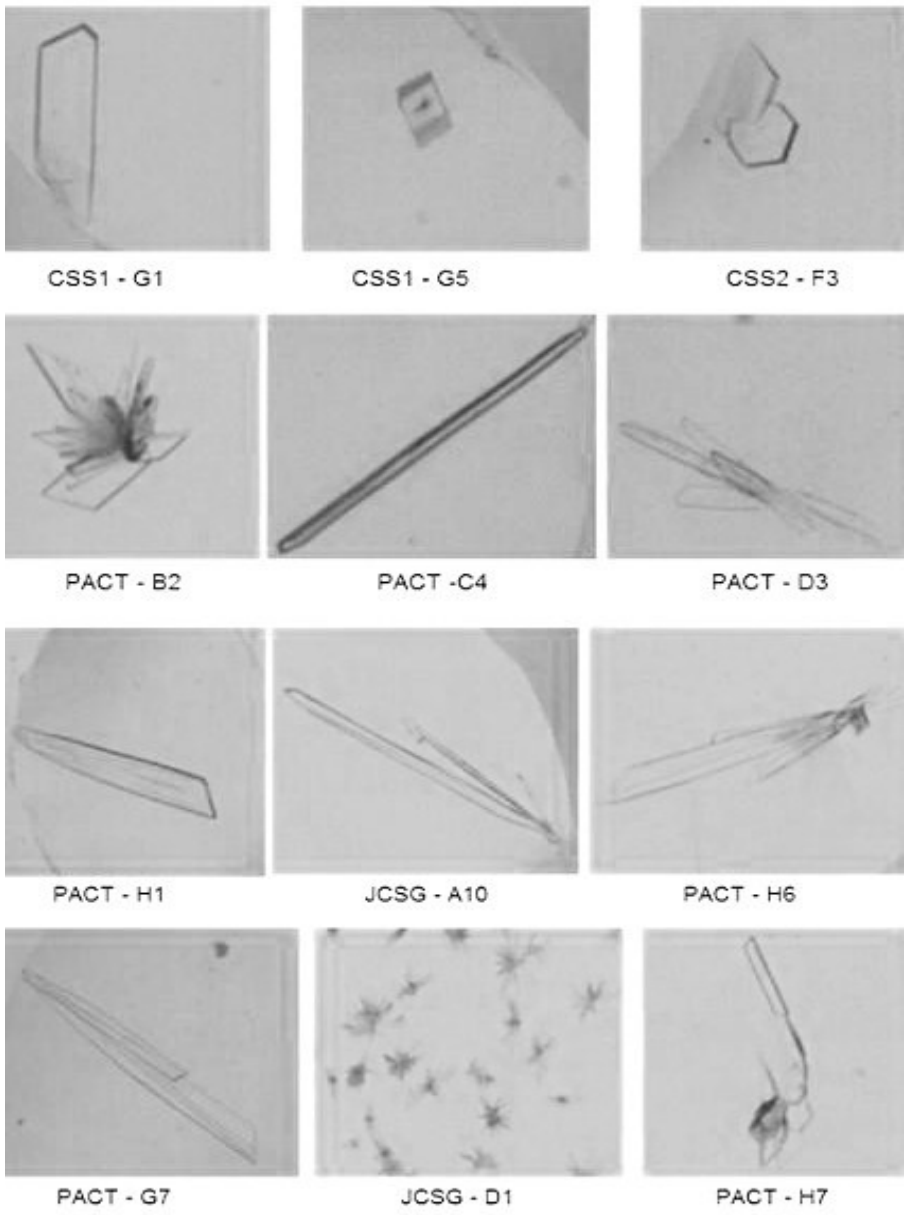

Figure 4.8: *Some of the primary CDTa' crystals. CSS – Clear strategy Screen, PPS- Pact Premier Screen. JCSG – JCSG plus screen.*

At least 20 out of 480 screened crystallisation conditions resulted in thin plate like crystals of similar morphology within 24 hours. This number went up to over 50 within a week's time (Figure 4.8).

Eight conditions were then chosen and these thin plate crystals were reproduced by the hanging drop vapour diffusion method. Crystals from two different conditions were chosen to test in the X-ray beam at Station I03 of Diamond Light Source, United Kingdom. These crystals were protein crystals but diffracted poorly to 5Å resolution.

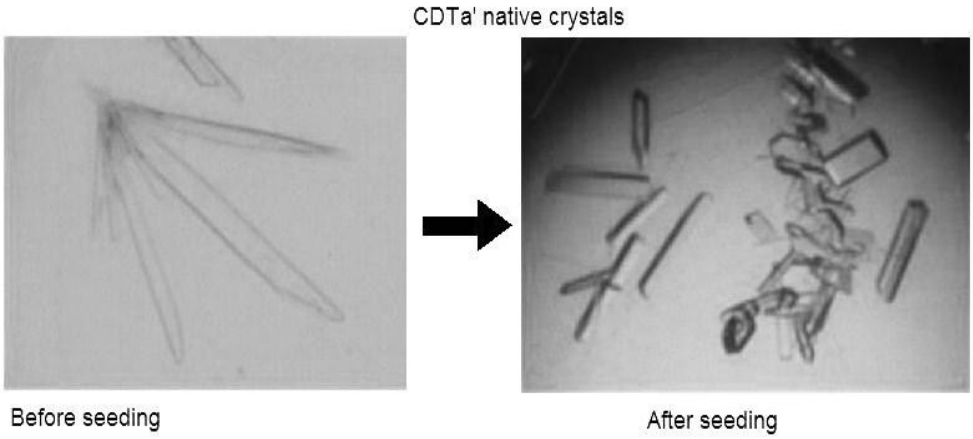

Figure 4.9: *The effect of seeding on CDTa' crystal morphology.*

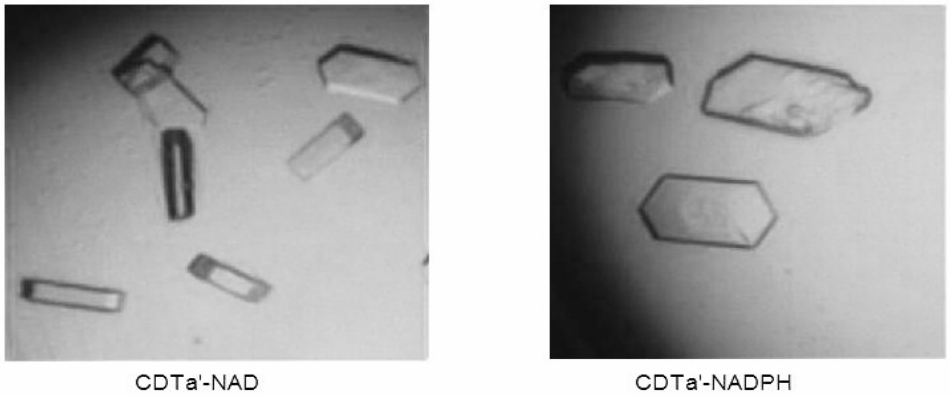

Figure 4.10: *The CDTa' crystals in complex with NAD (left) and NADPH (right) grown by streak seeding.*

Both of the tested conditions were selected for final crystal optimisation by employing the streak seeding technique. Seeding improved crystal morphology significantly and consequently the diffraction quality of crystals was improved. Figure 4.9 illustrates the effect of streak seeding on crystal

morphology. Co-crystallisations of the protein with its donor substrates were also set up. Seeding resulted in diffraction quality crystals of good morphology in co-crystallisations also (Figure 4.10). X-ray diffraction data collection and structure determination of CDTa′ are discussed in the chapter 5.

Concentration and Crystallisation of CDTb′ and CDTb″

CDTb′ protein was concentrated successfully up to 7 mg/ml. A small amount of precipitate was observed in CDTb″ samples during the concentration process. The concentrated protein samples were centrifuged at 10,000 rpm for 10 minutes at 4^{0}C prior to setting up crystallisation in order to remove the precipitated protein.

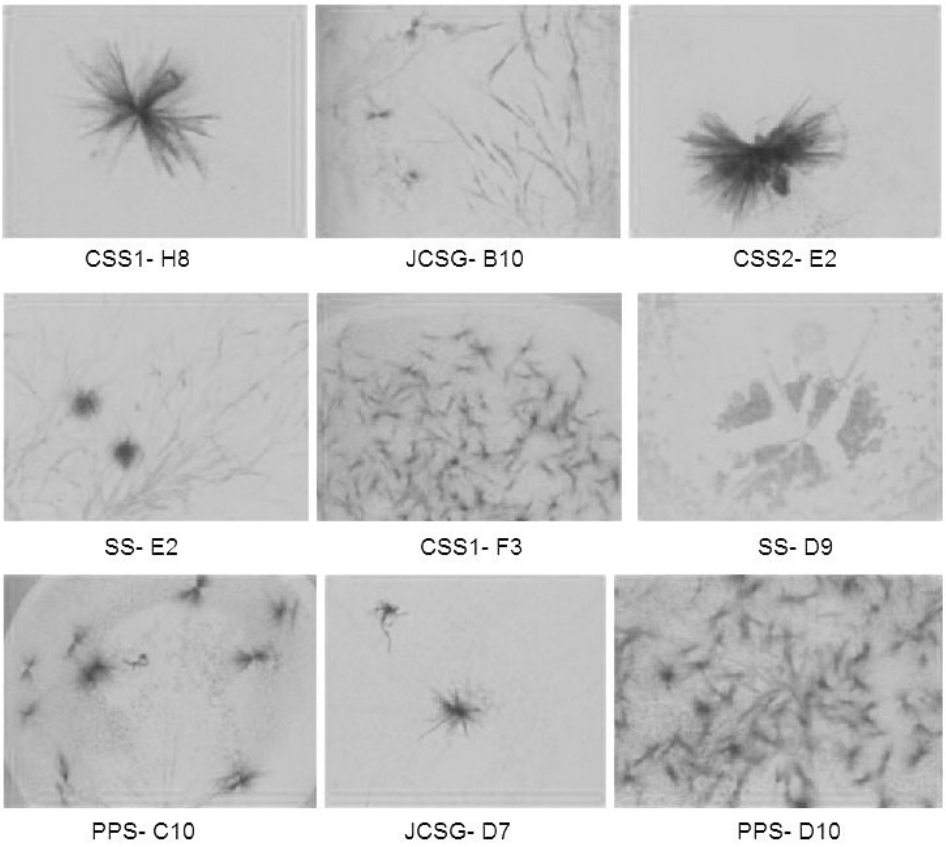

Figure 4.11: *Some of the primary crystallisation hits obtained for CDTb′ in the presence of 0.2% tween-20. SS – Structure screens 1 and 2, CSS – Clear strategy Screen, PPS- Pact Premier Screen. JCSG – JCSG plus screen.*

In the presence of Tween-20 several crystallisation conditions (Table 4.3) produced very thin needle crystals which were not suitable for data collection. These crystals of similar morphology were grown for CDTb′ as well

as for CDTb″ (Figures 4.11 and 4.12). Optimisation of these conditions in 24 well plates also did not result in any significant improvement.

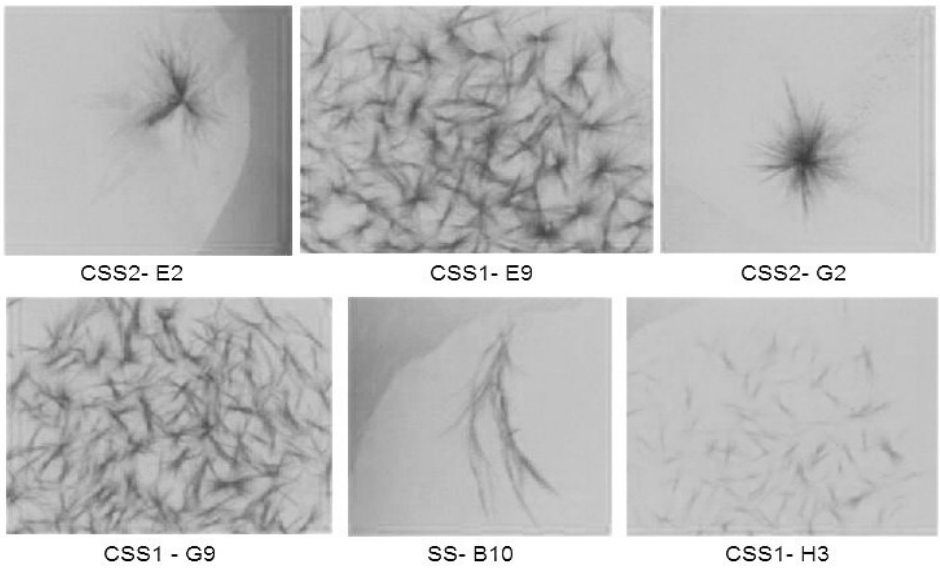

Figure 4.12: *Some of the primary crystallisation hits obtained for CDTb″ in the presence of 0.2% tween-20. SS – Structure screens 1 and 2, CSS – Clear strategy Screen, PPS- Pact Premier Screen. JCSG – JCSG plus screen.*

However, in the absence of Tween-20, much better diamond shaped crystals for CDTb′ were grown in few conditions (Table 4.3 and Figure 4.13). These crystals were not big in size but they appeared to be better than the crystals grown in the presence of Tween-20 (Figure 4.11).

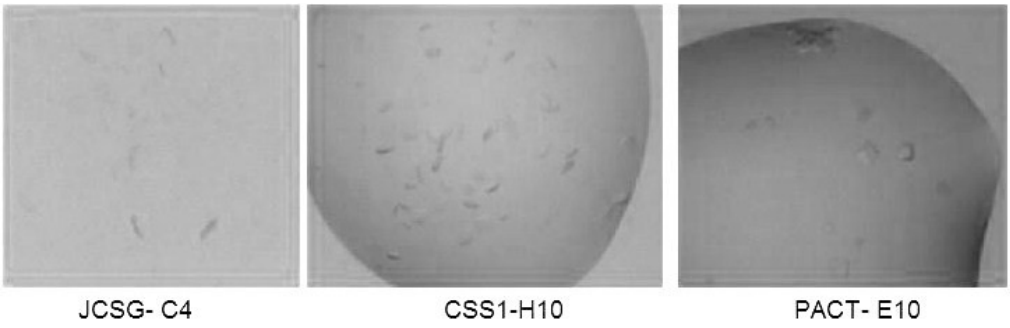

Figure 4.13: *Some of the primary crystallisation hits obtained for CDTb′ in the absence of tween-20. SS – Structure screens 1 and 2, CSS – Clear strategy Screen, PPS- Pact Premier Screen. JCSG – JCSG plus screen.*

In the absence of Tween-20, the Pact premier screen condition number E10 produced the best looking crystals for CDTb′ (Figure 4.13, right most panel). These crystals could be reproduced in 24 well plates by varying the

concentration of salt and precipitant in the reservoir solution as well as the protein to reservoir solution ratio in the drop (Table 4.5 and Figure 4.14). Optimisation of these conditions so far has not improved the size of crystals significantly. None of these crystals showed diffraction spots at Diamond Light Source.

Table 4.5: *Variation of condition for the optimisation of CDTb' crystals in a 24 well plate in the absence of Tween-20.*

Basic condition – 0.2 M K thiocyanate, 20% PEG 3350	
Parameter	**Variation**
Protein concentration (in drop)	3 mg/ml – 4 mg/ml
Precipitant (PEG) concentration)	16% – 20%
Salt concentration	0.0 M to 0.2 M
Protein : reservoir solution in drop	1:1, 1:2, 2:1

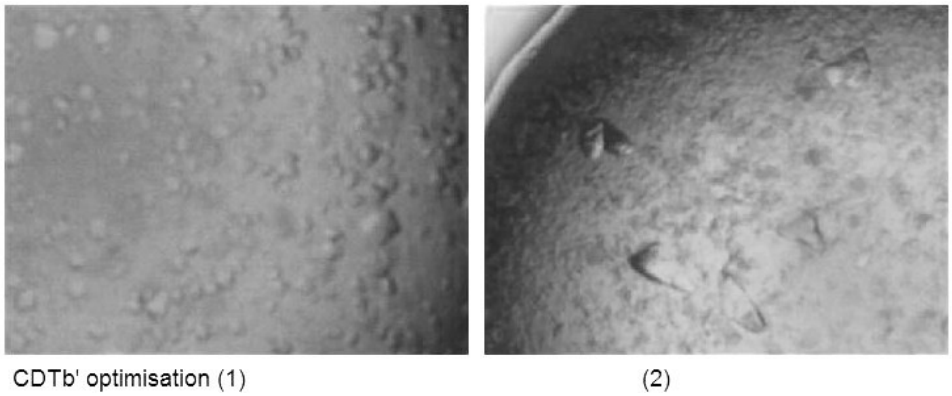

Figure 4.14: *CDTb' crystals grown in 24 well plates in hanging drop vapour diffusion method in the absence of Tween-20. (1) – 16% PEG3350, protein (3 mg/ml) : reservoir solution =1:2; (2) - 20% PEG3350, protein (3 mg/ml) : reservoir solution =1:2*

These crystals (Figure 4.14), however, could be reproduced in a wide range of concentrations of salt and precipitating agent (Table 4.5). Further optimisation of these crystals using the additive screen from Hampton research was carried out. Figure 4.15 displays some of the crystals grown in the crystallisation conditions from the additive screen. Unfortunately, the size of crystals still could not be improved. Further optimisation of different conditions in order to grow large diffraction quality crystals is underway.

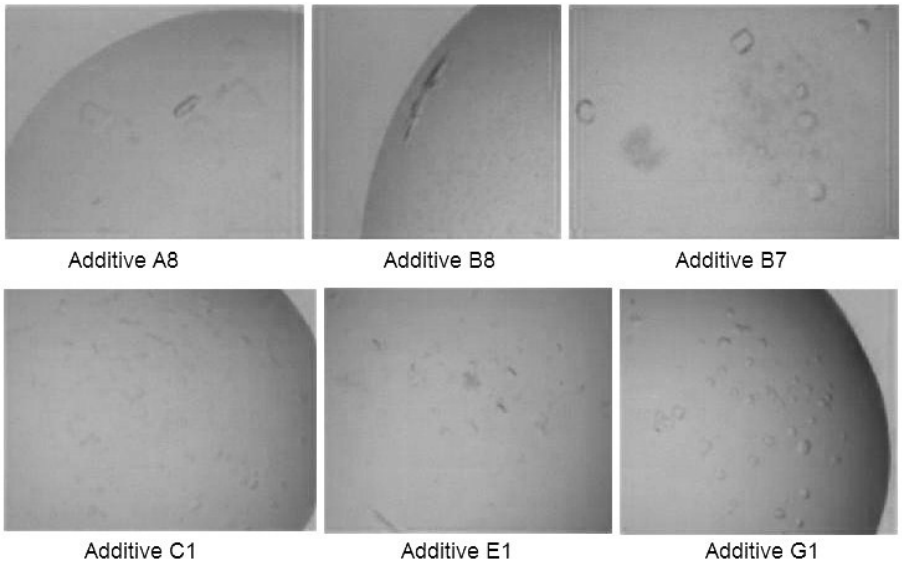

Figure 4.15: *The CDTb′ crystals grown in the absence of tween-20 using additive screen with basic condition – 20% PEG 3350*

SUMMARY

Chymotrypsin mediated activation of precursor CDTb (i. e. CDTb′) resulted in a fully functional activated protein fragment (CDTb′#). Results of the cell cytotoxicity experiments proved that the expressed and purified proteins (CDTa′, CDTb′ and CDTb″) were active and correctly folded. In addition to that, the cell cytotoxicity tests indicated that complete CDT has the potential to kill cells in isolation and hence should have a definite role in *C. difficile* infection.

The preliminary experimental results showed the formation of oligomeric CDTb complex in solution under acidic conditions. However, the intensity of the protein band was very low on an SDS-PAGE and these conditions need to be optimised further.

Crystallisation trials for CDTa′ resulted in diffraction quality crystals of CDTa in its native form as well as in complex with the ADP-ribose donor substrate i. e. NAD and NADPH (discussed in chapter 5 in details). Preliminary success was achieved for CDTb′ and CDTb″ crystallisations. Small but good morphology crystals for CDTb′ were grown in the absence of Tweeen-20. Optimisation of these preliminary crystallisation hits is underway.

CHAPTER – V

CRYSTAL STRUCTURE OF ENZYMATIC COMPONENT OF *C. DIFFICILE* BINARY TOXIN: CDTa

Structural Analysis of Known ADPRTs

ADP ribosylating toxins (ADPRTs) are a large superfamily that has been divided into four different classes based on their substrate specificity (Table 2.2). All ADPRTs share a common active site fold (Han and Tainer, 2002; Domenighini and Rappuoli, 1996). Three dimensional structures for representative members of each of the 4 classes of ADPRTs have been determined. These include - Diphtheria toxin (1TOX) from *Corynebacterium diphtheriae* (Bell and Eisenberg, 1996), Pseudomonas exotoxin A (1AER) from *Pseudomonas aeruginosa* (Li *et al.*, 1996), Pertussis toxin (1PRT) from *Bordetella pertussis* (Stein *et al.*, 1994), Cholera toxin (1XTC) from *Vibrio cholerae* (Zhang *et al.*, 1995), *Escherichia coli* heat labile enterotoxin (1LTS) (Sixma *et al.*, 1993), *Clostridium perfringens* Iota toxin (1GIQ) (Tsuge *et al.*, 2003), *Clostridium botulinum* C2 toxin (2J3V) (Schleberger *et al.*, 2006), Vegetative insecticidal protein (1QS1) from *Bacillus cereus* (Han *et al.*, 1999) and the C3-like toxins, C3Bot (1G24) from *Clostridium botulinum* (Han *et al.*, 2001) and C3stau (1OJZ) from *Staphylococcus aureus* (Evans *et al.*, 2003). Based on the ADP-ribose donor substrate binding pattern, ADPRTs have been classified into two classes:

1- **DT type,** which is based on active site architecture and NAD binding features that are present in diphtheria toxin (Bell and Eisenberg, 1996). And,

2- **CT type,** where the NAD binding features are similar to that is observed in cholera toxin (Zhang *et al.*, 1995). The CT type toxins include C3Bot, VIP2, pertussis toxin, Iota toxin and CDT from *C. difficile*.

The ADP-ribose donor (i. e. NAD) binds to the catalytic cleft in a high energy, closed conformation in all ADPRTs irrespective of their class and interacting residues. The NAD binding cleft in all ADPRTs comprises of a similar mixed α/β core structure. The cleft is positioned between a β-stranded framework and either an α-helix (examples - C3Bot, C3stau, VIP2 and Iota) or a variable length active site loop (such as in pertussis, cholera, diphtheria and exotoxin A). A sequence alignment of different ADPRTs reveals the presence of several conserved residues that form catalytically important motifs in the 3-dimensional structures (Figure 5.1).

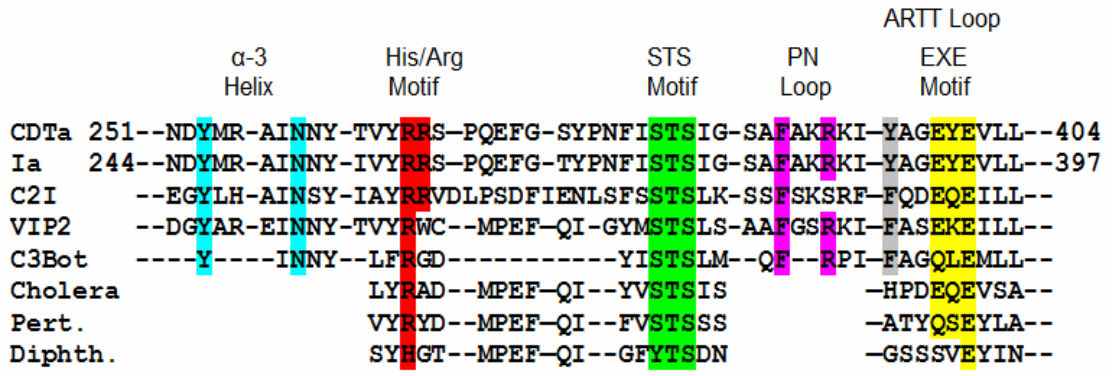

Figure 5.1: *The sequence alignment of different classes of ADPRTs showing conserved catalytically important motifs (adopted from Holbourn et al., 2006). Pert. = Pertussis toxin, Diphth. = diphtheria toxin.*

An aromatic residue followed by a conserved Arg/His has been found in all ADPRTs till date (Domenigini and Rappuoli, 1996). The DT class is characterised by a Tyr-His whereas the motif present in the CT class is Val/Leu-X-Arg (Figure 5.1, where X is an aromatic residue). This particular motif is designated as the Arg/His motif and is found to be involved in the ligand binding but not in catalysis (Holbourn *et al.*, 2006). However, mutational studies suggest that the loss of this conserved residue results in almost complete loss of transferase activity in several ADPRTs (Lobet *et al.*, 1991; Cieplak *et al.*, 1988; Burnette *et al.*, 1988; Burnette *et al.*, 1991; Tsuge *et al.*, 2003; Wilde *et al.*, 2002).

The STS motif of ADPRTs is supposed to act as an anchor to hold the NAD binding site together. In C3Bot, the first Ser residues of the motif forms hydrogen bond with the catalytic Glu beneath the cleft to hold it in the correct position to mediate NAD cleavage (Holbourn *et al.*, 2006). However, mutational studies on iota toxin and dipththeria toxin (Bell and Eisenberg, 1996; Nagahama *et al.*, 2000) suggest that while the STS motif stabilises the catalytic site, it is not essential in all ADPRTs. In diphtheria toxin the STS motif is replaced with the YTS motif where the Tyr residue is shown to be crucial for activity of the toxin (Carroll and Collier, 1988; Carroll and Collier, 1984; Carroll and Collier, 1987).

A loop of varying length known as the ARTT (ADP-ribosylating turn turn) loop is common in all CT type ADPRTs. The ARTT loop contains key catalytic residues in the form of an EXE or QXE motif which have been suggested to play a decisive role in ligand binding as well as in the transfer of

ADP-ribose to the substrate (Han *et al.*, 2001; Han and Tainer, 2002; Domenighini and Rappuoli, 1996). An aromatic residue at the centre of the ARTT loop in class 3 and class 4 ADPRTs is another conserved residue. In C3Bot, this residue (Phe) has been shown to be essential for substrate binding (Wilde *et al.*, 2002). However, an aromatic residue (Tyr) at an equivalent position in actin-ADPRTs such as iota toxin has not been assigned any such function to date (Tsuge *et al.*, 2003; Tsuge *et al.*, 2008).

The PN loop forms an essential part of the ligand binding machinery of ADPRTs. In C3Bot, the PN loop has been reported to undergo a large movement upon NAD binding. A similar movement of the loop, however, has not been observed in iota toxin. The PN loop of two different classes of ADPRTs (class 3 and class 4) has been found to contribute to the NAD binding with an Arg residue that interacts with NAD directly. An aromatic residue (Phe) in both classes has been suggested to stabilise ligand binding by stacking interactions (Tsuge *et al.*, 2003; Menetrey *et al.*, 2002; Holbourn *et al.*, 2006).

A 15 residue active site loop that is present in class 1 and class 2 of ADPRTs has been found missing in class 3 and class 4 (Bell and Eisenberg, 1996, Sixma *et al.*, 1991). This loop has been suggested to be involved in substrate recognition (O'Neal *et al.*, 2005). In class 3 and class 4 of ADPRTs, this loop has been replaced by an α-helix (named as α-3 helix) which packs itself tightly against the NAD binding cleft (Tsuge *et al.*, 2003; Han *et al.*, 1999; Han *et al.*, 2001; Evans *et al.*, 2003). This helix contains 3 conserved catalytically important residues amongst all class 3 and class 4 ADPRTs. This part of the NAD binding cleft is thought to be involved in holding the ADP-ribose component of NAD after cleavage of the N-glycosidic bond until it is transferred to the acceptor molecule (Holbourn *et al.*, 2006).

C. difficile binary toxin (CDT) along with *C. perfringens* iota toxin and *C. botulinum* C2 toxin belongs to the class 4 of ADPRTs. In this chapter, high resolution crystal structures of the enzymatic component of CDT in its native form under three different pH conditions as well as in complex with its donor substrates, i. e. NAD and NADPH are presented. The mode of donor substrate recognition by CDTa is compared with that of Iota A with a possible explanation of the ARTT loop stability upon ligand binding in CDTa. Based on the structural data presented, it seems that the ADP-ribosylation reaction by CDTa prefers to proceed via an S_N1 mechanism of catalysis rather than S_N2.

MATERIALS AND METHODS

Data Collection and Data Processing

A single crystal was mounted in a nylon loop and flash frozen at 100K temperature in a stream of nitrogen gas. Diffraction data sets for native CDTa crystals (Figures 4.9 and 4.10) grown in three different crystallisation conditions (Table 4.2) were collected at Stations I02 and I04 of Diamond Light Source, Didcot, UK. Single crystal diffraction data sets for CDTa in complex with its donor substrates, NAD and NADPH were also collected in similar fashion at Station I02. Each of the stations was equipped with a Quantum-4 CCD detector (Area Detector Systems Corp.). X-ray wavelengths used for data collection are provided in Table 5.1. Twenty percent glycerol was used as cryoprotectant for CDTa-NAD and CDTa-NADPH complex crystals. All native crystals were mounted without any cryoprotectant. Two hundred images for each crystal were collected using the rotation method of data collection with an oscillation range $\Delta\Phi = 1^0$. Raw data images were indexed and scaled with the HKL2000 package (Otwinowski *et al.*, 1997) and the scaled intensities were truncated to amplitudes using TRUNCATE (French *et al.*, 1978) from the CCP4 suite (CCP4, 1994). Detailed data collection and data processing statistics are given in table 5.1.

Structure Solution and Refinement

To solve phases for the CDTa-8.5 structure, the search model was derived from the coordinates of enzymatic component of iota toxin (Tsuge *et al.*, 2003) (PDB entry-1GIQ and 1GIR). Initial phases for structure solution were obtained using the molecular replacement routines of the MOLREP program (Vagin *et al.*, 1997). Data in the range of 50.0 to 3.0 Å was used for the molecular replacement step. The resultant model was refined using REFMAC5 (Murshudov *et al.*, 1997) of the CCP4 suite. Five percent of reflections were separated as R_{free} set and used for cross validation (Brunger, 1992). After an initial round of rigid-body refinement, iterative rounds of restrained refinement with electron density map calculations and manual adjustments of the model using COOT (Emsley and Cowtan, 2004) were carried out. On the basis of 2*F*o–*F*c electron density, side-chain atoms were omitted at some positions. Water molecules were added at positions where *F*o–

*F*c electron density peaks exceeded 3σ and potential hydrogen bonds could be made.

A similar approach was adopted to solve other CDTa structures. The refined CDTa-8.5 structure was used as phasing model to obtain initial phases for all other structures. The atomic coordinates for NAD and NADPH were obtained from studies of Tsuge and co-workers (Tsuge *et al.,* 2003).

Model validation was conducted with the aid of programs PROCHECK (Laskowski *et al.*, 1993) and MOLPROBITY (Davis *et al.*, 2007). Estimations of main chain Φ-Ψ torsion angles were obtained from Ramachandran plot (Ramachandran *et al.,* 1963). Figures were drawn with PyMOL (DeLano Scientific, San Carlos, CA, USA). Validated structure coordinates for all five structures were deposited with the Protein Data Bank (PDB) under entries – 2WN4, 2WN5, 2WN6, 2WN7 and 2WN8 (Sundriyal *et al.,* 2009).

RESULTS AND DISCUSSION

Data Collection and Data Processing

CDTa without its signal peptide (CDTa′) was crystallised in its native form in three different crystallisation conditions (Table 4.2). Two of the conditions were virtually identical except for the pH (9.0 and 4.0 respectively) (Table 4.2). Crystals of CDTa′ in complex with NAD and NADPH were grown in high pH condition at pH 9.0 (Table 4.2).

Figure 5.2 displays a typical X-ray diffraction image for native CDTa. Indexing of data sets suggested that all of the crystals belong to monoclinic system. All data sets were indexed and scaled in both, P2 and $P2_1$ space groups. However, a clear molecular replacement solution was obtained in $P2_1$ space group (Table 5.2). Hence, $P2_1$ was the correct space group with slight variations in cell parameters for different crystals (Table 5.1). Calculation of Matthew's coefficient (Matthews, 1968) indicated that all of the crystals contained one protein molecule per asymmetric unit with about 40-50 % solvent content in different crystals (Table 5.1).

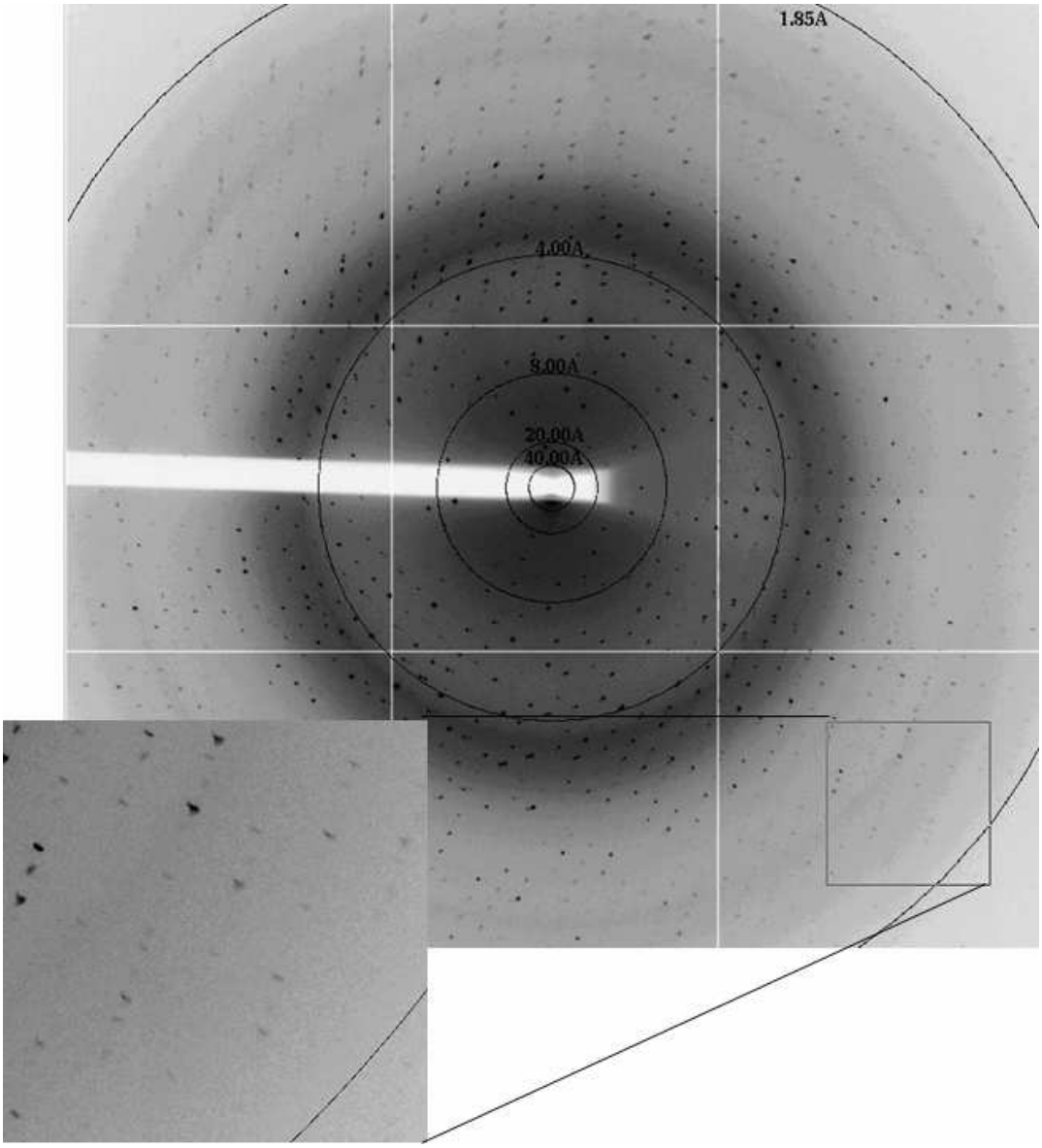

Figure 5.2: *A typical X-ray diffraction image for native CDTa crystal (CDTa-8.5).*

Structure Solution and Refinement

The CDTa-8.5 structure was determined at 1.85 Å resolution whereas CDTa-4.0, CDTa-9.0, CDTa-NAD and CDTa-NADPH structures were determined at 2.0 Å, 1.9 Å, 2.25 Å and 1.95 Å respectively. The molecular replacement process for CDTa-8.5 structure determination yielded one unique solution (Table 5.2.) in $P2_1$ space group. Molecular packing in the unit cell for this solution was free of any unfavourable intermolecular contacts confirming it to be the correct solution. Figure 5.3 shows the packing of CDTa molecules in the unit cell.

The resulting model was subjected to 20 cycles of rigid body refinement followed by 10 cycles of restrained refinement. A marked reduction in R_{cryst}

(from 49.41 to 31.78 %) and R_{free} (from 47.93 to 36.41 %) values and a significant increase in the figure of merit (from 39.8 to 70.1 %) were observed indicating the success of refinement steps (Table 5.3).

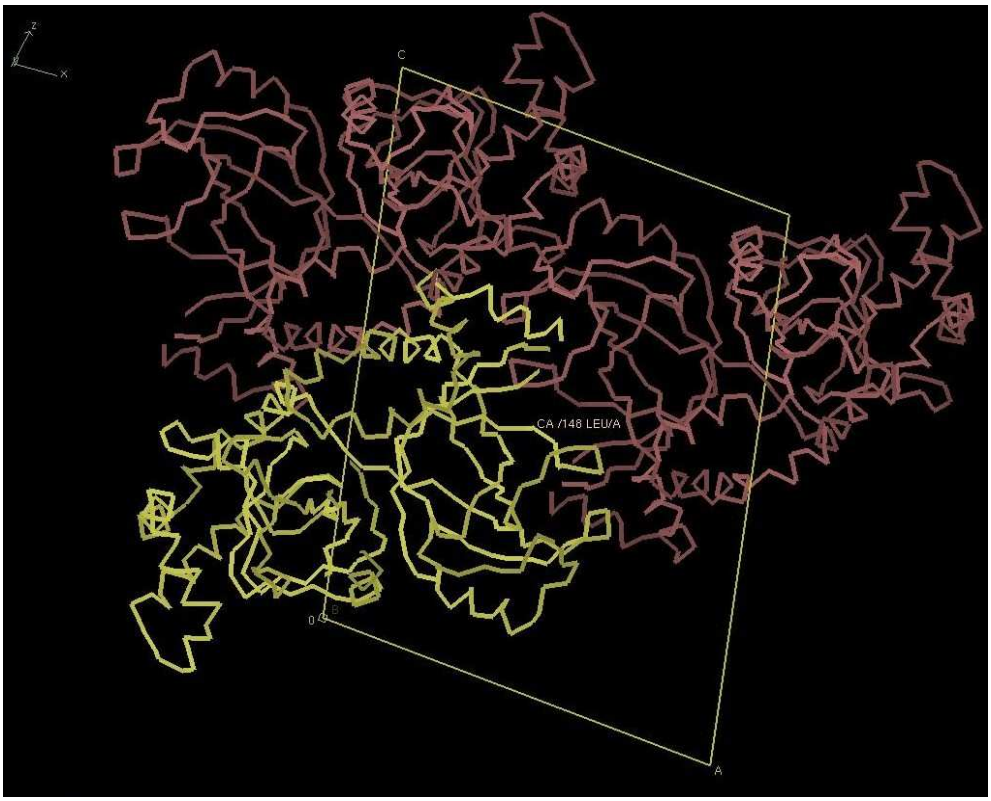

Figure 5.3: *The arrangement of CDTa molecules in the crystal unit cell (for CDTa-8.5 crystal). The two fold axis is perpendicular to the plane of the paper.*

The presence of bound ligand was confirmed in the electron density maps of respective structures immediately after the first round of restrained refinement. Figure 5.4 displays the observed electron density for unmodelled NAD in the CDTa-NAD structure at an initial stage of refinement.

The N terminal of CDTa was found to be highly disordered and electron density for few of the N terminal residues (1-27 for CDTa-8.5; 1-24 for CDTa 4.0; 1-23 for CDTa-9.0; 1-26 for CDTa-NAD and 1-16 for CDTa-NADPH) was not visible in any of the structures. However, the modelled length of the N-terminal varied among different structures which indicated the presence of these residues in the protein as all five crystals were grown from the same batch of the protein.

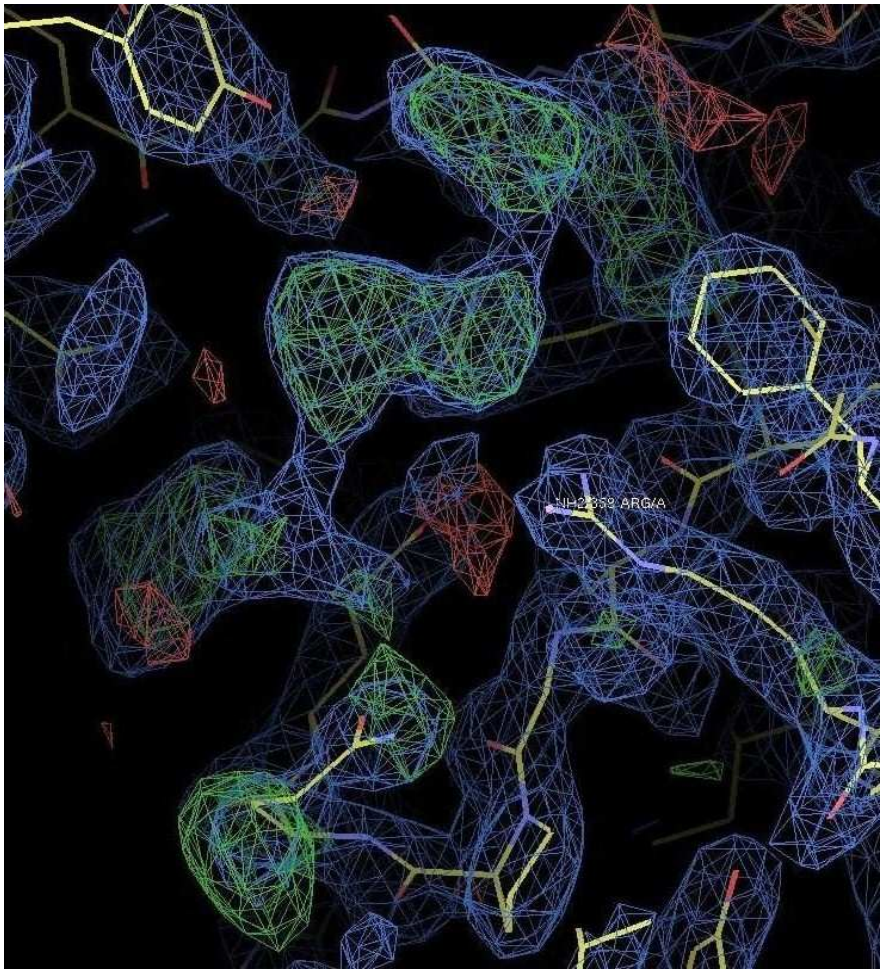

Figure 5.4: *Electron density for the bound NAD (unmodelled) seen after the first cycle of restrained refinement. The continuous density with green pieces of positive density corresponds to the bound NAD molecule. The 2Fo-Fc and Fo-Fc maps are shown at 1σ and 3σ contour level respectively.*

The loop regions -178-181, 382-384 for CDTa 4.0; 383-384 for CDTa-9.0 and 179-181 for CDTa-NAD, could also not be modelled. Figure 5.5 shows the final electron density for the ARTT loop (discussed in a later section) and the bound NAD in the CDTa-NAD structure.

There were no residues in the disallowed region of the Ramachandran plot for any of the structures (Figures 5.6 to 5.10). Table 5.4 summarises important structure refinement statistics for all five structures. A tight geometry for all molecules was maintained throughout the process of refinement and model building (Table 5.4).

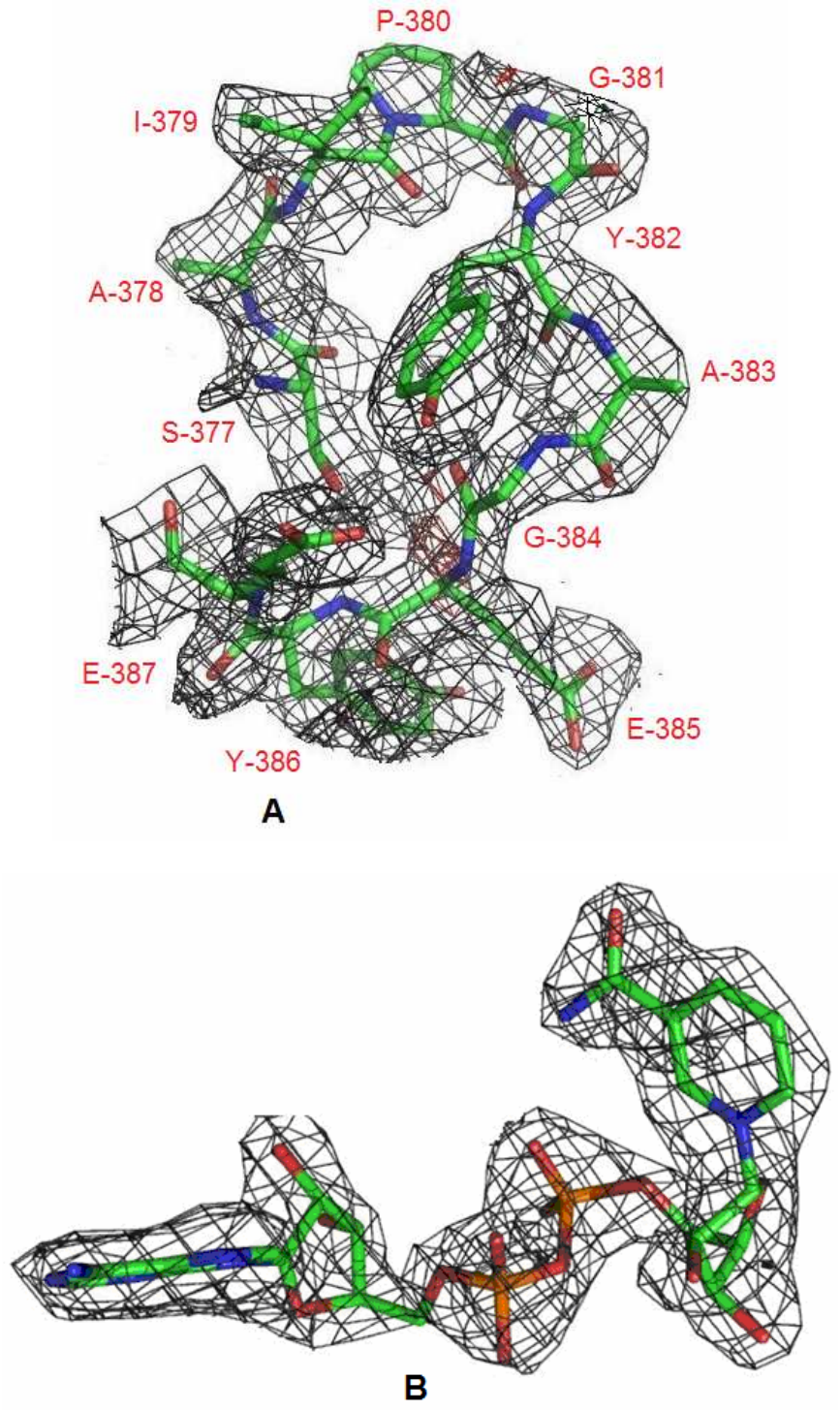

Figure 5.4: *The final electron density for (A) – the ARTT loop from CDTa-NAD structure, and (B) – the bound NAD in CDTa-NAD structure at 2.25 Å resolution. The chemical structure of NAD is given in Figure 2.7 (page 47). 2Fo-Fc and Fo-Fc maps are shown at 1σ and 3σ contour level respectively.*

Table 5.1: *The data collection and processing statistics for all five crystals.*

All crystals belong to **$P2_1$** space group	**CDTa-8.5 (pH = 8.5)**	**CDTa-4.0 (pH = 4.0)**	**CDTa-9.0 (pH = 9.0)**	**CDTa-NAD complex**	**CDTa-NADPH complex**
Wavelength of X-ray (Å)	0.9795	1.3625	1.3625	0.9795	0.9795
Exposure time per image	2 Seconds			4 Seconds	3 Seconds
Ligand / Substrate	-	-	-	NAD	NADPH
Cell parameters	57.9, 44.5, 78.0Å, β=102.8°	57.1, 42.7, 77.1Å, β=102.5°	57.4, 44.0, 78.5Å, β=102.5°	62.1, 46.8, 77.7Å, β=97.7°	60.8, 46.4, 77.5Å, β=98.4°
Maximum Resolution (Å)	1.85	2.0	1.90	2.25	1.95
Matthew's coefficient	2.18	2.04	2.12	2.48	2.40
Solvent content (%)	43.60	39.70	42.11	50.51	48.86
R_{symm} (%) Overall/ outermost shell	8.1 / 24.8	8.5 / 31.8	11.1 / 33.1	11.1 / 27.3	7.3 / 21.3
Completeness (%) Overall/ outermost shell	96.0 / 76.3	94.0 / 75.3	96.3 / 75.5	95.8 / 74.9	96.3 / 74.0
I / σI Overall / outermost shell	13.12 / 3.86	11.90 / 1.94	9.11 / 3.31	10.85 / 3.0	13.60 / 4.18
Data Multiplicity Overall / outermost shell	3.7 / 2.5	3.5 / 1.8	3.7 /2.7	3.7 / 2.1	3.8 / 2.6
No. of reflections Total / Unique	317439 / 33361	364537 / 24824	325716 / 29974	209467 / 21369	359671 / 30820

Table 5.2: *The molecular replacement solution statistics for CDTa-8.5 structure. The best solution is highlighted.*

S_ RF	TF	theta	phi	chi	tx	ty	tz	TFcnt	Rfac	Scor
S___1__1	1	83.22	-91.80	5.07	0.071	0.000	0.268	31.88	0.485	0.392
S___2__1	2	0.00	0.00	0.71	0.082	0.000	0.255	3.21	0.581	0.142
S___3__1	3	6.95	5.39	179.23	0.156	0.000	0.205	5.69	0.590	0.099
S___9__2	4	160.22	-5.33	128.87	0.769	0.000	0.306	1.53	0.595	0.092
S___4__5	5	34.23	102.54	174.58	0.798	0.000	0.289	2.41	0.598	0.083
S__12_13	6	52.98	26.26	90.42	0.156	0.000	0.284	2.02	0.586	0.081
S___7__5	7	157.78	133.21	160.85	0.407	0.000	0.171	1.54	0.598	0.075
S__10__7	8	48.00	-1.11	112.91	0.367	0.000	0.278	1.07	0.595	0.069
S__11__9	9	137.44	-5.69	56.61	0.055	0.000	0.244	1.58	0.591	0.062
S___5__4	10	60.41	-176.68	53.70	0.098	0.000	0.308	1.89	0.594	0.061
S___8__4	11	40.89	39.47	177.22	0.443	0.000	0.254	2.07	0.597	0.060
S___6__5	12	41.12	43.55	165.57	0.220	0.000	0.268	1.38	0.600	0.050

Table 5.3: *Refinement statistics for the first round of restrained refinement (10 cycles) for CDTa-8.5 structure. The starting and ending values are highlighted in cyan and yellow colours respectively.*

Ncyc	Rfact	Rfree	FOM	-LL	-LLfree	rmsBOND	zBOND	rmsANGL	zANGL	rmsCHIRAL
0	0.4941	0.4793	0.398	177378.	9471.3	0.0050	0.208	1.230	0.507	0.094
1	0.4341	0.4565	0.482	173289.	9318.0	0.0215	0.965	1.473	0.704	0.101
2	0.4005	0.4328	0.553	170723.	9224.7	0.0169	0.723	1.561	0.718	0.103
3	0.3796	0.4161	0.592	168959.	9156.3	0.0166	0.689	1.654	0.760	0.108
4	0.3646	0.4017	0.621	167647.	9102.5	0.0172	0.720	1.739	0.799	0.116
5	0.3525	0.3914	0.643	166603.	9061.2	0.0178	0.740	1.805	0.825	0.123
6	0.3426	0.3827	0.660	165762.	9028.1	0.0186	0.772	1.865	0.850	0.130
7	0.3347	0.3753	0.674	165047.	9002.4	0.0193	0.800	1.907	0.870	0.135
8	0.3281	0.3704	0.685	164467.	8980.2	0.0199	0.824	1.955	0.894	0.140
9	0.3224	0.3662	0.694	163974.	8963.4	0.0205	0.851	2.002	0.917	0.145
10	0.3178	0.3640	0.701	163593.	8949.3	0.0211	0.870	2.040	0.934	0.149

Table5.4: *The structure refinement statistics for all five CDTa structures.*

	CDTa-8.5	**CDTa-4.0**	**CDTa-9.0**	**CDTa-NAD**	**CDTa-NADPH**
R_{cryst} / R_{free} (%)	19.97 / 23.91	21.98 / 27.96	20.97 / 25.95	20.46 / 26.50	20.46 / 25.99
Ramachandran plot (%) Allowed / Generously allowed	99.70 / 0.30				99.40 / 0.60
RMSD bond angles (°)	0.94	1.08	1.00	1.15	1.06
RMSD bond length (Å)	0.007			0.008	0.007
Number of Protein atoms	3135	3137	3178	3157	3247
Number of Water molecules	424	154	231	156	259
Average B factor (Å^2)-					
protein atoms	19.39	34.15	25.35	32.35	30.99
main chain / side-chain	18.95 / 19.84	33.55 / 34.74	24.67 / 26.02	31.91 / 32.78	30.50 / 31.49
Water	28.99	39.39	32.58	41.07	39.86
Ligand (NAD/NADPH)	--	--	--	43.10	44.10
Glycerol	--	--	--	58.70	53.60
PDB ID	2WN4	2WN8	2WN5	2WN7	2WN6

- $R_{symm} = \Sigma h \Sigma i[|Ii(h) - \langle I(h)\rangle| / \Sigma h \Sigma i Ii(h)]$, where Ii is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$. $R_{cryst} = \Sigma h|F_o - F_c| / \Sigma h Fo$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively. R_{free} is equal to R_{cryst} for a randomly selected 5% of reflections not used in the refinement.

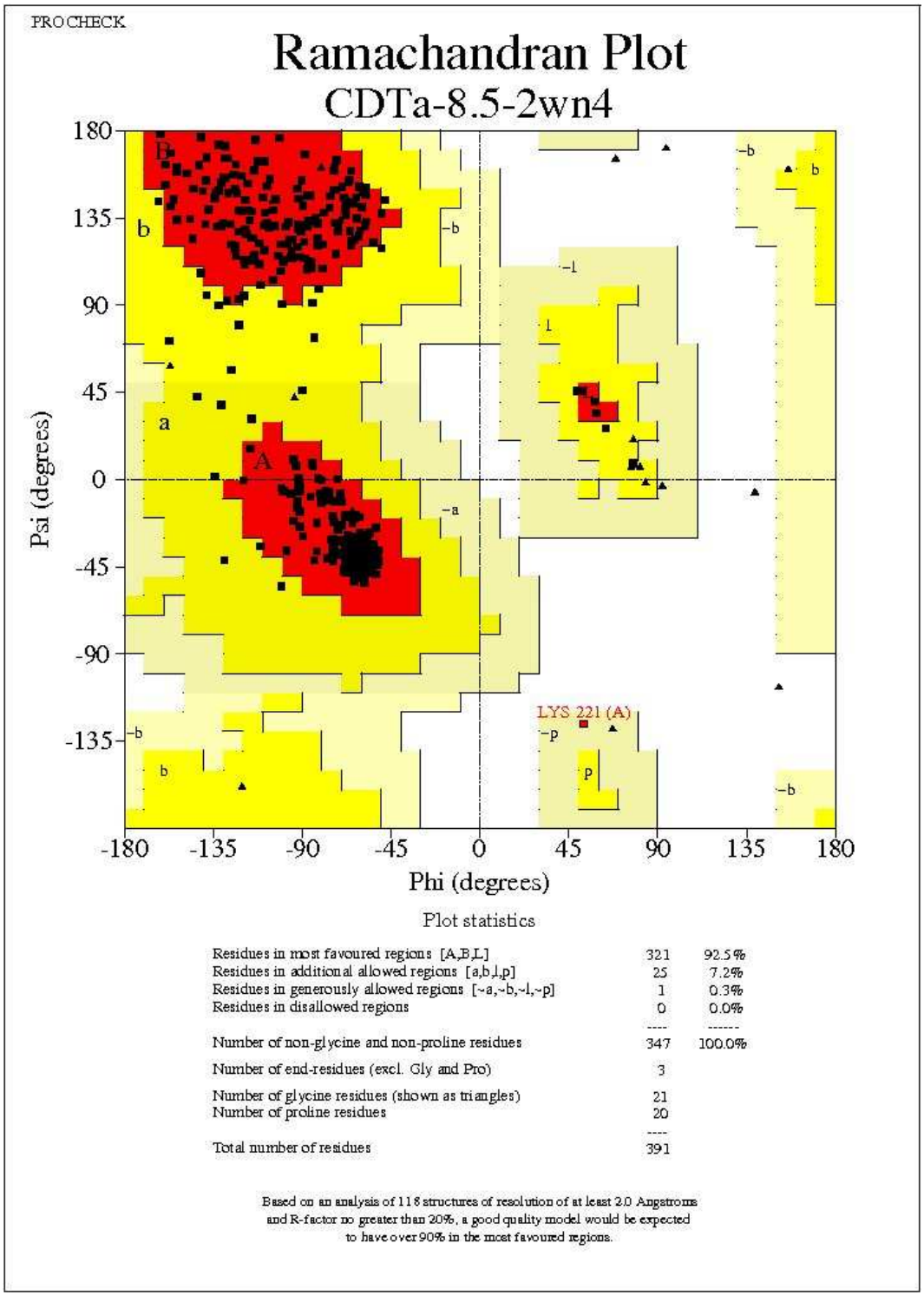

Figure 5.6: *The Ramachandran plot for CDTa-8.5 structure.*

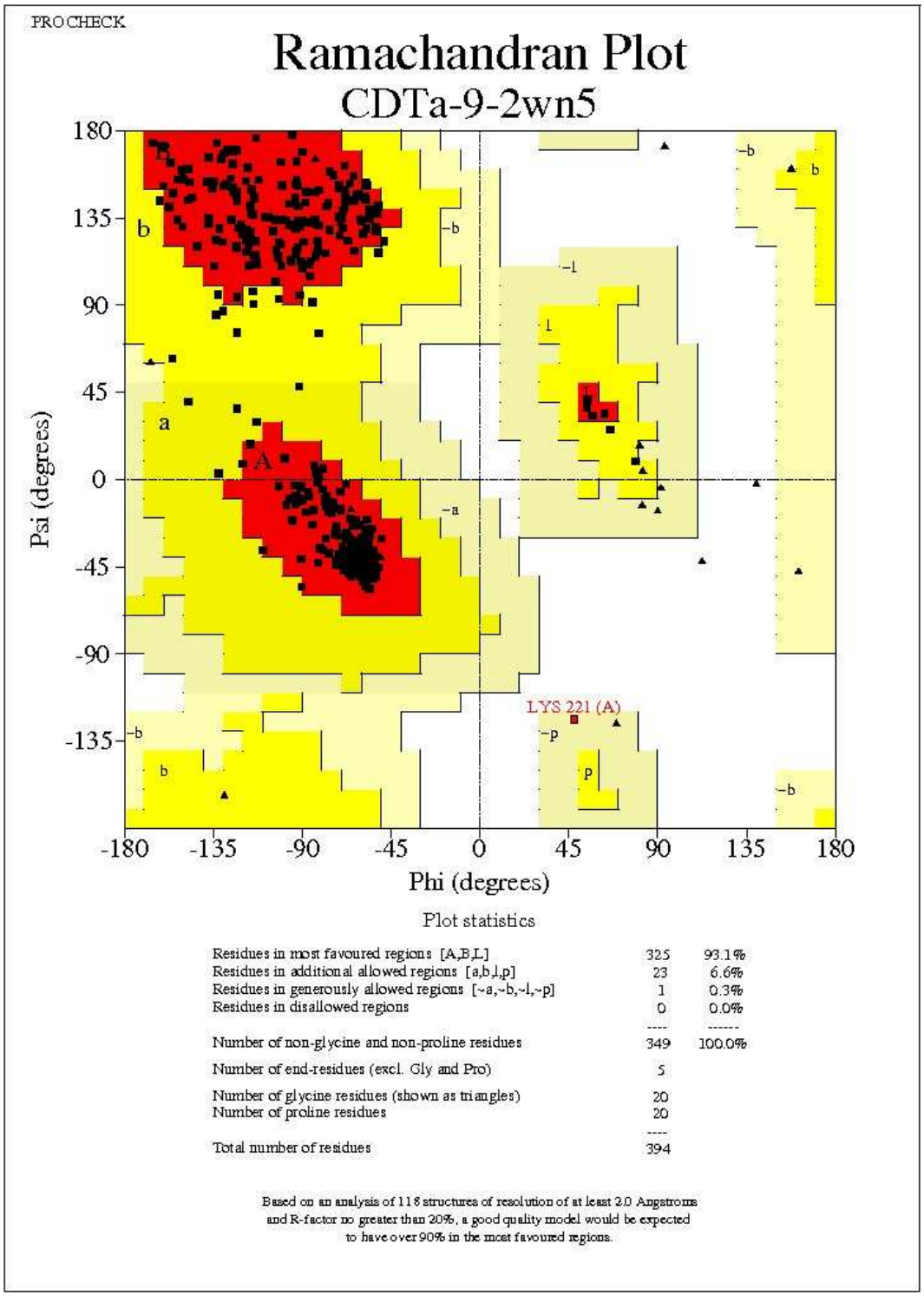

Figure 5.7: *The Ramachandran plot for CDTa-9.0 structure.*

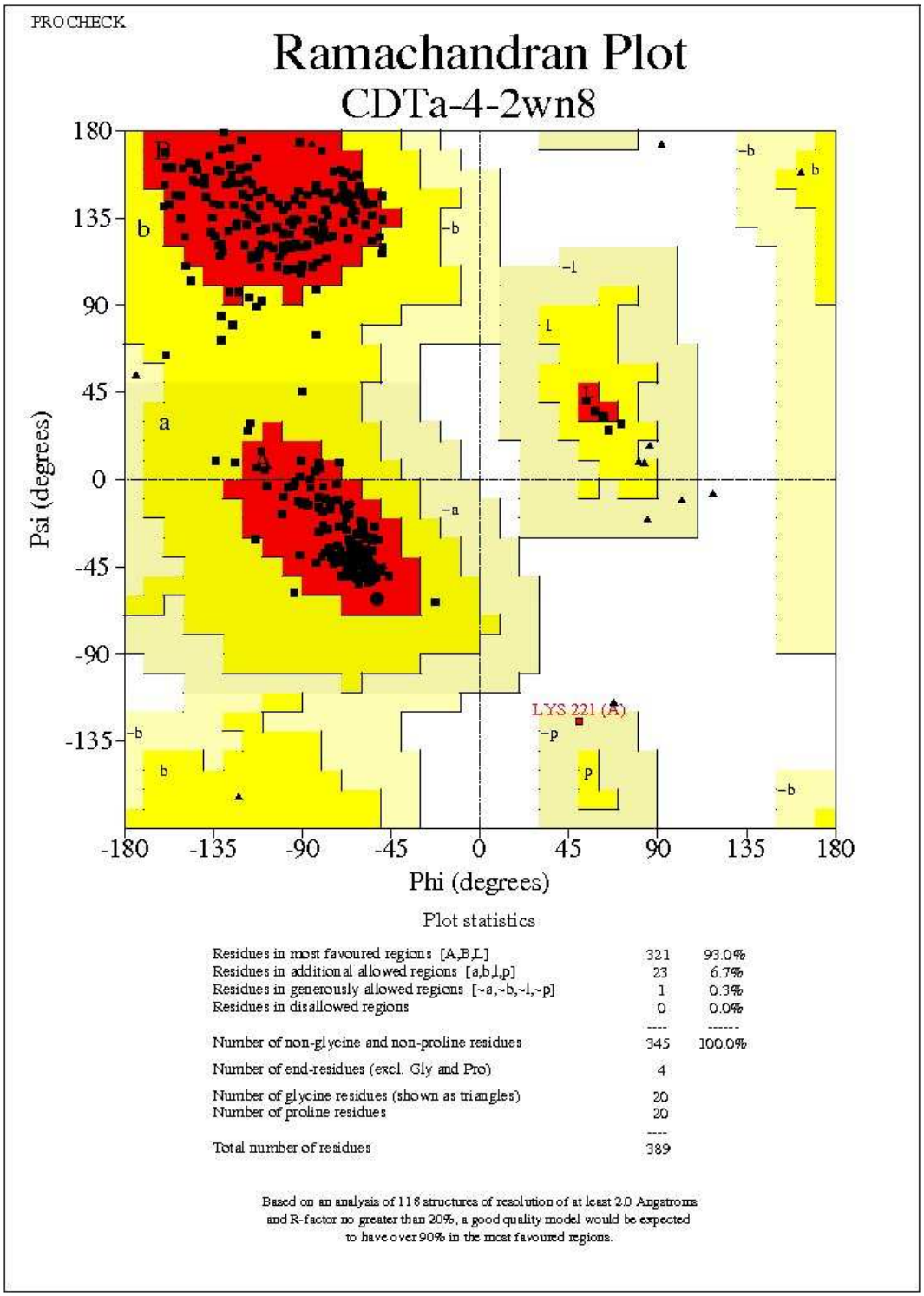

Figure 5.8: *The Ramachandran plot for CDTa-4.0 structure.*

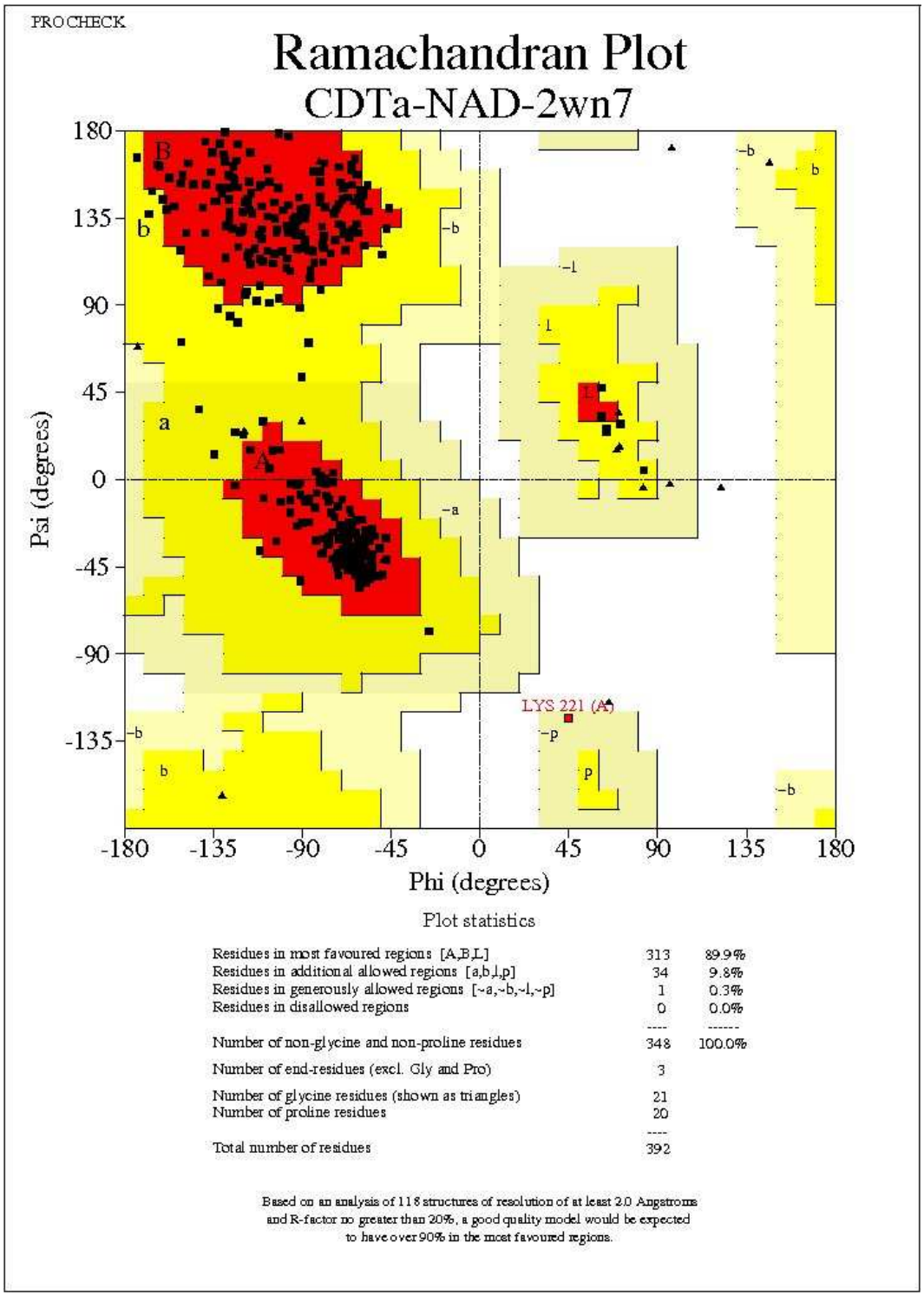

Figure 5.9: *The Ramachandran plot for CDTa-NAD structure.*

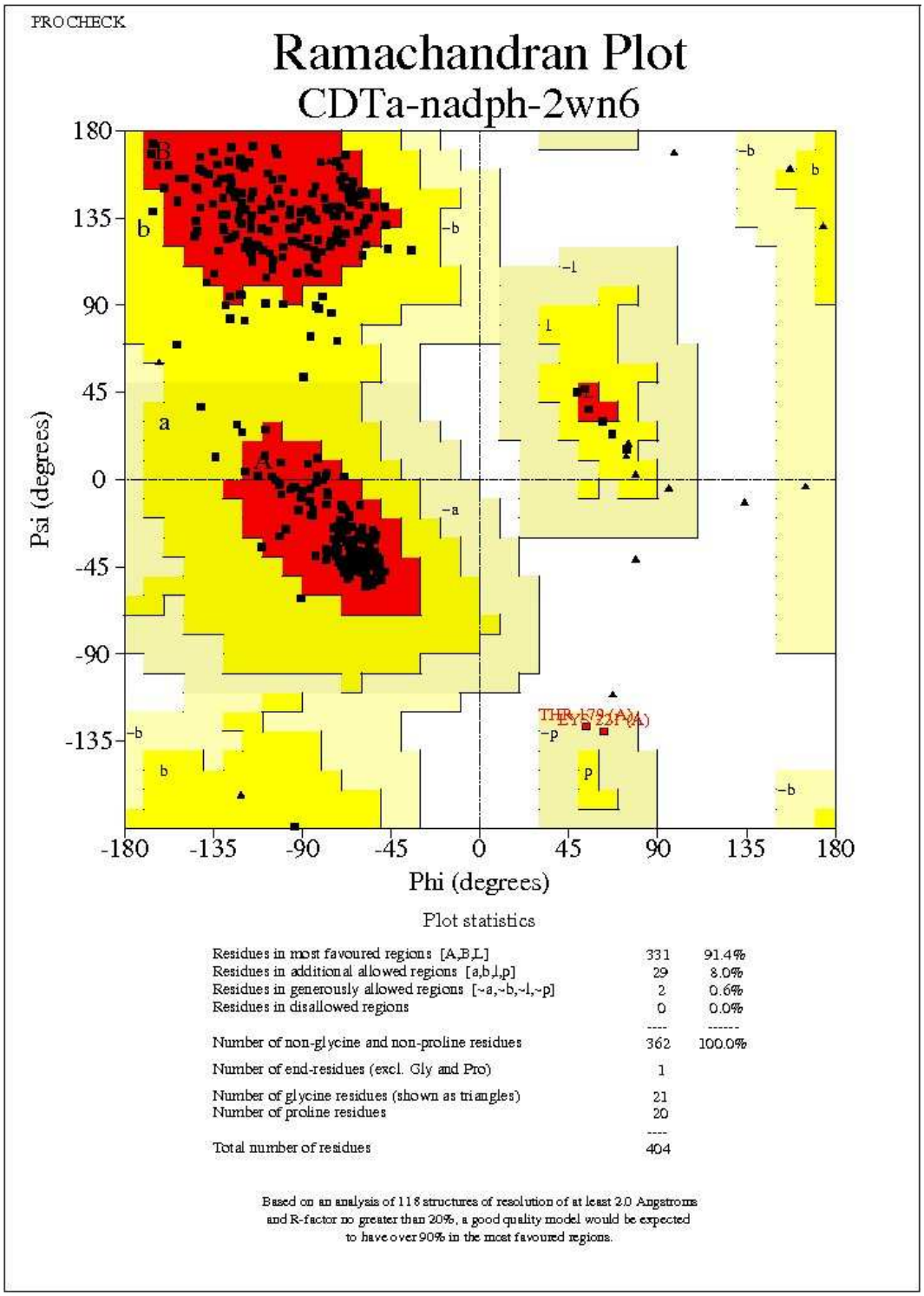

Figure 5.10: *The Ramachandran plot for CDTa-NADPH structure.*

Overall Structure of CDTa

C. difficile binary toxin (CDT) belongs to the class 4 of ADPRT superfamily. Toxins from this class target monomeric actin molecules in the target cell and are known as Actin-ADPRTs (Popoff and Boquet, 1988). CDTa is the enzymatic component of CDT and ADP-ribosylates all three isoforms of actin. CDTa shares about 84% sequence identity with the enzymatic component of *C. perfringens* Iota toxin (Ia). The sequence identity between CDTa and the enzymatic component of *C. botulinum* C2 toxin (C2I) is about 40% (Barth 2004).

The crystal structure of Ia in complex with NAD and NADPH but not in its native form has been reported by Tsuge and co-workers (Tsuge *et al.*, 2003). The crystal structure of C2I in its native form but not in ligand bound forms has recently been determined by Schleberger and co-workers (Schleberger *et al.*, 2006). Since both of these toxins belong to two different classes (Table 2.1) of Clostridial actin-ADPRTs (Mauss, 1990), it is not possible to compare them at the atomic level.

Irrespective of the variable sequence homology between different Actin-ADPRTs, they all possess similar three dimensional fold (Holbourn *et al.*, 2006). A high degree of sequence conservation is reflected at the structural level when the three dimensional structure of CDTa is compared with those of previously reported crystallographic results on Ia from *C. perfringens* (Tsuget *et al.*, 2003), C2I from *C. botulinum* (Schleberger *et al.*, 2006) and the ADPRT component of vegetative insecticidal protein, VIP2 from *Bacillus cereus* (Han *et al.*, 1999) (Table 5.5 and Figure 5.11).

Table 5.5: *The structural comparison of CDTa with the known homologues (The r.m.s.d. values are shown in Å). The aligned length of the protein (number of C^{α} atoms) is shown in brackets.*

	CDTa-NAD	**Ia-NAD**	**C2I**
CDTa-NAD	--	--	--
Ia-NAD (1GIQ)	1.02 (392)	--	--
C2I (2J3X)	2.75 (382)	2.86 (401)	--
VIP2-NAD (1QS2)	2.79 (378)	2.96 (390)	3.90 (386)

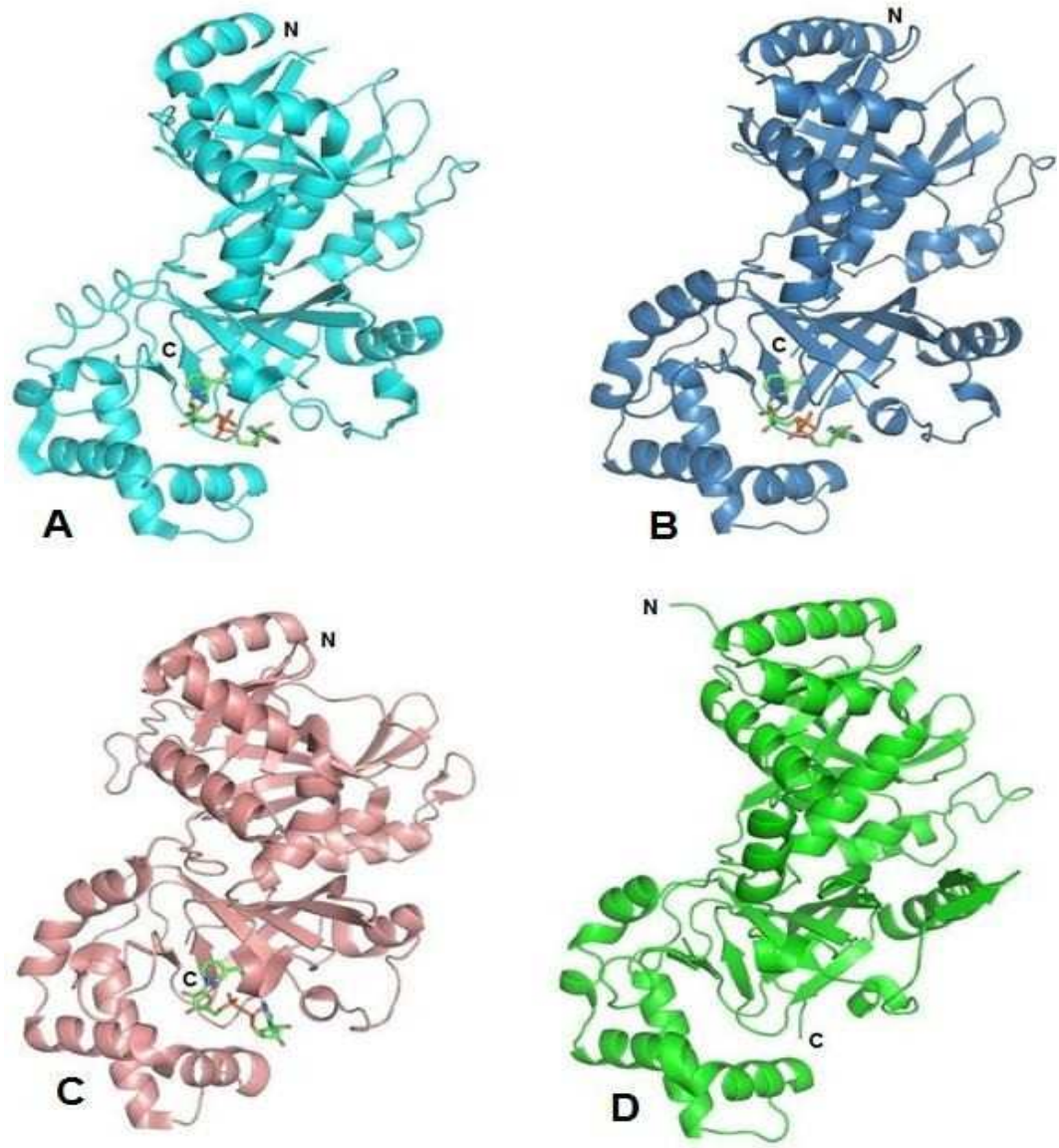

Figure 5.11: *Crystal structures of the enzymatic component of different Actin-ADPRT binary toxins indicating overall three-dimensional fold of the molecule. (A) – CDTa (PDB ID - 2WN7) (Sundriyal et al., 2009), (B) – Ia (PDB ID - 1GIQ) (Tsuge et al., 2003), (C) – VIP2 (PDB ID - 1QS2) (Han et al., 1999), and (D) – C2I (PDB ID - 2J3X) (Schleberger et al., 2006). Bound NAD is shown in sticks.*

However, substrate specificity (Table 2.1) of these toxins can not be explained from their structures. Perhaps the answer lies in the way these toxins interact with their ADP-ribose acceptor substrate i. e. actin. γ smooth muscle actin differs from the other two isoforms (α and β) of actin at the N-terminal only and therefore it was suspected that perhaps this region is primarily responsible for substrate recognition by different Clostridial binary toxins (Vendekerckhove and Weber, 1979). The crystal

structure of Ia in complex with actin has been determined (Tsuge *et al.*, 2008) but no structure for C2I-actin complex is available to compare with it.

All five CDTa structures superimpose well on Ia, C21 and VIP2 (Table 5.5). The overall structure of CDTa matches extremely well with that of Ia (Table 5.5, r.m.s.d. = 1.02 Å) except the ADP ribosyl turn turn (ARTT) loop region (discussed in detail later). Enzymatic components of Clostridial binary toxins (Figure 5.11 A, B and D) are composed of two mixed alpha-beta globular domains (Han *et al.*, 1999; Tsuge *et al.*, 2003; Schleberger *et al.*, 2006). In CDTa, the N terminal domain extends from 1 to 215 residues whereas the C terminal domain is from 224-420. The two domains of the protein are linked by a loop that stretches from residue 216 to 223 (Figure 5.12).

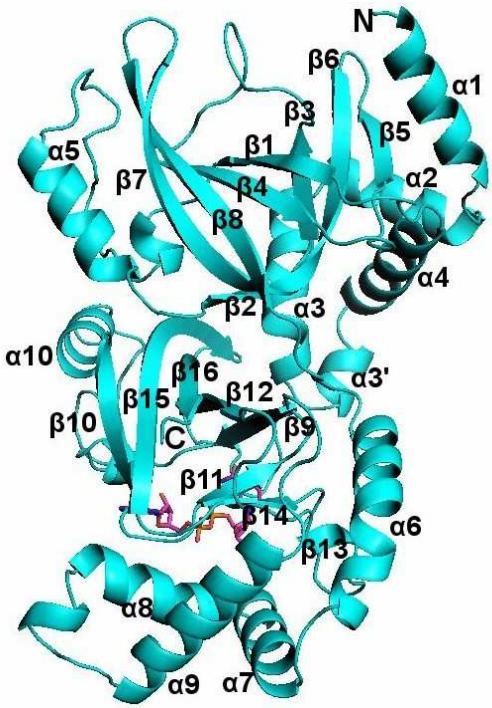

Figure 5.12: *Overall structure of CDTa (cartoon representation) with NAD bound to the catalytic cleft (shown in sticks)*

The N-terminal domain of CDTa consists of 5 alpha helices and 8 beta strands and is believed to interact with its translocation partner (i. e. CDTb in this case) during the process of translocation (Tsuge *et al.*, 2003). The C-terminal domain of the protein also comprises of 5 alpha helices and 8 beta strands and accommodates catalytic machinery of the

protein (Figure 5.12). Both domains are arranged almost perpendicular to each other but facing their clefts towards the same face of the protein similar to their organisation in VIP2, Ia or C2I (Han *et al.*, 1999; Tsuge *et al.*, 2003; Schleberger *et al.*, 2006). Numbering of secondary structure elements in CDTa (Figure 5.12) follow the secondary structure assignment as in Ia (Tsuge *et al.*, 2003).

As in other Actin-ADPRTs, both domains of the protein adopt a similar fold despite very low sequence identity (18% in case of CDTa) between them and this has been suggested to be a result of a gene duplication effect (Han *et al.*, 1999). In CDTa, both domains superimpose onto each other with an r.m.s.d. of 2.62 Å (Figure 5.13).

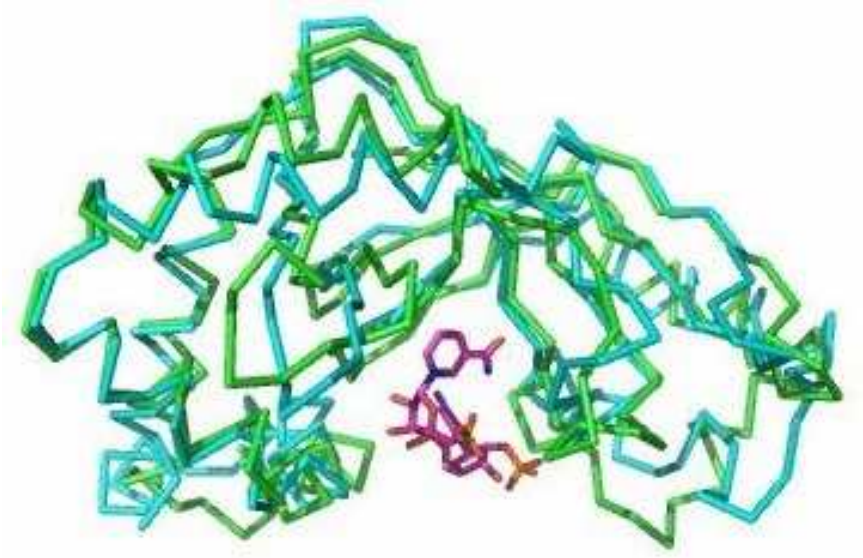

Figure 5.13: *Superimposition of the N-terminal (Green) and the C-terminal (Cyan) domains of CDTa on each other with bound NAD to the C-terminal domain.*

Catalytic Cleft and Binding of NAD and NADPH

The enzymatic component of *C. perfringens* iota toxin (Ia) is the closest homologue of CDTa, sharing about 84% sequence identity between them. Amino acid residues that have been suggested essential for the ADP ribosylating activity of Ia (Arg-295, Arg-296, Arg-352, Gln-300, Asn-335, Glu-378 and Glu-380) (van Damme *et al.*, 1996) are well conserved in CDTa (Arg-302, Arg-303, Arg-359, Gln-307, Asn-342, Glu-385 and Glu-387) (Table 5.6).

Our present structural analysis has shown that both, NAD and NADPH bind to the catalytic cleft of CDTa in a 'closed conformation' interacting with residues Arg-302, Arg-303, Arg-359, Gln-307, Asn-342 and Ser-345 (Figure 5.14). This is analogous to the structural observations made with Ia (Figure 5.14) with the NAD molecule interacting with residues Glu-380, Arg-295, Arg-296, Arg-352, Gln-300, and Asn-335 (Tsuge *et al.*, 2003).

Table 5.6: *The positional conservation of catalytically important residues in Ia and CDTa. Blue – residues directly interacting with NAD/ NADPH, Red – suggested residues to interact with Actin, Black – other important residues in the active site.*

Motif	Ia	CDTa
	Y-246	Y-253
	Y-251	Y-258
Arg/His	R-295	R-302
	R-296	R-303
	Q-300	Q-307
	E-301	E-308
	N-335	N-342
STS	S-338	S-345
P-N loop	F-349	F-356
	R-352	R-359
	N355	N362
EXE (ARTT loop)	Y-375	Y-382
	E-378	E-385
	E-380	E-387

Based on these observations (Figure 5.14 and Table 5.6) it is interesting to note that in CDTa, Glu-387, which corresponds to Glu-380 in Ia does not seem to interact either with NAD or with NADPH (Table 5.6, Figures 5.14 and 5.15). However, Ser-345 in CDTa seems to be an important residue in the catalytic site and makes direct interactions with the ligand in both NAD and NADPH complex structures. These observations point out that even between these two close homologues (CDTa and Ia), the mode of ligand recognition is significantly different (Table 5.6 and Figure 5.14).

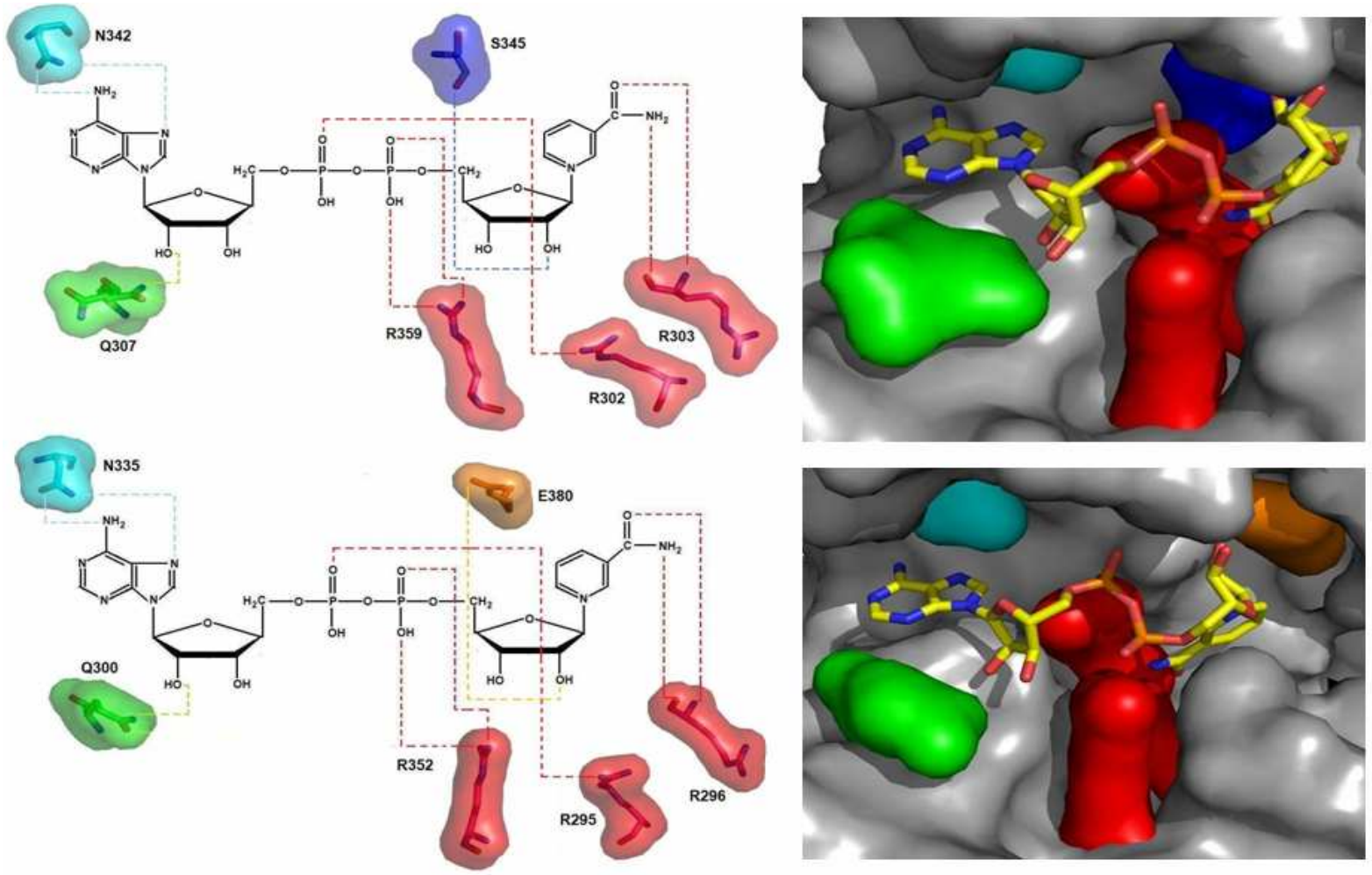

Figure 5.14: *A schematic representation of Hydrogen bonding of NAD to CDTa (top) and Ia (bottom).*

Gln-307 is an important residue in the catalytic cleft and makes direct interactions with both ligands i. e. NAD and NADPH in their respective complex structures. Gln-307 adopts similar orientation in all native CDTa structures with its side chain leaning towards the catalytic cleft. Dual conformation of Gln-307 was observed in the CDTa-NAD complex (Figures 5.14 and 5.15 A).

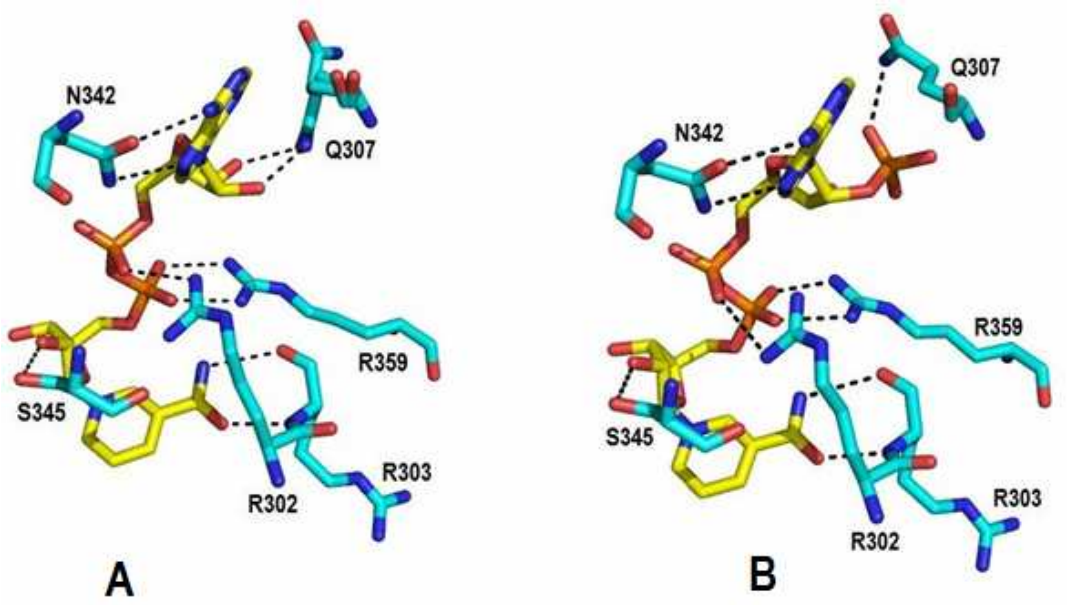

Figure 5.15: *Binding of (A) – the NAD and (B) – the NADPH to CDTa. The broken black lines show possible hydrogen bonds (based on distances).*

It seems that Gln-307 moves towards and away from the cleft and that its interaction with NAD in CDTa is not static but dynamic in nature (Figure 5.15 A). In CDTa-NADPH complex, Gln-307 has been pushed permanently away from the cleft by the phosphate group of NADPH which accommodates itself but still making direct interaction with it to stabilise the complex (Figure 5.15 B). Thus, Gln-307 seems to be one of the key residues for ligand-enzyme complex stability. A similar displacement of equivalent Gln residue (Gln-300) side chain has been reported in Ia-NADPH structure but its dual conformation has not been observed in Ia-NAD structure (Tsuge *et al.*, 2003). Authors could not compare this movement of Gln-300 with native Ia because of non-availability of crystals of Ia in its native form. Table 5.7 lists all hydrogen bond interactions to compare the binding of NAD and NADPH to CDTa.

Table 5.7: *The hydrogen bond interactions of CDTa with NAD and NADPH.*

	CDTa-NAD		**CDTa-NADPH**	
Bonded residues (Atoms)	**Length (Å)**	**Angle (⁰)**	**Length (Å)**	**Angle (⁰)**
R302 (NH1) – NADPH (O1A)	-	-	3.38	149.2
R302 (NH2) – NAD / NADPH (O1A)	2.92	158.7	3.34	151.5
R303 (N) – NAD / NADPH (O7N)	2.67	150.9	2.68	161.8
R303 (O) – NAD / NADPH (N7N)	3.09	-	3.08	-
S345 (OG) – NAD / NADPH(O2D)	3.12	139.1	3.0	148.9
N342 (OD1) – NAD / NADPH (N6A)	3.06	-	2.98	-
R359 (NH1)– NAD / NADPH (O1N)	2.71	164.1	2.40	154
R359 (NH2) – NAD / NADPH (O2N)	2.82	154.0	2.89	127.8
Q307 (N) – NAD / NADPH (O3X)	-	-	2.48	160.9
Q307 (NE2) – NAD (O3B) / NADPH (O1X)	2.60	94.1	2.71	103.5

Ligand Binding and ARTT Loop

It has been well established that in ADPRTs the ADP-ribosyl turn-turn (ARTT) loop is important for substrate binding and ADP-ribosylation even though the length of the loop varies among these proteins (Holbourn *et al.*, 2006). The ARTT loop in Ia spans from residue 370 to 380 (Tsuge *et al.*, 2003). In CDTa, this loop (connecting strands β13 and β14) spans from residues 377 to 387 and consists of two sharp turns as in Ia or C3Bot (Tsuge *et al.*, 2003; Han *et al.*, 2001).

Conformational changes in the ARTT loop induced by NAD binding have been reported for C3Bot toxin (Menetery *et al.*, 2002). These conformational changes in the loop, however, have not been claimed with confidence in Ia due to the non-availability of the same crystal form for native Ia. Authors (Tsuge *et al.*, 2003) suggested that in Ia, it was possible to have similar conformational changes in the ARTT loop as a result of NAD binding and that these conformational changes in the loop

possibly disturbed the molecular packing in the crystal and prevented the authors from having native Ia and Ia-NAD crystals in the same form.

With CDTa, we have overcome this problem and have determined the crystal structure of CDTa in its native form as well as in complex with NAD and NADPH in the same crystal forms (Tables 4.2 and 5.1). Hence, a direct comparison between native CDTa structures at acidic (CDTa-4.0) as well as at basic pH (CDTa-9.0) with ligand bound structures was possible.

The ARTT loop in CDTa is found to be associated with significant disorder and high conformational flexibility in all three native structures as observed from their electron density maps. However, upon ligand (NAD/NADPH) binding, the loop adopts a highly ordered structure (Figure 5.16) associated with some critical changes in the orientation of side-chains in the catalytic site when compared with Ia.

Electron density for both of the proposed catalytically important residues (Glu-385 and Glu-387) of the EXE motif in the ARTT loop was well defined in all five structures. The EXE motif adopts similar orientation in all structures (Figure 5.17). Ligand binding seems to stabilise the loop and electron density for the whole loop was clearly visible (Figure 5.16). This finding from two different ligand bound structures (CDTa-NAD and CDTa-NADPH) suggests that although the ligand binding stabilises the loop, it does not induce any specific large conformational changes in the loop as suggested for C3Bot (Menetrey *et al.*, 2002) or proposed for Ia (Tsuge *et al.*, 2003).

C3Bot belongs to the Rho-ADPRT superfamily that targets Rho proteins. Phe-209, the conserved aromatic residue in the ARTT loop of C3Bot has been shown to be essential for substrate binding (Han *et al.*, 2001). This residue corresponds to Tyr-375 in Ia, Tyr-382 in CDTa, Phe-423 in VIP and Phe-384 in C2I but its functional implications have not been discussed for any of these structures.

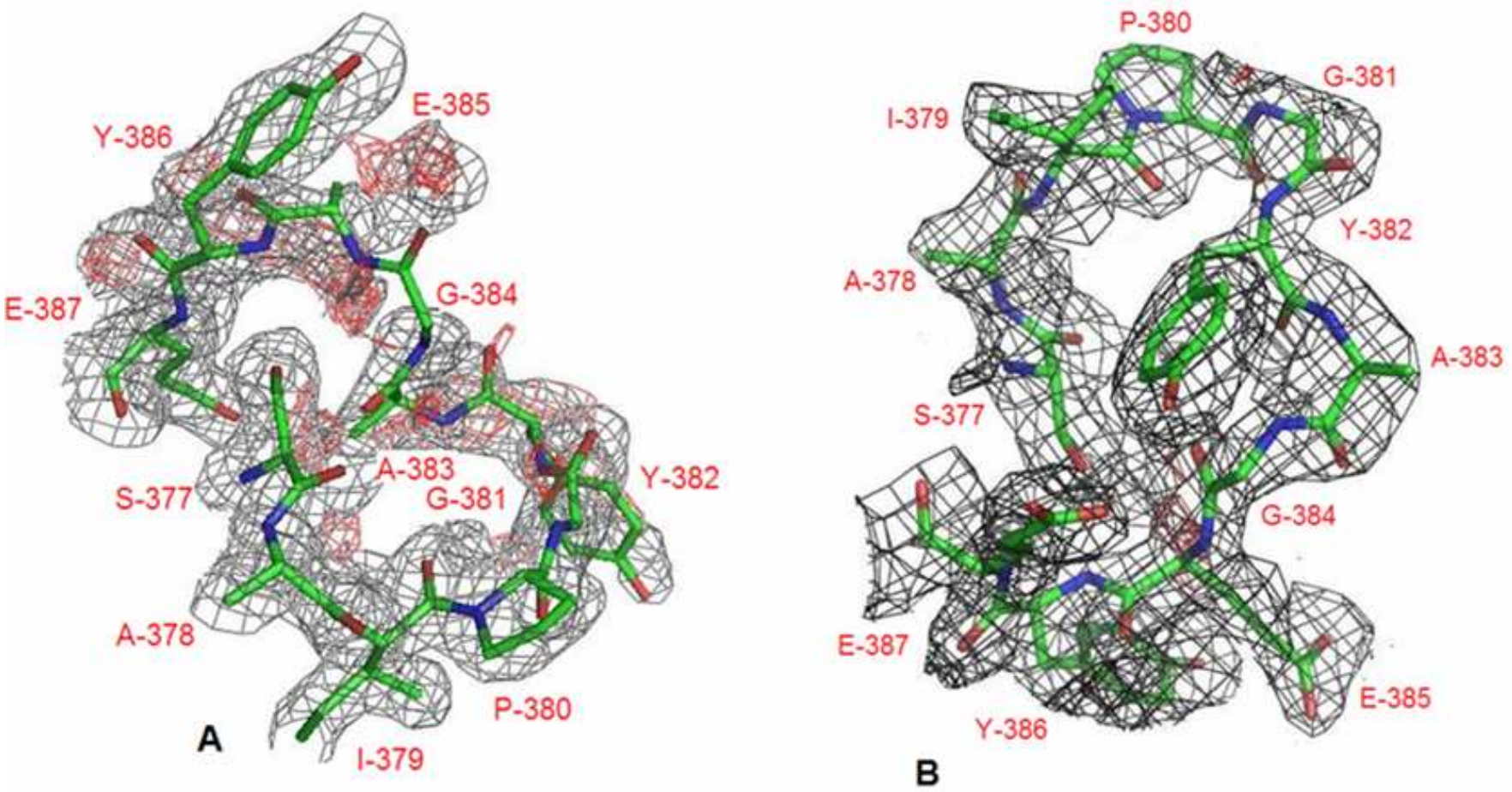

Figure 5.16: *Electron density around the ARTT loop in (A) – CDTa 9.0, and (B) – CDTa-NAD structure. Disorder in the loop region can be seen clearly in the form of breaks in the electron density and the noise (red coloured density). 2Fo-Fc and Fo-Fc maps are shown at 1σ and 3 σ contour level respectively.*

Han and co-workers (Han *et al.*, 2001) have suggested that the solvent exposed side chain of Phe-209 in C3Bot may have a possible role in Rho protein binding to the enzyme. They further suggested that the absence of any other hydrophobic residue near Phe-209 in the protein 3-dimensional structure will lead to significant conformational changes in the protein in order to bury Phe-209. In the crystal structure, the authors have reported that Phe-209 of the protein interacts hydropobically with Phe-49, Trp-58 and Ile-124 of the non-crystallographic symmetry related molecule in the crystal in order to stabilise the structure.

The side chain of Tyr-382 (a conserved critical aromatic residue known to be important in ADPRTs) in the ARTT loop was not visible in the CDTa-8.5 and CDTa-4.0 structures. It could be modelled in the CDTa-9.0, CDTa-NAD and CDTa-NADPH structures. However, interestingly, it adopts a different orientation in the native (CDTa-9.0) and ligand bound forms (Figures 5.17 and 5.18) which seems to be crucial for stabilisation of the protein-ligand complex. In the native CDTa structure, Tyr-382 stacks itself with Phe-126 of the symmetry related molecule similar to that seen in C3Bot. In the ligand bound structures, Tyr-382 flips inside towards the catalytic cleft and adopts a similar orientation in both of the complexes (CDTa-NAD and CDTa-NADPH) and interacts with Glu-387 (of EXE motif) which is considered an important catalytic residue (Figure 5.17).

This movement of Tyr-382 would make it unavailable for an interaction with the substrate molecule unlike in C3Bot. Tyr-375 of Ia was also found in an inward flipped orientation in Ia-NAD structure (Figure 5.18). However, a recent report on Ia-Actin complex structure reveals that Tyr-375 of Ia in the complex adopts a similar inward flipped side chain orientation and does not interact with its substrate i. e. Actin (Tsuge *et al.*, 2008).

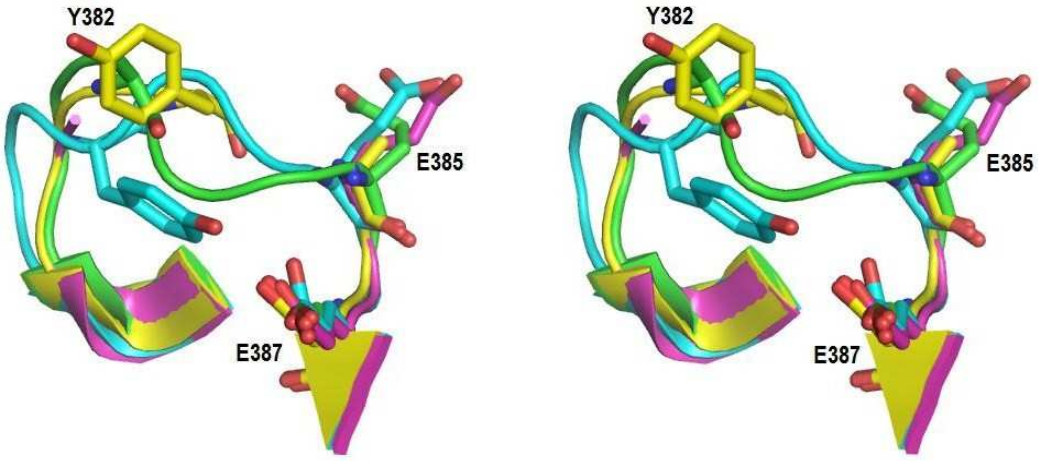

Figure 5.17: *A stereo view of the orientation of ARTT loop in CDTa in native and NAD bound form. Green – CDTa-8.5, Yellow – CDTa-9.0, Magenta – CDTa-4.0, Cyan – CDTa-NAD. Green – CDTa-8.5, Yellow – CDTa-9.0, Cyan – CDTa-NAD, Magenta – Ia-NAD. (The residue numbering is according to CDTa. The corresponding residues in Ia are Tyr-375, Glu-378 and Glu-380.*

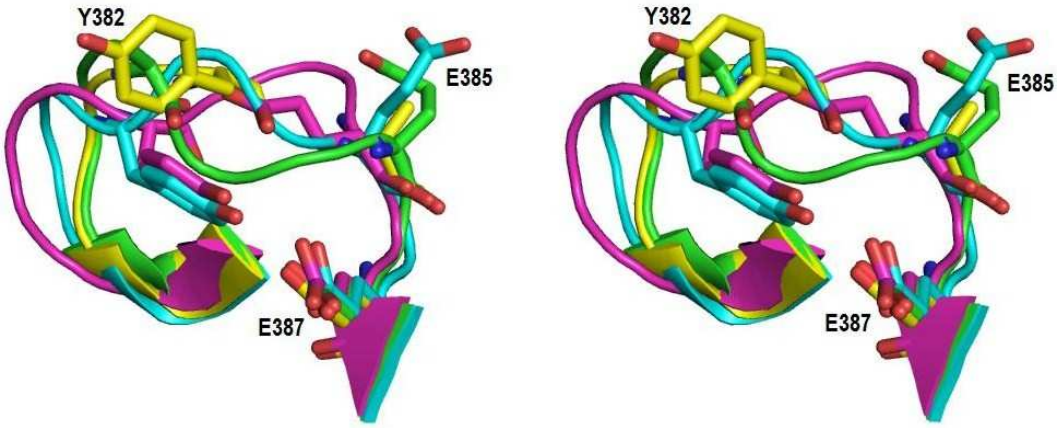

Figure 5.18: *A stereo view of the ARTT loop in CDTa (native and NAD bound form) and Ia (NAD bound form). Green – CDTa-8.5, Yellow – CDTa-9.0, Cyan – CDTa-NAD, Magenta – Ia-NAD. (The residue numbering is according to CDTa. The corresponding residues in Ia are Tyr-375, Glu-378 and Glu-380.*

EXE Motif and STS Motif

The EXE motif present in the ARTT loop has been considered important for ligand binding (Han *et al.*, 2001; Han and Tainer, 2002). Glu-378 and Glu-380 form the EXE motif in Ia and correspond to Glu-385 and Glu-387 in CDTa.

Site-directed mutagenesis of Glu-378 and other catalytically important residues in Ia have been studied in detail by Nagahama and

co-workers (Nagahama *et al.*, 2000). Results of their study suggest that Glu-378 plays a crucial role in stabilising substrate-enzyme complexes and catalysis. Its mutation to Ala resulted in the complete loss of NADase, ARTase, cytotoxic and lethal activity of Ia indicating that the carboxylic group of Glu-378 is essential for these activities. However, the kinetic analysis suggests that Glu-378 is essential for catalytic activity of Ia but not required for binding to NAD. Mutagenesis data from the same study suggests that Glu-380 is also not required for NAD binding in Ia.

Glu-380 has been shown to interact directly with NAD in the Ia-NAD complex whereas both residues (Glu-378 and Glu-380) are at hydrogen-bonding distance from NADPH in the Ia-NADPH complex (Tsuge *et al.*, 2003). The binding of NADPH to Ia has not been discussed by the authors. In CDTa, however, the structurally equivalent Glu residues (Glu-385 and Glu-387) are not involved in direct interaction with either NAD or NADPH. (Figures 5.14 and 5.15, Tables 5.6 and 5.7).

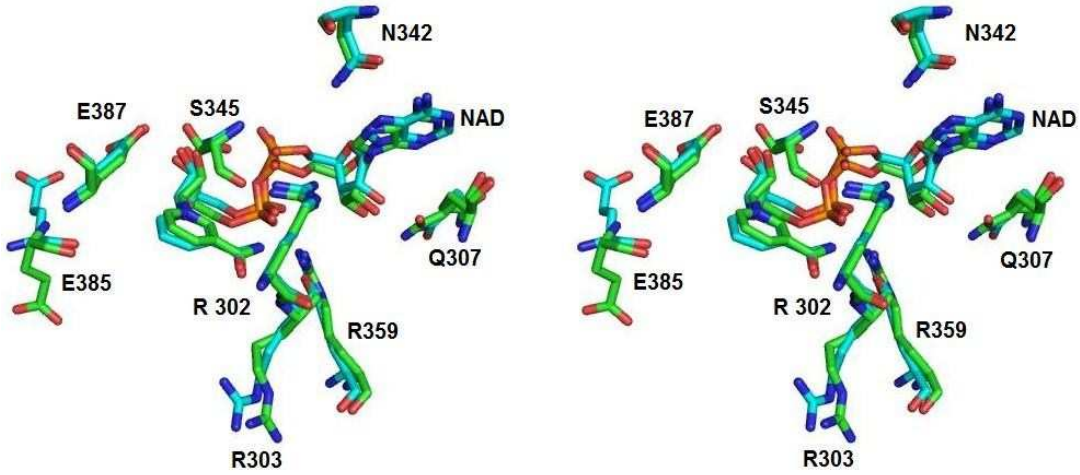

Figure 5.19: *A stereo representation of superimposition of the catalytic machinery of CDTa and Ia with bound NAD. Green – CDTa-NAD, Cyan – Ia-NAD. The residues numbering is according to CDTa.*

In addition, Glu-385 (corresponding to Glu-378 in Ia) adopts different orientation all together in CDTa (Figure 5.18). In Ia, the side chain of Glu-378 points towards the ligand binding cleft whereas in all five CDTa structures determined so far, it points away from the cleft eliminating possibilities of its interaction with any of the two studied ligands (Figure 5.19). No interaction of these two residues of CDTa (Glu-385 and Glu-387) with NAD or NADPH still resulting in stable complex

formation suggests that the EXE motif is perhaps not necessary for the ligand binding and stabilisation of the complex in CDTa. This finding agrees with the results of mutational studies by Nagahama and co-workers (Nagahama *et al.*, 2000) on Ia.

Ser-345, Thr-346 and Ser-347 together constitute the STS motif in CDTa. This motif corresponds to Ser-338, Thr-339 and Ser-340 in Iota toxin (Ia). Replacement of Ser-338 to Ala or Cys in Ia did not result in the complete loss of activity and suggests that the hydroxyl group of Ser-338 is not essential for catalytic activity (Nagahama *et al.*, 2000). However, its replacement to amino acids with a larger side chains such as Phe results in complete loss of ADPase activity (Nagahama *et al.*, 2000). Ser-345 in CDTa occupies the equivalent position of Ser-338 in Ia. In CDTa Ser-345 is situated very close to the active site cleft. Based on the structural observation it is clear that (as in Ia) the replacement of Ser-345 with a larger residue would abrogate substrate binding by not allowing the ADP-ribose donor to sit into the cleft properly.

Furthermore, in all CDTa structures, Ser-345 and Glu-387 sit in close proximity to each other and form a strong hydrogen bond (2.4-2.7Å). Ser-345 makes a direct hydrogen bond with both NAD/NADPH in their respective complex structures. This is a significant difference observed based on the structural data from Ia where Glu-380 makes direct interaction with the ligand rather than Ser-338 (Tsuge *et al.*, 2003). However, in Ia-NAD structure, Ser-338 of the STS motif is also positioned at a hydrogen bonding distance from the NAD molecule (Figure 5.19) but its implications have not been discussed by the authors (Tsuge *et al.*, 2003). Based on these structural results and in the light of results from the study of Nagahama and co-workers (Nagahama *et al.*, 2000) on Ia, it is tempting to suggest that Ser-345 in CDTa appears to have a crucial role in ligand binding and perhaps in catalysis as speculated by Tsuge *et al.* (Tsuge *et al.*, 2003).

Effect of Ligand Binding on ARTT loop Stability

A crucial difference between the ligand binding pattern of Ia and CDTa is the involvement of Ser-345 of CDTa in ligand binding. In CDTa, OG atom of Ser-345 makes a hydrogen bond interaction with O2D atom

of NAD/ NADPH (Table 5.7) whereas, in Ia, Glu-380 interacts with the same atom of the ligand (Figure 5.14).

Ser-345 and Glu-387 of CDTa (Ser-338 and Glu-380 of Ia) are positioned at close proximity with a strong hydrogen bond between them in all five CDTa structures (Figure 5.20). Thus the side chain of Glu-387 of ARTT loop is held from one end by Ser-345 in all native as well as ligand bound CDTa structures (Figure 5.20). However, in this situation, the side chain of Glu-387 is still free to move in the ligand binding cleft in a hinge-like motion in all native structures. This freedom is perhaps translated throughout the loop exhibiting the observed flexibility in the loop region (Figure 5.16 A) in the absence of ligand.

On ligand binding, Try-382, a conserved aromatic residue at the centre of the ARTT loop, flips towards the catalytic cleft to form a hydrogen bond with Glu-387 from the other side of its side chain (Figure 5.20). Fixing the side chain of Glu-387 from both sides restricts its movement in the cleft which otherwise could have abrogated the ligand binding. This restricted movement of Glu-387 could be the possible reason for the improved stability in the ARTT loop region upon ligand binding (Figure 5.16 B).

Other Important Residues

Glu-301, Tyr-246, Asn-255 and Phe-349 have also been suggested to play an important role in the enzymatic activity of Ia (Tsuge *et al.*, 2003). In CDTa, Glu-308 (Glu-301 in Ia) does not seem to participate in ligand binding directly but stays close to Arg-302 (Arg-295 in Ia) which interacts with the ligand directly (Figure 5.15). Replacement of Glu-301 to Ala in Ia resulted in the complete loss of NADase and ARTase activity of enzyme (Nagahama *et al.*, 2000).

Our structural analysis shows that Glu-308 holds Arg-302 in position by hydrogen bonding to form an optimal interaction with the ligand (Figure 5.20). A similar role can be attributed for residues Tyr-253 and Asn-262 (Tyr-246 and Asn-255 in Ia). Tyr-253 forms a hydrogen bond with Asn-262 and thus restricts its movement. Asn-262 further restricts the movement of Asn-342 (Asn-335 in Ia) and places it optimally for interaction with NAD/NADPH (Figures 5.15 and 5.20).

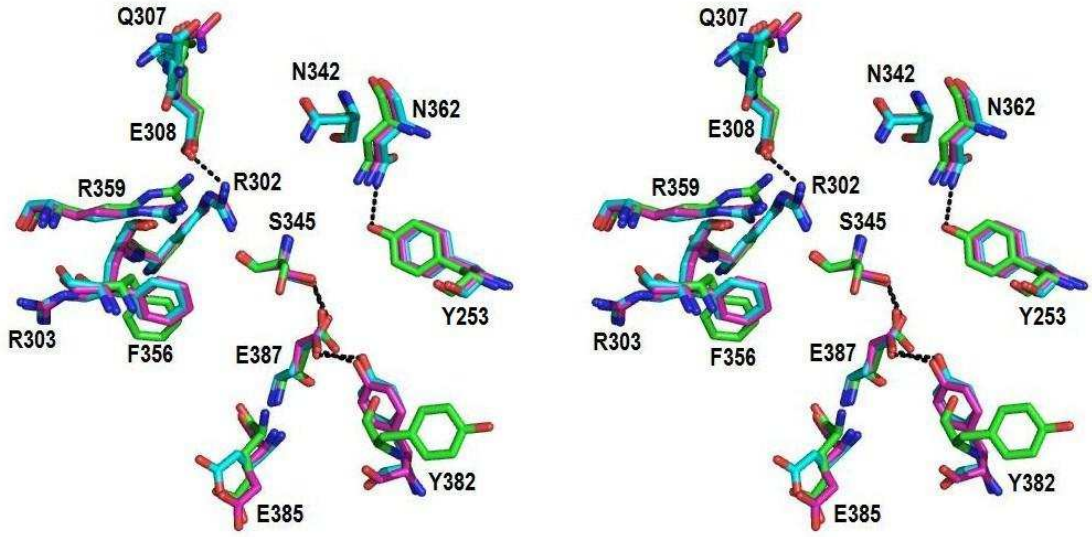

Figure 5.20: *The arrangement of residues in the CDTa catalytic cleft. Green – CDTa-9.0, Cyan – CDTa, Magenta – CDTa-NADPH. The residues numbering is according to CDTa.*

Phe-356 (Phe-349 in Ia) adopts similar orientation in all five CDTa structures. The side-chain of Phe-356 is relatively mobile in the three native structures. However, in the ligand bound structures its orientation is rearranged and provides stacking interactions against the nicotinamide ring of the ligand thus preventing its (nicotinamide ring's) rotation in the plane. This fixed rotation of nicotinamide ring is further stabilised by Arg-303 through hydrogen bonding similar to the observations made with Ia (Figure 5.15). This network of interactions facilitates tight binding of the ligand at the active site.

pH Induced Catalytic Site Flexibility

In order to understand the active site flexibility in CDTa, the native structures were determined at three different pH levels- 4.0, 8.5 and 9.0. It is suggested that the highly acidic pH of the endosomal compartment (~4.0) induces a drastic conformational change in CDTa which facilitates its translocation into the cytosol through the heptameric CDTb pore (Barth *et al.* 2000; Simpson, 1989). It was thought that the crystal structure at low and high pH levels (CDTa-9.0 and CDTa-4.0) under the identical conditions of crystal growth would help in analysing the pH induced conformational changes within the protein.

All native CDTa structures superimpose well with an rmsd of 0.63Å (Table 5.8). However, clear 'conformational flexibility' was observed among these structures in the active site (i. e., functionally important part). This was confined to the ARTT loop between strands β13 and β14 (Figure 5.17) and the loop between strand β9 and helix α10 named 'loop 304' (Figure 5.21).

Table 5.8: *The structural comparison of all different CDTa structures.*

Protein/Protein	CDTa-8.5	CDTa-4.0	CDTa-9.0	CDTa-NAD
CDTa-8.5	--	--	--	--
CDTa-4.0	0.63 (386)	--	--	--
CDTa-9.0	0.35 (389)	0.54 (389)	--	--
CDTa-NAD	0.71 (391)	0.85 (387)	0.65 (390)	--
CDTa-NADPH	0.62 (391)	0.88 (389)	0.74 (394)	0.37 (392)

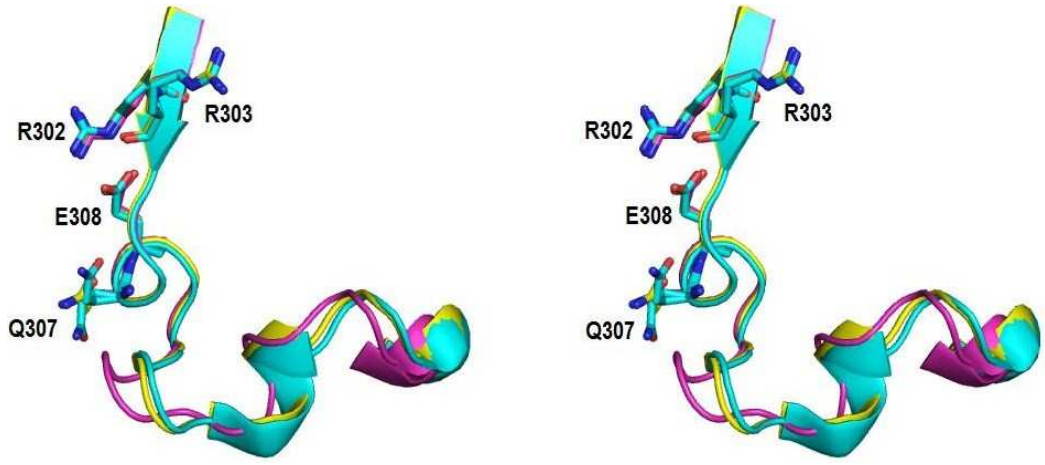

Figure 5.21: *The orientation of loop 304 which shows differences between CDTa-4.0 and other CDTa structures. Yellow – CDTa-9.0, Magenta – CDTa-4.0, Cyan – CDTa-NAD.*

This flexibility was more pronounced in the CDTa-4.0 structure and was consistent with the analysis of crystallographic temperature factors (Table 5.9) which provides an opportunity to obtain a relatively unbiased picture of the mobility of different parts of the structure. Indeed, these regions adopt a more stable structure at a higher pH (e. g. 8.5 and 9.0) and NAD/NADPH complex structures of CDTa. Although this region is clearly influenced by the conditions required to obtain crystals (which are identical for the CDTa-4.0 and CDTa-9.0 structures except for the pH

of the crystallisation buffers) the innate flexibility may be important in the translocation of the enzymatic component (CDTa) of the toxin into the cytosol via receptor mediated early endosomal pathway.

Table 5.9: *Average B factors for the two flexible loop regions in different CDTa structures.*

Region (residue number)	**Average B factor**				
	CDTa-8.5	**CDTa-9.0**	**CDTa-4.0**	**CDTa-NAD**	**CDTa-NADPH**
Whole protein	19.39	25.35	34.15	32.35	30.99
ARTT loop* (377-387) [Atoms]	25.88 [74]	30.56 [68]	40.37 [60]	40.53 [81]	40.49 [81]
Loop 304 (304-325) [Atoms]	21.05 [175]	34.02 [175]	41.61 [171]	34.41 [175]	35.77 [175]

** The modelled length of ARTT loop varies in different structures*

However, no appreciable conformational changes could be observed as a result of pH change in different structures and all three CDTa structures- CDTa-4.0, CDTa-8.5 and CDTa-9.0 superimpose well. Similar studies with the enzymatic component of *C. botulinum* C2 toxin (C2I) also did not show any such pH induced conformational changes (Schleberger *et al.*, 2006). It is possible that these changes take place at acidic pH and it is quite likely that the presence of the translocation partner may be required to facilitate these conformational changes.

Mechanistic Implications

Currently available structural and biochemical data on ADPRTs, i. e., the conservation of catalytic site apparatus and NAD binding suggest a common catalytic mechanism based on nuclephilic substitution (S_N) – either an S_N1 or S_N2 type reaction (Tsuge *et al.*, 2003; Tsuge *et al.*, 2008).

A nucleophilic substitution reaction involves an electron pair donor (the nucleophile, Nu) with an electron pair acceptor (the electrophile) where a sp^3-hybridised electrophile must have a leaving group (X) in order for the reaction to take place. The nucleophilic substitution reactions can proceed via two mechanisms –

An S_N1 (substitution nucleophilic order 1) reaction is a first order chemical reaction where the attack by the nucleophile and the departure of the leaving group occurs in two separate steps. An S_N1 reaction proceeds via formation of a planar carbenium ion in the first step, which is then, in second step, attacked by the nucleophile (Figure 5.22 A). The rate limiting step in an S_N1 reaction is the first step i. e. the formation of carbenium ion. A higher stability of carbenium ion is a favourable condition for a reaction to take place via the S_N1 mechanism.

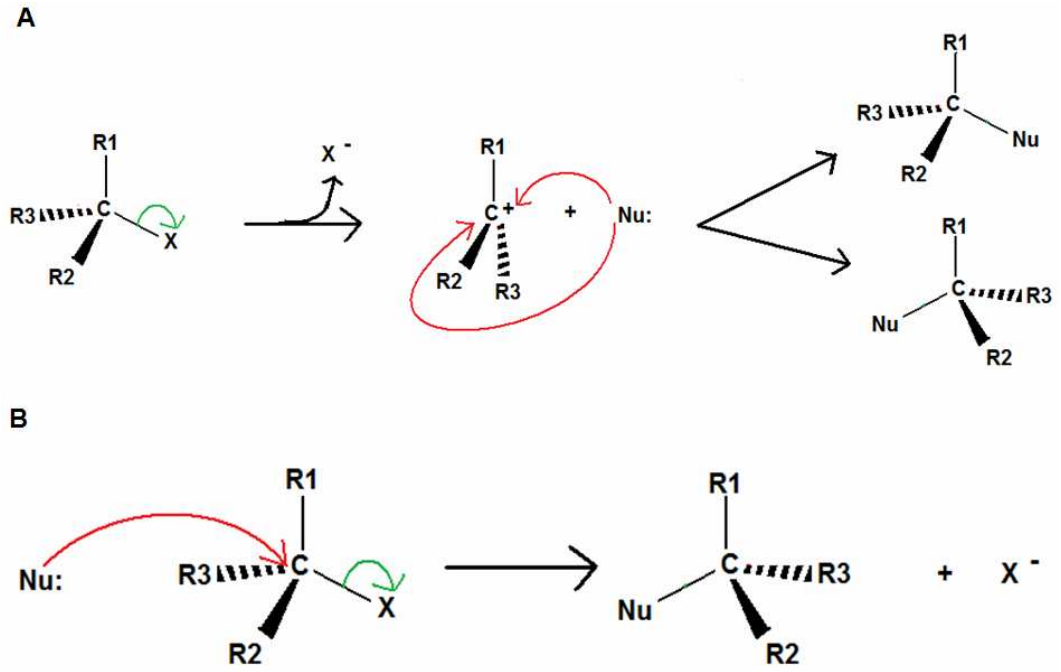

Figure 5.22: *A schematic representation of the progression of an S_N1 reaction (A) and an S_N2 reaction (B).*

On the other hand, an S_N2 (substitution nucleophilic order 2) reaction is a second order reaction where the departure of the leaving group (formation of carbenium ion) takes place simultaneously with the backside attack by the nucleophile (Figure 5.22 B) and hence the reaction completes in one step. The rate of the S_N2 reaction is determined by the ease of simultaneous nucleophilic attack and the departure of the leaving group. However, these are not the only factors determining the rate of the reaction.

The S_N2 reaction mechanism in class 4 ADPRTs has been proposed based on the structural analysis of VIP2 (Han et al., 1999). In the case of Ia, progression of the S_N2 reaction has been postulated via

two possible ways. In the first hypothesis, guanidium group of Arg-177 of actin has been suggested to act as the nucleophile following its deprotonation by Glu-378 of the toxin.

However, in the structure of Ia-Actin complex, it has been shown that Arg-177 of actin is positioned at a considerably long distance (7.0 Å) from Glu-378 of the toxin or the nicotinamide ring of NAD (Tsuge *et al.*, 2008). This eliminates the possibility of both – deprotonation of Arg-177 by Glu-378 and a direct nucleophilic attack on ADP-ribose+ oxocarbenium ion by Arg-177 (Figure 5.23). Formation of the ADP-ribose+ oxocarbenium ion has been suggested to be a spontaneous process driven by the specific highly folded conformation of NAD (Figure 5. 4 B) in the catalytic cleft (Tsuge *et al.,* 2003; Tsuge *et al.,* 2008).

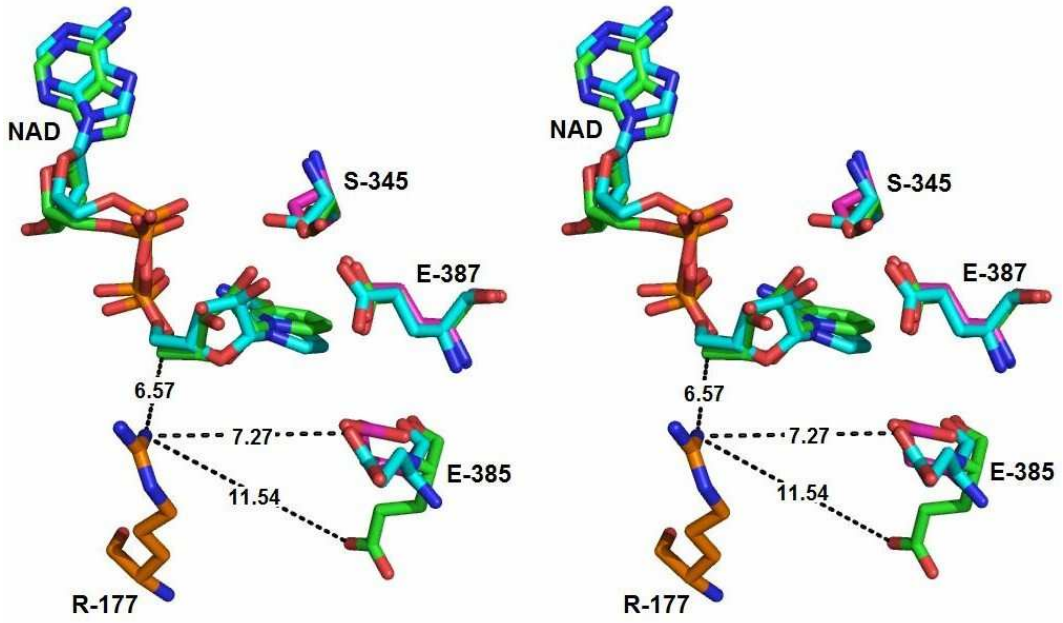

Figure 5.23: *A stereo representation of distances between the catalytic centre (C1D of NAD) and deprotonting Glu (Glu-385 in CDTa, Glu-378 in Ia). Green – CDTa-NAD, Cyan – Ia-NAD, Magenta – Ia of Ia-Actin, Orange – Actin of Ia-Actin. All distances shown are in Å units.*

Superimposition of CDTa-NAD and Ia-NAD complexes on Ia-Actin complex reveals that the toxin (Ia) does not undergo any large conformational change upon actin binding. It indicates that in the case of CDTa also, the S_N2 reaction via direct nucleophilic attack by Arg-177 of actin would not be possible (Figure 5.23).

Alternatively, for Ia, it was suggested that a water molecule that was present near the NMN (Nicotinamide mono nucleotide) ring (~4.0 Å)

could be a possible nucleophile. However, this water molecule could be modelled only in one of the two molecules in the asymmetric unit with a high temperature factor (Tsuge et al., 2003). These findings rule out its role in mediating the S_N2 reaction for Ia.

In CDTa, in complex with either NAD or NADPH, there are at least two water molecules with reasonably low temperature factors near the nicotinamide ring. One of the water molecules seems to be important as it bridges NAD, Ser-345 and Tyr-253 in the complex (Figure 5.24). However, this water molecule which is closest to the reaction centre (C1D of NAD/NADPH) is present at a considerably large distance of 5.45 Å in the CDTa-NAD (Figure 5.24) and at 4.25 Å in the CDTa-NADPH complex structure. These observations make the S_N2 mechanism of catalysis less preferred in the case of CDTa.

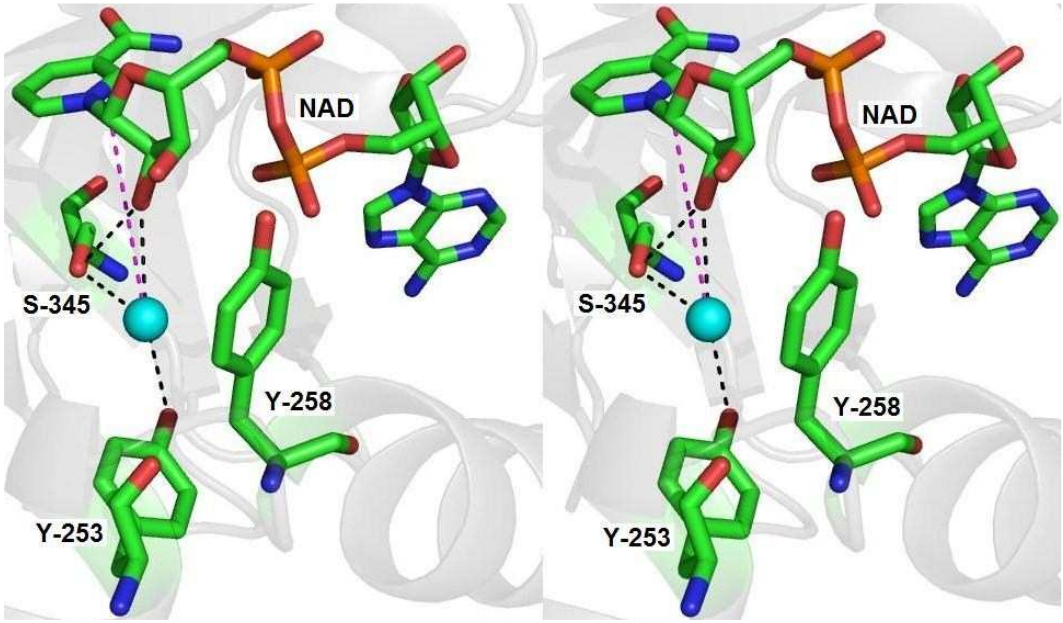

Figure 5.24: *The position of nearest water molecule in the catalytic centre and hydrogen bonding network. The bound NAD is shown in green colour. The water molecule is shown as sphere. Hydrogen bonds are shown using black broken lines. Distance of the catalytic centre (C1D of NAD) from the water molecule (5.50 Å) is shown using red broken line.*

For the ADP-ribosylation reaction to proceed via an S_N1 mechanism, the formed oxocarbenium ion (ADP-ribose+) must be highly stable. In the case of Actin-ADPRTs, the S_N1 reaction mechanism would involve formation of an isolated positively charged oxocarbenium intermediate with the direct stabilising electrostatic interactions from the negatively

charged carboxylate group of catalytic glutamate (Glu-380 in Ia) or hydroxyl group of serine (Ser-345 in CDTa).

In Ia, the S_N1 reaction mechanism has been proposed via two reaction intermediates where rotation of the primary oxocarbanium ion, resulting in the formation of a secondary cation has been suggested (Tsuge *et al.*, 2008). In Ia-actin complex structure, loop II of Ia (between α7 and α8) undergoes significant conformational changes. As a result, Gly-249 of the loop seems to interact directly with Arg-177 (acceptor residue) of actin. These changes in the loop rearrange Tyr-246 and Tyr-251 in Ia (Tyr-253 and Tyr-258 in CDTa) also. Tyr-251 in Ia is suggested to play a role in transferring the rotated positively charged ADP-ribose intermediate cation to the substrate (Tsuge *et al.*, 2008). Previous mutational studies on both of these residues (Tyr-246 and Tyr-251) in Ia have been shown to have adverse effects on NADase as well as ARTase activity of the protein (Tsuge *et al.*, 2003).

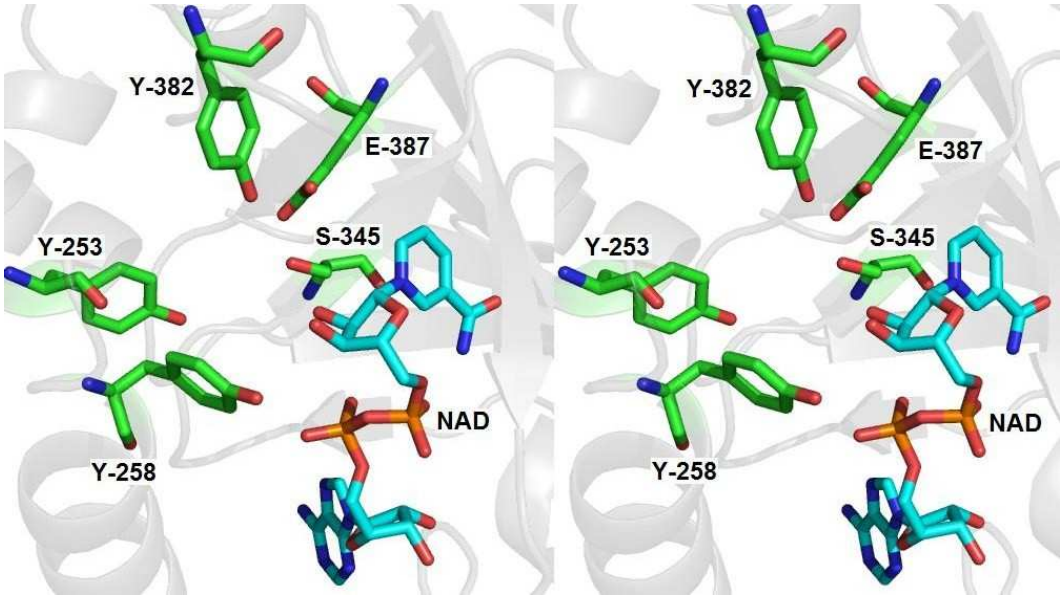

Figure 5.25: *A stereo representation of negatively charged residues surrounding the catalytic centre (C1D) of NAD. These residues probably contribute towards the stability of formed oxocarbenium ion.*

In CDTa, a similar S_N1 mechanism could be followed. Based on our structural data it is clear that Ser-345 interacts with both of the ligands (NAD or NADPH) directly which is further surrounded by Glu-387, Tyr-253, Tyr-258 and Tyr-382. The negatively charged environment created by these residues could play a crucial role in stabilising the formed

positively charged oxocarbenium ion, which is a favourable condition for an S_N1 reaction to take place (Figure 5.25).

We propose that in CDTa it is Ser-345 that stabilises the oxocarbenium ion (Figure 5.26, step A) by direct interactions and facilitates its transfer to Tyr-258 following its rotation (Figure 5.26, step B and C) in a similar way as it has been proposed for Ia (Tsuge *et al.*, 2008).

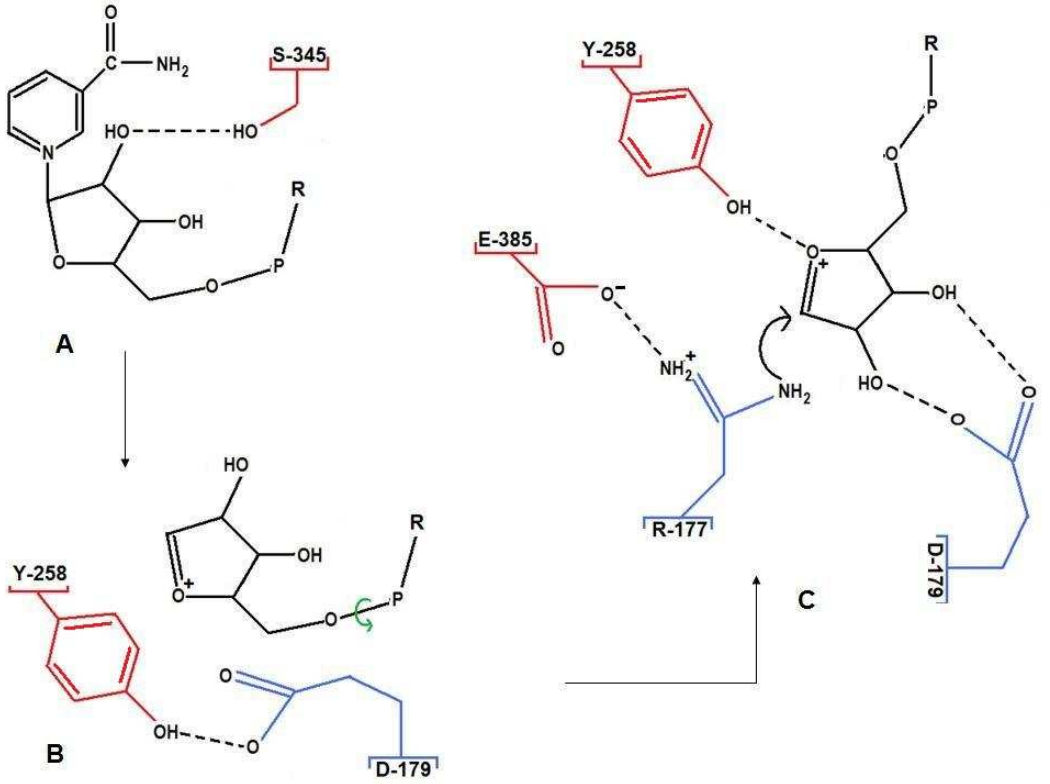

Figure 5.26: *The proposed S_N1 mechanism of ADP-ribosylation of actin by CDTa (Adopted from Tsuge et al., 2008). Colour coding – Black-NAD, Red-CDTa, Blue-Actin.*

Suggested rotation of the primary oxocarbenium ion (Figure 5.26 step B to step C) overcomes two difficulties. Firstly, NAD binds in the catalytic cleft in a highly compact conformation which is a high energy state. By rotation around the P-O bond, the formed primary oxocabanium ion moves to a relaxed, low energy state and thus becomes more stable. Secondly, rotation of the primary oxocarbenium ion would bring it closer to the surrounding negatively charged residues (Figure 5.25) and thus the stability of the secondary cation would be enhanced resulting in the S_N1 mechanism favouring conditions.

In Ia, an S_N1 mechanism is further proposed to be progressed via rearrangement in Arg-177 of actin (Tsuge *et al.*, 2008). This rearrangement in actin would bring Arg-177 of actin very near to Glu-378 (another conserved Glu of EXE motif) of Ia (Figures 5.26 and 5.27) (Tsuge *et al.*, 2008). Glu-378 thus participates in the ADP-ribose transfer reaction by deprotonating the guanidium group of Arg-177. In addition to that, Glu-378 holds Arg-177 of actin in the catalytic centre.

When compared with Ia, Glu-385 of CDTa (equivalent to Glu-378 in Ia), adopts a different orientation and sticks away from the catalytic cleft (Figures 5.18 and 5.27). In this orientation, rearrangement of Arg-177 of actin would still leave both of the residues at a considerable distance of about 7.0 Å from each other (Figure 5.27). How this different orientation of Glu-385 still mediates the catalysis is an open question to investigate.

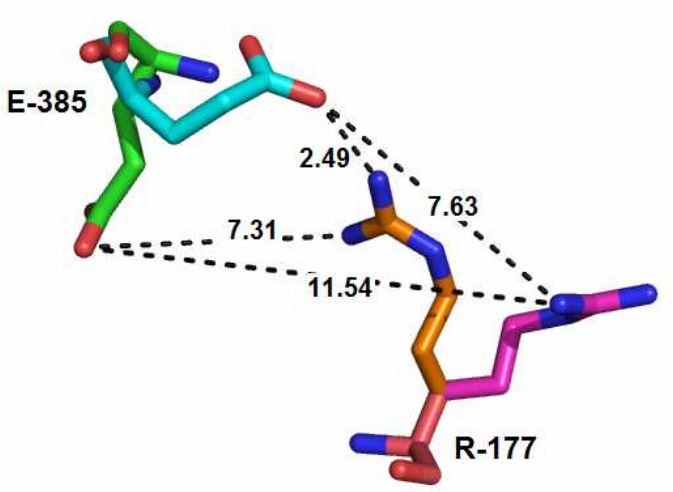

Figure 5.27: *The representation of distances between catalytic Glu-385 (378) of CDTa (Ia) and Arg-177 of Actin before and after the proposed rearrangement of Arg-177 (Tsuge et al., 2008). The figure was produced by superimposing CDTA-NAD structure on Ia-Actin complex structure. Cyan – Ia (Gul-378), Green – CDTa (Gul-385), Magenta – Actin (Arg-177) before rearrangement), Orange – Actin (Arg-177) after rearrangement. Distances (in Å units) are shown using broken lines.*

Our structural data shows that the ARTT loop is not directly involved in the ligand binding in any of the complex CDTa structures (Figures 5.14 and 5.15), and is free to rearrange itself further. It is tempting to suggest based on two different complex crystal structures (CDTa-NAD and CDTa-NADPH) that once the donor substrate (NAD/ NADPH) is cleaved followed by the formation of oxocarbenium ion, further

rearrangement of the ARTT loop can not be ruled out considering its high flexibility.

The presence of a large open cavity near the active site cleft as observed in the Ia-actin complex (Tsuge *et al.*, 2008) also supports the hypothesis of ARTT loop rearrangements upon actin binding. This rearrangement in the ARTT loop would bring Glu-385 of CDTa into the reaction centre to proceed with the transfer of ADP-ribose moiety to Arg-177 of actin from Ser-345 via Tyr-258 (Figure 5.22, step C). However, this hypothesis needs to be validated by direct structural evidence of CDTa in complex with actin, in combination with functional dissection of key residues by site-directed mutagenesis.

SUMMARY

CDTa and Ia belong to the actin-ADPRT family that irreversibly modify monomeric actin molecules by transferring the ADP ribose moiety of NAD/NADPH to Arg-177 of actin. Based on our structural data, despite high homology at primary sequence, structural as well as functional level, the mode of donor substrate recognition in Ia and CDTa appears to be different.

The enzymatic components of Actin-ADPRTs have been suggested to undergo a low pH induced conformational changes during the process of translocation from the early endosome to the cytosol. The observed conformational flexibility and enhanced level of disorder in two of the catalytically important loop regions of CDTa at low pH state provide preliminary evidence for these conformational changes. However, to understand the exact mechanism of translocation of CDTa as well as the transfer of ADP-ribose to actin by CDTa, additional input in terms of mutational studies and structures (such as CDTa-actin complex) are required.

DIRECTIONS OF FUTURE WORK

Understanding of *C. difficile* binary toxin (CDT) is still in the initial stage. This thesis is a step towards the structural, functional and biological characterisation of *C. difficile* binary toxin.

In this thesis, we report cloning, expression and purification methods for both of the components (enzymatic as well as transport) of CDT. Purification methods described (Chapter III) resulted in milligram quantities of proteins of high purity. The cell cytotoxicity effect of CDT were shown on Vero cells (Chapter IV). Various combinations of CDTa and CDTb concentrations were tested including two different versions of recombinantly expressed CDTb (named as CDTb′ and CDTb″) and their chymotrypsin activated fragments. It is clear from the results that both of the purified components are active and the complete CDT has a definite cell killing potential. However, there are at least two questions yet to be answered.

1- Variation in the concentrations of CDTa or CDTb did not yield in any observable changes in the number of dead cells. It is still not clear whether the concentration of CDTa or the chymotrypsin activated CDTb is the key step in the process of cell death by CDT.
2- Recombinantly expressed mature fragment of CDTb (CDTb″) resulted in relatively poor cell death (in combination of CDTa) when compared with its chemotrypsin activated fragment. The length of CDTb″ during the expression was decided based on a report by Perelle and co-workers (1997). Our experimental data does not reveal why CDTb″ is less active. It may possible that the N terminal part of active CDTb is important for its function. Furthermore, CDTb″ was expressed as GST-CDTb″ fusion protein and cleavage of the tag would still leave 4 to 5 undesired residues from the PreScission protease recognition sequence at the N terminal of the mature protein. Could these

residues interfere with the activity of the protein, bearing in mind that chymotrypsin activation of CDTb″ improves the number of dead cells significantly (Chapter IV). Another possibility could be that the activation site predicted by Perelle and co-workers may not be precise and we have, in reality, expressed a larger fragment than the required. This issue can however be resolved by the N terminal sequencing of CDTb″ and chymotrypsin activated CDTb′ and then aligning both of the sequences against each other. The expressed CDTb″ can not be shorter than the required mature fragment due to the fact that chymotrypsin activation of CDTb″ improves cell death count.

High resolution crystal structure of CDTa has been determined in its native from at low and high pH states as well as in ligand bound forms (Chapter V). The CDTa structure shows crucial differences in the donor substrate recognition pattern when compared with the closest homologue i. e. the enzymatic component of iota toxin (Ia) from *C. perfringens*. In CDTa, the crystallographic data suggests that it is Ser-345 and not Glu-387 that plays a key role in the protein-ligand complex stabilisation. On the other hand, in Ia, the analysis of crystallographic data (Tsuge et al., 2003) indicates that Ser-338 and Glu-380 may play interchangeable roles in the protein-ligand complex stabilisation as both of these residues seem to interact directly with ligand. However, the authors have not discussed the binding of Ser-338 with NAD.

In CDTa, mutational studies are required to assign definitive functional roles to Ser-345 or Glu-387 in the ligand binding. Several sets of primers for point mutations in CDTa have been designed for this purpose (S345A, S345R, S345Y, S345F, E387A, E387R, E387F and E387D). Positive clones for S345A and S345F

have been constructed successfully. These primers can be used to produce double mutants as well (such as S345A/E387A) which would be advantageous to study the interchangeable role of S-345 and E-387 in the ligand binding.

NAD and NADPH are the donor substrates for CDTa and other similar toxins. The ADP-ribose part of NAD/NADPH is transferred to monomeric Actin by the action of these ADPRTs. The crystal structure of Ia in complex with actin at 2.8 Å has been reported by Tsuge and co-workers in 2008. Ia does not seem to undergo any significant conformational changes except in one of the loops which brings G-249 of Ia at a hydrogen bonding distance from R-177 of actin. It has been postulated that E-378 residue in the ARTT loop of Ia mediates transfer of the ADP-ribose to R-177 of actin. The corresponding residue in CDTa is E-385. However, when compared, we see that the side chain of E-385 of CDTa adopts entirely different orientation and points away from the catalytic cleft unlike in Ia (Chapter V). This difference in orientation leaves E-385 of CDTa at a longer distance from R-177 of actin when superimposed on the Ia-Actin structure. Though, owing to the flexibility in the ARTT loop, rearrangement/s in the loop can not be ruled out.

The crystal structure of CDTa in complex with actin could provide a definitive answer regarding how this side chain orientation of E-385 still carries out an identical function in CDTa. Preliminary experiments towards achieving this goal are in progress. In addition, site directed mutagenesis studies of E-385 could throw some light on this issue. Different sets of primers (E385A, E385R, E385F and E385D) have been designed for this purpose.

CDTb, like the transport components of other binary toxins is believed to form a homo-heptameric complex upon activation by chymotrypsin. The crystal structure of the transport component of

C2 toxin from *C. botulinum* in monomeric form (Schleberger et al., 2006) has been reported. However, at present, structural information about any of the Clostridial binary toxin transport components in the heptameric form is not available. In this regard, a well established protocol for the expression and purification of CDTb has been developed. Preliminary crystallisation hits for CDTb′ have produced crystals which are currently being optimised.

In the long term, it would be exciting to be able to characterise the CDTb homo-heptameric complex alone and the CDTb homo-heptamer in complex with the bound CDTa. Till date there is no information available about the amino acid residues of CDTb which play a role in the heptamer formation as well as about the residues of CDTb and CDTa which facilitate the binding of CDTa to the CDTb heptamer. However, when the chymotrypsin activated CDTb′ was incubated overnight at 4^0C at low pH condition, a faint but clearly visible protein band corresponding to the oligomeric form was visible on the SDS-PAGE (Chapter IV). Further, the structure of CDTa-CDTb complex could be useful to understand the pH induced conformational changes in CDTa which have been considered important for the translocation of CDTa from the endosome to the cytosol.

Further research on all thesis topics at the molecular level will be of great academic, therapeutic as well as industrial interest towards the development of treatments against *C. difficile* infection. Answers to these questions will enhance our understanding of *C. diffiicile* binary toxin which will be helpful in the elucidation of general principles in protein-protein recognition involving similar binary toxins such as *C. perfringens* iota toxin and *C. botulinum* C2 toxin.

BIBLIOGRAPHY

1. Aktories K. & Wegner A. (1989). ADP-ribosylation of actin by Clostridial toxins. *J. Cell. Biol.* **109,** 1385-1387.

2. Aktories K., Barmann M., Ohishi I., Tsuyama S., Jakobs K. H. & Habermann E. (1986). Botulinum C2 toxin ADPribosylates actin. *Nature* **322,** 390-392.

3. Aktories K., Weller U. & Chhatwal G. S. (1987). Clostridium botulinum type C produces a novel ADP-ribosyltransferase distinct from botulinum C2 toxin. *FEBS Lett.* **212,** 109-113.

4. Aktories, K. & Wegner K. (1992). Mechanisms of the cytopathic action of actin-ADP-ribosylating toxins. *Mol. Microbiol.* **6,** 2905-2908.

5. Allured V. S., Collier R. J., Carroll S. F. & McKay D. B. (1986). Structure of exotoxin A of Pseudomonas aeruginosa at 3.0 Angstrom resolution. *Proc. Natl. Acad. Sci. USA* **83,** 1320-1324.

6. Barth H., Aktories K., Popoff M. R. & Stiles B. G. (2004). Binary bacterial toxins: biochemistry, biology, and applications of common *Clostridium* and *Bacillus* proteins. *Microbiol. Mol. Biol. Rev.* **68,** 373-402.

7. Barth H., Blocker D., & Aktories K. (2002). The uptake machinery of Clostridial actin ADP-ribosylating toxins--a cell delivery system for fusion proteins and polypeptide drugs. *Naunyn. Schmiedebergs Arch. Pharmacol.* **366,** 501-512.

8. Barth H., Blocker D., Behlke J., Bergsma-Schutter W., Brisson A., Benz R. & Aktories K. (2000). Cellular uptake of *Clostridium botulinum* C2 toxin requires oligomerization and acidification. *J. Biol. Chem.* **275,** 18704-18711.

9. Barth H., Pfeifer G., Hofmann F., Maier E., Benz R. & Aktories K. (2001). Low pH-induced formation of ion channels by *Clostridium difficile* toxin B in target cells. J. *Biol. Chem.* **276,** 10670-10676.

10. Bartlett, J. G. & Perl T. M. (2005). The new *Clostridium difficile*--what does it mean? *N. Engl. J. Med.* **353,** 2503-2505.

11. Bell C. E. & Eisenberg D. (1996). Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry* **35,** 1137-1149.

12. Blocker D., Behlke J., Aktories K. & Barth H. (2001). Cellular uptake of the binary *Clostridium perfringens* iota-toxin. *Infect. Immun.* **69,** 2980-2987.

13. Blow D. (2002). Outline of crystallography for biologists. *Oxford University Press.*

14. Blundell T. L. & Johnson L. N. (1976). Protein crystallography. *Academic Press, London.*

15. Brünger A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472-475.

16. Brünger A. T., Adams P. D., Clore G. M., DeLano W. L., Gros P., Grosse-Kunstleve R. W., Jiang J.-S., Kuszewski J., Nilges M., Pannu N. S., Read R. J., Rice L. M., Simonson T. & Warren G. L. (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Cryst.* **D54,** 905-921**.**

17. Burnette W. N., Cieplak W., Mar V. L., Kaljot K. T., Sato H. & Keith J. M. (1988). Pertussis toxin S1 mutant with reduced enzyme activity and a conserved protective epitope. *Science* **242,** 72-74.

18. Burnette W. N., Mar V. L., Platler B. W., Schlotterbeck J. D., McGinley M. D., Stoney K. S., Rohde M. F. & Kaslow H. R. (1991). Site-specific mutagenesis of the catalytic subunit of cholera toxin: substituting lysine for arginine 7 causes loss of activity. *Infect. Immun.* **59,** 4266-4270.

19. Carroll S. F. & Collier R. J. (1984). NAD binding site of diphtheria toxin: identification of a residue within the nicotinamide subsite by photochemical modification with NAD. *Proc. Natl. Acad. Sci. USA* **81,** 3307-3311.

20. Carroll S. F. & Collier R. J. (1987). Active site of Pseudomonas aeruginosa exotoxin A. Glutamic acid 553 is photolabeled by NAD and shows functional homology with glutamic acid 148 of diphtheria toxin. *J. Biol. Chem.* **262,** 8707-8711.

21. Carroll S. F. & Collier R. J. (1988). Amino acid sequence homology between the enzymic domains of diphtheria toxin and Pseudomonas aeruginosa exotoxin A. *Mol. Microbiol.* **2,** 293-296.

22. CCP4. (1994). The CCP4 suite: Programs for protein crystallography. *Acta Cryst.* **D50,** 760-763.

23. Chardin P., Boquet P., Madaule P., Popoff M. R., Rubin E. J. & Gill D. M. (1989). The mammalian G protein rhoC is ADP-ribosylated by Clostridium botulinum exoenzyme C3 and affects actin microfilaments in Vero cells. *EMBO J.* **8,** 1087-1092.

24. Cieplak W., Burnette W. N., Mar V. L., Kaljot K. T., Morris C. F., Chen K. K., Sato H. & Keith J. M. (1988). Identification of a region in the S1 subunit of pertussis toxin that is required for enzymatic activity and that contributes to the formation of a neutralizing antigenic determinant. *Proc. Natl. Acad. Sci. USA* **85,** 4667-4671.

25. Collier R. J. & Cole H. A. (1969). Diphtheria toxin subunit active in vitro. *Science* 164, 1179-1181.

26. Collier R. J. (1975). Diphtheria toxin: mode of action and structure. *Bacteriol. Rev.* **39,** 54-85.

27. Considine R. V. & Simpson L. L. (1991). Cellular and molecular actions of binary toxins possessing ADPribosyltransferase activity. *Toxicon* **29,** 913-936.

28. Dauter, Z. (1999). Data collection strategies. *Acta cryst.* **D55,** 1703-1717

29. Davis I. W., Murray L. W., Richardson J. S. & Richardson D. C. (2007). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32,** 615-619.

30. Deng, Q. & Barbieri, J. T. (2008). Molecular Mechanisms of the Cytotoxicity of ADP-Ribosylating Toxins. *Annu Rev Microbiol.* **62,** 271-288.

31. Domenighini M. & Rappuoli R. (1996). Three conserved consensus sequences identify the NAD-binding site of ADP-ribosylating enzymes, expressed by eukaryotes, bacteria and T-even bacteriophages. *Mol. Microbiol.* **21,** 667-674.

32. Dove C. H., Wang S. Z., Price S. B., Phelps C. J., Lyerly D. M., Wilkins T. D. & Johnson J. L. (1990). Molecular characterization of the Clostridium difficile toxin A gene. *Infect. Immun.* **58,** 480-488.

33. Dupuy, B. & Sonenshein A. L. (1998). Regulated transcription of *Clostridium difficile* toxin genes. *Mol. Microbiol.* **27,** 107-120.

34. Eckhardt M., Barth H., Blocker D. & Aktories K. (2000). Binding of *Clostridium botulinum* C2 toxin to asparagine-linked complex and hybrid carbohydrates. *J. Biol. Chem.* ***275,*** 2328-2334.

35. Egerer M., Giesemann T., Jank T., Satchell K. J. & Aktories K. (2007). Auto-catalytic cleavage of Clostridium difficile toxins A and B depends on cysteine protease activity. *J Biol Chem.* **282,** 25314-25321.

36. Elliott B., Chang B. J., GolledgeC. L. & Riley T. V. (2007). Clostridium difficile-associated diarrhoea. *Internal Med. J.* **37,** 561-568.

37. Emsley P. & Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Cryst.* **D60,** 2126-2132.

38. Evans H. R., Sutton J. M., Holloway D. E., Ayriss J., Shone C. C. & Acharya K. R. (2003). The crystal structure of C3stau2 from Staphylococcus aureus and its complex with NAD. *J. Biol. Chem.* **278,** 45924-45930.

39. Evans P. R. (1999). Some notes on choices in data collection. *Acta Cryst.* **D55,** 1771-1772.

40. Faust C., Ye B. & Song K.-P. (1998). The enzymatic domain of *Clostridium difficile* toxin A is located within its N-terminal region. *Biochem. Biophys. Res. Commun.* **251,** 100–105.

41. Fernie D. S., Knights J. M., Thomson R. O. & Carman R. J. (1984). Rabbit enterotoxaemia: purification and preliminary characterization of a toxin produced by *Clostridium spiroforme*. *FEMS Microbiol. Lett.* **21,** 207–211.

42. Fernie D. S., Knights J. M., Thomson R. O. & Carman R. J. (1984). Rabbit enterotoxinaemia: purification and preliminary characterization of a toxin produced by *Clostridium spiroform. FEMS Microbiol. Letters.* **21,** 207-211.

43. Finkelstein R. A., Burks M. F., Zupan A., Dallas W. S., Jacob C. O. & Ludwig D. S. (1987). Epitopes of the cholera family of enterotoxins. *Rev. Infect. Dis.* **9,** 544-561.

44. Florin, I. & Thelestam M. (1983). Internalization of *Clostridium difficile* cytotoxin into cultured human lung fibroblasts. *Biochem. Biophys. Acta.* **763,** 383-92.

45. Fox G. E., Stackebrandt E., Hespell R. B., Gibson J., Maniloff J., Dyer T. A., Wolfe R. S., Balch W. E., Tanner R. S., Magrum L. J., Zablen L. B., Blakemore R., Gupta R., Bonen L., Lewis B. J., Stahl D. A., Luehrsen K. R., Chen K. N. & Woese C. R. (1980). The phylogeny of prokaryotes. *Science* **209,** 457–463.

46. French G. S. and Wilson K. S. (1978). On the treatment of negative intensity observations. *Acta Cryst.* **A34,** 517-525.

47. Friedlander, A. M. (1986). Macrophages are sensitive to anthrax lethal toxin through an acid-dependent process. *J. Biol. Chem.* **261,** 1723-1726.

48. Frisch C., Gerhard R., Aktories K., Hofmann F. & Just I. (2003). The complete receptor-binding domain of *Clostridium difficile* toxin A is required for endocytosis. *Biochem. Biophys. Res. Commun.* **300,** 706-711.

49. Gerding D. N. (2004). Clindamycin, cephalosporins, fluoroquinolones, and Clostridium difficile-associated diarrhea: this is an antimicrobial resistance problem. *Clin. Infect. Dis.* **38,** 646–648.

50. Geric B., Johnson S., Gerding D. N., Grabnar M. & Rupnik M. (2003). Frequency of binary toxin genes among *Clostridium difficile* strains that do not produce large Clostridial toxins, J. of Clinic. Microbiol. **41,** 5227-5232.

51. Giesemann T., Egerer M., Jank T. & Aktories K. (2008). Processing of Clostridium difficile toxins. *J Med Microbiol.* **57,** 690-696.

52. Gill D. M., Clements J. D., Robertson D. C. & Finkelstein R. A. (1981). Subunit number and arrangement in Escherichia coli heat-labile enterotoxin. *Infect. Immun.* **33,** 677-682.

53. Gluke I., Pfeifer G., Liese J., Fritz M., Hofmann F., Aktories K. & Barth H. (2001). Characterization of the enzymetic component of the ADP ribosyltransferase toxin CDTa from *Clostridium difficile. Infect. Immun.* **69,** 6004-6011.

54. Govind R., Vediyappan G., Rolfe R. D. & Fralick J. A. (2006). Evidence that *Clostridium difficile* TcdC is a membrane-associated protein. *J. Bacteriol.* **188,** 3716-3720.

55. Greco A., Ho J. G., Lin S. J., Palcic M. M., Rupnik M. & Ng K. K. (2006). Carbohydrate recognition by Clostridium difficile toxin A. *Nat Struct Mol Biol.* **13,** 460-461.

56. Green D. W., Ingram V. M. & Perutz M. F. (1954). The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. *Proc. Roy. Sci.* **225,** 287-307.

57. Hachmann J. P. & Amshey J. W. (2005). Models of proteinmodification in Tris-glycine and neutral pH Bis-tris gels during electrophoresis: effects of gel pH. *Anal. Biochem.* **342,** 237-245.

58. Hall A. (1990). The cellular functions of small GTP-binding proteins. *Science* **249,** 635-640.

59. Hall I. C. & O'Toole E. (1935). Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, *Bacillus difficile*. *Amer. J. Dis. Child.* **49,** 390-402.

60. Hammond G. A., & Johnson J. L. (1995). The toxigenic element of *Clostridium difficile* strain VPI 10463. *Microb. Pathog.* **19,** 203-213.

61. Han S. & Tainer J. A. (2002). The ARTT motif and a unified structural understanding of substrate recognition in ADP-ribosylating bacterial toxins and eukaryotic ADPribosyltransferases. *Int. J. Medical. Microbiol.* **291,** 523-529.

62. Han S., Arvai A. S., Clancy S. B. & Tainer J. A. (2001). Crystal structure and novel recognition motif of rho ADPribosylating C3 exoenzyme from Clostridium botulinum: structural insights for recognition specificity and catalysis. *J. Mol. Biol.* **305,** 95-107.

63. Han S., Craig J. A., Putnam C. D., Carozzi N. B. & Tainer J. A. (1999). Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. *Nat. Struct. Biol.* **6,** 932-936.

64. Haug G., Aktories K. & Barth H. (2003). The host cell chaperone Hsp90 is necessary for cytotoxic action of the binary iota-like toxins. *Infect. Immun.* **72,** 3066-3068.

65. Haug G., Leemhuis J., Tiemann D., Meyer D. K., Aktories K. & Barth H. (2003). The host cell chaperone Hsp90 is essential for translocation of the binary *Clostridium botulinum* C2 toxin into the cytosol. J. Biol. Chem. **278,** 32266-32274.

66. Helliwell J. R. (1992). Macromolecular crystallography with synchrotron radiation. *Cambridge University Press, Cambridge, UK*.

67. Helliwell J. R. (1997). Overview of synchrotron radiation and macromolecular crystallography. *Methods Enzymol.* **276,** 203-217.

68. Hendrickson W. & Ogata C. (1997). Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **276,** 494–523.

69. Ho J. G., Greco A., Rupnik M. & Ng K. K. (2005). Crystal structure of receptor-binding C-terminal repeats from *Clostridium difficile* toxin A. *Proc. Natl. Acad. Sci. USA* **102,** 18373-18378.

70. Hofmann F., Busch C., Prepens U., Just I. & Aktories K. (1997). Localization of the glucosyltransferase activity of *Clostridium difficile* toxin B to the N-terminal part of the holotoxin. *J. Biol. Chem.* **272,** 11074-11078.

71. Holbourn K., Shone C. C. & Acharya K. R. (2006). A family of killer toxins Exploring the mechanism of ADP ribosylating toxins. *FEBS Journal* **273,** 4579-4593.

72. Hurley B. W. & Nguyen C. C. (2002). The spectrum of pseudomembranous enterocolitis and antibiotic-associated diarrhea. *Arch. Intern. Med.* **162,** 2177-2184.

73. Hwang J., Fitzgerald D. J., Adhya S. & Pastan I. (1987). Functional domains of Pseudomonas exotoxin identified by deletion analysis of the gene expressed in E. coli. *Cell* **48,** 129-236.

74. Jank T. & Aktories K. (2008). Structure and mode of action of Clostridial glucosylating toxins: the ABCD model. *Trends Microbiol.* **16,** 222-229.

75. Jones T. A., Zou J. Y., Cowan S. W. & Kjeldgaard M. (1991). Improved methods for building models in electron density maps and the location of errors in these models. *Acta cryst.* **A47,** 110-119.

76. Just I. & Gerhard R. (2004). Large Clostridial cytotoxins. *Rev. Physiol. Biochem. Pharmacol.* **152,** 23-47.

77. Just I., Hofmann F. & Aktories K. (2000). Molecular mode of action of the large Clostridial cytotoxins. *Curr. Top. Microbiol. Immunol.* **250,** 55-83.

78. Just I., Selzer J., Wilm M., von Eichel-Streiber C., Mann M. & Aktories K. (1995). Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* **375,** 500-503.

79. Just I., Wilm M., Selzer J., Rex G., von Eichel-Streiber C., Mann M & Aktories K. (1995). The enterotoxin from *Clostridium difficile* (Toxin A) monoglucosylates the Rho proteins. *J. Biol. Chem.* **270,** 13932-13936.

80. Kaiser E., Haug G., Hliscs M., Aktories K. & Barth H. (2006). Formation of a biologically active toxin complex of the binary *Clostridium botulinum* C2 toxin without cell membrane interaction. *Biochemistry* **45,** 3361-13368.

81. Kleywegt G. J. & Jones T. A. (1998). Databases in protein crystallography. *Acta Cryst.* **D54,** 1119-1131.

82. Klimpel K. R., Molloy S. S., Thomas G. & Leppla S. H. (1992). Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin. *Proc. Natl. Acad. Sci. USA* **89,** 10277–10281.

83. Krivan, H. C., Clark, G. F., Smith, D. F. & Wilkins, T. D. (1986). Cell surface binding site for Clostridium difficile enterotoxin: evidence for a glycoconjugate containing the sequence Gal alpha 1-3Gal beta 1-4GlcNAc. *Infect. Immun.* **53,** 573-581.

84. Kubo K. (1994). Effect of incubation of solutions of proteins containing dodecyl sulphate on cleavage of peptide bonds by boiling. *Anal. Biochem.* **225,** 351-353.

85. Lacy D. B., Wigelsworth D. J., Melnyk R. A., Harrison S. C. & Collier R. J. (2004). Structure of heptameric protective antigen bound to an anthrax toxin receptor: a role for receptor in pH-dependent pore formation. *Proc Natl Acad Sci U S A.* **101,** 13147-13151.

86. Lamzin V. S. & Wilson K. S. (1993). Automated refinement of protein models. *Acta Cryst.* **D49,** 129-147.

87. Laskowski R. A., MacArthur M. W., Moss D. S. & Thornton J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283-291.

88. Leslie A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 ESF-EAMCB newsletter on Protein crystallography* no. **26**.

89. Li M., Dyda F., Benhar I., Pastan I. & Davies D. R. (1996). Crystal structure of the catalytic domain of Pseudomonas exotoxin A complexed with a nicotinamide adenine dinucleotide analog: implications for the activation process and for ADP ribosylation. *Proc. Natl. Acad. Sci.* USA **93,** 6902-6906.

90. Liu, S. & Leppla S. H. (2003). Cell surface tumor endothelium marker 8 cytoplasmic tail-independent anthrax toxin binding, proteolytic processing, oligomer formation, and internalization. *J. Biol. Chem.* **278,** 5227-5234.

91. Lobet Y., Cluff C. W, & Cieplak W. Jr. (1991). Effect of site-directed mutagenic alterations on ADP-ribosyltransferase activity of the A subunit of Escherichia coli heatlabile enterotoxin. *Infect. Immun.* **59,** 2870-2879.

92. Lyras D., O'Connor J. R., Howarth P. M., Sambol S. P., Carter G. P., Phumoonna T., Poon R., Adams V., Vedantam G., Johnson S., Gerding D. N. & Rood J. I. (2009). Toxin B is essential for virulence of *Clostridium difficile. Nature* **458,** 1176-1179.

93. Madshus I. H., Stenmark H., Sandvig K. & Olsnes S. (1991). Entry of diphtheria toxin-protein A chimeras into cells. *J. Biol. Chem.* **266,** 17446-17453.

94. Matthews B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491-497.

95. Mauss S., Chaponnier C., Just I., Aktories K. & Gabbiani G. (1990). ADP-ribosylation of actin isoforms by *Clostridium botulinum* C2 toxin and *Clostridium perfringens* iota toxin. *Eur. J. Biochem.* **194,** 237-241.

96. McCoy A. J., Grosse-Kunstleve R. W., Adams P. D., Winn M. D., Storoni L. C. & Read R. J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658-674.

97. McDonald L. C., Killgore G. E., Thompson A., Owens R. C. Jr., Kazakova S. V., Sambol S. P., Johnson S. & Gerding D. N. (2005). An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N. Engl. J. Med.* **353,** 2433-2441.

98. McFarland L. V., Mulligan M. E., RY Kwok R. Y. & Stamm W. E. (1989). Nosocomial acquisition of *Clostridium difficile* infection. *N. Engl. J. Med.* **320,** 204-210.

99. McPherson, A. (1999). Crystallization of biological macromolecules. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.*

100. McPherson A. (2004). Introduction to protein crystallization. *Methods* **34,** 254-265.

101. McRee D. E. (1993). Practical protein crystallography. *Practical Protein Crystallography, Academic Press Inc.*

102. Menetrey J., Flatau G., Stura E. A., Charbonnier J. B., Gas F., Teulon J. M., Le Du M. H., Boquet P. & Menez A. (2002). NAD binding induces conformational changes in Rho ADP-ribosylating Clostridium botulinum C3 exoenzyme. *J. Biol. Chem.* **277,** 30950-30957.

103. Mitchell M. J., Laughon B. E. & Lin S. (1987). Biochemical studies on the effect of *Clostridium difficile* toxin B on actin in vivo and in vitro. *Infect. Immun.* **55,** 1610-1615.

104. Morris R. E., Gerstein A. S., Bonventre P. F. & Saelinger C. B. (1985). Receptor-mediated entry of diphtheria toxin into monkey kidney (Vero) cells: electron microscopic evaluation. *Infect. Immun.* **50,** 721-727.

105. Mukherjee S., Keitany G., Li Y., Wang Y., Ball H. L., Goldsmith E. J. & Orth K. (2006). *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science* **312,** 1211–1214.

106. Murshudov G.N., Vagin A. A. & Dodson E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53,** 240-255.

107. Nagahama M., Sakaguchi Y., Kobayashi K., Ochi S. & Sakurai J. (2000). Characterization of the enzymatic component of Clostridium perfringens iota-toxin. *J. Bacteriol.* **182,** 2096-2103.

108. Navaza J. (1994) AmoRe: an automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.

109. O'Neal C. J., Jobling M. G., Holmes R. K. & Hol W. G. (2005). Structural basis for the activation of cholera toxin by human ARF6-GTP. *Science* **309,** 1093-1096.

110. Ohishi I. & Yanagimoto A. (1992). Visualizations of binding and internalization of two nonlinked protein components of botulinum C2 toxin in tissue culture cells. *Infect. Immun.* **60,** 4648-4655.

111. Ohishi I. (1983). Lethal and vascular permeability activities of botulinum C2 toxin induced by separate injections of the two toxin components. Infect. Immun. **40,** 336-339.

112. Ohishi I. (1987). Activation of *botulinum* C2 toxin by trypsin, *Infect. Immun.* **55,** 1461-1465.

113. Ohishi I., Iwasaki M. & Sakaguchi G. (1980). Purification and characterisation of two components of *botulinum* C2 toxin. *Infect. Immun.* **30,** 668-673.

114. Otwinowski Z. & Minor W. (1997). Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

115. Perelle S., Domenighini M. & Popoff M. R. (1996). Evidence that Arg-295, Glu-378 and Glu-380 are active-site residues of the ADP-ribosyltransferase activity of iota toxin. *FEBS Lett.* **395,** 191-194.

116. Perelle S., Gibert M., Boquet P. & Popoff M. R. (1993). Characterization of *Clostridium* perfringens iota-toxin genes and expression in Escherichia coli. *Infect. Immun.* **61,** 5147-5156.

117. Perelle S., Gibert M., Bourlioux P., Corthier G. & Popoff M. R. (1997). Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect. Immun.* **65,** 1402-1407.

118. Petosa C., Collier R. J., Klimpel K. R., Leppla S. H. & Liddington R. C. (1997). Crystal structure of the anthrax toxin protective antigen. *Nature* **385,** 833-838.

119. Pfeifer G., Schirmer J., Leemhuis J., Busch C., Meyer D. K., Aktories K. & Barth, H. (2003). Cellular uptake of *Clostridium difficile* toxin B: translocation of the N-terminal catalytic domain into the cytosol of eukaryotic cells. *J. Biol. Chem.* **278,** 44535-44541.

120. Popoff M. R. & Boquet P. (1988). Clostridium spiroforme toxin is a binary toxin which ADP-ribosylates cellular actin. *Biochem. Biophys. Res. Commun.* **152,** 1361-1368.

121. Popoff M. R. (2000). Molecular biology of actin-ADP-ribosylating toxins. *In* K. Aktories and I. Just (ed.), *Handbook of experimental pharmacology*. Bacterial protein toxins. Springer-Verlag KG, Berlin, Germany, vol. **145,** 275-306.

122. Popoff M. R., Rubin E. J., Gill D. M. & Boquet P. (1988). Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain, *Infect. Immun.* **56,** 2299-2306.

123. Pothoulakis C., Gilbert R.J., Cladaras C., Castagliuolo I., Semenza G., Hitti Y., Montcrief J.S., Linevsky J., Kelly C.P., Nikulasson S., Desai H.P., Wilkins T.D. & LaMont J.T. (1996). Rabbit sucrase-isomaltase contains a functional intestinal receptor for Clostridium difficile toxin A. *J. Clin. Invest.* **98,** 641-649.

124. Pruitt R. N., Chagot B., Cover M., Chazin W. J., Spiller B. & Lacy D. B. (2009). Structure-function analysis of inositol hexakisphosphate-induced autoprocessing in Clostridium difficile toxin A. *J Biol Chem.* **284,** 21934-21940.

125. Qa'Dan M., Spyres L. M. & Ballard J. D. (2000). pH-induced conformational changes in *Clostridium difficile* toxin B. *Infect. Immun.* **68,** 2470-2474.

126. Ramachandran G., Ramakrishnan C. & Sasisekharan V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7,** 95-99.

127. Read T. D., Peterson S. N., Tourasse N., Baillie L. W., Paulsen I. T., Nelson K. E., Tettelin H., Fouts D. E., Eisen J. A., Gill S. R., Holtzapple E. K., Okstad O. A., Helgason E., Rilstone J., Wu M., Kolonay J. F., Beanan M. J., Dodson R. J., Brinkac L. M., Gwinn M., DeBoy R. T., Madpu R., Daugherty S. C., Durkin A. S., Haft D. H., Nelson W. C., Peterson J. D., Pop M., Khouri H. M., Radune D., Benton J. L., Mahamoud Y., Jiang L., Hance I. R., Weidman J. F., Berry K. J., Plaut R. D., Wolf A. M., Watkins K. L., Nierman W. C., Hazen A., Cline R., Redmond C., Thwaite J. E.,

White O., Salzberg S. L., Thomason B., Friedlander A. M., Koehler T. M., Hanna P. C., Kolstø A. B. & Fraser C.M. (2003). The genome sequence of *Bacillus* anthracis Ames and comparison to closely related bacteria. *Nature* **423,** 81-86.

128. Regnier F. E. (1983). High-performance liquid chromatography of biopolymers. *Science* **222,** 245-252.

129. Reineke J., Tenzer S., Rupnik M., Koschinski A., Hasselmayer O., Schrattenholz A., Schild H. & von Eichel-Streiber C. (2007). Autocatalytic cleavage of Clostridium difficile toxin B. *Nature* **446,** 415-419.

130. Reinert D. J., Jank T., Aktories K. & Schulz G. E. (2005). Structural basis for the function of Clostridium difficile toxin B. *J Mol Biol.* **351,** 973-981.

131. Reuner K. H, Presek P., Boschek C. B. & Aktories K. (1987). Botulinum C2 toxin ADP-ribosylates actin and disorganizes the microfilament network in intact cells. *Eur. J. Cell. Biol.* **43,** 134-140.

132. Rhodes G. (2000). Crystallography Made Crystal Clear. *Elsevier Science (USA).*

133. Richard J. F., Mainguy G., Gibert M., Marvaud J. C., Stiles B. G. & Popoff M. R. (2002). Transcytosis of iota-toxin across polarized CaCo-2 cells. *Mol. Microbiol.* **43,** 907-917.

134. Riley T. V. (1998). Clostridium difficile: a pathogen of the nineties. *Eur. J. Clin. Microbiol. Infect. Dis.* **17,** 137-141.

135. Rossmann M. & Blow D. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15,** 24-31.

136. Rossmann M. G. & van Beek C. G. (1999) data processing. *Acta Cryst.* **D55,** 1631-1640.

137. Rupnik M., Grabnar M. & Geric B. (2003). Binary toxin producing *Clostridium difficile* strains. *Anaerobe* **9,** 289-294.

138. Rupnik M., Pabst S., Rupnik M., von Eichel-Streiber C., Urlaub H. & Soling H. D. (2005). Characterization of the cleavage site and function of resulting cleavage fragments after limited proteolysis of *Clostridium difficile* toxin B (TcdB) by host cells. *Microbiology* **151,** 199-208.

139. Sandkvist M., Bagdasarian M. & Howard S. P. (2000). Characterization of the multimeric Eps complex required for cholera toxin secretion. *Int. J. Medical Microbiol.* **290,** 345-350.

140. Sandvig K. & Olsnes S. (1980). Diphtheria toxin entry into cells is facilitated by low pH. *J. Cell. Biol.* **87,** 828-832.

141. Santelli E., Bankston L. A., Leppla S. H. & Liddington R. C. (2004). Crystal structure of a complex between anthrax toxin and its host cell receptor. *Nature* **430,** 905-908.

142. Schering B., Barmann M., Chhatwal G. S., Geipel U. & Aktories K. (1988). ADP-ribosylation of skeletal muscle and non-muscle actin by *Clostridium perfringens* iota toxin. *Eur. J. Biochem.* **171,** 225-229.

143. Schiavo G., Benfenati F., Poulain B., Rossetto O., Polverino de Laureto P., DasGupta B. R. & Montecucco C. (1992). Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **359,** 832-835.

144. Schleberger C., Hochmann H., Barth H., Aktories K. & Schulz G. E. (2006). Structure and action of the binary C2 toxin from *Clostridium* botulinum. *J. Mol. Biol.* **364,** 705-715.

145. Schmidt G., Sehr P., Wilm M., Selzer J., Mann M. & Aktories K. (1997). Gln 63 of Rho is deamidated by *Escherichia coli* cytotoxic necrotizing factor-1. *Nature* **387,** 725-729.

146. Schuettelkopf A. W. & van Aalten D. M. F. (2004). PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Cryst.* **D60,** 1355--1363.

147. Sehr P., Joseph G., Genth H., Just I., Pick E. & Aktories K. (1998). Glucosylation and ADP ribosylation of Rho proteins: effects on nucleotide binding, GTPase activity, and effector coupling. *Biochemistry* **37,** 5296-5304.

148. Sekine A., Fujiwara M. & Narumiya S. (1989). Asparagine residue in the rho gene product is the modification site for botulinum ADP-ribosyltransferase. *J. Biol. Chem.* **264,** 8602-8605.

149. Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A., Shiba T., Ogasawara N., Hattori M., Kuhara S. & Hayashi H. (2002). Complete genome sequence of *Clostridium* perfringens, an anaerobic flesh-eater. *Proc. Natl. Acad. Sci. USA* **99,** 996-1001.

150. Simpson L. L., Stiles B. G., Zepeda H. & Wilkins T. D. (1989). Production by Clostridium spiroforme of an iotalike toxin that possesses mono (ADP-ribosyl) transferase activity: identification of a novel class of ADP-ribosyltransferases. *Infect. Immun.* **57**, 255-261.

151. Simpson, L. L. (1982). A comparison of the pharmacological properties of *Clostridium* botulinum type C1 and C2 toxins. *J. Pharmacol. Exp. Ther.* **223,** 695-701.

152. Sixma T. K., Kalk K. H., van Zanten B. A , Dauter Z., Kingma J., Witholt B. & Hol W. G. (1993). Refined structure of Escherichia coli heat-labile enterotoxin, a close relative of cholera toxin. *J. Mol. Biol.* **230,** 890-918.

153. Sixma T. K., Pronk S. E., Kalk K. H., Wartna E. S., van Zanten B. A., Witholt B. & Hol W. G. (1991). Crystal structure of a cholera toxin-related heat-labile enterotoxin from E. coli. *Nature* **351,** 371-377.

154. Smyth M. S. & Martin J. H. J. (2000). X ray crystallography. *Mol. Pathol.* **53,** 8-14.

155. Stein P. E., Boodhoo A., Armstrong G. D., Cockle S. A., Klein M. H. & Read R. J. (1994). The crystal structure of pertussis toxin. *Structure* **2,** 45-57.

156. Stiles B. G. & Wilkins T. D. (1986). Purification and characterization of Clostridium perfringens iota toxin: dependence on two nonlinked proteins for biological activity. *Infect. Immun.* **54,** 683-638.

157. Stiles B. G. (1987). Purification and characterization of *Clostridium perfringens* iota toxin. *Ph.D. thesis*, Virginia Polytechnic Institute and State University, Blacksburg.

158. Stiles B. G., Hale M. L., Marvaud J.-C. & Popoff M. R. (2000). *Clostridium perfringens* iota toxin: binding studies and characterization of cell surface receptor by fluorescence-activated cytometry. *Infect. Immun.* **68,** 3475-3484.

159. Stiles B. G., Hale M. L., Marvaud J.-C. & Popoff M. R. (2002). *Clostridium perfringens* iota toxin: characterization of the cell-associated iota b complex. *Biochem. J.* **367,** 801-809.

160. Stubbs S., Rupnik M., Gilbert M., Brazier J., Duerden B. & popoff M. (2000). Production of actin specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile. FEMB Micriobiol. Letters* **186,** 307-312.

161. Sugii, S. & Kozaki S. (1990). Hemagglutinating and binding properties of botulinum C2 toxin. *Biochim. Biophys. Acta* **1034,** 176-179.

162. Sundriyal A., Roberts A. K., Shone C. C. and Acharya K. R. (2009). Structural Basis for Substrate Recognition in the Enzymatic Component of the ADP-ribosyltransferase Toxin CDTa from *Clostridium difficile*. *J. Biol. Chem.* **284,** 28713-28719.

163. Tang J. S. C-C. (1997). Modification of the Laemmli sodium dodecyl sulphate polyacrylamide gel electrophoresis procedure to eliminate artifacts on reducing and non reducing gels. *Anal. Biochem.* **246,** 146-148.

164. Taylor, G. (2003). The phase problem. *Acta cryst.* **D59,** 1881-1890.

165. Tedesco F. J., Barton R. W. & Alpers D. H. (1974). Clindamycin-associated colitis: a prospective study. *Ann. Intern. Med.* **81,** 429-433.

166. Thelestam, M. & Chaves-Olarte E. (2000). Cytotoxic effects of the *Clostridium difficile* toxins. *Curr. Top. Microbiol. Immunol.* **250,** 85-96.

167. Tsuge H., Nagahama M., Nishimura H., Hisatsune J., Sakaguchi Y., Itogawa Y., Katunuma N. & Sakurai J. (2003). Crystal structure and site-directed mutagenesis of enzymatic components from Clostridium perfringens iota-toxin. *J. Mol. Biol.* **325,** 471-483.

168. Tsuge H., Nagahama M., Oda M., Iwamoto M., Utsunomiya H., Marquez V. E., Katunuma N., Nishizawa M. & Sakurai J. (2008). Structurl basis of actin recognition and arginine ADP-ribosylation by *Clostridium perfringens* i-toxin. *Proc. Natl. Acad. Sci.* **105,** 7399-7404.

169. Tucker K. D. & Wilkins T. D. (1991). Toxin A of *Clostridium difficile* binds to the human carbohydrate antigens I, X, and Y. *Infect. Immun.* **59,** 73-78.

170. Vagin A. & Teplyakov A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30,** 1022-1025.

171. van Damme J., Jung M., Hofmann M., Just I., Vandekerckhove J & Aktories K. (1996). Analysis of the catalytic site of the actin ADP-ribosylating *Clostridium* perfringens iota toxin. *FEBS. Lett.* **380,** 291-295.

172. Van Ness B. G., Howard J. B. & Bodley J. W. (1980). ADPribosylation of elongation factor 2 by diphtheria toxin. Isolation and properties of the novel ribosyl-amino acid and its hydrolysis products. *J. Biol. Chem.* **255,** 10717-10720.

173. Vandekerckhove J. & Weber K. (1979). The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle, and rabbit slow skeletal muscle. A protein-chemical analysis of muscle actin differentiation. *Differentiation* **14,** 123–133.

174. Vandekerckhove J., Schering B., Barmann M. & Aktories K. (1987). *Clostridium* perfringens iota toxin ADP-ribosylates skeletal muscle actin in Arg-177. *FEBS. Lett.* **225,** 48-52.

175. Vandekerckhove J., Schering B., Barmann M. & Aktories K. (1988). Botulinumb C2 toxin ADP-ribosylates cytoplasmic beta/gamma –actin in arginine 177. *J Biol Chem.* **263,** 696-700.

176. von Eichel-Streiber C., Boquet P., Sauerborn M. & Thelestam M. (1996). Large Clostridial cytotoxins - A family of glycosyltransferases modifying small GTP-binding proteins. *Trends Microbiol.* **4,** 375-382.

177. von Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S., Schulze J., & Sauerborn M. (1990). Cloning of *Clostridium difficile* toxin B gene and demonstration of high N-terminal homology between toxin A and B. *Med. Microbiol. Immunol.* **179,** 271-279.

178. von Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S., Schulze J., & Sauerborn M. (1992). Comparative sequence analysis of the *Clostridium difficile* toxins A and B. *Mol. Gen. Genet.* **233,** 260-268

179. Voth D. E. & Ballard J. D. (2005). *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin. Microbiol. Rev.* **18,** 247-263.

180. Werner G., Hagenmaier H., Drautz H., Baumgartner A. & Zahner H. (1984). Metabolic products of microorganisms. 224. Bafilomycins, a new group of macrolide antibiotics. Production, isolation, chemical structure and biological activity. *J. Antibiot.* **37,** 110-117.

181. Wilde C. & Aktories K. (2001). The Rho-ADP-ribosylating C3 exoenzyme from Clostridium botulinum and related C3-like transferases. *Toxicon* **39,** 1647-1660.

182. Wilde C., Just I. & Aktories K. (2002). Structure–function analysis of the Rho-ADP-ribosylating exoenzyme C3stau2 from Staphylococcus aureus. *Biochemistry* **41,** 1539-1544.

183. Wilson B. A. & Collier R. J. (1992). Diphtheria toxin and Pseudomonas aeruginosa exotoxin A: active-site structure and enzymic mechanism. *Curr. Top. Microbiol. Immunol.* **175,** 27-41.

184. Wren B. W. (1991). A family of Clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. *Mol. Microbiol.* **5,** 797-803.

185. Zhang R. G., Scott D. L., Westbrook M. L., Nance S., Spangler B. D., Shipley G. G. & Westbrook E. M. (1995). The three-dimensional crystal structure of cholera toxin. *J. Mol. Biol.* **251,** 563-573.

APPENDIX I

AMINO ACID SEQUENCES OF *C. DIFFICILE* BINARY TOXIN COMPONENTS

Amino Acid Sequence of Full Length Enzymatic Component (CDTa) of *C. difficile* Binary Toxin (CDTa) (Refer to Figure 2.5)

MKKFRKHKRISNCISILLILYLTLGGLLPNNIYAQDLQSYSEKVCNTTY
KAPIESFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEIS
KYSQTRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPE
KFAFNKVIRTENQNEISLEKFNEFKETIQNKLFKQDGFKDISLYEPGK
GDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIEQGYSIKIDKIVRIVI
DGKHYIKAEASVVNSLDFKDDVSKGDSWGKANYNDWSNKLTPNELA
DVNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIENALKREPIPTN
LTVYRRSGPQEFGLTLTSPEYDFNKLENIDAFKSKWEGQALSYPNFIS
TSIGSVNMSAFAKRKIVLRITIPKGSPGAYLSAIPGYAGEYEVLLNHGS
KFKINKIDSYKDGTITKLIVDATLIP

Amino Acid Sequence of Functionally Mature Fragment of Enzymatic Component (CDTa') of *C. difficile* Binary Toxin (Refer to Figure 2.5)

KVCNTTYKAPIESFLKDKEKAKEWERKEAERIEQKLERSEKEALESY
KKDSVEISKYSQTRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMY
VYYFESPEKFAFNKVIRTENQNEISLEKFNEFKETIQNKLFKQDGFKD
ISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIEQGYSIKI
DKIVRIVIDGKHYIKAEASVVNSLDFKDDVSKGDSWGKANYNDWSN
KLTPNELADVNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIENA
LKREPIPTNLTVYRRSGPQEFGLTLTSPEYDFNKLENIDAFKSKWEGQ
ALSYPNFISTSIGSVNMSAFAKRKIVLRITIPKGSPGAYLSAIPGYAGEY
EVLLNHGSKFKINKIDSYKDGTITKLIVDATLIP

Amino Acid Sequence of Full Length Transport Component (CDTb) of *C. difficile* Binary Toxin (Refer to Figure 2.5)

MKIQMRNKKVLSFLTLTAIVSQALVYPVYAQTSTSNHSNKKKEIVNED
ILPNNGLMGYYFSDEHFKDLKLMAPIKDGNLKFEEKKVDKLLDKDKS
DVKSIRWTGRIIPSKDGEYTLSTDRDDVLMQVNTESTISNTLKVNMK
KGKEYKVRIELQDKNLGSIDNLSSPNLYWELDGMKKIIPEENLFLRDY
SNIEKDDPFIPNNNFFDPKLMSDWEDEDLDTDNDNIPDSYERNGYTI
KDLIAVKWEDSFAEQGYKKYVSNYLESNTAGDPYTDYEKASGSFDK
AIKTEARDPLVAAYPIVGVGMEKLIISTNEHASTDQGKTVSRATTNSKT
ESNTAGVSVNVGYQNGFTANVTTNYSHTTDNSTAVQDSNGESWNTG
LSINKGESAYINANVRYYNTGTAPMYKVTPTTNLVLDGDTLSTIKAQE
NQIGNNLSPGDTYPKKGLSPLALNTMDQFSSRLIPINYDQLKKLDAGK
QIKLETTQVSGNFGTKNSSGQIVTEGNSWSDYISQIDSISASIILDTEN
ESYERRVTAKNLQDPEDKTPELTIGEAIEKAFGATKKDGLLYFNDIPID
ESCVELIFDDNTANKIKDSLKTLSDKKIYNVKLERGMNILIKTPTYFTN
FDDYNNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIPMSELKPYKR
YVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTEKD
SSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIY
FADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATLQYSGFK
DIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYKKLRIYAIT
PDDRELLVLSVD

Amino Acid Sequence of Signal Peptide less Fragment of Transport Component (CDTb') of *C. difficile* Binary Toxin (Refer to Figure 2.5)

EIVNEDILPNNGLMGYYFSDEHFKDLKLMAPIKDGNLKFEEKKVDKL
LDKDKSDVKSIRWTGRIIPSKDGEYTLSTDRDDVLMQVNTESTISNTL
KVNMKKGKEYKVRIELQDKNLGSIDNLSSPNLYWELDGMKKIIPEEN
LFLRDYSNIEKDDPFIPNNNFFDPKLMSDWEDEDLDTDNDNIPDSYE
RNGYTIKDLIAVKWEDSFAEQGYKKYVSNYLESNTAGDPYTDYEKAS
GSFDKAIKTEARDPLVAAYPIVGVGMEKLIISTNEHASTDQGKTVSRA
TTNSKTESNTAGVSVNVGYQNGFTANVTTNYSHTTDNSTAVQDSNG
ESWNTGLSINKGESAYINANVRYYNTGTAPMYKVTPTTNLVLDGDTLS
TIKAQENQIGNNLSPGDTYPKKGLSPLALNTMDQFSSRLIPINYDQLK

KLDAGKQIKLETTQVSGNFGTKNSSGQIVTEGNSWSDYISQIDSISASI
ILDTENESYERRVTAKNLQDPEDKTPELTIGEAIEKAFGATKKDGLLY
FNDIPIDESCVELIFDDNTANKIKDSLKTLSDKKIYNVKLERGMNILIKT
PTYFTNFDDYNNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIPMSE
LKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEF
ETTEKDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIM
DAHKIYFADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATL
QYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYK
KLRIYAITPDDRELLVLSVD

Amino Acid Sequence of Functionally Mature Fragment of Transport Component (CDTb'') of *C. difficile* Binary Toxin (Refer to Figure 2.5)

LMSDWEDEDLDTDNDNIPDSYERNGYTIKDLIAVKWEDSFAEQGYK
KYVSNYLESNTAGDPYTDYEKASGSFDKAIKTEARDPLVAAYPIVGVG
MEKLIISTNEHASTDQGKTVSRATTNSKTESNTAGVSVNVGYQNGFT
ANVTTNYSHTTDNSTAVQDSNGESWNTGLSINKGESAYINANVRYYN
TGTAPMYKVTPTTNLVLDGDTLSTIKAQENQIGNNLSPGDTYPKKGLS
PLALNTMDQFSSRLIPINYDQLKKLDAGKQIKLETTQVSGNFGTKNSS
GQIVTEGNSWSDYISQIDSISASIILDTENESYERRVTAKNLQDPEDKT
PELTIGEAIEKAFGATKKDGLLYFNDIPIDESCVELIFDDNTANKIKDSL
KTLSDKKIYNVKLERGMNILIKTPTYFTNFDDYNNYPSTWSNVNTTNQ
DGLQGSANKLNGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKIK
AKEEKTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGSGTTYLDNLSI
TELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMY
FAPTQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQ
PKTNYVNLRSYFTGGENIMTYKKLRIYAITPDDRELLVLSVD