BIOCHEMICAL STUDIES AND BIOMOLECULAR HOMOLOGY MODELING (STRUCTURAL BIOINFORMATICS) OF BIOLOGICALLY IMPORTANT TOXIN PROTEASES FROM *VIBRIO CHOLERAE*

THESIS
Submitted To

THE FACULTY OF LIFE & ENVIROMENTAL SCIENCES
UNIVERSITY OF PESHAWAR

In fulfillment for the degree

of

DOCTOR OF PHILOSOPHY

by

FARHAT AMIN

CENTER OF BIOTECHNOLOGY & MICROBIOLOGY
UNIVERSITY OF PESHAWAR
2009

CENTER OF BIOTECHNOLOGY & MICROBIOLOGY
UNIVERSITY OF PESHAWAR
SEPTEMBER 2009

This Thesis Submitted
by

FARHAT AMIN

Titled

BIOCHEMICAL STUDIES AND BIOMOLECULAR HOMOLOGY MODELING (STRUCTURAL BIOINFORMATICS) OF BIOLOGICALLY IMPORTANT TOXIN PROTEASES FROM *VIBRIO CHOLERAE*

In partial fulfillment of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

In

BIOTECHNOLOGY

DEDICATION

Of Research

To My Respected and Compassionate Parents

To My Brothers and Sisters

&

To everyone who supported me

CONTENTS

ACKNOWLEDGEMENT

I take this marvelous moment, as an opportunity to thanks Almightly ALLAH, it is all through ALLAH's wish and will, giving me the strength and courage to carry out this study.

I take this opportunity to express my sincere gratitude and thanks to supervisior Prof. Dr. Ghosia Lutfulla, Centre of Biotechnology & Microbiology, University of Peshawar & co-supervisor Dr. Asmat Salim ,Dr.Panjwani Centre For Molecular medicine & Drug Research,University of Karachi, for their invaluable concerns, continuous help, valuable advices and guidance throughout the course of my study which enabled me to bring my work to completion. It would have been not possible to accomplish my task successfully without their continuous support.

I am extremely greatful to Prof. Dr. Bashir Ahmed, Director, Centre of Biotechnology & Microbiology, University of Peshawar, for his great help, encouragement and for providing me all the facilities of research work at centre.

I would like to express my deep appreciation, regards and thanks to Professor Dr. M. Iqbal Choudary, Director Intenational Center of Chemical & Biological Sciences, University of Karachi, for permitting me to work as collaborative student under the supervision of Dr. Asmat salim and providing me all the required facilities.

Sincere expressions of deep regards are due to all my teachers and staff members of this centre.

Many thanks and regards are due to Dr. Shamshad Zarina, Department of Biochemistry, University of Karachi, Dr. Kamran Azim, International Centre for Chemical & Biological Sciences, University of Karachi, Dr. Siddiqa, Department of Biochemistry, university of Karachi, and Dr.Shabana, International Centre for Chemical & Biological Sciences, University of Karach, for their fruitful discussions and valuable suggestions.

I wish to acknowledge Prof. Dr. Lutfullah Kakakhel for his professional guidance & technical assistance.

I must never forget a few other names as without them it would have been an uphill task for me to complete my work. I am really grateful to Khalida Shoukat, Yasmeen Rashid, Kawnal Hanif, Nadia Naeem, Han'a Iqbal, Humera Jehan, Nazia & Afsheen, Dr.Panjwani Centre For Molecular medicine & Drug Research,University of Karachi, who in every step helped me a great deal and made in-roads for me. I might not have done my task that easily without their support and encouragement.

Special words of gratitude for all my research fellows, Mrs. Noreen Azhar, Miss Seema Gul, Mr. zahid Khan, Mr. Roshan Ali, and Mr.Ashfaq Ahmed for constant encouragement and moral support.

Finally I would like to forward a special word of gratitude and to salute my very supportive parents, who have consistently encouraged & motivated me for my every educational accomplishments, for their boundless spiritual and material support permit me to complete my research work ambitiously. Also, the long-lasting support and encouragement of my siblings especially my very supportive brother, Mr.Tariq Amin khan, enabled me to continue and finished my thesis.

In the end I must admit that Mr. Mohd Humayun Khan also gave me every possible help and encouraged me at every moment of despair and frustration during my task. I am extremely grateful to him.

Last but not least my thanks goes to our lab staff Mr. Shahid Shakoor, Dr.Panjwani Centre For Molecular medicine & Drug Research,University of Karachi, Mr.Adnan,Mr.Amir,Mr.Fayyaz, Mr.waqas & Mr. Aftab for all their help.

Farhat Amin.

SUMMARY

At the molecular level, the pathogenesis of cholera is a multifactorial process. Several genes encoding virulence factors aid the pathogen in its colonization, coordinated expression of virulence factors, and toxin action such as cholera toxin (CT) and the toxin-coregulated pillus (TCP). Cholera Toxin is primarily responsible for the profuse secretory diarrhea while the toxin-coregulated pillus is an essential colonization factor and receptor for the phage encoding CT. In addition to potent CT, there are other putative toxic factors such as accessory cholera enterotoxin (Ace), zonula occludens toxin (Zot), hemolysin/cytolysin and hemagglutinin/protease (Vibriolysin). Neuraminidase, a sialidase, is also considered a virulence factor. It is a key glycoside hydrolase that catalyzes the conversion of higher-order gangliosides to monosialoganglioside, which is the cholera toxin receptor. It has been hypothesized that neuraminidase may locally produce high concentration of the GM1 receptor for CT, thereby enhancing the binding and internalization of CT and resulting in greater efflux of water and electrolytes.

The present study comprises the analysis of toxin proteases of *V. cholerae* in order to have a better understanding of pathogenesis using molecular biology and bioinformatics tools as the signaling cascades involved in the induction and regulation of mucosal inflammatory responses to infection by *V.cholerae* are still largely unknown. The proposed work is carried out in two phases. The first phase deals with the mRNA expression of proinflammatory, anti-inflammatory and various other cytokines in intestinal tissue and bone marrow cells upon infection with purified Cholera Toxin (CT) alone and in combination with purified neuraminidase from *Vibrio cholerae* in order to elucidate the mechanisms of the *V.cholerae* induced proinflammatory response as well as

the effects of Cholera Toxin in CT-induced immune response for the development of safe, life attenuated vaccines. The transcriptional profiling upon CT infection in the intestinal tissue and bone marrow cells at 2, 6 18 and 48 hrs was carried out by RT-PCR. Among the 15 cytokines (IL-1α, IL1β, IL-2, IL3, IL-4, IL-5, IL-6, IL-7, IL-10, IFNγ, TNFα, GM-CSF, TGFβ, MCP-1, and HIF) involved in proinflammatory, anti inflammatory and antigen-specific immune responses, the mRNA of the eight cytokines (IL-1α, IL1β, IL-7, IFNγ, TNFα, TGFβ, MCP-1 and HIF) were found to be expressed in intestinal tissue and bone marrow cells upon infection whereas rest of these cytokines were either down regulated or not expressed upon infection.

The second phase of the study deals with the prediction of 3D models of toxin proteases of *V. cholerae* i.e. Accessory cholera enterotoxin (Ace), zonula occludens toxin (Zot), and hemagglutinin/protease (Vibriolysin; HA/P) based on comparative homology modeling in order to have better understanding of pathogenesis using bioinformatics tools. The predicted structure is discussed with special reference to the active site and cofactor binding site. The enzyme-inhibitor model was also constructed separately in order to identify important residues that contribute to the active site interaction of the enzyme with the ligand. The homology model was constructed in the presence of inhibitor HPI [N-(1-carboxy-3-phenylpropyl)-phenylalanyl-alpha-aspargine] based on its sequence homology with Pseudomonas aeruginosa elastase (PAE). Comparison of the 3D structures of PAE and HA/P reveals a remarkable similarity having a conserved a + b domain. The inhibitor shows similar binding features as seen in other metalloproteases of M4 peptidase family. The study also highlights the key catalytic residues as well as the residues at the S1 and S'1 binding sub-sites. The similarities between the two proteins

provide support for the hypothesis that the two enzymes have similar three-dimensional structures and a common mechanism of action. The fact that both enzymes are secreted as zinc-containing proteases, led us to further hypothesize that they may play similar role in pathogenesis.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

CT	Cholera Toxin
CTA	Cholera Toxin A (Subunit)
CTB	Cholera Toxin B (Subunit)
LPS	Lipopolysacchride
TCP	Toxin-Coregulated Pilus
VPI	Vibrio Pathogenicity Island
Zot	Zonula Occludens Toxin
Ace	Accessory Cholera Enterotoxin
cep	Core-Encoded Pilin
HA/P	Hemagglutinin/Protease
HlyA	Hemolysin/Cytolysin
GAPDH	Glyceraldehyde Phosphate Dehydrogenase
IL-1α	Interleukin-1 α
IL-1β	Interleukin-1 β
IL-2	Interleukin-2
IL3	Interleukin-3
IL-4	Interleukin-4
IL-5	Interleukin-5
IL-6	Interleukin-6
IL-7	Interleukin-7
IL-10	Interleukin-10
MCP-1	Monocyte Chemotactic Protein-1

GM-CSF	Granulocyte-macrophage colony-stimulating factor
TNF-α	Tumor Necrosis Factor -α
IFNγ	Inteferon- γ
TGFβ	Transforming Growth Factor- β
HIF-1α	Hypoxia Inducing Factor-1α
SCF	Stem Cell Factor
PMNs	Polymorphonuclear Leukocytes
PCR	Polymearase Chain Reaction
RT-PCR	Reverse Transcriptase-PCR
PBS	Phosphate Buffer Saline
β-ME	β-Mercaptoethanol
EDTA	EthylinediamineTetraacetic acid
TE Buffer	Tris EDTA
TBE Buffer	Tris-Borate EDTA
DEPC	Diethylpyrocarbonate
3D Three	Dimentional
PDB	Protein Data Bank
BLAST	Basic Local Alignment Search Tool
FASTA	Fast Alignment
PIR	Protein Information Reseourses
CATH	Class,Architecture,Topology and Homologous superfamily.
SCOP	Structural Classification Of Protein

Amino Acid	Three-Letter-Code	One- Letter-Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

List of Publications

1. Ghosia Lutfullah,Farhat Amin, Zahid Khan, Noreen Azhar, M. Kamran Azim, Sajid Noor, Khalida Shoukat. **Homology Modeling of Hemagglutinin/Protease [HA/P (vibriolysin)] from Vibrio Cholerae: Sequence Comparision, Residue Interactions and Molecular Mechanism.** Protein J, **(2008)**, 27:105–114.DOI 10.1007/s10930-007-9113-0.
2. Farhat Amin, Ghosia Lutfullah. **Structure Prediction of Zona occludens toxin (Zot) and Accessory cholera enterotoxin (Ace) from *V. Cholerae* by Comparative Modeling Techniques.** Submitted in Internation Jornal of Chemical Society Pakistan.
3. Farhat Amin, Asmat Salim, Ghosia Lutfullah. **Transcriptional profiling of different cytokines in rat intestinal tissues and bone marrow cells upon Cholera Toxin (CT) infection.** To be published
4. Farhat Amin, Ghosia Lutfullah, Asmat Salim.**Effect of Purified Cholera Toxin on expression of Proinflammatory Cytokines in in rat intestinal tissues and bone marrow cells.** To be published.

CHAPTER

1.0

INTRODUCTION

1.1. INTRODUCTION

Vibrio Cholerae (also *Kommabacillus*) is a highly motile, facultatively anaerobic, non-invasive Gram negative (1-2.) asporogenous curved-rod shaped bacterium with a polar flagellum (3-4.) which belongs to the *Vibrionaceae* family (5).

The foremost assistance to recognized cholera was made by physician John Snow who originated the link between cholera and contaminated drinking water in 1854 during the fourth pandemic. In the same year the bacteria was first isolated, described and termed *V.cholerae* by Filippo Pacini (5), but his discovery was not widely known until Robert Koch, working independently thirty years later, exposed the knowledge and identified *V. Cholerae* as "comma bacillus" in 1883 (4). Cholera Toxin (CT), the causative protein factor was first discovered by Robert Koch in 1884 (5).

V.Cholerae is a well defined organism on the basis of biochemical tests (3-4) and DNA homology studies (4). It is subdivided into serogroups on the basis of somatic O antigen (3). A serotyping scheme, the conventional method used to classify *V.Cholerae* strains, was first studied by Gardner and Venkatraman (6). This method is based on epitopic difference of cell surface lipopolyscaccharide (LPS) (7). To date, *V.Cholerae* has been classified into about more than 200 serogroups (1,3) on the basis of its somatic antigens (O antigens) (3,8), of which only two serogroups O1 and O139 strains have been shown to cause endemic of cholera, a severe diarrheal disease (7-9). The major difference between the two serogroups is the occurrence of diverse lipopolysaccharide (LPS) antigens and a capsular polysaccharide (10).

All strains that were identified by biochemical tests as *V.Cholerae* that do not agglutinate with O-antiserum were collectively referred to as non-O1 *V. Cholerae*. They were also called non-cholera vibrios or non-agglutinable vibrios (3, 4,6). 154 serogroups of non-O1 *vibrios* are known. The biochemical and morphological characteristics, possessed by these *vibrios*, are analogous to those of the cholera *vibrio* but are nonagglutinable with O1 antiserum. These *vibrios* are agglutinable by their own antisera (4). The non-O1 serogroups, despite the fact that are not related to cholera incidence, can be pathogenic, and are rarely related with small outbreaks of diarrhoeal disease (3).

The O1 serogroup exists as two biotypes, **Classical** and **El Tor** (3,8) that can be discriminated by applying different biochemical tests i.e. assays of haemolysis, haemagglutination, phage and polymyxin B sensitivity. In addition, presence of biotype-specific genes (eg, *tcpA*, *rtxC*), is the latest approach to distinguish between the two (3). These can be further subdivided into two major serotypes, **Ogawa** and **Inaba**, which have different lipopolyscacchride (LPS) structures (3,11-12). Strains of the serotype, Ogawa are said to produce A and B antigens and a small amount of C antigen, whereas Inaba strains express only A and C antigens. A third serotype, designated Hikojima, express all the three antigens, but it is exceptional and unstable (3,11). It has been revealed by the analysis of the chemical structure of the O antigen that it is composed of a homopolymer, have the amino sugar D perosamine substituted with 3-deoxy-L-glycerotetronic acid. Both Inaba and Ogawa serotypes contain this structure and thus it may be the A antigen while the nature of B and C antigens is unknown (Fig.1.1) (11).

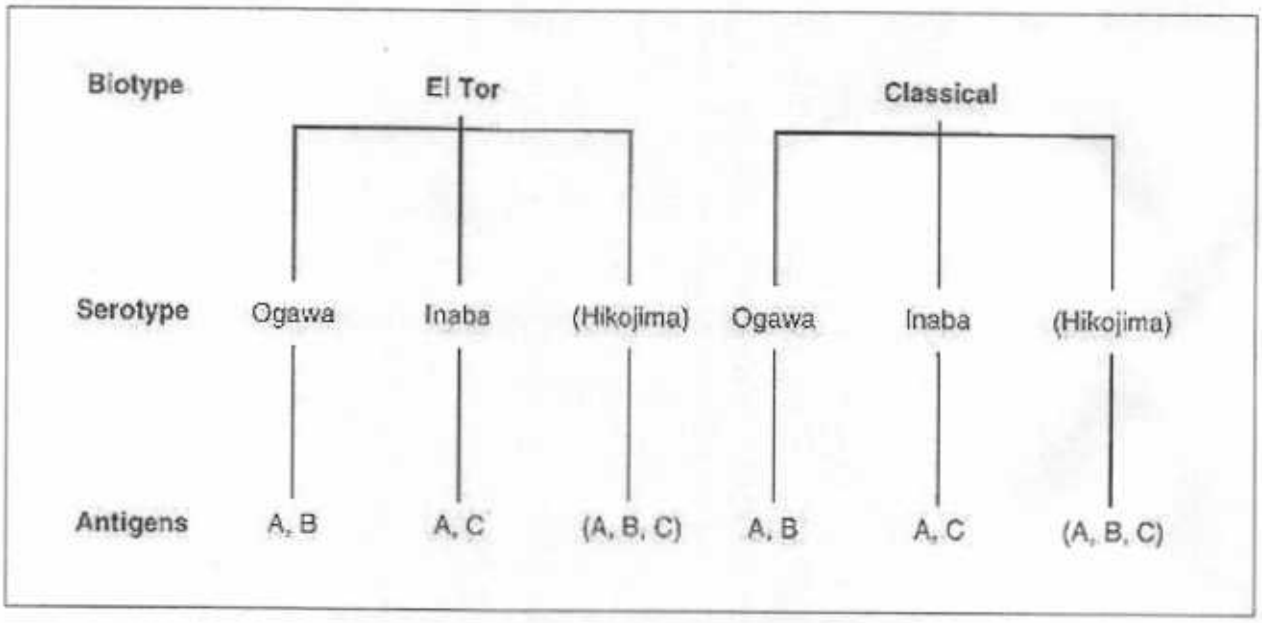


Fig.1.1. A serotyping scheme to classify *V.Cholerae* strains on the basis of O antigen.

The classical biotype was responsible for the first six epidemics while the El Tor biotype was responsible for the seventh epidemic which began in 1961 and is enduring. Since 1992 in Southeast Asia, apart from O1, the new cholera pathogen *V. cholerae* O139 has caused epidemics. But, presently the El Tor biotype is still the main pathogen causing epidemics (8).

It was assumed earlier, that for cholera pandemics and all major epidemics only *V. cholerae* O1 strains were responsible and the non-O1 strains were non-pathogenic and responsible for only periodic cases of gastroenteritis and extra-intestinal infections. But then there was an outbreak of cholera in India and Bangladesh in 1992. It was reported that it was caused by a novel non-O1 strain: *V. cholerae* O139 Bengal (7).

Two significant properties of *V. Cholerae* are taken into account with respect to virulence; (i) The production of cholera toxin (CT), which is responsible for the severe diarrhea, and (ii) the presence of the O1 or O139 antigen, which acts as a marker of epidemic potential. On the other hand, molecular analysis has revealed that in addition to CT encoding genes, all strains competent of causing cholera also carry genes for a colonization factor known as toxin-coregulated pilus (TCP) and a regulatory protein, ToxR, which coregulates the expression of CT and TCP. The synergistic effect of a number of pathogenic factors produced by toxigenic *V. Cholerae* is responsible for the pathogenesis of cholera. (4).

1.2. HABITAT

Majority of *Vibrio* species are everywhere in estuarine and aquatic environments and are also found in fresh water. *V. Cholerae* is extensively dispersed in tropical and moderate aquatic environments. Various abiotic factors including the organic and inorganic contents of water and sediments, variations in oxygen tension, fluctuating temperature, salinity, pH, and exposure to the sunlight affect its distribution (13).

V. Cholerae, a marine microorganism (14), is an autochthonous inhabitant of coastal marine environment (1,15), and is capable of that transmitting a disease to humans through infected water and uncooked or partially cooked seafood (14). This *Vibrio* specie not only stays alive in riverine, estuarine, and coastal waters around the world but also survive and multiplies in marine environments by adhering with zooplankton and phytoplankton (3). Its capacity to form biofilms on biotic and abiotic surfaces plays important roles in environmental survival though it is still not understandable that to what

extent these mechanisms are established distinctively in toxigenic O1 and O139 strains. Various environmental aspects have been associated with cyclic cholera epidemics, but the correlations of genetic and biochemical determinants are still unidentified (16). As this human pathogen is principally an inhabitant of the marine environment, water plays an imperative role in the spread and epidemiology of cholera (17). *V. cholerae* colonizes the gastrointestinal tract, where it sticks to villi. Furthermore, it produces two different proteases called chitinase and mucinase, which facilitate the attachment of *V.cholerae* to aquatic organisms. *V.cholerae* has the ability of to enter copepods with the help of chitinase. The assistance to entry into the human gastro-intestinal tract is facilitated by a non-specific protease, mucinase (13).

1.3. PATHOGENESIS AND VIRULENCE FACTORS:

After oral ingestion of contaminated water or food the organism must pass through the acid barrier of the stomach and colonize the epithelium of the small intestine. Colonization is aided by filamentous protein structures called the toxin-coregulated pili (TCP) and probably by other colonization factors such as the accessory colonization factor, different haemagglutinins, and core-encoded pilus. Adherent *vibrios* produced cholera enterotoxin (CT) that is secreted across the bacterial outer membrane into the extracellular environment and perturb ion transport by intestinal epithelial cells (13). The uptake of cholera toxin by the intestinal epithelium results in the activation of adenylate cyclase which in turn results in the increased secretion of Cl^- and eventually loss of as much as 20 liters of water and electrolytes per day (18). The exact mechanism of

V.Cholerae colonization and the interaction of pili with host are still not clearly understood (2).

At the molecular level, the pathogenesis of cholera is a multifactorial process. For establishing a productive infection by *V. cholera*, colonization is a prerequisite. Several genes encode virulence factors that aid the pathogen in its colonization, coordinated expression of virulence factors, and toxin action [2]. The main virulence genes in *V.cholerae* are present in clusters (4). Genetic analyses have shown that there are at least two important genetic elements of the *V.cholerae* chromosome in which genes encoding the virulence factors are clustered together. These include the CTX genetic element, which is the genome of a lysogenic bacteriophage, designated CTXΦ that carries the genes encoding cholera toxin, (3) and several other toxins and accessory virulence factors, (19) and large Vibrio pathogenicity island (VPI) that encodes a toxin-coregulated pilus (TCP) gene cluster that functions as basic colonization factor and acts as CTXΦ receptor (4). Other factors that have been associated with the pathogenesis of cholera include Zonula occludens toxin (Zot), accessory cholera enterotoxin (Ace), Hemagglutinins, neuraminidase (nanH), outer membrane protein (ompU), RTX toxin, Vac a-like toxin (20) and a ToxR regulatory protein. (21).

CTXΦ genome (~7Kb) has been extensively characterized which revealed that it has a modular structure composed of two functionally distinctive genomes, the core and RS2 regions (18). The CTXΦ genome had been thought originally as a transposon-like genetic element. A dynamic 4.5kb core region, termed the virulence cassette encodes for both cholera toxin (CT) and the genes involved in phage morphogenesis. It is known to carry at least six genes, including *ctxAB* (encoding the A and B subunits of CT), *zot* (encoding

zonula occludens toxin), *ace* (encoding accessory cholera enterotoxin), *cep* (encoding core-encoded pilin), and *orfU* (encoding a product of unknown function). The core region is flanked by one or more copies of a 2.5kb repetitive sequence designated as RS2 region that encodes genes required for replication (*rstA*), integration (*rstB*), and regulation (*rstR*) of CTXΦ (4,13).

Vibrio pathogenicity island (VPI), a 39-kb DNA region, has been shown to contain the *tcp* gene cluster, which contains a gene that encodes for the ToxT regulatory protein as well as several other clusters of genes that encodes products of known and unknown functions (19). The presence of groups of virulence genes, a regulator of virulence gene, a transposase gene, and specific (*att*-like) attachment sites flanking each end of the island are some important structural features of the VPI. The CTXΦ uses TCP as the receptor during infection. At least 11 accessory proteins are required for the biogenesis of TCP; the majority of their genes are located in the TCP operon (3). It has been revealed that the major subunit of TCP is encoded by the *tcpA* gene; the products of a number of other genes that are required for the formation and function of the pilus assembly are located on a large DNA region referred to as the TCP pathogenicity island, which includes the *tcp* and *acf* gene clusters (4). In brief, the *V.cholerae* pathogenicity island (VPI) contains genes related with virulence (TCP-ACF cluster), the regulation of virulence (toxT and tcpP/H), the regulation of chemotaxis (tcpI or acfB), and mobility (int and ofrI) (21). The TCP–encoding genes are clustered within a region known as both the vibrio pathogenicity island (VPI) and the TCP pathogenicity island (18).

The genome of one *V.cholerae* strain has been sequenced completely (20). Its genome comprises two chromosomes, of which chromosome I, the larger chromosome, harbors

all virulence factors, including CT and the toxin-co-regulated pilus (TCP) (5). Between the two chromosomes there is marked irregularity in the distribution of genes known to be essential for pathogenesis and growth. Chromosome I encodes genes for DNA replication and repair, transcription, translation, cell-wall biosynthesis and a variety of central catabolic and biosynthetic pathways. Most genes known to be essential in bacterial pathogenicity are also located on chromosome I i.e. those encoding toxin co-regulated pilus, cholera toxin, lipopolysacharide and the extracellular protein secretion machinery. Genes that are thought to be essential for normal cell function are found only on chromosome II. Furthermore, chromosome II encodes many intermediates of metabolic pathways (22).

1.4. IMPORTANT TOXIN PROTEINS

Cholera Toxin (CT)

The most essential virulence property of *V.cholerae* is the production of cholera toxin (CT) and the ability to adhere to and colonize the small intestine of the host (8). The 84-kDa CT belongs to the super family of AB toxins and is an oligomeric protein composed of a heterodimeric catalytically active A subunit (CTA) and five B subunits (CTB) encoded by *ctx*A and *ctx*B, respectively (8,23-24). The 11.6-kDa CTB is a nontoxic protective antigen responsible for mediating toxin binding to the monosialoganglioside (GM1) receptor on host intestinal epithelial cells. The 28-kDa CTA activates the cellular adenylate cyclase and accelerates the secretion of chloride and bicarbonates from the mucosal cells to the intestinal lumen (8,23). Cholera toxin travels from the plasma membrane of intestinal cells to the endoplasmic reticulum (ER) where a portion of the A-subunit, the A1 chain, crosses the membranes into the cytosol (25). *V. cholerae* colonizes

the small bowel using TCP and interacts with receptors on the intestinal epithelium. Once attached, the bacterium secretes its toxin, which is accompanied by the release of hemagglutinin/protease (HA/protease). This extracellular HA/protease is responsible for nicking the CT-A subunit at Arg192, yielding discrete CT-A_1 and CT-A_2 subunits, which are solely connected by a single disulfide bond. This post-translational modification is critical for full activity of the toxin, leading to an increase in cAMP-production. This causes massive secretion of electrolytes and water into the intestinal lumen, paralleled by excretion of the bacteria (5).

Neuraminidase

V.Cholerae neuraminidase, a sialidase, is also considered a virulence factor (26). It is a key glycoside hydrolase (27) that catalyzes the conversion of higher-order gangliosides to monosialoganglioside, which is the cholera toxin receptor. It has been hypothesized that neuraminidase may produce locally high concentration of the GM1 receptor for CT, thereby enhancing the binding and internalization of CT and resulting in greater efflux of water and electrolytes (24, 26, 28).

Hemagglutinins

V. Cholerae produces several hemagglutinins, which may be involved in adherence of the vibrios to the human gut (29). Out of several hemagglutinins produced by *V. cholerae*, one of them is the soluble Zn-dependent metalloprotease, which was originally identified as a hemagglutinin/protease (HA/P) encoded by the gene *hapA* (30-32). All HA/Ps have both hemagglutination and proteolytic activities (33). HA/P can proteolytically activate CTA subunit by nicking it into two (A_1 and A_2) peptide fragments [34, 35] and increasing its biological activity (35). It can also activate the E1 Tor

cytolysin and hemolysin (31,35,36) and is capable of hydrolyzing several other physiologically important proteins such as mucin, fibronectin and lactoferin (30,36,37). Microarray studies of the global transcription pattern of *V. cholerae* have shown that the hemagglutinin protease (*hap*) gene is one of the 12 genes that belong to the pathogenesis functional group (38).

Hemolysin

V.Cholerae hemolysin/cytolysin (HlyA), an extracellular membrane damaging protein with a native monomeric weight of 65kda, belongs to a unique class of dimorphic proteins that can exist in two stable states, a water-soluble monomer and an oligomeric integral membrane protein. These proteins are commonly known as pore-forming toxins (PFTs) for their ability to destroy target eukaryotic cells by punching holes in the plasma membrane and are released as water-soluble monomeric proteins by a wide variety of pathogenic bacteria (39,40). HlyA is synthesized as an 82-kDa preprohemolysin by *V.cholerae* and exported as the 79-kDa prohemolysin. Proteolytic removal of the 132-residue N-terminal stretch generates the mature 65-kDa HlyA with a specific hemolytic activity (40, 41).

Zonula occludens toxin (Zot) and Accessory cholerae enterotoxin (Ace)

V. cholerae also produces two other putative toxins known as zonula occludens toxin (Zot) and accessory cholerae enterotoxin (Ace) (42). Zot increases the permeability of the small intestinal mucosa by affecting the structure of the intercellular tight junction, or zonula occludens. ORF of the *Zot* gene is of 1.3-kb, which encodes a 44.8-kDa polypeptide and is located immediately upstream of the *ctxA* gene (4). *Zot* gene is known to be involved in phage morphogenesis (21,43). Ace has been described as a third

enterotoxin, which increases potential difference across the intestinal epithelium and affects ion transport (7). Even though Ace is thought to be a minor coat protein of CTXΦ, its toxicity is not dependent on incorporation into CTXΦ. The process by which Ace is independently released from *V. cholerae* has not been characterized (18).

1.5. CHOLERA INFECTION:

Cholera is an acute bacterial water-borne infection of the intestine [30] that results in an estimated 100,000 to 200,000 deaths annually worldwide [23]. Cholera still remains an imperative worldwide problem, especially in developing countries (2). It is characterized by the loss of significant fluid, voluminous stool and dehydration and leads to death if left untreated (44). The customary factors that are involved in the transmission of cholera are contaminated water supply, poor sanitation, indiscriminate defaecation and poor personal hygiene (45).

Symptoms are generally rapid and consist of watery diarrhoea and vomiting after an incubation period of about 18 h and 5 days. The distinguishing feature of cholera is the painless eradication of voluminous stools resembling rice-water, having a fishy odor. In general the vomit is usually clear, watery, alkaline fluid. In adults the rate of diarrhoea may promptly exceed up to 500–1000 mL/h with severe cholera and leading to severe dehydration. Symptoms of severe dehydration include undetectable blood pressure, absent or low-volume peripheral pulse, sunken eyes, poor skin turgor, and wrinkled hands and feet. In the early days, patients are fidgety and enormously thirsty; they become lazy and may lose consciousness as the shock progresses. The loss of the fluid may be so rapid that within a few hours of infection the patient is at risk of death, and most deaths occur during the first day. Conversely, the patient may live on for the time

being if rehydration fluids are available in inadequate quantities. In general, from the improper treatment of cholera infection several complications can occur. If insufficient fluids are given, acute renal failure from prolonged hypotension may occur. If the intravenous fluids are not in appropriate concentrations, electrolyte imbalance can occur, especially hypokalaemia. As a dilemma of shock and poor perfusion of the placenta, miscarriage or premature delivery can occur in pregnant women. In addition to these symptoms, severe muscle cramps of arms and legs are common due to the electrolyte imbalance. Replacing water and electrolytes as fast as they are being lost is the simple treatment. An electrolytes solution is given to patients intravenously followed by an oral rehydration solution (ORS). As it is known that sugars are contransported with Na^+ over the apical membrane of the enterocytes therefore for more effective reabsorption of electrolytes the ORS is supplemented with carbohydrates (46).

1.6. EFFECT OF CHOLERA TOXIN (CT) ON INFLAMMATORY CYTOKINES:

A crucial part of the host response to infection with microbial pathogens is the stimulation of proinflammatory cytokine genes and other genes with immunomodulatory function. Initially it has been documented that cytokines were produced exclusively by cells of the immune system; though now, it is evident that non immune cells, including epithelial cells, also produce cytokines and present an essential first-line defense. (47) Multiple cytokines can be produced by individual cells and can be classified into different groups, such as interleukins (IL-1 to IL-23), interferons (IFN-α, IFN-γ), colony-stimulating factors, tumor necrosis factors (TNFs), tumor growth factors (TGF-β), and chemokines (MCP-1, MIP-1) (48) .

The initiation of an early inflammatory response in the intestinal epithelial cells is a distinguishing characteristic of *V. cholerae* induced pathology (49-50). Though cholera is traditionally considered to be a non-inflammatory diarrheal disease, there are various reports suggesting that there is an inflammatory component in the pathogenesis of the disease (49,51). The presence of lymphocytes and mononuclear cells, congested blood vessels and the infiltration of polymorphonuclear leukocytes (PMNs) in the lamina propria as well as elevated level of LTB_4, high serum titers of IL-6, localized release of tumor necrosis factor (TNF-α), (52-53) tissue damage, and the neutrophil chemo-attractant protein macrophage inhibitory protein (MIP-2) by accessory toxins of *V. cholerae* provide significant evidence of existence of inflammatory component (49-50). Reports have also documented the induction of proinflammatory cytokines in intestinal epithelial cells upon *V.cholerae* infection such as IL-1α, IL-1β, IL-6, IL-8, MCP-1, GM-CSF, TNF-α and epithelial neutrophil-activating peptide-78 (ENA-78) with reduction of anti-inflammatory cytokines transforming growth factor-beta (TGF-β) (49).

The role of *V.cholerae* in initiating and supporting the innate inflammatory response in intestinal epithelial cells as well as the potential contribution of specific *V.cholerae* components to cytokine induction is still largely unknown. The specific components include lipopolysaccharide (LPS), any secreted protein of *V.cholerae*, including CT, and the major surface proteins of *V.cholerae*. It has been reported that in addition to certain strains of *V.Cholerae*, CT may also stimulate a modest intestinal inflammatory response (54). The adjuvanticity of CT is associated with its ability to provoke the local production of proinflammatory cytokines from various types of cells that create an immunological microenvironment suitable for the induction of systemic and mucosal immune responses

(55). Cholera toxin has shown significant immuno-modulatory properties and stimulates Th1/Th2 cells with a prejudice towards Th2 cells. Thus, the mechanisms of cholera toxin adjuvanticity can be elucidated by studying the cholera toxin pro-inflammatory activity.

In addition to inducing an acute influx of PMNs, *V.cholerae* infected intestinal epithelial cells may play a significant role to coordinate the initiation of mucosal inflammatory and immune responses in which a number of different types of cells might participate (50). Localized production of pro-inflammatory cytokines is a general feature of the innate immune response to microbial pathogens, leading to the influx and activation of neutrophils and macrophages at the site of infection, resulting in inflammation (56). For intestinal pathogens, epithelial cells are the first site of entry and provide the underlying mucosa with early signals via the release of pro-inflammatory and inflammatory cytokines for the acute mucosal inflammatory response. The intestinal response upon infection by pathogens characterizes a complex interaction between non-specific inflammatory mechanisms and immunologically specific adaptive events. Epithelial cells produce pro-inflammatory cytokines after exposure to infection that can recruit and activate inflammatory cells. Epithelial cells express some accessory molecules, in addition to producing soluble factors that stimulate the adherence and activation of immune and inflammatory cells. Thus, epithelial cells regulate the host response by transducing signals from luminal pathogens to adjacent immune and inflammatory cells (57).

1.6. AIM OF THE PRESENT STUDY:

The present project aims towards the study of toxin proteases of *V. cholerae* in order to have better understanding of pathogenesis using biochemical and bioinformatics techniques as the signaling cascades involved in the induction and regulation of mucosal inflammatory responses to infection by *V.cholerae* are still largely unknown. The proposed work is carried out in two phases. The first phase deals the determination of mRNA expression of proinflammatory, anti-inflammatory and other cytokines in intestinal tissue and bone marrow cells upon infection with purified Cholera Toxin (CT) alone and in combination with purified neuraminidase from *vibrio cholerae* in order to elucidate the mechanisms of the *V.cholerae* induced proinflammatory response.

The transcriptional profiling of different cytokines i.e IL-1α, IL1β, IL-2, IL3, IL-4, IL-5, IL-6, IL-7, IL-10, IFNγ, TNFα, GM-CSF, TGFβ, MCP-1 and HIF, upon CT infection in the intestinal and one marrow cells for 2, 6 18 and 48 hrs is carried out by Reverse Transcriptase-PCR (RT-PCR).

CHAPTER

2.0

MATERIAL & METHODS

2.1. REAGENTS:

All reagents used were of molecular biology grade. The details of each reagent and buffer preparation are given in Appendix I.

2.2. ANIMALS:

Wistar rats of both sexes weighing 200-220g were used in all experiments. Rats were obtained from The Animal Facility of Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi. The rats were housed in an environmental room at 22^0C and 58-65% relative humidity with free access to water. The rats were deprived of food for 2, 6, 18 and 48hrs after immunization. All experiments were performed with the prior approval from the local research ethical committee.

2.3. IMMUNIZATION:

Animals were divided into two groups. Rats of group I were orally immunized with 1mL of sterile solution containing 10ug of CTA and 5ug of CTB subunits by blunt end feeding needle. Rats of group II were immunized with 1mL solution containing 10ug of CTA, 5ug CTB and 44ul (1mg Protein/1ml) of neuraminidase. Rats immunized with or without sterile PBS served as controls.

2.4. ISOLATION OF RNA FROM INTESTINAL TISSUE AND BONE MARROW CELLS:

a) INTESTINAL TISSUE

Rats were killed by cervical dislocation at intervals of 2hrs, 6hrs, 18hrs and 48hrs. The abdomen was sterilized with 70% ethanol. An incision was made in the midline of the

abdomen. The complete length of small intestine was removed and flushed with PBS using 3mL syringe with 25-gauge needle for at least 5-8 times in a petri dish. Small intestine was slit open longitudinally and cut into pieces of 2-3 mm lengths. The samples were immediately put in the RNA Later Reagent (500-1000mg /100ul) and stored at -20^{0}C until used further.

The total RNA from the intestinal tissue was extracted using guanidium thiocyanate/phenol/choloroform method which is the modified method of "single-step RNA isolation method developed by Chomczynski and Sacchi", (1987). Total RNA was isolated from the RNAlater-preserved intestinal tissues. Tissue samples (50-100mg) were homogenized in 1mL of TRIzol reagent using tissue homogenizer. The supernatant was collected by centrifugation at 10,000rpm for 10mins at 2-8^{0}C. Samples were incubated for 5mins at 15-30^{0}C to permit the complete dissociation of nucleoprotein complexes. 0.2mL of chloroform per 1mL of TRIzol reagent was added to the samples. Tubes were capped tightly and shaken vigorously for 15sec and incubated at 15-30^{0}C for 2-3mins. The mixture was then centrifuged at 10,000rpm for 15mins at 2-8^{0}C and separated into a lower red, phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. RNA remained exclusively in the aqueous phase. The aqueous phase was then transferred to a new 2mL eppendorf tube. The RNA was precipitated from the aqueous phase by adding 0.5mL isopropyl alcohol. Samples were incubated at 15-30^{0}C for 10mins and centrifuged at 10,000rpm for 10mins at 2-8^{0}C. The RNA was observed in the form of gel-like pellet on the side and bottom of the tube. RNA pellet was washed once with 1mL 75% ethanol and mixed by vortexing and then centrifuged at 8,000rpm for 5mins at 2-

8^0C. The RNA pellet was dried briefly for 5-10mins. The RNA pellet was dissolved in 30-50 ul of RNase-free water and stored at -20^0C.

b) BONE MARROW CELLS

Bone marrow cells were isolated from tibia and femur of Wistar rats by flushing the bones with PBS using a 3mL syringe with a 25-gauge needle. Cells were pipetted up and down to bring them into single-cell suspension and centrifuged at 1400 rpm for 6 mins. The cell pellet was colleted for RNA isolation.

Total RNA was extracted from the bone marrow cells using the RNeasy mini Kit (Cat No#74124. Qiagen Inc., Valencia, CA, USA) using spin column method according to the manufacturer's instructions. Briefly, 1 x 10^7 cells were lysed in 350ul RLT buffer. The lysate was directly pipetted onto a QIAshredder mini spin column assembly (supplied with kit) and centrifuged for 2 min at maximum speed. 350 µl of 70% ethanol was then added to the lysate, and mixed well by pipetting. 700 µl of the sample was transferred to an RNeasy mini column placed in a 2ml collection tube, centrifuged for 15sec at 10,000rpm. Subsequent steps involve addition of 700 µl of Buffer RW1 and 500 ul of Buffer RPE to the RNeasy column and spinning the column each time for 15sec at 10,000 rpm. 500 µl of Buffer RPE was again added to the column and centrifuged for 2mins at 10,000 rpm. The column was centrifuged for 1min at full speed to remove all traces of ethanol. Finally, 30-50 µl RNase-free water was directly pipetted onto RNeasy column to elute RNA and centrifuged for 1 min at 10,000 rpm. Eluted RNA samples were stored at -20^0C until further use.

QUANTIFICATION OF ISOLATED RNA:

The concentration and purity of RNA samples were determined spectrophotometrically. 3ul of RNA samples were diluted up to 600ul (1:200) with deionized water. The concentration of RNA was determined by absorbance at 260nm.

$$[RNA]\ ug/ml = A_{260} \times Dilution \times *40.0$$

Where: A_{260} = absorbance at 260nm.

* An absorbance of 1 unit at 260 nm corresponds to 40 ug of RNA per ml.

The absorbance ratio (A_{260}/A_{280}) of all preparations of purified RNA was between 1.8 and 2.0.

2.5. FIRST STRAND cDNA SYNTHESIS:

First strand cDNA was synthesized from 2ug samples of total RNA using RevertAid™ first strand cDNA synthesis kit (Fermentas-Life Sciences) with oligo$(dT)_{18}$ primer (0.5ug/ul), according to the manufacturer's instructions. 2ug of total RNA was used for cDNA synthesis.

The following reaction mixture was prepared.

- Total RNA 2ug
- oligo$(dT)_{18}$ primer(0.5ug/ul) 2ul
- DEPC-Treated water 24ul

The mixture was incubated at 70^0C for 5mins, chilled on ice and collected by brief centrifugation.

Following components were then added to the mixture:

5X reaction buffer	8ul
RiboLock™ Ribonuclease inhibitor (20u/ul)	2ul
10mM dNTP mix	4ul
(10mM of each dGTP, dATP, dTTP, dCTP)	

The reaction mixture was incubated at 37^0C for 5mins and then the 2ul RevertAid™ M-MuLV Reverse Transcriptatse (200u/ul) was then added. The reaction mixture was incubated at 42^0C for 60min. The reaction was stopped by heating at 70^0C for 10mins and then chilled on ice.

2.6. PCR PRIMERS AND AMPLIFICATION CONDITIONS:

PCR PRIMERS: Primers used in this study along with their, sequence, annealing temperatures and product sizes are shown in Table 2.1:

The oligonucleotide primers for each of the selected cytokines were designed using program Primer3 (http://frodo.wi.mit.edu/primer3/). Oligonucleotide primers were synthesized by GenLink Inc. 2ul of cDNA was amplified in a 50ul reaction volume using GoTaq Flexi DNA Polymerase Kit (Promega, Madison, USA), containing 5X Green GoTaq Flexi Buffer (10mM Tris-HCl, pH8.3, 50mM KCl), 25mM $MgCl_2$, 10mM of dNTP mix and Go*Taq* DNA polymerase (500U; 5u/ul). Glyceraldehydes-3-phosphate dehydrogenase (GAPDH) was used as a reference gene in all experiments. Hot start RT-PCR was used to increase specificity of amplification.

PRIMERS RECONSTITUTION: The volume of the TE buffer to be added was calculated by multiplying the "nmoles" of each primer by 10 to get 100pmols/ul of stock solution. The primers were diluted 10 fold to prepare a 10pmols/ul. 20pmol/ul primer was used for PCR reaction.

PCR reactions were carried out as follows:

Component	**per sample**	**Final Concentration**
5X Green GoTaq		
Flexi Buffer	10ul	1X
$MgCl_2$ solution, (25mM)	3ul	1.5mM
PCR Nucleotide Mix, 10mM each	1ul	0.2mM each
Upstream Primer	1ul	4pmols
Downstream Primer	1ul	4pmols
Template DNA	2ul	<0.5ug (0.2-0.3ug)/50ul
Go*Taq* DNA		
Polymearase (5u/ul)	0.25ul	1.25ul
Nuclease-Free water	31.75ul	
Final Volume	50ul	

Table.2.1. Primers used in this study along with their, sequence, annealing temperatures and product sizes.

PCR primers*	Sequences	Annealing Temperatures (°C)	Product sizes (BP)
IL-1α (F)	5'-GTCAGTCTTTTGCCCTCCTG-3'	60	179
IL-1α (R)	5'-CAAATTTGGAACCCAGAGGA-3'		
IL-1β (F)	5'-AGGCTTCCTTGTGCAAGTGT-3'	57	230
IL-1β (R)	5'-TGAGTGACACTGCCTTCCTG-3'		
IL-2 (F)	5'-AAACTCCCCATGATGCTCAC-3'	54	227
IL-2 (R)	5'-TCATCGAATTGGCACTCAAA-3'		
IL-3 (F)	5'-ATGCCCACCATTTACTCAGG-3'	60	302
IL-3 (R)	5'-GCAGCTGCAGGAATACAACA-3'		
IL-4 (F)	5'-TCCTTACGGCAACAAGGAAC-3'	57	209
IL-4 (R)	5'-CAGTGTTGTGAGCGTGGACT-3'		
IL-5 (F)	5'-GTTGACGAGCAATGAGACGA-3'	56	250
IL-5 (R)	5'-CATCACGCCAAGGAACTCTT-3'		
IL-6 (F)	5'-GCCCTTCAGGAACAGCTATG-3'	55	240
IL-6 (R)	5'-CAGAATTGCCATTGCACAAC-3'		
IL-7 (F)	5'-GTGAACTGCACAAGCAAGGA-3'	57	203
IL-7 (R)	5'-GAAACTTCTGGGAGGGTTCC-3'		
IL-10 (F)	5'-GAATTCCCTGGGAGAGAAGC-3'	55	219
IL-10 (R)	5'-CGGGTGGTTCAATTTTTCAT-3'		
IFN-γ (F)	5'-AGGAAAGAGCCTCCTCTTGG-3'	58	245
IFN-γ (R)	5'-GGATCTGTGGGTTGTTCACC-3'		

TNF-α (F) TNF-α (R)	5'-ACGATGCTCAGAAACACACG-3' 5'-CAGTCTGGGAAGCTCTGAGG-3'	58	244
TGF-β (F) TGF-β (R)	5'-TTGCTTCAGCTCCACAGAGA -3' 5'-TGGTTGTAGAGGGCAAGGAC-3'	53	182
MCP-1 (F) MCP-1(R)	5'-TGTCTCAGCCAGATGCAGTT-3' 5'-TTCCTTATTGGGGTCAGCAC-3'	56	179
HIF-1α (F) HIF-1α (R)	5'-CCCAATGGATGATGATTTCC-3' 5'-TGGGTAGAAGGTGGAGAGC-3'	54	243
GM-CSF (F) GM-CSF (R)	5'-GGCCCTGGAAGCATGTAGAT-3' 5'-GGTAGTGGCTGGCTATCATG-3'	58	221
GAPDH (F) GAPDH (F)	5'-GGAAAGCTGTGGCGTGATGG-3' 5'-GTAGGCCATGAGGTCCACCA-3'	60	414

PCR AMPLIFICATION CONDITIONS:

PCR amplification conditions were as follows: 30 cycles of PCR were performed

Step	Temperature	Time
Initial denaturation	95^0C	1min
Denaturation	95^0C	1min
Annealing	56-67^0C*	1min
Extension	72^0C	1min
Final Extension	72^0C	10min

* Annealing temperature for each primer set was based on Tm of the primer, which was calculated as $T_m = 4\ (G+C) + 2\ (A+T)$

2.7. AGAROSE GEL ELECTROPHORESIS:

PCR products were separated by 1 % agarose gel electrophoresis containing adding 4 ul 10mg / ml ethidium bromide (0.4 ug / ml) in 1X TBE buffer. 10ul samples of the PCR products were applied onto the gel. Electrophoresis was carried out at 70Volts for 90mins.The gel was visualized, scanned and recorded using a gel documentation system (Model-GeneGenius S.No. SYDR/2138. Syngene Synoptics Co.UK). The integrated density of each band was normalized to the corresponding GAPDH band.

CHAPTER

3.0

RESULTS

3.1. TRANSCRIPTIONAL PROFILING OF CYTOKINES:

The mRNA levels of different cytokines were analyzed in intestinal tissue upon toxin infection at various time points by Reverse Transcriptase-PCR (RT-PCR). These cytokines belong to varied functional categories i.e. interleukins (IL-1α, IL-6), chemokines (MCP-1), interferons (IFN-γ), tumor necrosis factors (TNF- α). The concentration of all RNA samples was determined by the absorbance at 260nm. The absorbance ratio (A_{260}/A_{280}) of all preparations of purified RNA samples was between 1.8 and 2.0. For each sample, 2ug of total RNA was used for cDNA synthesis. <0.5ug of the template DNA was amplified in a 50ul reaction mixture.

The gene expression profiles were analyzed for the time periods of 2, 6, 18 and 48h of CT infection alone or in combination with neuraminidase. The products were run on a 1% agarose gel and stained with ethidium bromide and visualized and scanned by gel documentation system (Model-GeneGenius S.No. SYDR/2138. Syngene Synoptics Co.UK). The integrated density of each band was compared with the corresponding GAPDH band. We could not quantify the expression of expressed cytokines as we analyzed their expression by Reverse-Transcriptase PCR rather the Real-Time PCR. The level of expression of each mRNA was determined in accordance to the level of expression of GAPDH mRNA.

3.1.1. Effect of CT and Neuraminidase infection on the expression of cytokines in rat intestinal tissue:

a) GAPDH:

GAPDH was used as internal standard and was analyzed in rat intestinal tissues after various toxin treatments as well as in control. The little variation in density of these bands shows that the initial amount loaded on to the gel was not uniform (Fig.3.1.1).

b) Interleukin-1 α (IL-1 α):

The pro-inflammatory IL-1α mRNA expression was detected as early as 2h following CT infection while very low expression was observed at 6h, 18h and 48h after infection. Similar expression profile was observed following CT infection in combination with neuraminidase (Fig.3.1.2).

c) Interleukin-1β (IL-1β) :

Significant up-regulation of IL-1β mRNA was detected at 2h after CT infection. The expression increased at 6 and 18h after-infection and then declined at 48h. Similar pattern for IL1-β mRNA expression was observed following CT infection in combination with neuraminidase (Fig.3.1. 3).

d) Interleukin-7 (IL-7):

Rats treated with CT alone and in combination with neuraminidase induced IL-7 mRNA expression in the intestinal tissue uniformly at all time points i.e. 2, 6 18 and 48 h (Fig.3.1.4).

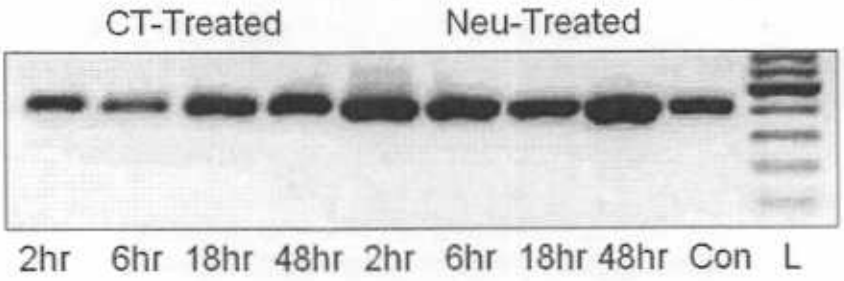


Fig.3.1.1. GAPDH mRNA expression in intestinal tissue in control and following CT and neuraminidase infection.

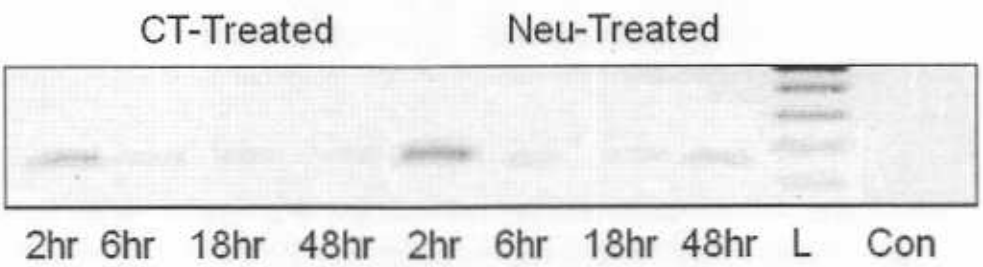


Fig.3.1.2. IL-1α mRNA expression in intestinal tissue following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of IL-1α after CT infection alone and in combination with neuraminidase.

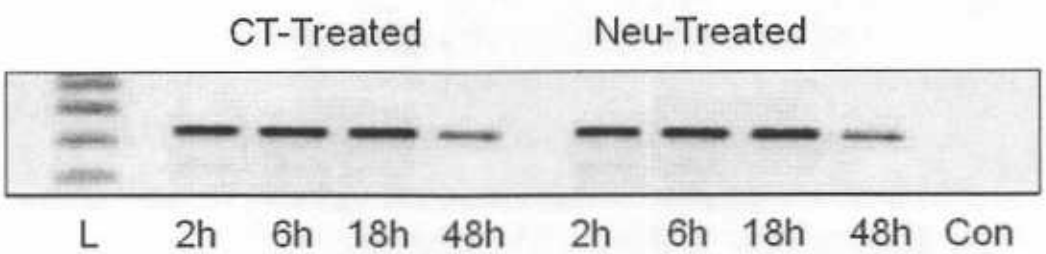


Fig.3.1.3. IL-1β mRNA expression in intestinal tissue following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of IL-1β after CT infection alone and in combination with neuraminidase.

e) Inteferon-γ (IFN- γ):

The mRNA level of immuno-regulatory cytokine, IFN-γ was markedly increased at 2, 18 and 48h after CT infection while the low expression was observed at 6h after infection. The expression was decreased at 2h and 48h while it was up-regulated at 6 and 18h after infection when CT was administered in combination with neuraminidase (Fig.3.1.5).

f) Tumor Necrosis Factor (TNF-α):

TNF-α showed increased expression at 6, 18 and 48h after CT infection. Similar cytokine expression profile was observed upon infection with CT in combination with neuraminidase (Fig.3.1.6).

g) Transforming Growth Factor-β (TGF-β):

The mRNA expression of anti-inflammatory cytokine TGF-β was decreased at all time points after infection with CT alone or in combination with neuraminidase as compared to control (Fig.3.1.7).

h) Monocyte chemotactic protein-1 (MCP-1):

Increased expression of C-C chemokine, MCP-1 mRNA was observed at all time points upon CT infection except 18h after CT infection where the expression was slightly decreased. When CT was administered in combination with neuraminidase, lower expression was observed 2 and 48h after infection and up-regulation in expression 6 and 18h after infection (Fig.3.1.8).

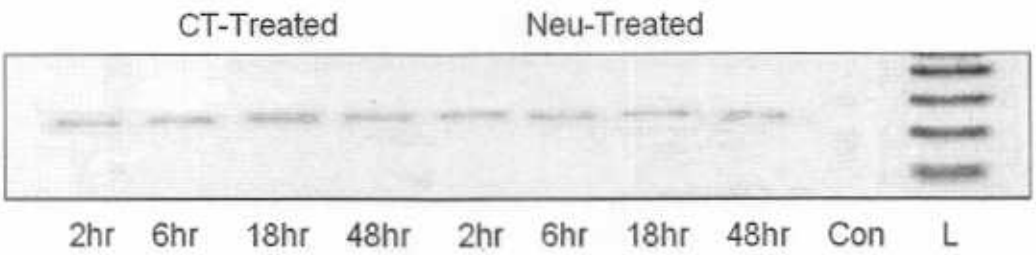


Fig.3.1.4. IL-7 mRNA expression in intestinal tissue following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of IL-7 as determined by RT-PCR after CT infection alone and in combination with neuraminidase.

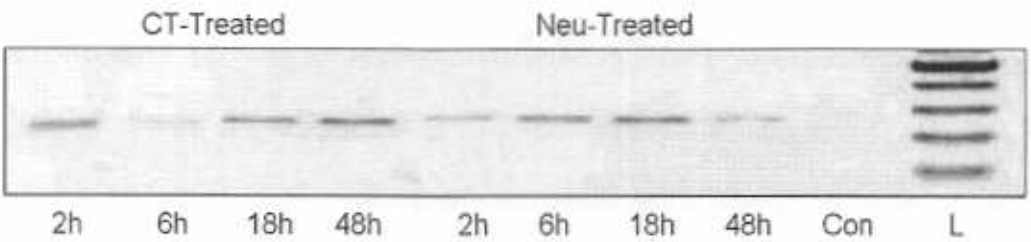


Fig.3.1.5. IFN-γ mRNA expression in intestinal tissue following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of IFN-γ after CT infection alone and in combination with neuraminidase

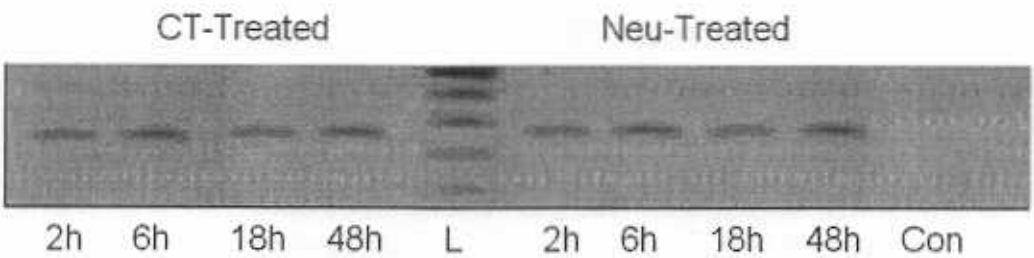


Fig.3.1.6. TNF-α mRNA expression in intestinal tissue following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of TNF-α after CT infection alone and in combination with neuraminidase.

i) Hypoxia Inducing Factor-1α (HIF-1α):

Up-regulation of HIF-1α mRNA was observed as early as 2h after infection. Thereafter uniform decline at 6, 18 and 48h after infection was observed. Similar expression profile was observed upon CT infection when treated in combination with neuraminidase (Fig.3.1.9).

3.2. Effect of CT and Neuraminidase infection on the expression of cytokines in rat bone marrow cells:

To obtain a more detailed analysis of the toxin induced cytokine network upon infection, the expression of cytokines belonging to various functional categories was analyzed further in bone marrow cells following infection alone with CT and in combination with neuraminidase. The mRNA levels of different cytokines were analyzed in bone marrow cells upon toxin infection at various time points by Reverse Transcriptase-PCR (RT-PCR). These cytokines belong to varied functional categories i.e. interleukins (IL-1α, IL-6), chemokines (MCP-1), interferons (IFN-γ), tumor necrosis factors (TNF- α). The concentration of all RNA samples was determined by the absorbance at 260nm. The absorbance ratio (A_{260}/A_{280}) of all preparations of purified RNA samples was between 1.8 and 2.0. For each sample, 2ug of total RNA was used for cDNA synthesis. <0.5ug of the template DNA was amplified in a 50ul reaction mixture.

The gene expression profiles were analyzed for the time periods of 2, 6, 18 and 48h of CT infection alone or in combination with neuraminidase. The products were run on a 1% agarose gel and stained with ethidium bromide and visualized and scanned by gel documentation system (Model-GeneGenius S.No. SYDR/2138.Syngene Synoptics

Co.UK). The integrated density of each band was compared with the corresponding GAPDH band. We could not quantify the expression of expressed cytokines as we analyzed their expression by Reverse-Transcriptase PCR rather the Real-Time PCR. The level of expression of each mRNA was determined in accordance to the level of expression of GAPDH mRNA.

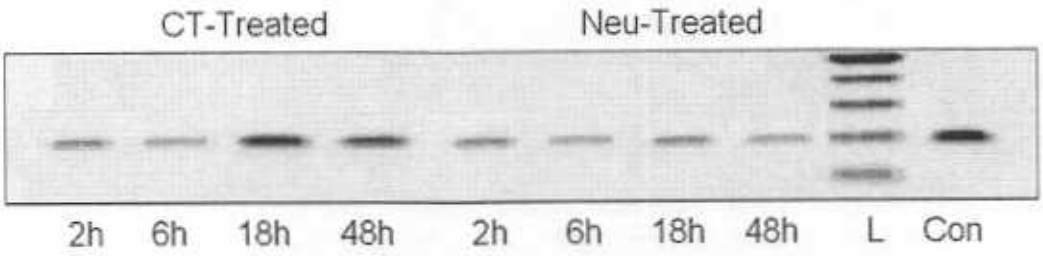


Fig.3.1.7. TGF-β mRNA expression in intestinal tissue following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of TGF-β after CT infection alone and in combination with neuraminidase.

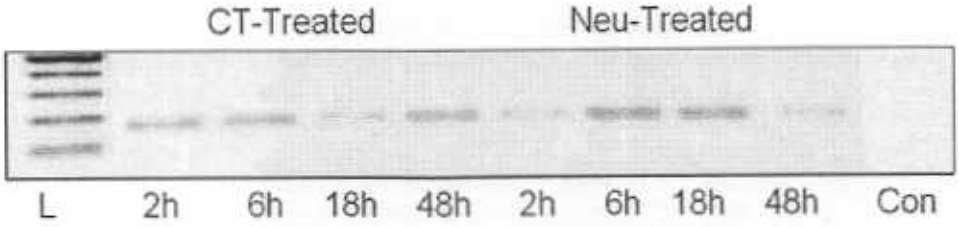


Fig.3.1.8. MCP-1 mRNA expression in intestinal tissue following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of MCP-1 after CT infection alone and in combination with neuraminidase.

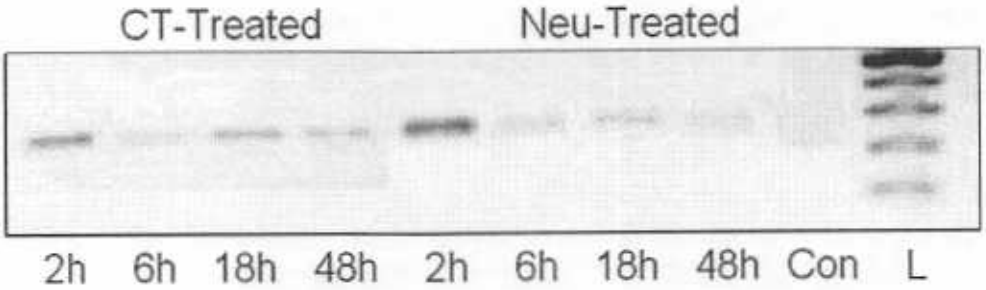


Fig.3.1.9. HIF-1α mRNA expression in intestinal tissue following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of HIF-1α after CT infection alone and in combination with neuraminidase.

a) GAPDH:

GAPDH was used as internal standard and was analyzed in rat bone marrow samples after various treatments as well as in control. The little variation shows in density of these bands shows that the initial amount loaded on to the gel was not uniform. (Fig.3.2.1).

b) Interleukin-1 α (IL-1 α):

The pro-inflammatory IL-1α mRNA expression was detected as early as 2h following CT infection while very low expression was observed at 6h, 18h and 48h after infection. Similar expression profile was observed following CT infection in combination with neuraminidase (Fig.3.2.2).

c) Interleukin-1 β (IL-1 β):

Significant up-regulation of pro-inflammatory cytokine IL-1β mRNA was observed at 2h after infection. The expression decreased at 6 and 18h after infection and then slightly increased at 48h after infection. In contrast, when rats were treated with CT in combination with neuraminidase, IL1-β mRNA showed uniform up-regulation at 2, 6 and 18h after infection while the expression was decreased at 48h after infection (Fig.3.2.3).

d) Interleukin-7 (IL-7):

Rats treated with CT alone or in combination with neuraminidase slightly induced IL-7 mRNA expression in the bone marrow cells uniformly at all time points i.e. 2, 6 18 and 48 h (Fig.3.2.4).

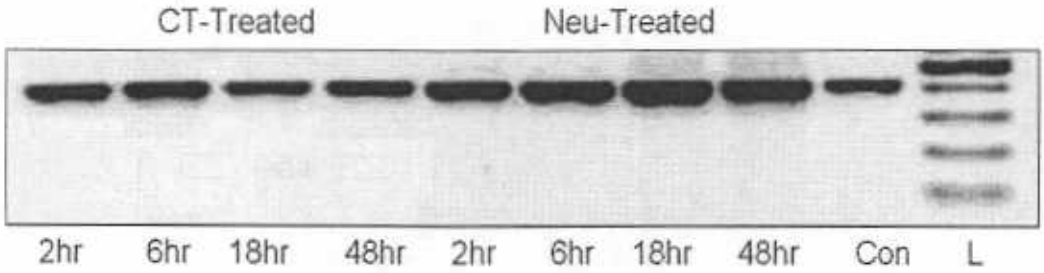


Fig.3.2.1.GAPDH mRNA expression in rat bone marrow cells in control and following CT and neuraminidase infection as analyzed by RT-PCR.

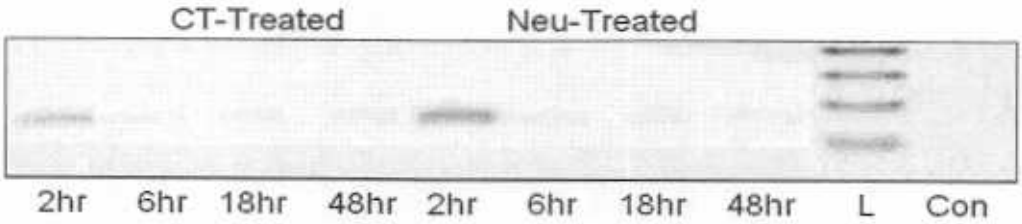


Fig.3.2.2. IL-1α mRNA expression in bone marrow cells following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of IL-1α after CT infection alone and in combination with neuraminidase

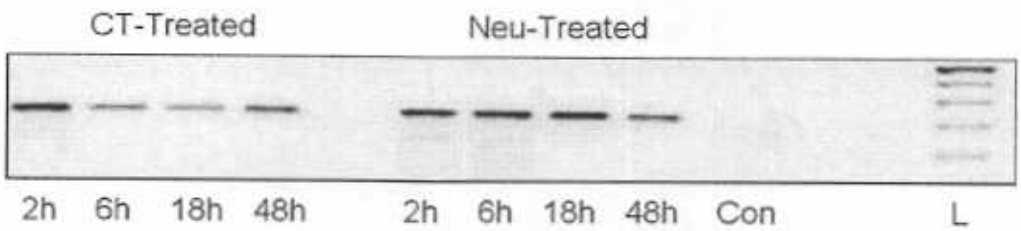


Fig.3.2.3. IL-1β mRNA expression in bone marrow cells following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of IL-1β after CT infection alone and in combination with neuraminidase.

e) Inteferon-γ (IFN- γ):

The mRNA level of immuno-regulatory cytokine IFN-γ was markedly increased upon infection with CT at 2, 6 and 18h after infection while it decreased at 48h after infection. In contrast, when rats were treated with CT in combination with neuraminidase, the expression was increased at 2h and increased further at 6 and18h after infection. It decreased again at 48h (Fig.3.2.5).

f) Tumor Necrosis Factor (TNF-α):

TNF-α showed increased expression in bone marrow cells after 2h of CT infection and then declined thereafter at 6, 18 and 48h after infection. When treated with CT in combination with neuraminidase, the mRNA of TNF-α was up-regulated at 2h, increased further at 6h and 18h and decreased slightly at 48h after infection (Fig.3.2.6).

g) Transforming Growth Factor-β (TGF-β):

The mRNA levels of anti-inflammatory cytokine TGF-β was almost similar at all time points as compared to control when treated with CT alone or in combination with neuraminidase (Fig.3. 2.7).

h) Monocyte chemotactic protein-1 (MCP-1):

Relatively low expression of MCP-1 mRNA was observed after in CT and neuraminidase infection. MCP-1 expression was slightly increased at 2h following CT infection and then decreased thereafter at all time points. When treated with CT along with neuraminidase, decreased expression after 2 and 48h and increased expression after 6 and 18h was observed (Fig.3. 2.8).

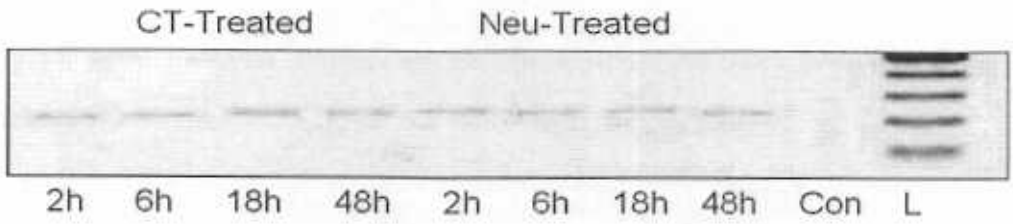


Fig.3.2.4 .IL-7 mRNA expression in bone marrow following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of IL-7 as determined by RT-PCR.

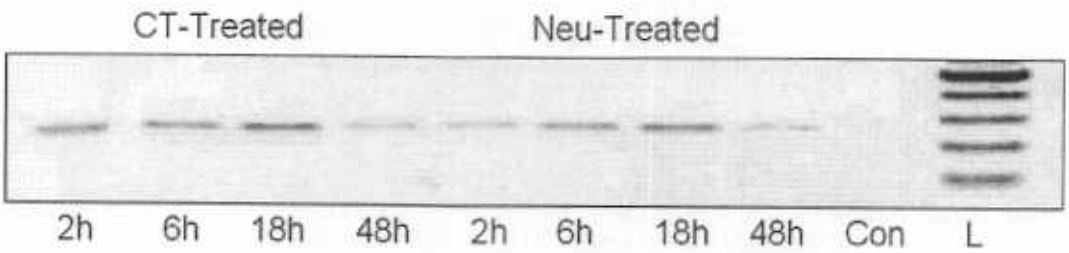


Fig.3.2.5. IFN-γ mRNA expression in bone marrow following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of IFN-γ.

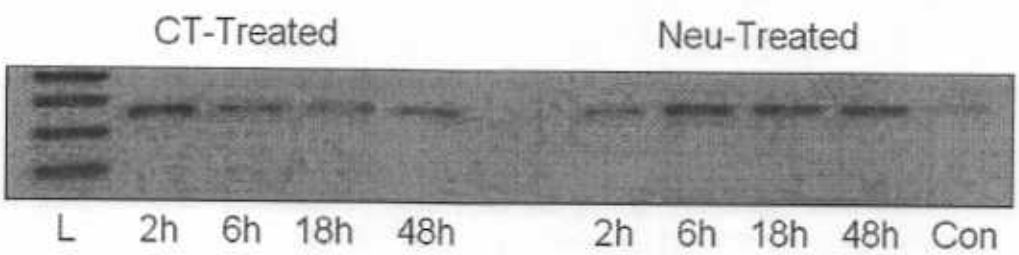


Fig.3.2.6.TNF-α mRNA expression in bone marrow cells following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of TNF-α after CT infection alone and in combination with neuraminidase.

i) Hypoxia Inducing Factor-1α (HIF-1α):

HIF-1α was detected at 2h following CT infection and then decrease slightly thereafter at all time points. In contrast, increased expression was observed at after 6 and 18h after infection as compared to 2h and 48h when CT was administered along with neuraminidase (Fig.3.1.9).

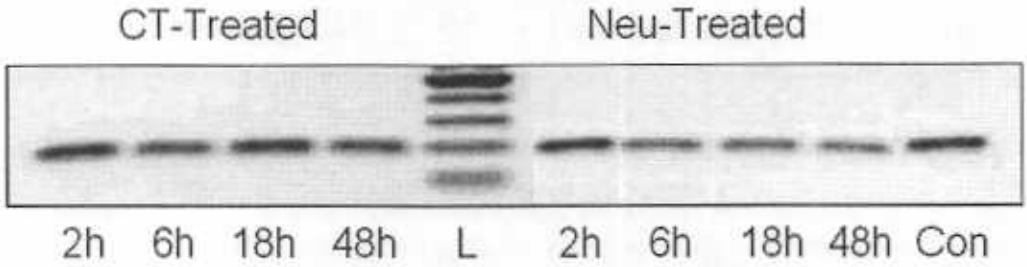


Fig.3.2.7. TGF-β mRNA expression in bone marrow cells following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of TGF-β after CT infection alone and in combination with neuraminidase

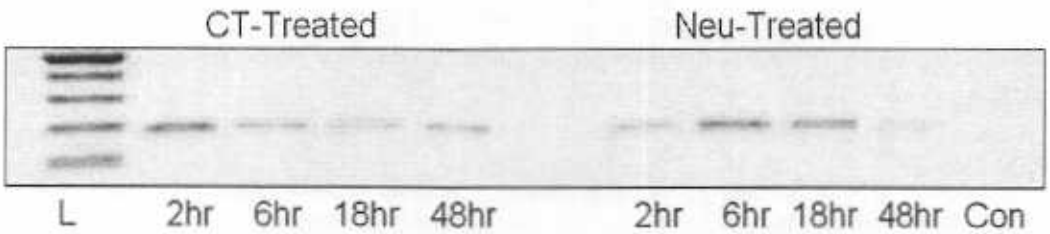


Fig.3.2.8. MCP-1 mRNA expression in bone marrow cells following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of MCP-1 after CT infection alone and in combination with neuraminidase.

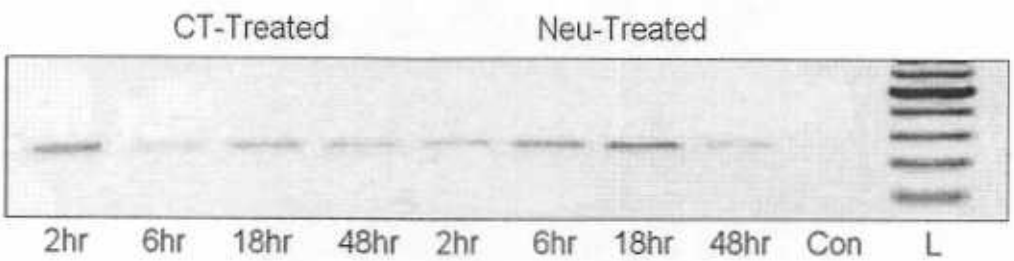


Fig.3.1.9. HIF-1α mRNA expression in bone marrow cells following CT infection. A representative agarose gel analysis shows 2, 6, 18 & 48h expression of HIF-1α after CT infection alone and in combination with neuraminidase.

3.3. No expression of IL-2, IL-3, IL-4, IL-5, IL-6, IL-10, SCF & GM-CSF:

The level of mRNA expression of IL-2, IL-3, IL-4, IL-5, IL-6, IL-10, SCF & GM-CSF could not detected upon infection with CT as well as when CT was administered along with neuraminidase in both intestinal tissues and bone marrow cells (Fig.3.3.1 & Fig.3.3.2)

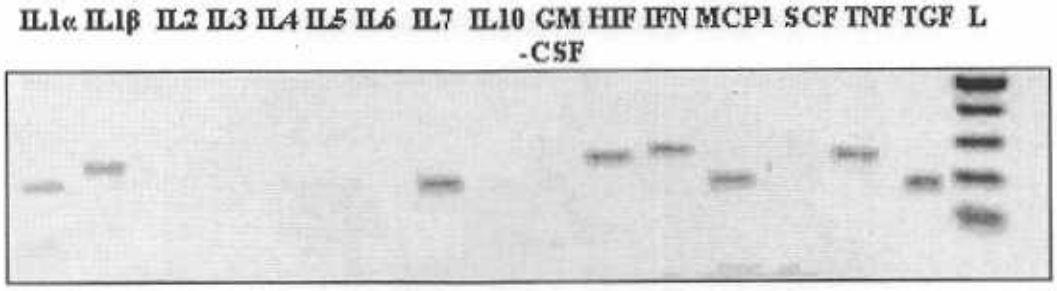


Fig.3.3.1. mRNA expression of all cytokines studied in intestinal tissue following CT infection. A representative agarose gel analysis shows 6h expression of all cytokines after CT infection alone and in combination with neuraminidase.

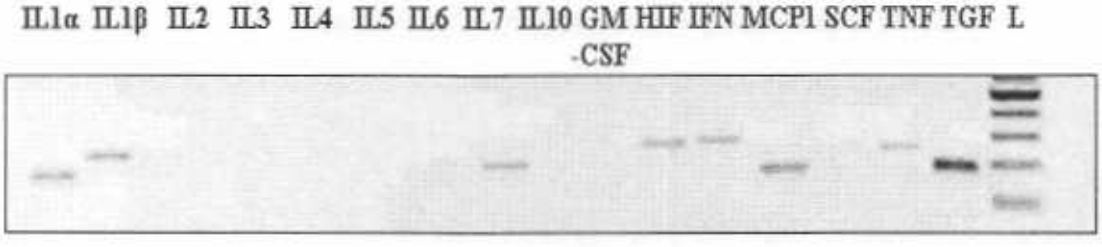


Fig.3.3.2. mRNA expression of all cytokines studied in bone marrow cells following CT infection. A representative agarose gel analysis shows 6h expression of all cytokines after CT infection alone and in combination with neuraminidase.

CHAPTER

4.0

DISCUSSION

4.1. EXPRESSION OF CYTOKINES/CHEMOKINES IN RESPONSE TO CT AND NEURAMINIDASE.

Epithelial cells are thought to characterize an integral component of mucosal immune system, as they provide the underlying mucosa with the early signals for an infection through the acute inflammatory response via the release of inflammatory and pro-inflammatory cytokines (50,54). The intestinal mucosal epithelial cells therefore act as the first barrier for intestinal pathogens. The overall response of intestinal cells involves an array of interactions between inflammatory and immunologic events through functional cytokines. Various reports have demonstrated that the production and secretion of inflammatory cytokines are stimulated by various *V. cholerae* strains as well as CT, thus initiating a modest intestinal inflammatory response (54). The exact mechanism of how individual *V. Cholerae* initiate and sustains the innate inflammatory response as well as how its components exert their effect through cytokine expression in intestinal epithelial cells is still largely unknown (49-50,54). It has been hypothesized Cholera toxin bind to the monosialoganglioside (GM1) receptor on host intestinal epithelial cells and neuraminidase may produce locally high concentration of the GM1 receptor for CT, thereby enhancing the binding and internalization of CT (24,26,28). We therefore, attempted to analyze the transcriptional profile of pro-inflammatory cytokines in rat intestinal tissue after ingestion of purified CT alone or in combination with neuraminidase in order to identify the specific cytokines that are involved in the process as well as to elucidate the effect of neuraminidase in inducing the cytokine expression when administered with CT. We studied the cytokine expression in bone marrow cells as well after ingestion of the toxin as these cells could contribute towards the immunologic

response or substantiate the role of intestinal cells in its inflammatory response towards infection. The expression of the cytokines belonging to varied functional groups such as IL-1α, IL-1β, IL-6, IL-7, TNF-α, IFN-γ, GM-CSF (pro-inflammatory), MCP-1 (chemokine), IL-2, IL-5 (Acquired specific immune response), IL-4(anti-inflammatory, Acquired specific immune response), IL-10, and TGF- β (anti-inflammatory).Our studies demonstrate that among the 15 cytokines involved in proinflammatory, anti inflammatory and antigen-specific immune responses, the mRNA of the seven cytokines IL-1α, IL1β, IL-7, IFNγ, TNFα, MCP-1, and HIF were expressed in the intestinal tissue and bone marrow cells whereas rest of the cytokines were either down regulated or not expressed. The induction of these cytokines was also time dependent. In case of expression of the cytokines, the untreated control showed no expression, suggesting that this effect is not a general inflammatory response but is due to CT infection. The expression of inflammatory group of cytokines in rat intestinal epithelial tissue and bone marrow cells was significant emphasizing that these cytokines have a major contribution in the initiation and amplification of the inflammatory response due to toxin infection.

4.1.1 Effect on the expression pro-inflammatory cytokines IL-1α & IL1β:

The mRNA of IL-1α was moderately expressed upon CT administration alone or in combination with neuraminidase both in intestinal tissue as well as in bone marrow cells. The expression was maximal at 2h after infection and then there was a slight reduction in the expression of IL-1α at all time points. It can therefore be suggested that CT is the moderate inducer of IL-1α. It can also be suggested that in addition to CT, other factors/components of *V. Cholerae* are also required for its expression.

On the other hand, production and release of IL-1β in the intestinal epithelial tissue and bone marrow cells were markedly stimulated upon toxin infection. The expression profile of IL-1β revealed that it was maximally induced in intestinal tissue upon infection with CT alone as well as in combination with neuraminidase whereas in bone marrow cells the expression was intermediate upon CT infection but maximal upon CT and neuraminidase administration. Again the expression in both samples was almost constant; it sustained till 18 h and by 48 h it subsided. These results clearly reveal that CT is an effective inducer of IL-1β but the possibility of induction of IL-1β by other factors/components could not be ruled out. The CT induction of IL-1β is in agreement with earlier reports.

Pro-inflammatory cytokines IL-1α and IL-1β are produced by many cell types including fibroblasts, macrophages, PBMC, and intestinal epithelial cells. IL-1 (IL-1α & IL1β) can induce the production of many additional cytokines, including IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-12, TGF-β, TNF-α, GM-CSF that mediate additional effects, which serve to further amplify an inflammatory response (55). IL-1β, a potent immuno-modulator, is known to mediate a wide range of inflammatory and immune response i.e. the activation of B and T cells (56). IL-1β is effective for chemotaxis of leukocytes which is crucial for inflammatory response in the intestine (52).

4.1.2 Effect on the expression of IL-7:

IL-7 is a hematopoietic cytokine, which regulates development, maintenance and proliferation of T lymphocytes within the immune system. IL-7 influences cell differentiation at extrathymic sites (59). The present study shows that the IL-7 mRNA was slightly induced in both intestinal epithelial tissues and bone marrow cells upon infection with CT alone as well as in combination with neuraminidase at all the time

points. Additional sites that induce the production of IL-7 include keratinocytes, intestinal epithelium, fetal liver, adult liver, dendritic cells and follicular dentdritic cells (60).Production of IL-7 by intestinal epithelial cells is considered to give rise to a population of γδ (gamma-delta) intraepithelial T cells that produce growth factors involved in healing epithelial cells damaged by infection or inflammation in addition to fight pathogens and kill infected cells directly (59). TGF-β and IL-7 share a reciprocal relationship wherein each is capable of down-regulating the expression of the other (60). Thus, the present results revealed that CT is the moderate inducer of IL-7.

4.1.3 Effect on the expression of pro-inflammatory cytokines, TNF-α & IFN-γ:

A multipotential proinflammatory cytokine, tumor necrosis factor-α (TNF-α) is produced by monocytes, lymphocytes, macrophages and neutrophils. TNF-α has been implicated in the defense of bacterial infections and promoting the bactericidal activity of leukocytes (49,58,61).

The present study demonstrates the induction and transcription of TNF-α in response to the CT stimulus. The expression profile of TNF-α revealed that it was maximally induced in both intestinal epithelial tissue and bone marrow cells upon infection with CT alone as well as in combination with neuraminidase at all time points. TNF-α can augment the inflammatory response by activating neutrophils, mononuclear phagocytes, and other cell types such as eosinophils (58,62). Although CT exhibits no systemic manifestation of an acute-phase reaction, some of the symptoms could be due to the production of TNF-α, because as a proinflammatory cytokine, it initiates the acute-phase reaction such as fever, malaise, myalgia and cellular hypermetabolism (61).

Our study shows that the expression of IFN-γ was intermediate but uniform throughout the entire period of analysis in both the intestinal tissue and bone marrow cells. IFN-γ is produced predominantly by T-lymphocytes as part of the innate immune response. Our results indicate that though moderate, CT induces the immune response through IFN-γ which might be one of several other factors that take part in this process.

4.1.4 Effect on the expression of anti-inflammatory cytokine, TGF-β:

The present results demonstrate that the expression of the anti-inflammatory cytokine, TGF-β was down-regulated upon infection with CT. In contrast to the pro-inflammatory cytokines, synthesis of TGF-β, a cytokine which can exhibit potent immunosuppressive and anti-inflammatory activities (53-54) was down-regulated in intestinal epithelial tissue in response to CT infection and therefore it aids in the inflammatory response by the respective cytokines. In contrast to intestinal epithelial tissue, the expression profile of TGF-β in bone marrow cells showed almost no change in TGF-β expression.

4.1.5 Effect on the expression of C-C Chemokine MCP-1:

MCP-1 is a potent chemotactic agent for monocytes/macrophages, eosinophils, and subpopulations of T-cells (49,53-54). Chemokines are important biochemical mediators of the inflammatory response of the intestinal mucosa (63) The expression profile of MCP-1 revealed that it was moderately induced in both intestinal tissue and bone marrow cells upon infection with CT alone as well as in combination with neuraminidase at all time points investigated. MCP-1 is a potential regulator of monocyte recruitment in inflammations (58). It has been documented that the production of chemokines is controlled by pro-inflammatory cytokines or by Th1 and Th2 cytokines. IL-1β, IL-6 and TNF-α, the primary pro-inflammatory cytokines, generate a second wave of cytokines

including chemokines (64). The present results revealed that cytokines such as IL-1β, IFNγ and TNF-α may induce chemokines such as MCP-1, which in turn trigger and recruit cells from the monocyte/macrophage lineage to induce inflammatory response.

4.1.6. Effect on the expression of Transcriptional Factor HIF-1α:

Hypoxia inducible factor-1α (HIF-1) is a proangiogenic transcription factor stabilized and activated under hypoxia (65). The present study demonstrates that HIF-1α was expressed upon infection with CT. Though moderate, the HIF-1α was almost constant through the all time points both upon CT infection alone and in combination with neuraminidase in intestinal epithelial tissue as well as in bone marrow cells. HIF-1α expression is generally induced by oxidative stress caused by a reduced cellular oxygen level (66). The results are suggestive of the fact that there might be a direct or indirect role of the toxin in the modulating the cellular oxygen level following infection.

4.1.7 Effect on the expression of IL-2, IL-3, IL-4, IL-5, IL-6, IL-10 and GM-CSF:

In the present study we could not detect the expression of several cytokines such as IL-2, IL-3, IL-4, IL-5, IL-6, IL-10 and GM-CSF. These cytokines are commonly involved with antigen-specific acquired immunity. The present results suggest that CT play either no significant role in initiation and regulation of antigenic-specific mucosal immune response or it requires additional pathogen component to exert its effect.

As cytokine signaling involves a network of effects between cells, one group of cytokines (i.e IL-1α, TNF) induces production of other cytokines/chemokines by those cells that mediate additional effects (58). IL-1β and TNF-α are known to inhibit IL-6 induction (67) while IL-1α is known to induce the production of IL-6 and GM-CSF (55,68). The

present results demonstrate the low expression of IL-1α which may suggest the reason why IL-6 and GM-CSF were not expressed. Similarly, IFN-γ is known to inhibit the production of IL-4, an important cytokine associated with the Type 2 response.

CONCLUSIONS:

Following conclusions can be drawn from the present study:

- CT alone or in combination with neuraminidase induce the expression of cytokines/chemokines in intestinal tissue and bone marrow cells that appear to be involved in immunoregulatory, proinflammatory and stress response. These include IL-1α, IL-1β, IL-7, TNF-α, IFN-γ, MCP-1 and HIF1-α.
- According to the present study, CT is the effective inducer of IL-1β and moderate inducer of IL-1α, IL-7, TNF-α, IFN-γ, MCP-1 and HIF1-α.
- The expression of specific array of pro-inflammatory cytokines suggests that these cytokines act alone or in coordination with each other in order to attract and activate neutrophils and monocytes/macrophages, key cell types of inflammatory response thus providing an important early signaling system for the initiation and implication of the mucosal inflammatory response.
- The expression of pro-inflammatory cytokines appears to be an important component of the cellular response to cholera infection and provides further evidence towards the existence of the inflammatory component in *V.Cholerae.*
- The expression of specific class of cytokines by intestinal tissue shows that the toxin play a more significant role in initiating and regulating the innate mucosal inflammatory response relatively than antigen-specific mucosal immune response.
- In contrast to the pro-inflammatory cytokines, reduced expression of the anti-inflammatory cytokine, TGF-β and no expression of another major anti-inflammatory cytokine, IL-10 provide evidence towards the inflammatory response.

- It may be possible that for the expression of IL-2, IL-3, IL-4, Il-5, IL-6 and GM-CSF, that besides CT some other component(s) of the pathogen should be present that would lead to expression.
- The study cannot delineate the roles of separate areas of mucosal surface or villi and crypts that may lead to distinct response towards the expression of cytokines/chemokines and help in the interpretations of their specific functions as the study was done with the whole intestinal tissue and bone marrow.
- Increased expression of TNF-α & IFN-γ (Th-1 cytokines) and no expression of IL-4, IL-5, IL-6 & IL-10 (Th-2 cytokines) suggests Th-1 type cellular response. For a Th-2 type of immune response, other components might be needed.
- As the adjuvant activity of CT has been associated with its ability to induce local production of cytokines from various cell types. IFN- γ and IL-7 are known to modulate immune response in infectious diseases or tumor models, and therefore may play a key role in the CT adjuvanticity.
- The pro-inflammatory action of cholera toxin can expose important approaching towards the mechanisms of CT adjuvanticity.
- A striking similarity was observed in the nature of the cytokine expression pattern following CT infection in the intestinal tissues and bone marrow cells when CT was administered with neuraminidase suggesting that neuraminidase play no significant role in the expression of cytokines studied.

CHAPTER

1.0

INTRODUCTION

1.1. HOMOLOGY MODELING OR PROTEIN STRUCTURE MODELING

The knowledge of the tertiary structure of a protein can be dealt much more successfully to understand the protein's function, mechanism of action, and its structure-function relations, than sequence alone (69). Understanding of the 3D structure of proteins is also crucial for engineering their properties or for the rational design of site directed mutations and drugs interfering with their biological functions such as antigenic properties or binding specifities (69-70).

Two very well known experimental methods for obtaining 3D structures are X-ray crystallography & NMR spectroscopy. These methods make available comprehensive and precise structural information. In fact X-ray diffraction and NMR spectroscopy have enhanced the quality and speed of structural studies (71) but the problem still lies that both these techniques are difficult, and require expensive equipment and complicated technical procedures. In addition, many proteins fail to crystallize, a requirement for x-ray crystallography, or are inadequately soluble for NMR studies (69). Yet, current statistics still show that the known protein sequences hugely rising than the available protein structures (71).

Structural Biology now faces the laborious assignment of characterizing the shapes and dynamics of the encoded proteins to make possible the perceptive of their functions and mechanisms of action (71). Millions of proteins that have been sequenced to date come into sight to have domains that are limited in the topology they become recognizable to possibly only a few thousand folds (72). Domains with similar sequences and structures can be clustered into a relatively small number of families on the basis of evolutionary

relationships, which enable the use of computational methods to predict the structures of protein sequences such as threading and comparative protein structure modeling. In fact, many structural genomics efforts merge computational modeling techniques and experimental structure determination methods to determine enough suitably selected structures so that a good number of other known protein sequences can be placed within modeling distance of at least one known structure (71). Comparative modeling, also called homology modeling, takes advantages of topologically comparable experimental structures and generally is the best method for obtaining a more accurate model of a protein (72).

Comparative protein modeling is playing a key role in protein and peptide sciences due to the improvements in modeling methods and the huge amount of biological data becoming available. Modeling tools can often used to predict the structure, function and its underlying mechanism. It can also approaching to design experiments and recommend potential lead for drug design and used to learn biochemical processes, such as enzyme reactions and electron transfer. The entire illustration can be depicting by using computer simulations that bridge the space between experimental data and theoretical models. It is extensively documented that comparative protein modeling is a crucial component of modern molecular biology (73). The critical objective of protein modeling is to predict a structure from its sequence with an accuracy that is analogous to the best experimental results. In addition, if experimental techniques fail, protein modeling is the only way to obtain structural information (74).

1.1.1. METHODS OF 3D STRUCTURE PREDICTION:

The most commonly used and best-developed methods for tertiary structure prediction are of two main types (73).

a. **Ab-Initio Methods:** This method predicts a protein structure based on physic-chemical principles directly (75).

b. **Template Based Methods:** This method use known protein structures as templates. Template based methods include.

i. Fold Recognition via Threading.

ii. Homology or Comparative Modeling.

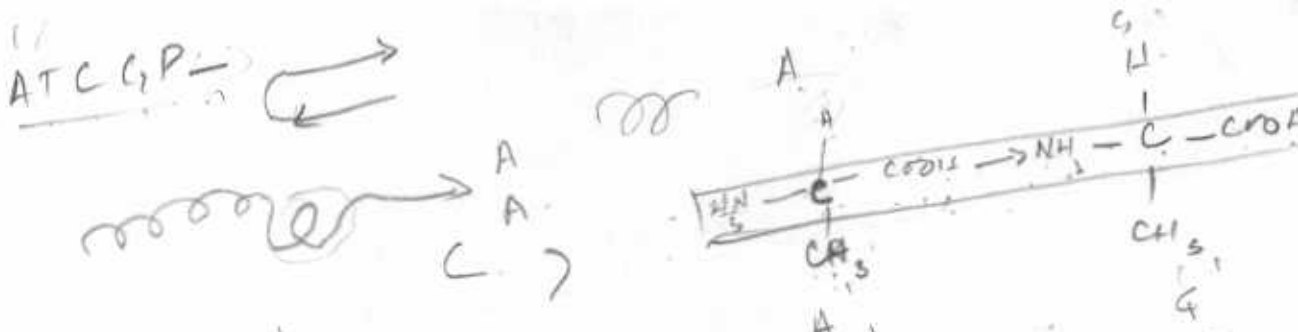
Homology Modeling:

Homology modeling is the easiest knowledge based method among the three major approahes to predict the three-dimensional structure. It is based on two major obsevations:

1. The structure of a protein is uniquely determined by its amino acid sequence.
2. At least in theory knowing the sequence should suffice to obtain the structure (74).

1.1.2. Principles of Homology Modeling:

Model building on the basis of a known 3D structure of a homologous protein in the absence of an experimental structure for a protein is documented as Knowledge-based modeling or Comparative modeling (69). It depends on the fact that the structure is more stable and changes much slower than the sequence during evolution, so that similar sequences adopt practically identical structures and distantly related sequences still fold into similar structures (74).



In 1960's the initial homology models were built using wire and plastic models of bonds and atoms. In 1969 the first available homology model structure was of the small globular protein α-lactalbumin, which was built on the basis of the 3D structure of hen egg white lysozyme as the template proteins, these two proteins shared 39% sequence identity. Later on when the X-ray crystallized structure of α-lactalbumin was solved, apart from the carboxy-terminal end the model turned out to be essentially accurate, which is different in structure to that of lysozyme (69).

The model-building procedure of more closely related sequences of the target and the template consists of simple steps and results in more precise model. The accuracy of the model tends to increase if both sequences have greater than 90% sequence identity and massive errors are often made if the sequence identity falls below 25-30% (69,71,74). In general, the additional experimental data is prerequisite requirement for modeling procedure that can help in determining the 3D structure. In homology modeling two most important suppositions are made. First, it is supposed that template and target have identical spatial coordinates if the regions of polypeptide backbone are conserved between them. Even though this assumption is satisfactory for modeling, but it is not rigorously accurate, because usually the conserved regions have non-identical coordinates. Further, it is hypothesized that insertions and deletions of the sequence alignment will fall mainly in loop regions. The loops permitted to be more elastic in the model than the conserved core and have differed conformations between target and template. There are some additional secondary structure elements inserted in loop regions even for close homology (69).

1.2. STEPS IN HOMOLOGY MODELING:

Homology modeling can be defined as the step-wise conversion of the structure of one protein into the structure of another protein.

1.2.1. Template Searching and Initial Target-Template alignment:

The initial starting step is the searching of structural homology to the target protein. Protein Data Bank (PDB) (76) contained the most experimentally solved structures. The accuracy of the model depends on the sequence alignment between the protein to be modeled-the target structure and the protein that is used as the template (72). The sequence-search programs such as BLAST (77) or FASTA (78) align the target sequence with all the sequences in the PDB database of known structures. The structure with the highest similarity score and the highest sequence identity is chosen as template (78).

1.2.2. Alignment Correction:

After selecting the suitable template protein, an optimal alignment between the target and template sequence is needed further to construct a 3D model of the target. The alignment is the most important step in homology modeling because the model of a protein is the 3D representation of the sequence alignment (72). The misalignment will build the low quality 3D structure. Even the misplacement of single residue in alignment can result to incorrect interpretation. The alignment can be create better if the insertions and deletions should not fall in the middle of secondary structure elements (69).

1.2.3. Backbone Generation:

The modeling of the backbone and side chains occurs simultaneously in most homology modeling programs. By copying the backbone coordinates of the template model onto the

target model the backbone of the target model is generated. The side chain can also be copied to the model when a residue is conserved (79). The backbone conformation should be inspected for regions of insertions or deletions, and regions that were considered to be difficult to align (69).

1.2.4. Loop Modeling:

The existence of a loop results when the model contains more residues than the model. There are several approaches to create loops. Some of them use libraries of loops with the same residues, created out of all structures in the PDB. Other approaches try to build completely new loops. The other major procedure when a good alignment has been achieved, is modeling the loops. These regions usually contain insertions or deletions and are the most variable regions in the sequence (69, 74).

1.2.5. Side Chain Modeling:

In comparative modeling of protein structure the side chain modeling on a fixed main chain commonly arises, where the initial main chain conformation for the target structure is obtained from copying the main chain coordinates of a template structure (79). In more than 90% of the cases the side chains have the same conformations in the two homologous proteins. In some cases **rotamer** libraries are used to predict the side-chain conformation, if the aligned residues are not identical. (69)

1.2.6. Model Optimization:

Many errors are introduced during the generation of the model. To improve local geometries, atomic interactions and to alleviate bad bond angles global energy

minimization is performed (69). Several sets of energy function parameters have been developed including CHARMM and AMBER (73).

1.2.7 Model Validation:

Evaluation of the predicted model involved analysis of geometry, stereochemistry, and energy distributions in the models. It is essential to verify the validation of the model, and to predict the resemblance and magnitude of potential error (69,74).Reliability of the homology model can be carrying out by the programs PROCHECK (80) and What-IF (81). The result of the complete process is a model of an unknown structure that can be used for structural analysis mutation analysis, experimental design, etc.

1.3. TOOLS FOR COMPARATIVE OR PROTEIN MODELING:

1.3.1. SEQUENCE ASSESSMENT:

1.3.1.1. Pairwise Sequence Alignment:

To find the possible homologs for a protein in sequence databases such as *SWISS-PROT* (82), pairwise sequence alignment is the most important step. A pairwise sequence alignment compares two protein sequences. BLAST (77) and FASTA (78) are the widely used programs for detecting the sequence similarities. BLAST (**B**asic **L**ocal **A**lignment **S**earch Tool) is the most extensively used local alignment tool. BLAST gives an expectation value, which calculates approximately the chance such alignment can be expected. The BLAST results permit to determine the quality of protein alignment quantitatively (77). FASTA maximize number of aligned residues by allocate insertions or deletions during the alignment phase and works for global alignment (78).

1.3.1.2. Multiple Sequence Alignment:

To attain the excellent alignment among sequences, a multiple sequence alignment aligns several sequences. In multiple sequence alignment the conserved regions (motifs) often have biological importance in requisites of function and structure. Progressive method is the broadly used algorithm for global alignment. It constructs a tree on the basis of pairwise similarity scores. This algorithm has been executed in the most accepted program for global multiple sequence alignment i.e CLUSTALX (83).

1.3.1.3. Sequence Family and Domain Parsing:

Multiple sequence alignment can classified protein sequences into families. A family affiliation often specifies a functional, structural, and evolutionary relationship. Classification of protein family can be based either on the alignment of long sequence domains or small conserved motifs (84-85). Pfam, ProDom methods mainly concentrate on the alignment of long sequence domains (86). The classifications of protein sequences in PROSITE (87), PRINTS (86), and BLOCKS (86) databases are based on fingerprints of small conserved motifs in sequences.

1.3.1.4. Phylogenetic Classification:

In different organisms phylogenetic relationaship among proteins is possibly be drawn from the protein sequences. The evolutionary relationship among a family of proteins can be illustrated by a tree structure of proteins. The most popular are *minimum distance*, *maximum parsimony*, and *maximum likelihood* trees. A protein tree, used to predict the phylogenetic relationship that is build on the basis to reduce the total pairwise sequence distance of adjacent tree nodes is recognized as "A minimum distance method." The other two maximum parsimony and maximum likelihood methods are based on multiple

sequence alignments (73). The widely used computer tools available for protein tree constructions are PHYLIP (88)

1.3.2. SEQUENCE INTERPRETATIONS:

1.3.2.1. Hydrophobicity Profile:

Hydropathy scale is a physiochemical property that computes the hydrophobicity of an amino acid. Hydrophobicity profile plots are accessible in several commercial protein modeling packages, i.e. Insight-II package (89).

1.3.2.2. Prediction of Signaling Sites:

In signaling proteins, the signaling sites illustrate particular patterns surrounding the active sites and at the boundaries of the active sites. The widely used SignalP server (90) and Signal3L (91) predict signal peptides in secretory proteins and their cleavage sites.

1.3.2.3. Search of Possible Active Sites:

Patterns extracted from motif databases such as PROSITE (87) and PRINTS (86) helps in investigating possible active sites. Some patterns are related to known protein functions, for this reason, a match to a pattern may recommend a function of the target protein.

1.3.2.4. Secondary Structure Prediction:

α-helices, β-sheet, and coils are the major types of secondary structure that inspected for sequence variation. A correlation between amino acid sequence and secondary structure is the basic assumption in all secondary structure prediction. PHD (92) and The Consensus Secondary Structure Prediction Server (93) are the most commonly used secondary structure prediction programs.

1.3.3. COMPARATIVE PROTEIN STRUCTURE MODELING TOOLS:

To construct a 3D model from given templates and alignments different comparative modeling methods use different approaches. SWISSMODEL (94) WHATIF (81), and MODWEB (71) are few comparative modeling tools. The most widely used comparative modeling tool is MODELLER (71). This program is completely automated and is capable of generating energy minimized protein models by satisfying spatial restraints on bond distances, bond length, bond angle, dihedral angles and non-bonded interatomic distances extracted from the template PDB file. The final 3D model is optimized through the combination of conjugate gradients and molecular dynamics with simulated annealing.

1.3.4. STRUCTURE ANALYSIS:

The quality and accuracy of the predicted model can be assessed by various modeling tools for analysis based on protein structures.

1.3.4.1. Visualization Tools:

Structural visualization presents a suitable way to study spatial associations of atoms, residues, secondary structures, domains, and subunits. There are a number of software's, both free and commercial, which help in visualizing biomolecules. VMD (95) has been used as a tool for interactive visualization and analysis of biomolecules. It can read standard PDB files and display the structure. RasMol (74) is a molecular graphics program proposed to visualize the 3D structure of macromolecules. GRASP (96) is used for visualization of the protein surface color-coded with electrostatic potential or geometry properties. Swiss-pdbviewer (94) makes high quality images using ray tracing

method. Insight-II (89) is a commercial package for protein modeling normally includes visualization tools with broad features.

1.3.4.2. Geometry investigation:

Further information associated to the conformation and energetics, as well as the excellence of model of a given protein structure can be provided through geometry analysis. These analyses can be perceived using programs such as, PROCHECK (80) and WHAT_CHECK (81).

1.3.4.3. Structure association and Structure Family:

According to structural and evolutionary relationships the association between the proteins in a structure database can be classified at diverse hierarchical levels. A broadly used classification comprises family, superfamily, and fold. CATH (97) is a hierarchical classification of protein domain structures. SCOP (98) uses improved manual classification with the hierarchical levels of class, fold, superfamily, and family of close homologs.

1.3.4.4. Molecular Dynamics, Quantum Mechanics, and Electrostatics:

Molecular dynamics simulation is a well conventional method to study a dynamic process of protein. CHARMM (99), XPLOR (100) and AMBER (101) are accessible programs for molecular dynamics simulation. Small conformational change and energetics such as free energy differences between two protein states can be calculated by using molecular dynamics simulation.

1.4 OBJECTIVES OF THE PRESENT STUDY:

The present study deals with the prediction of 3D models of toxin proteases of *V. cholerae* i.e. Accessory cholera enterotoxin (Ace), zonula occludens toxin (Zot), and hemagglutinin/protease (Vibriolysin) based on comparative homology modeling in order to have better understanding of pathogenesis using bioinformatics tools. The study has been carried out with the special reference to the active site and cofactor binding site. The enzyme-inhibitor model has been constructed separately in order to identify important residues that contribute to the active site interaction of the enzyme with the ligand. The present project aims towards the better understanding of various aspects of structural features of these toxin proteases.

CHAPTER

2.0

METHODOLOGY

2.1 Homology Modeling Methods:

In this study the protein structure prediction was carried out with the comparative protein structure modeling. Comparative modeling results in the most accurate, comprehensive, and unambiguous models of protein structure.

Comparative modeling is an efficient way to obtain information about the protein of interest such as interpretation of the existing functional data, identifying active, antigenic properties or binding specifities, design of ligands, rational design of site directed mutations and drugs, and construction of mutants and chimeric proteins for testing new functional hypotheses. Comparative protein structure modeling of a target sequence consists of four basic steps (*i*) Identification of known structures related to the target sequence (templates), (*ii*) Alignment of the templates with the target sequence, (*iii*) Building a model based on the alignment, and (*iv*) Evaluation of the model (71).

The present study represents structure prediction studies of Hemagglutinin/proteinase (Vibriolysin), Zona occludens toxin (Zot) and Accessory cholera enterotoxin (Ace) from *V. Cholerae* by comparative or Homology modeling techniques.

All steps of homology modeling including sequence alignments, model building and evaluations were executed on a Pentium IV PC. Model building was done by MODELLER. A succession of outcome of various programs was obtained from World Wide Web servers. The alignments are presented in Appendix II. The software WebLab viewer, Ds-Viewer and Astex viewer for graphical display, structural inspection and visualization, were used.

2.2. Sequences Retrieval.

During the course of current studies primary sequences of *V. Cholerae* proteases i.e Hemagglutinin/proteinase (Vibriolysin) (SwisProt AC: P24153), Zona occludens toxin (SwisProt AC: P38442) and Accessory cholera enterotoxin (SwisProt AC: P38441) were retrieved from the SWISSPROT (http://www.us.expasy.org) (82) and PIR (http://www.georgetown.edu) (102) database.

2.3. Searching Template:

Sequences homologous to the target sequences were extracted from PDB (76) databases. BLAST (Psi-BLAST) algorithm (77) against Protein Data Bank (PDB) was used to carry out the sequence homology searches.

In this study the templates were chosen on the basis of highest sequence identities between the targets and templates shown in (Appendix II).

2.4. Protein 3D Structure Databases:

The crystal structure coordinates of 1EZM, 1U4G, 1PXT and 1CKV has been selected as templates for constructing the comparative models of *V. Cholerea* proteases. The 3D coordinates of templates were extracted from the Brookhaven Protein Data Bank (http://www.rcsb.org/pdb).

2.5. Multiple Sequence Alignment:

In order to identify the functionally important regions, building sequence profile, protein family classification and phylogenetic reconstruction, multiple sequence alignment was carried out using the program CLUSTAL X (Version: 1.81) (83) with default parameters and manually adjusted where necessary. Sequences homologous to the target sequences were extracted from SWISS-PROT (82).

2.6. Phylogenetic Analysis:

Once the multiple sequence alignment has been constructed, a matrix of pairwise similarity scores is calculated and employs to construct a tree to express the structural relationship among the protein sequences in the family. Phylogenetic lineage was calculated using the phylogeny Inference program Phylip (Version 3.6b) (88).

2.7. Secondary Structure Prediction:

The Secondary structure prediction of target sequences was carried out by submitting the sequences to the Consensus Secondary Structure Prediction Sever at http://pbil.ibcp.fr/NPSA/npsa_npsa.html (103).

2.8. Model Building and Refinement:

Three dimensional homology models of *V. Cholerea* proteases were constructed using the crystal coordinates of templates on the basis of alignment between target and template sequences. All steps of homology modeling and refinement were carried out by the program MODELLER (Version 9 (9v8)) (71). The model building process was completed in two steps.

a. Target-template alignment.

b. Model Building.

a. Target-Template Alignment:

To obtain the target-template alignment for constructing homology modeling, the template structure and target sequence were realigned by the use of the ALIGN2D command of MODELLER.

This step is optional, a homology models can also be construct without using the first step by providing an alignment file as an input to the MODELLER.

MODELLER Input Files For ALIGN2D:

There are three kinds of input files for running ALIGN2D command of modeler.

- PDB atom coordinates file.
- Alignment file.
- Script or command file.

○ PDB Atom Coordinates File:

It is the Brookhaven Protein Data Bank atom file that contains the coordinates of known structure or template. Each atom file is named code.pdb where code is a short protein code preferably the PDB code e.g. 1EZM.pdb.

○ Alignment File:

The alignment file contains the alignment of the template structure with the target sequence. The preferred format for comparative modeling is related to the PIR database format. A sample of alignment file is shown in Fig. 2.1

The first line contains the sequence code, in the format >P1; 1ezm. The second line with ten fields separated by colons generally contains information about the structure coordinates file. The rest of the file contains the sequence of 1ezm, with '*' making its end. The standard one-letter amino acid codes are used.

The alignment files for other two proteases Zona occludens toxin (Zot) and Accessory cholera enterotoxin (Ace) are presented in (Appendix II).

3. Script or Command File (align2d.py):

The script file contains commands for MODELLER in the Python language. This script file provides all the compulsory commands to the MODELLER to produce sequence-

structure alignment using the crystal coordinates of the known structure. The **align2d.py** file for vibriolysin is presented in Figure 2.2.

To run MODELLER with the script file **align2d.py**, the modeller command prompt was opened and typed the following command at the command prompt:

```
mod9v8 align2d.py
```

MODELLER Output Files:

The align2d command was executed to align the two sequences, finally, the alignment is written out in two formats PIR **(align2d.ali)** and PAP **(align2d.pap)**. The PIR format **(align2d.ali)** is used by MODELLER in the subsequent model building stage, while the PAP alignment format **(align2d.pap)** is easier to inspect visually.

b. Model Building:

Once the target-template is constructed, three-dimensional homology models of *V. Cholerae* proteases i.e Hemagglutinin/proteinase (Vibriolysin), Zona occludens toxin and Accessory cholera enterotoxin were constructed using crystal structure coordinates of 1EZM, 1PXT and 1CKV respectively by using program MODELLER (9v8). This program is completely automated and is capable of generating energy minimized protein models by satisfying spatial restraints on bond distances and dihedral angles extracted from the template PDB file. The crystal coordinates of templates were obtained from Brookhaven Protein Data Bank (PDB). Many runs of MODELLER were carried out in order to obtain the most reasonable model satisfying the stereo-chemical properties. The models build by the MODELLER, at each step, were checked by PROCHECK and the final model was selected that fits or nearly fits the standard of the best and accurate model.

```
C; A sample alignment in the PIR format; used in tutorial

>P1;1ezm
structureX:1ezm:1::295::elastase:Pseudomonas::
AEAGGPGGNQKIGKYTYGSDYGPLIVNDR----CEMDDGNVITVDMNSSTDDSKTTPFRF
ACPT----NTYKQVNGAYSPLNDAHFFGGVVFKLYRDWFGTSPLTHKLYMKVHYGRSVEN
AYWDGTAMLFGDGATMFYPLVSLDVAAHEVSHGFTEQNSGLIYRGQSGGMNEAFSDMAGE
AAEFYMRGKNDFLIGYDIKKGSGALRYMDQPSRDGRSIDNASQYYNGIDVHHSSGVYNRA
FYLLANSPGWDTRKAFEVFVDANRYYWTATSNYNSGACGVIRSAQNRNYSAADVTRAFST
VGV*
>P1;vib1
sequence:vib1:::::vibriolysin:Vibrio::
ATGTGPGGNQKTGRYEYGSNGLPGFTIDKTGTTCTMNNSAVKTVNLNGGT--SGSTAFSY
ACNNSTNYNSVKTVNGAYSPLNDAHFFGKVVFDMYQQWLNTSPLTFQLTMRVHYGNNYEN
AFWDGRAMTFGDGYTRFYPLVDINVSAHEVSHGFTEQNSGLVYRDMSGGINEAFSDIAGE
AAEYFMRGNVDWIVGADIFKSSGGLRYFDQPSRDGRSIDHASQYYSGIDVHHSSGVFNRA
FYLLANKSGWNVRKGFEVFAVANQLYWTPNSTFDQGGCGVVKAAQDLNYNTADVVAAFNT
VGV*
```

Fig.2.1. *Alignment.ali* file used to generate *align2d.ali* & *align2d.pap* files for comparative model building of Hemagglutinin/proteinase (Vibriolysin) from crystal structure of PAE (1ezm).

```
log.verbose()
env = environ()

env.libs.topology.read(file='$(LIB)/top_heav.lib')

# Read aligned structure(s):
aln = alignment(env)
aln.append(file='alignment.ali', align_codes='1ezm')
aln_block = len(aln)

# Read aligned sequence(s):
aln.append(file='alignment.ali', align_codes='vib1')

# Structure sensitive variable gap penalty sequence-sequence alignment:
aln.align2d(overhang=0, gap_penalties_1d=(-450, 0),
            gap_penalties_2d=(0.35, 1.2, 0.9, 1.2, 0.6, 8.6, 1.2, 0., 0.),
            align_block=aln_block)

aln.write(file='align2d.ali', alignment_format='PIR')
aln.write(file='align2d.pap', alignment_format='PAP',
          alignment_features='INDICES HELIX BETA STRAIGHTNESS ' + \
                             'ACCESSIBILITY CONSERVATION')
```

Fig.2.2. The script file (*align2d.py*) contains commands for MODELLER to produce sequence-structure alignment using the crystal coordinates of the known structure.

MODELLER Inputs Files for Model Building:

There are three kinds of input files for running model building command of MODELLER.

- PDB atom coordinates file.
- Alignment file.
- Script or command file.

1. PDB Atom Coordinates File:

It is the Brookhaven Protein Data Bank atom file that contains the coordinates of known structure or template. Each atom file is named code.pdb where code is a short protein code preferably the PDB code e.g 1EZM.pdb

2. Alignment File:

The alignment file contains the alignment of the template sequence with the target sequence. The preferred format for comparative modeling is related to the PIR database format. The PIR format **(align2d.ali)** resulted as an output file of align2d command is used by MODELLER in the subsequent model building stage. A sample of alignment file **(align2d.ali)** is shown in Fig. 2.3

The first line contains the sequence code, in the format >P1; 1ezm. The second line with ten fields separated by colons generally contains information about the structure coordinates file. The rest of the file contains the sequence of 1ezm; the standard one-letter amino acid codes are used. '/' separated the sequence form the water molecules and HETATM; '.' Represented the total number of water molecules and hetatom (Zn) with '*' making its end.

The alignment.ali files for all the three proteases i.e Hemagglutinin/proteinase (Vibriolysin), Zona occludens toxin (Zot) and Accessory cholera enterotoxin (Ace) are presented in (Appendix II).

3. Script or Command File (model-ligand.py):

This script file provide all the compulsory commands to the MODELLER to produce the homology models using the crystal coordinates of the known structure on the basis of alignment between the target and template sequences. In order to build all hydrogen atom models with water molecules and other non-protein atoms (HETATM), the following parameters were written in **model-ligand.py** file.

```
# Read in HETATM records from template PDBs
env.io.hydrogen = env.io.hetatm = env.io.water = True
```

This was used for reading the HETATM coordinates from the PDB file. The **model-ligand.py** file for vibriolysin is presented in Fig. 2.4

To run MODELLER with the script file **model-ligand.py**, the modeller command prompt was opened and typed the following command at the command prompt:

```
mod9v8 model-ligand.py
```

MODELLER Output Files:

As the MODELLER proceeds for building the homology model, a number of intermediary files were produced. The final target model is written to file vib1.B99990001.pdb.The MODELLER output includes

i. **Job.log:** A log file i.e. model-ligand.log

ii. **Job.ini:** Initial conformation for optimization i.e. vib1.ini

iii. **Job.rsr:** Restrain file i.e. vib1.rsr

iv. **Job.D00000001:** Progress for optimization i.e. vib1.D00000001

v. **Job.sch:** A schedule file for the variable target function optimization (VTFM) i.e. vib1.sch

vi. **Job.V99990001:** A file for violation profiles for the newly built model. i.e. vib1.V99990001

vii. **Job. B99990001:** PDB atoms file for the model of target sequence. i.e. vib1.B99990001.pdb

2.9. Inhibitor Modeling:

In order to identify the residues that contribute to individual specificity subsites and to understand the mechanism of its action and protein interactions with substrate and/or inhibitor, a 3D model of HA/P complexed with N-(1-carboxy-3-phenylpropyl)-phenylalanyl-alphaaspargine (HPI) was constructed by using program MODELLER. The 3D crystal structure of elastase complexed with N-(1-carboxy-3-phenylpropyl)-phenylalanyl alphaaspargine (HPI) (PDB id: 1U4G) was used as a template to model this inhibitor into the active site of HA/P.

```
C; A sample alignment in the PIR format; used in tutorial

>P1;1ezm
structureX:1ezm:1     : :453  : :elastase:Pseudomonas:-1.00:-1.00
AEAGGPGGNQKIGKYTYGSDYGPLIVNDR----CEMDDGNVITVDMNSSTDDSKTTPFRFACPTN--TYKQVNGA
YSPLNDAHFFGGVVFKLYRDWFGTSPLTHKLYMKVHYGRSVENAYWDGTAMLFGDGATMFYPLVSLDVAAHEVSH
GFTEQNSGLIYRGQSGGMNEAFSDMAGEAAEFYMRGKNDFLIGYDIKKGSGALRYMDQPSRDGRSIDNASQYYNG
IDVHHSSGVYNRAFYLLANSPGWDTRKAFEVFVDANRYYWTATSNYNSGACGVIRSAQNRNYSAADVTRAFSTVG
V/.....................................................................
.....................................................................
.........*

>P1;vib1
sequence:vib1:      : :      : :vibriolysin:Vibrio:-1.00:-1.00
ATGTGPGGNQKTGRYEYGSNGLPGFTIDKTGTTCTMNNSAVKTVNLNGGTSGSTAFSYACNNSTNYNSVKTVNGA
YSPLNDAHFFGKVVFDMYQQWLNTSPLTFQLTMRVHYGNNYENAFWDGRAMTFGDGYTRFYPLVDINVSAHEVSH
GFTEQNSGLVYRDMSGGINEAFSDIAGEAAEYFMRGNVDWIVGADIFKSSGGLRYFDQPSRDGRSIDHASQYYSG
IDVHHSSGVFNRAFYLLANKSGWNVRKGFEVFAVANQLYWTPNSTFDQGGCGVVKAAQDLNYNTADVVAAFNTVG
V/.....................................................................
.....................................................................
.........*
```

Fig. 2.3. *Alignment.ali* file used to contruct comparative model of Hemagglutinin/proteinase (Vibriolysin) from crystal structure of PAE (1ezm). The alignment file contains the alignment of the template sequence (1ezm) with the target sequence (vib1); the standard one-letter amino acid codes are used. '/' separated the sequence form the water molecules and HETATM; '.' Represented the total number of water molecules and hetatom (Zn) with '*' making its end.

```
# Homology modeling with ligand transfer from the template
from modeller import *                  # Load standard Modeller classes
from modeller.automodel import *        # Load the automodel class

log.verbose()    # request verbose output
env = environ()  # create a new MODELLER environment to build this model in

# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

# Read in HETATM records from template PDBs
env.io.hydrogen = env.io.hetatm = env.io.water = True

a = automodel(env,
              alnfile  = 'align2d.ali',  # alignment filename
              knowns   = '1ezm',              # codes of the templates
              sequence = 'vib1')              # code of the target
a.starting_model= 1                 # index of the first model
a.ending_model  = 60                # index of the last model
                                    # (determines how many models to calculate)
a.make()                            # do the actual homology modeling
```

Fig.2.4. The script file (model-ligand.py) contains all the compulsory commands of MODELLER to produce the homology models using the crystal coordinates of the known structure.

2.10 Model Visualization.

In order to peruse the reliability of the alignment and modeling of variable surface loops, structural investigations on the graphics screen using 3D visualization programs, Ds-Viewer (104), WebLab Viewer (105), Swiss-PdbViewer (94) and Astexviewer (106) were performed. Structural visualization helps to study spatial associations of atoms, residues, secondary structures, domains, and subunits.

2.11 Model Evalution:

Evaluation of the predicted model involved analysis of geometry, stereochemistry, and energy distributions in the models. The Energy Command of the MODELLER (71) carried out the reliability assessment of the comparative model.

Reliability of the homology model was carried out by the programs PROCHECK (80) (Version: 3.4) and WHATCHECK (81).

2.11.1 PROCHECK:

The principle of PROCHECK is to evaluate the overall stereo-chemical quality of a given model. It checks the geometry of the residues in a given protein structure as compared with well-refined structures at the same resolution. To evaluate a structure, the PROCHECK makes use of a number of parameters i.e. Ramachandran plot quality, Peptide bond planarity, Bad non-bonded interactions, C-alpha tetrahedral distortion, Overall G-factor, Main-chain hydrogen bond energy (80). The input to PROCHECK is a single file containing the coordinates of protein structure in Brookhaven file (PDB) format. For homology model-built structures, the resolution of the structure(s) on which the model has been built was used.

To run PROCHECK, the following command was typed at the window command prompt.

```
pro.bat
```

```
PRO vib1 2.0
```

The output of PROCHECK is a set of postscript files with graph.

2.11.2 ProSA:

The energetic architecture of protein folds was determined by using the tool ProSA (Protein Structure Analysis, Version.4) (107). It can assist in protein structure quality control. An energy graph calculated by ProSA provides diagnostic tools to mark problematic segments. Energy graphs demonstrate the energetic architecture of protein folds as a function of amino acid sequence position. It calculates a score for input-structure. Scores of native protein folds are in a distinctive range. If score is outside this range then structure may have problems.

The program ProSA was installed from http://www.came.sbg.ac.at/ after sending an installation request form and getting the license key. The PC system physical address was requisite for installation request which was noted form the window command prompt by giving the following command at the window command prompt.

```
ipconfig /all
```

The license key for installing ProSA was send via email which was saved in ASCII format and named the file as "LicenseDB".

The input files for ProSA are the target and template models containing the coordinates of protein structure in Brookhaven file (PDB) format. The generation of energy graphs

and calculation of z-scores were done by following the commands mentioned in ProSA's manual on page no. 17 & 18.

2.11.3 SUPERPOSITION:

In addition, the variability among the models was evaluated by superposition of C-alpha traces and backbones onto the template crystal structure from which the RMSD values for positional differences between equivalent atoms were calculated. The structural superposition was performed by using the SUPERPOSE command of MODELLER.

The MODELLER input files for running superpose.py command were as follows.

- Alignment file i.e. align2d.ali
- PDB co-ordinates atoms files of Target and Template models.
- Command file i.e. superpose.py

To run MODELLER with the script file **superpose.py**, the modeller command prompt was opened and typed the following command at the command prompt:

```
mod9v8 superpose.py
```

The MODELLER outputs are:

- .log log output produces by the MODELLER ren.
- .fit fitted protein structures in the PDB format.

2.12. PROTEINS–LIGAND INTERACTIONS:

Program Ligand Explorer (http://www.kukool.com/ligand) (108) and Ligplot (109) were used to study the Protein–Ligand interactions.

2.12.1. Ligand Explorer:

The Ligand Explorer is designed for quick examination of protein-ligand interactions, such as hydrophobic interactions, hydrophilic interactions and the interactions with

ordered H_2O molecules. Ligand Explorer has been incorporated into the RCSB Protein Data Bank (PDB). It can also be run as a separate application on desktop to access local structure files and files on PDB's servers. The input of program is a standard Protein Data Bank file.

2.12.2. LIGPLOT:

Ligplot was designed to make possible the rapid inspection of many enzyme complexes, but can also be used to show other types of interaction in proteins and nucleic acids. The program automatically generates schematic diagrams of protein-ligand interactions from standard 3D coordinates in a **PDB** file input. The output present a simple and informative representation of the intermolecular interactions and their strengths, including hydrogen bonds, hydrophobic interactions and atom accessibilities in the form of color, or black-and-white PostScript file.

CHAPTER

3.0

RESULTS

3.1. HEMAGGLUTININ/PROTEASE [HA/P (VIBRIOLYSIN)]

3.1.1. Sequence Analysis:

Sequence homology searches for the query hemagglutinin/protease HA/P (Vibriolysin) was carried out by using BLAST algorithm against Protein Data Bank (PDB).Crystal structure co-ordinates of *Pseudomonas Aeruginosa* Elastase (1EZM.pdb) was chosen as a template on the basis of highest sequence similarity score and lowest E-value for constructing the predicted 3D structure of vibriolysin. The BLAST result is presented in Appendix II.

The align2d command of MODELLER realigned the target and template sequences, as a result reduced the gaps in vibriolysin sequence while shifted the gaps between Asn61 and Thr62 in the template (PAE) sequence as shown in Fig.3.1.1.

3.1.2. Multiple Sequence Alignment:

Comparison of multiple sequences (18 sequences) shows considerable sequence similarity among primary structures of all known neutral proteases of M4 family. The active site and the metal binding residues are highly conserved. HEXXH is the putative signature of Zn-metalloproteases. This motif contains all of the residues that have been positively identified as part of the active site. HA/P also contains the typical HEXXH amino acid motif. The result of Clustal X presenting multiple sequence alignment is presented in Appendix II.

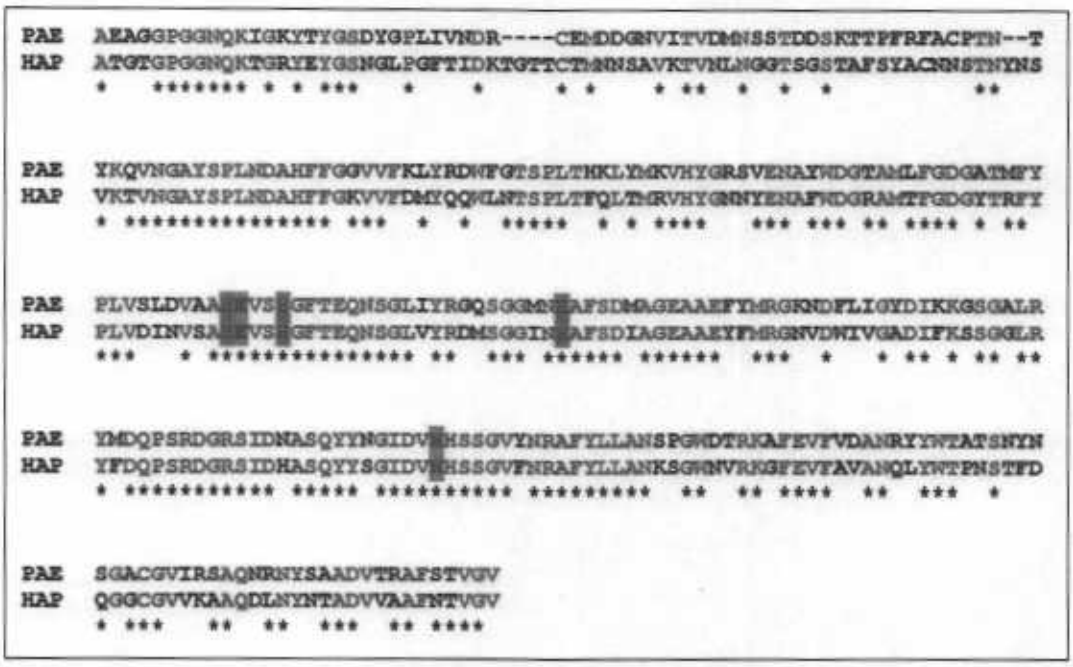


Fig.3.1.1: Target (HA/P)-Template (1EZM) alignment by using ALIGN2D command of Modeller (9v8). Conserved active site (red) and Zn-binding (green) residues are highlighted. The regions that are thought to be the active site and Zn-binding domains show strong amino acids homologies.

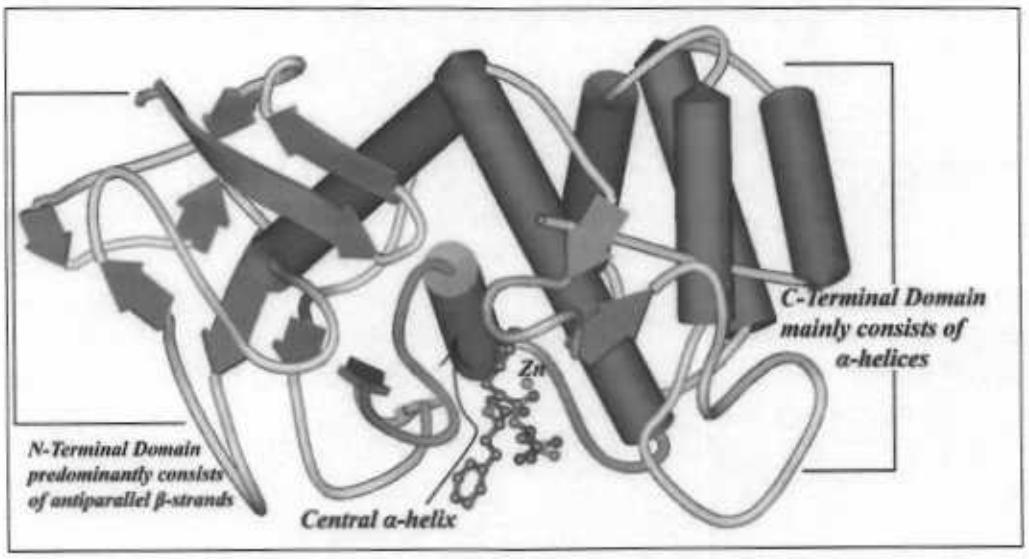


Fig.3.1.2. The schematic representation of homology model of vibriolysin showing arrangements of secondary structural elements The N-terminal (left) domain of the protein is constructed primarily of antiparallel β-strands, while the C-terminal (right) domain is α-helical. Central α-helix is shown in magenta, which is at the bottom of active site cleft with the catalytic Zn ion (yellow sphere).

3.1.3. Phylogenetic Analysis:

A Phylogenetic tree results show that vibriolysin from both *V.choelare* and elastase from *P.aeruginosa* are derived from common ancestors and are homologous to each other according to evolutionary point of view, it also shows that they belong to the same subfamily as shown in Appendix III.

3.1.4. Secondary Structure Prediction:

The output of consensus Secondary Structure Prediction server helps in identifying the secondary structural elements i.e. α-helices, β-sheets, and coils. The result of secondary structure prediction is presented in Appendix IV. The secondary structure of HA/P is closely resembled to its template structure.

3.1.5. Topology of Predicted Homology Model of HA/P (Vibriolysin):

The overall tertiary structure of the HA/P strongly resembles that of its template PAE and other neutral metalloproteases, whose structures are known. Fig. 3.1.2 depicts the schematic representation of the homology model of HA/P showing arrangement of secondary structural elements. The tertiary structure comprises two major domains with α + β structural topology. The N-terminal domain contains mainly β-sheets, whereas the C-terminal domain predominantly contains α-helices. The domains are connected by a central a-helix. The active site is located in the cleft between these two domains. The catalytically essential Zn-ion and the central helix are located at the bottom of the active site cleft.

3.1.6. Model Assessment:

Table 3.1.1a shows Ramachandran plot statistics of the HA/P models obtained with PROCHECK. Ramachandran plot indicates that 88.7% of the main-chain dihedral angles are found in the most favored regions, 10.5% in the additionally allowed, and 0.8% in the generously allowed while no residue is found in the disallowed regions. The structural conformations of the predicted HA/P model is similar to X-ray structure of PAE. Table 3.1.1b shows Ramachandran plot statistics of template PAE. The Ramachandran Plots of target and template are presented in Appendix V.

Fig.3.1.3 (a & b). Shows the energy graph of HA/P calculated by ProSA in term of Z-score (-7.89) which indicated the overall quality of protein structure. The energy graph for PAE (Template:1U4G) molecule (Z-score: -8.63) is presented in Appendix VI.

The overall description of similarities and differences derived from backbone superposition is given in Fig. 3.1.4. The RMSD value between template and the predicted model using all main-chain atoms was found to be $0.467A^0$ and $0.0916A^0$ in absence and presence of inhibitor HPI, respectively.

Table.3.1.1a. Table shows Ramachandran plot statistics of the HA/P models obtained with PROCHECK.

```
 +----------<<< P R O C H E C K    S U M M A R Y >>>----------+
 |                                                             |
 | vibl_lig36   2.5                                 301 residues |
 |                                                             |
+| Ramachandran plot:   88.7% core   10.5% allow     .8% gener     .0% disall |
 |                                                             |
 | Gly & Pro Ramach:     0 labelled residues (out of  43)      |
+| Chi1-chi2 plots:      1 labelled residues (out of 160)      |
 |                                                             |
+| Main-chain params:    5 better     0 inside      1 worse    |
 | Side-chain params:    5 better     0 inside      0 worse    |
 |                                                             |
*| Residue properties: Max.deviation:    10.7          Bad contacts:    9 |
*|                     Bond len/angle:   12.8    Morris et al class:  1  1  2 |
+|     2 cis-peptides                                          |
+| G-factors           Dihedrals:   -.04  Covalent:   -.89    Overall:   -.33 |
 |                                                             |
 | M/c bond lengths: 85.7% within limits  14.3% highlighted    |
*| M/c bond angles:  61.9% within limits  38.1% highlighted      7 off graph |
+| Planar groups:    94.9% within limits   5.1% highlighted      1 off graph |
 |                                                             |
 +-------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```

Table.3.1.1b. Table shows Ramachandran plot statistics of template PAE obtained with PROCHECK.

```
 +----------<<< P R O C H E C K    S U M M A R Y >>>----------+
 |                                                             |
 | 1u4g   1.5                                       295 residues |
 |                                                             |
+| Ramachandran plot:   87.6% core   12.0% allow     .4% gener     .0% disall |
 |                                                             |
 | Gly & Pro Ramach:     0 labelled residues (out of  42)      |
 | Chi1-chi2 plots:      0 labelled residues (out of 163)      |
 |                                                             |
 | Main-chain params:    6 better     0 inside      0 worse    |
 | Side-chain params:    5 better     0 inside      0 worse    |
 |                                                             |
*| Residue properties: Max.deviation:    10.1          Bad contacts:   16 |
*|                     Bond len/angle:    5.3    Morris et al class:  1  1  2 |
+|     2 cis-peptides                                          |
 | G-factors           Dihedrals:    .27  Covalent:    .55    Overall:    .39 |
 |                                                             |
 | M/c bond lengths:100.0% within limits    .0% highlighted    |
 | M/c bond angles:  92.9% within limits   7.1% highlighted    |
 | Planar groups:   100.0% within limits    .0% highlighted    |
 |                                                             |
 +-------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```

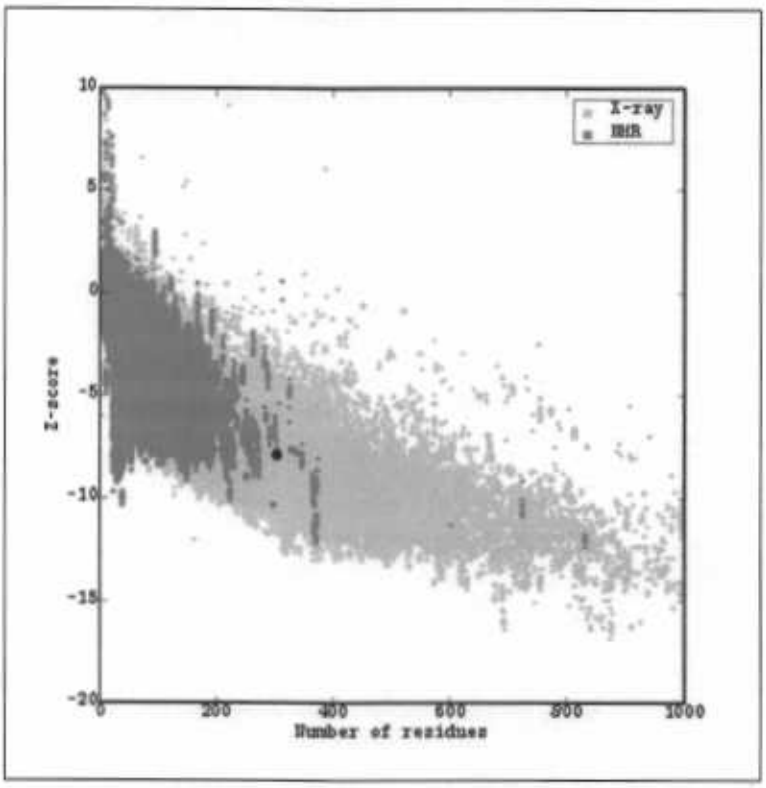


Fig.3.1.3a: Screenshot of ProSA z-score plot of HA/P showing the Z value < 0.

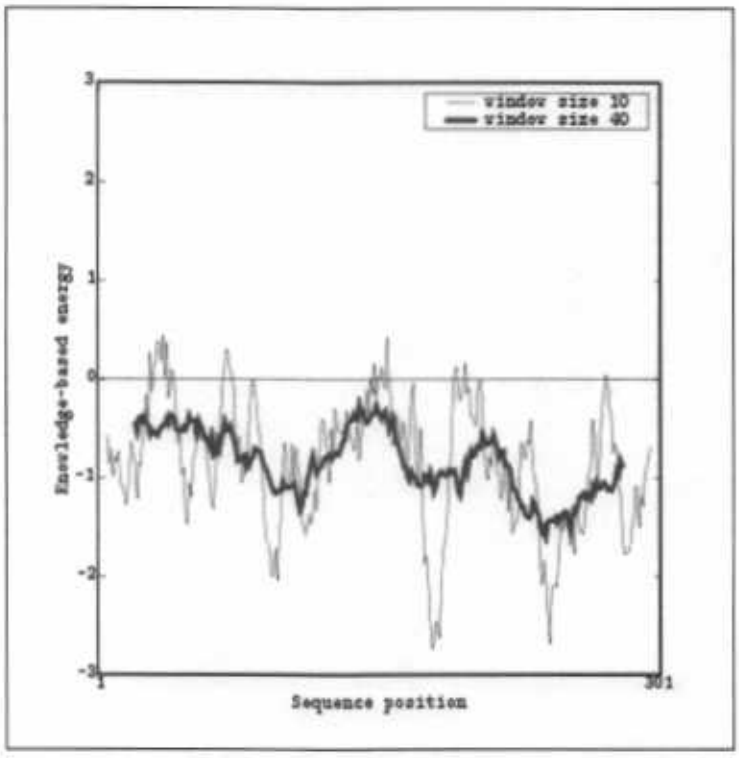


Fig.3.1.3b: Screenshot of ProSA plot of HA/P showing the energy graph of residue scores of a predicted protein structure.

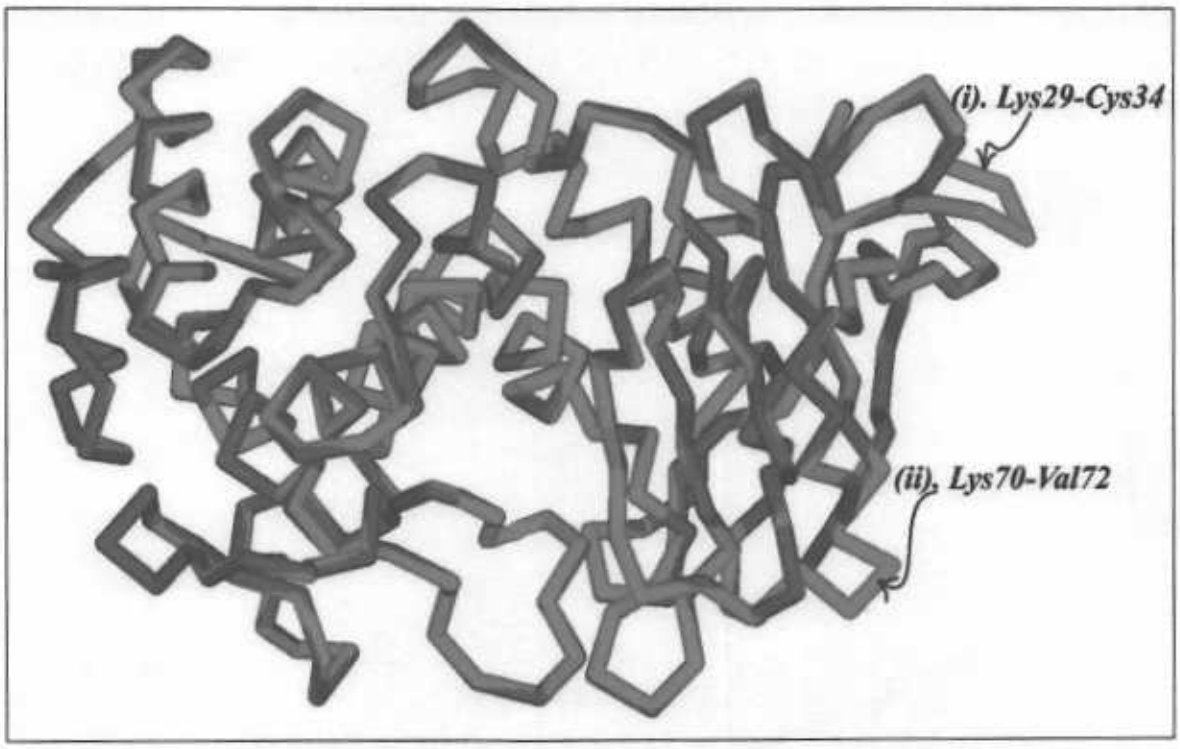


Fig. 3.1.4. Structural superposition of Cα atoms of HA/P (magneta) onto the crystal structure of PAE (blue). The structural differences due to insertions in loop regions are indicated by (i) & (ii).

3.2. Zona Occludens Toxin (Zot)

3.2.1. Sequence Analysis:

Sequence homology searches for the query Zona Occludens Toxin (Zot) was carried out by using **BLAST** algorithm against Protein Data Bank (PDB). Crystal structure co-ordinates of chain A of Peroxisomal 3-Ketoacyl-Coa Thiolase of *Saccharomyces Cerevisiae* (1PXT.pdb) was chosen as a template on the basis of highest sequence similarity score and lowest E-value for constructing the predicted 3D structure of Zot. The BLAST results show the alignment of only 108 residues out of 399 (APPENDIX II). The **align2d** command of MODELLER realigned the full length sequences of target and template, as first 27 residues are not crystallized in the crystal structure of 1PXT these were deleted from the template sequence, as a result gaps are reduced in both target (Zot) and template (1PXT) sequence as shown in Fig.3.2.1.

3.2.2. Multiple Sequence Alignment:

Comparison of multiple sequences (7 sequences) showed considerable sequence similarity among the family members. Little is known about the Zot functions and its biological activity that's why no residue is found to be conserved as well as no significant number of sequences is found to be closed homologue. The result of Clustal X presenting multiple sequence alignment is presented in Appendix II.

3.2.3. Phylogenetic Analysis:

A Phylogenetic tree results showed that Zot from both *V.choelare* and Peroxisomal 3-Ketoacyl-Coa Thiolase from *S.Cerevisiae* are derived from common ancestors and are homologous to each other according to evolutionary point of view, it also shows that they belong to the different subfamily as represent in Appendix III.

```
1PXT  28  LLEKRPEDVVIVAANRSAIGKGFKGAFKDVNTDYLLYNFLNEFIGRFPEPLRADLNLIEEVACGNVLN
ZOTV   1  SIFIHHGAPGSYKTSGALWLRLLPAIKSGRHIITNVRGLNLERMAKYLKMDVSDISIEFIDTDHPDGR
                                                    *           *

1PXT      VGAGATEHRAACLASGIPYSTPFVALNRQCSSGLTAVNDIANKIKVGQIDIGLALGVESMTNNYKNVN
ZOTV      LTMARFWHWARKDAFLFIDECGRIWPPRLTVTNLKALDTPPDLVAEDRPESFEVAFDMHRHHGWDICL
                *  *   *                 *     * *

1PXT      PLGMISSEELQKNREAKKCLIPMGITNENVAANFKISRKDQDEFAANSYQKAYKAKNEGLFEDEILPI
ZOTV      TTPNIAKVHNMIREAAEIGYRHFNRATVGLGAKFTLTTHDAANSGQMDSHALTRQVKKIPSPIFKMYA
              *           *                 * *     *

1PXT      KLPDGSICQSDEGPRPNVTAESLSSIRPAFIKDRGTTTAGNASQVSDGVAGVLLARRSVANQLNLPVL
ZOTV      STTTGKARDTMAGTALWKDRKILFLFGMVFLMFSYSFYGLHDNPIFTGGNDATIESEQSEPQSKATVG
              *       *           *     *                 *                 *   *

1PXT      GRYIDFQTVGVPPEIMGVGPAYAIPKVLEATGLQVQDIDIFEINEAFAAQALYCIHKLGIDLNKVNPR
ZOTV      NAVGSKAVAPASFGFCIGRLCVQDGFVTVGDERYRLVDNLDIPYRGLWATGHHIYKDTLTVFFETESG
                                      *                 *

1PXT      GGAIALGHPLGCTGARQVATILRELKKDQIGVVSMCIGTGMGAAAIFIKE 417
ZOTV      SVPTELFASSYRYKVLPLPDFNHFVVFDTFAAQALWVEVKRGLPIKTEND 391
              *                       *                 *
```

Fig. 3.2.1. Target (ZOT)-Template (1PXT) alignment by using ALIGN2D command of Modeller (9v8). * represents the conserved residues between the two sequences.

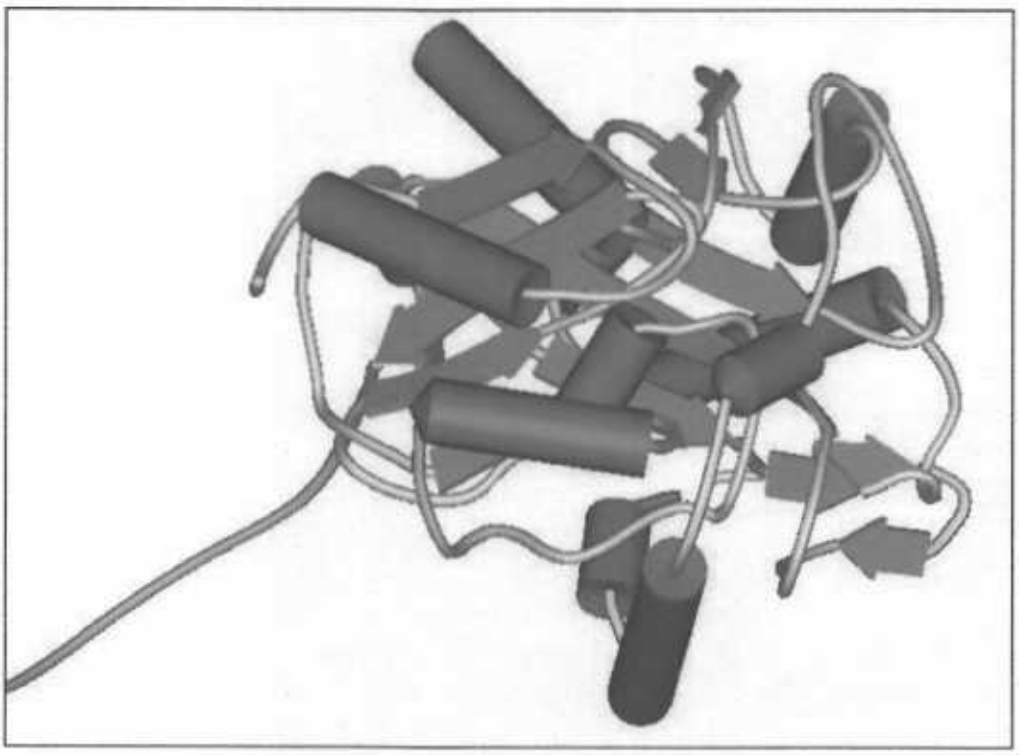


Fig.3.2.2. The schematic representation of the homology model Zot showing arrangement of secondary structural elements.

3.2.4. Secondary Structure Prediction:

The output of consensus Secondary Structure Prediction server helps in identifying the secondary structural elements i.e. α-helices, β-sheets, and coils. The result of secondary structure prediction is presented in Appendix IV. The secondary structure of Zot is closely resembled to its template structure.

3.2.5. Topology of Predicted Homology Model of Zona Occludens Toxin (Zot):

The overall tertiary structure of the Zot strongly resembles that of its template 1PXT and other proteases belong to superfamily thiolase-like, whose structures are known. Fig. 3.2.2 depicts the schematic representation of the homology model Zot showing arrangement of secondary structural elements. The tertiary structure comprises with α/β structural topology, mainly antiparallel β-sheets (beta-alpha-beta units).

3.2.6. Model Assessment:

Table 3.2.1a shows Ramachandran plot statistics of the Zot models obtained with PROCHECK. Ramachandran plot indicates that 85.9% of the main-chain dihedral angles are found in the most favored regions, 11.1% in the additionally allowed, 2.1% in the generously allowed while 0.9% (3 residues) is found in the disallowed regions. The structural conformations of the predicted Zot model is similar to X-ray structure of 1PXT. Table 3.2.1b shows Ramachandran plot statistics of template 1PXT. The Ramachandran Plots of target and template are presented in Appendix V.

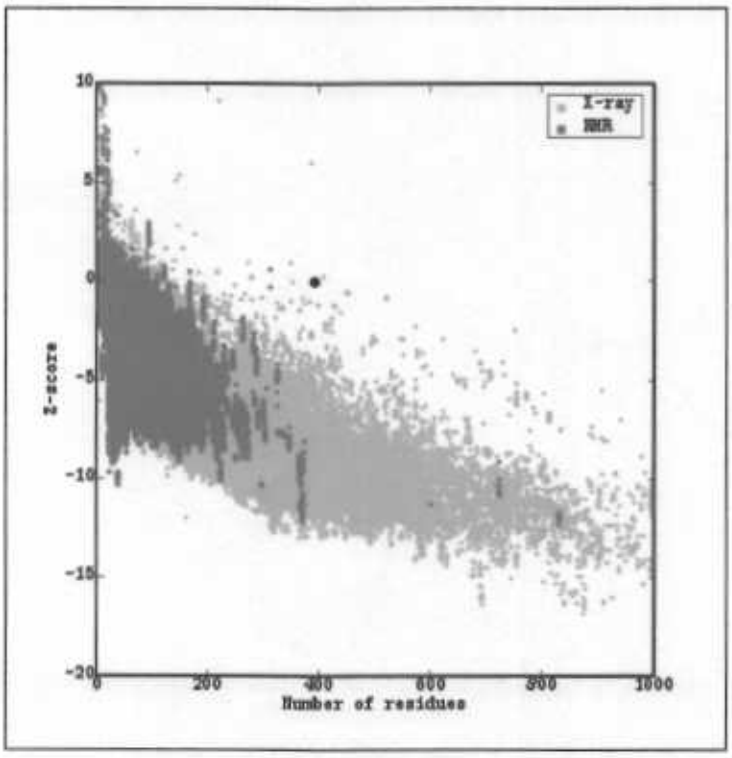
Fig.3.2.3 (a & b) Shows the energy graph of Zot calculated by ProSA in term of Z-score (-0.09) which indicated the overall quality of protein structure. The energy graph for 1PXT (Template) molecule (Z-score: -8.21) is presented in Appendix VI.

Table.3.2.1a. Table shows Ramachandran plot statistics of the Zot models obtained with PROCHECK.

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
|                                                            |
| ZOTV10  2.8                                    390 residues |
|                                                            |
*| Ramachandran plot:  85.9% core  11.1% allow  2.1% gener   .9% disall |
|                                                            |
*| Gly & Pro Ramach:    4 labelled residues (out of  47)      |
*| Chi1-chi2 plots:    11 labelled residues (out of 231)     |
|                                                            |
| Main-chain params:   6 better    0 inside    0 worse       |
| Side-chain params:   5 better    0 inside    0 worse       |
|                                                            |
*| Residue properties: Max.deviation:    5.7      Bad contacts:  10 |
*|                     Bond len/angle:  12.9   Morris et al class:  1  1  2 |
|                                                            |
+| G-factors           Dihedrals:  -.20  Covalent:  -.55  Overall:  -.32 |
|                                                            |
| M/c bond lengths: 79.2% within limits  20.8% highlighted   |
*| M/c bond angles:  62.2% within limits  37.8% highlighted    3 off graph |
+| Planar groups:    95.0% within limits   5.0% highlighted    1 off graph |
|                                                            |
+------------------------------------------------------------+
  + May be worth investigating further.  * Worth investigating further.
```

Table.3.2.1b. Table shows Ramachandran plot statistics of template 1PXT obtained with PROCHECK.

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
|                                                            |
| 1PXT  2.8                                      343 residues |
|                                                            |
*| Ramachandran plot:  79.0% core  18.5% allow  2.4% gener   .0% disall |
|                                                            |
| Gly & Pro Ramach:    0 labelled residues (out of  49)      |
+| Chi1-chi2 plots:     1 labelled residues (out of 200)     |
|                                                            |
| Main-chain params:   6 better    0 inside    0 worse       |
| Side-chain params:   5 better    0 inside    0 worse       |
|                                                            |
+| Residue properties: Max.deviation:    4.0      Bad contacts:   6 |
+|                     Bond len/angle:   4.7   Morris et al class:  1  2  2 |
|                                                            |
| G-factors           Dihedrals:  -.13  Covalent:   .30  Overall:   .05 |
|                                                            |
| M/c bond lengths: 97.5% within limits   2.5% highlighted   |
| M/c bond angles:  84.2% within limits  15.8% highlighted   |
+| Planar groups:    95.2% within limits   4.8% highlighted    1 off graph |
|                                                            |
+------------------------------------------------------------+
  + May be worth investigating further.  * Worth investigating further.
```



F ig.3.2.3a: Screenshot of ProSA z-score plot of Zot showing the Z value < 0.

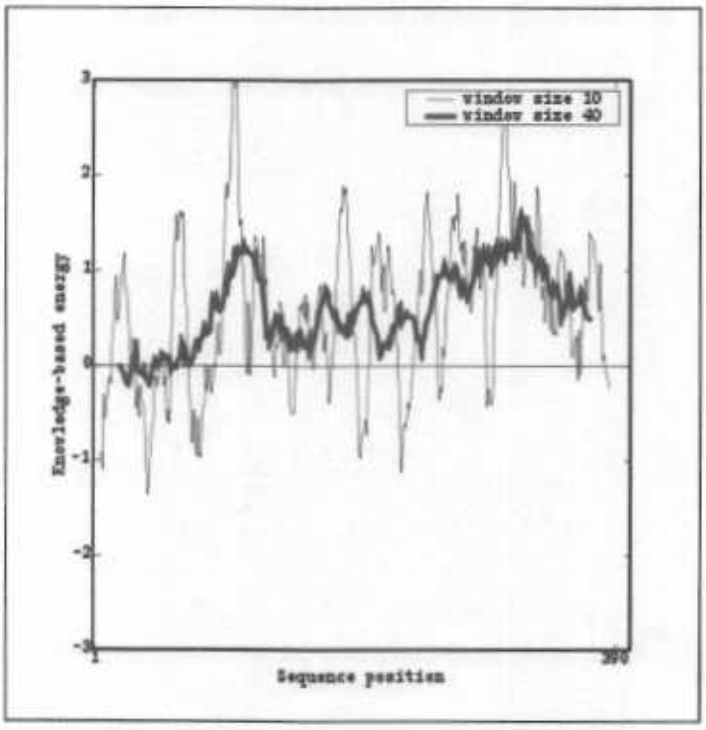


Fig.3.2.3b: Screenshot of ProSA plot of Zot showing the energy graph of residue scores of a predicted protein structure.

The overall description of similarities and differences derived from backbone superposition is given in Fig. 3.2.4. The RMSD value between template and the predicted model using all main-chain atoms was found to be 0.5432A °.

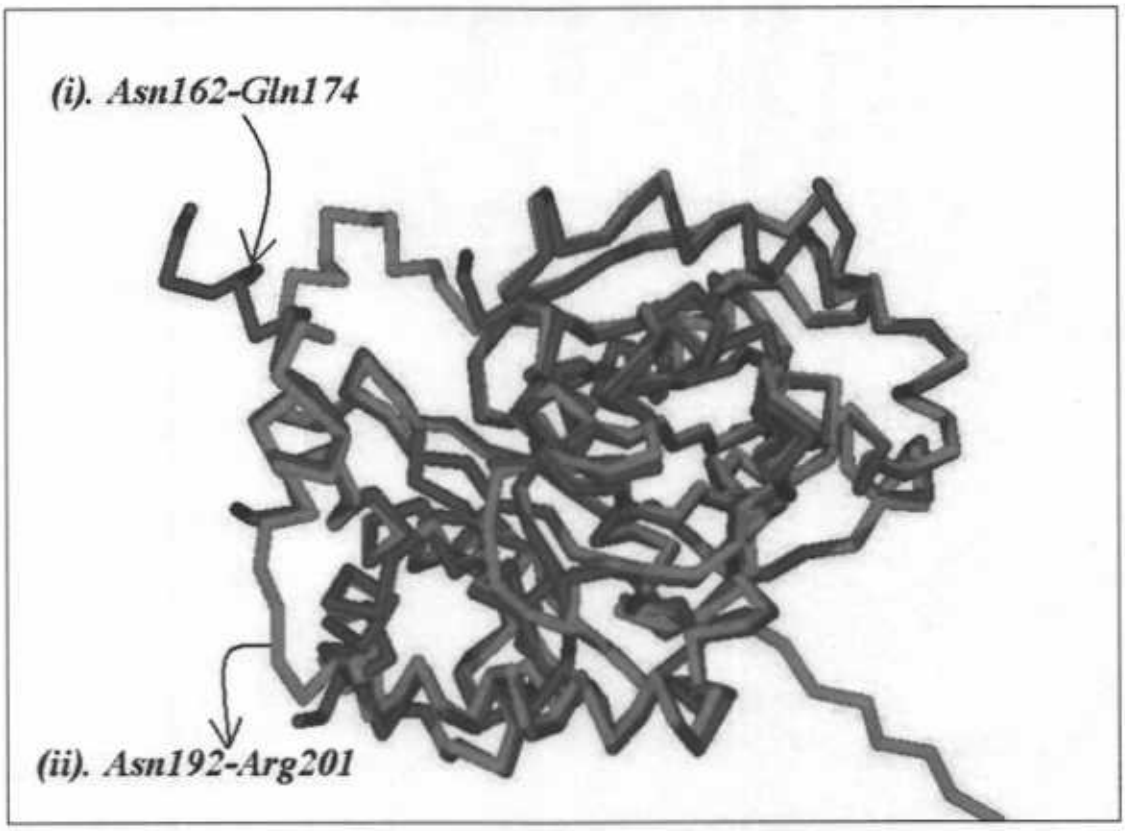


Fig.3.2.4. Structural superposition of Cα atoms of Zot (magneta) onto the crystal structure of 1PXT (blue). The structural differences due to missing regions are indicated by (i) & (ii)

3.3.Accessory Cholera Enterotoxin (Ace)

3.3.1. Sequence Analysis:

A sequence homology search for the query Accessory cholera enterotoxin (Ace) was carried out by using **BLAST** algorithm against Protein Data Bank (PDB). There was only a single crystal structure of the Soluble Methane Monooxygenase Regulatory Protein B (1CKV.pdb) of *Escherichia coli* that showed 31% sequence identity, which was chosen as a template on the basis of sequence similarity score and E-value for constructing the predicted 3D structure of Ace. The BLAST results show the alignment of only 35 residues out of 96 residues (APPENDIX II).

The **align2d** command of MODELLER realigned the full length sequences of target and template, as first 45 residues did not show any alignment with the target sequence that's why these were deleted from the template's crystallized structure as well from the sequence as shown in Fig.3.3.1.

3.3.2. Multiple Sequence Alignment:

Comparison of multiple sequences (4 sequences) showed considerable sequence similarity among the family members. Little is known about the Ace functions and its biological activity that's why no residue is found to be conserved as well as no significant number of sequences is found to be closed homologue. The result of Clustal X presenting multiple sequence alignment is presented in Appendix II.

3.3.3. Phylogenetic Analysis:

A Phylogenetic tree results show that Accessory cholera enterotoxin (Ace) from both *V.choelare* and Monoxygenase from *E.coli* are derived from common ancestors and are homologous to each other according to evolutionary point of view, it also shows that they belong to the different subfamily (Appendix III).

3.3.4. Secondary Structure Prediction:

The output of consensus Secondary Structure Prediction server helps in identifying the secondary structural elements i.e. α-helices, β-sheets, and coils. The result of secondary structure prediction is presented in Appendix IV. The secondary structure of Ace is closely resembled to its template structure.

3.3.5. Topology of Predicted Homology Model of ACE:

The overall tertiary structure of the ACE strongly resembles that of its template 1CKV and other proteases belong to superfamily hydroxylase regulatory proteins, whose structures are known. Fig.3.3.2 depicts the schematic representation of the homology model ACE showing arrangement of secondary structural elements. The tertiary structure comprises with α + β structural topology, mainly antiparallel β-sheets (segregated alpha & beta regions).

```
1CKV 46  DEINTFIEEILLTDYKKNVNPTVNVEDRAGYWWIKANGKIEVDCDEISELLGRQFNVYDFLVDVSSTI
ACEV  1  MLMMDTLYDWLIDGFTWLVIKLGIMWIESKIFVIQFFWEMSQKVIDMFTIYPLIQQAIDMLPPQYSGF
           *       *                *                                * *    *

1CKV   GRAYTLGNKFTITSELMGLDRKLEDYHA 141
ACEV   LFFLGLDQALAIVLQALMTRFALRALNL 96
         *     *          *
```

Fig. 3.3.1. Target (ACE)-Template (1CKV) alignment by using ALIGN2D command of Modeller (9v8). * represents the conserved residues between the two sequences.

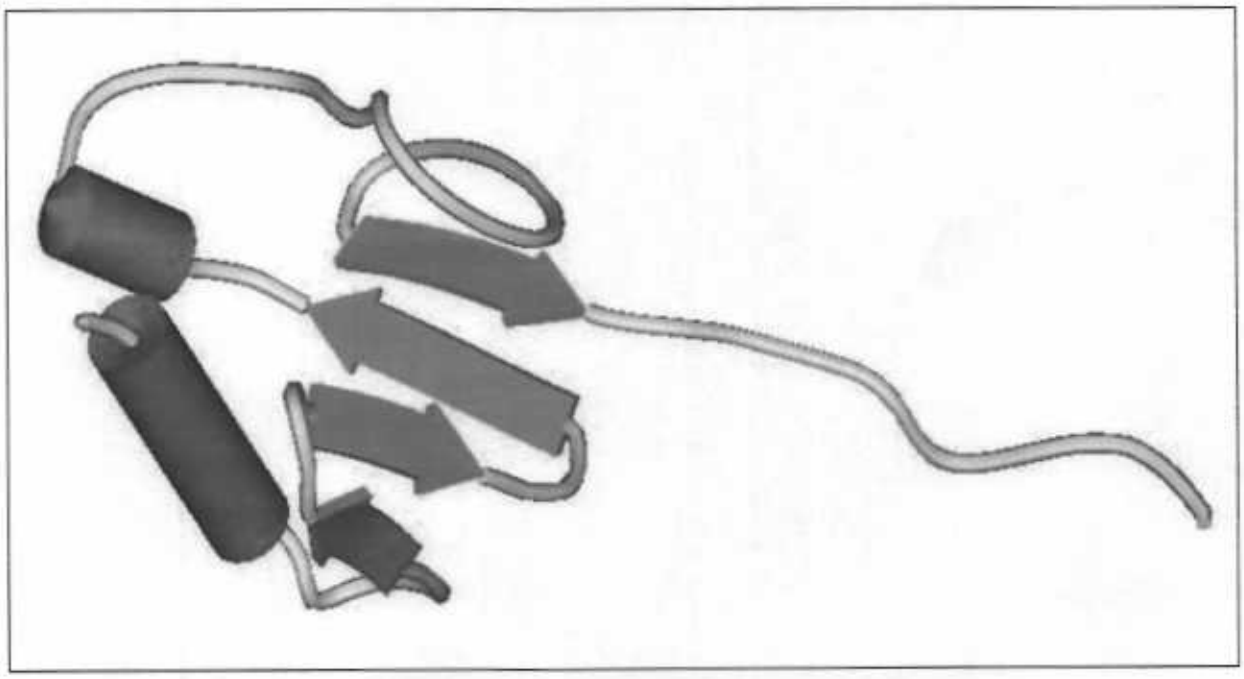


Fig.3.3.2. The schematic representation of the homology model ACE showing arrangement of secondary structural elements.

3.3.6 MODEL ASSESSMENT:

Table 3.3.1a shows Ramachandran plot statistics of the ACE models obtained with PROCHECK. Ramachandran plot indicates that 64.4% of the main-chain dihedral angles are found in the most favored regions, 27.6% in the additionally allowed, and 8.0% in the generously allowed while no residue is found in the disallowed regions. The structural conformation of the predicted ACE model is similar to X-ray structure of 1CKV. Table 3.3.1b shows Ramachandran plot statistics of template 1CKV.

Fig.3.3.3 (a & b) shows the energy graph of Ace calculated by ProSA in term of Z-score (-0.17) which indicated the overall quality of protein structure. The energy graph for 1CKV (Template) (Z-score: -4.28) molecule is presented in Appendix VI.

The overall description of similarities and differences derived from backbone superposition is given in Fig. 3.3.4. The RMSD value between template and the predicted model using all main-chain atoms was found to be 0.7577A °. Thus this value and the small variability between model and the template reflect the presence of strong restraints in most regions. As expected from the low sequence identity, the two structures are not well superimposed and but have similar folding topology and have different conformations at certain regions.

Table 3.3.1a. Table shows Ramachandran plot statistics of the ACE models obtained with PROCHECK

```
+----------<<< P R O C H E C K    S U M M A R Y >>>----------+
|                                                             |
| 1CKV   2.0                                        96 residues |
|                                                             |
*| Ramachandran plot:   64.4% core   27.6% allow    8.0% gener     .0% disall |
|                                                             |
*| Gly & Pro Ramach:     1 labelled residues (out of   7)      |
*| Chi1-chi2 plots:     10 labelled residues (out of  65)      |
|                                                             |
+| Main-chain params:    5 better     0 inside      1 worse    |
*| Side-chain params:    1 better     0 inside      4 worse    |
|                                                             |
*| Residue properties: Max.deviation:    20.6          Bad contacts:   13 |
|                     Bond len/angle:    1.5    Morris et al class:  3  4  3 |
|                                                             |
+| G-factors           Dihedrals:   -.66  Covalent:    .65    Overall:   -.15 |
|                                                             |
| M/c bond lengths:100.0% within limits    .0% highlighted    |
| M/c bond angles: 100.0% within limits    .0% highlighted    |
| Planar groups:   100.0% within limits    .0% highlighted    |
|                                                             |
+-------------------------------------------------------------+
  + May be worth investigating further.  * Worth investigating further.
```

Table 3.3.1b. Table shows Ramachandran plot statistics of template 1CKV obtained with PROCHECK.

```
+----------<<< P R O C H E C K    S U M M A R Y >>>----------+
|                                                             |
| ACEV23   2.0                                      96 residues |
|                                                             |
*| Ramachandran plot:   75.9% core   14.9% allow    9.2% gener     .0% disall |
|                                                             |
*| Gly & Pro Ramach:     1 labelled residues (out of   7)      |
*| Chi1-chi2 plots:      4 labelled residues (out of  72)      |
|                                                             |
| Main-chain params:    6 better     0 inside      0 worse    |
| Side-chain params:    5 better     0 inside      0 worse    |
|                                                             |
+| Residue properties: Max.deviation:    13.8          Bad contacts:    9 |
+|                     Bond len/angle:    4.9    Morris et al class:  1  1  3 |
|                                                             |
+| G-factors           Dihedrals:   -.33  Covalent:   -.54    Overall:   -.39 |
|                                                             |
| M/c bond lengths: 86.3% within limits  13.7% highlighted    |
| M/c bond angles:  71.2% within limits  28.8% highlighted    |
| Planar groups:   100.0% within limits    .0% highlighted    |
|                                                             |
+-------------------------------------------------------------+
  + May be worth investigating further.  * Worth investigating further.
```

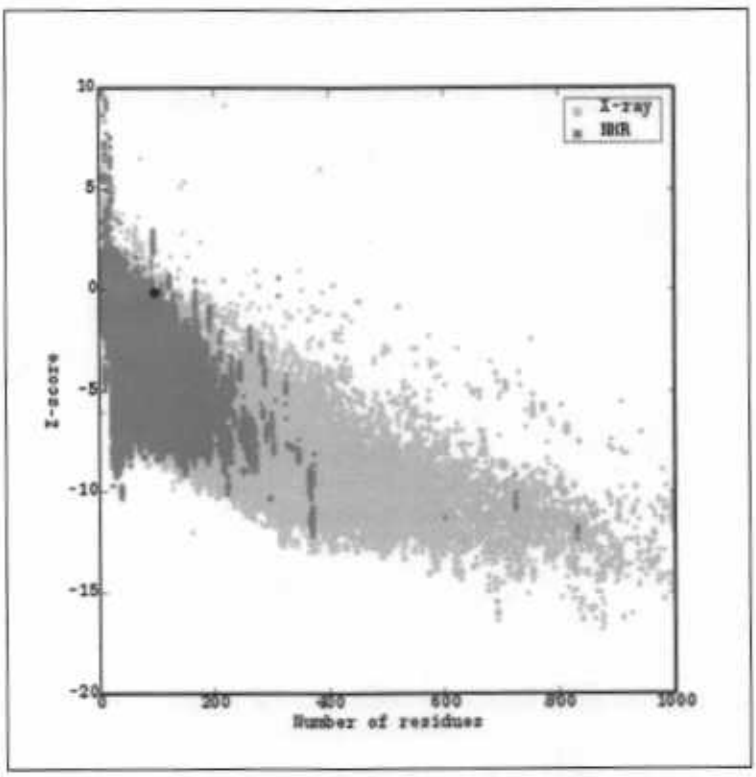


Fig.3.3.3a: Screenshot of ProSA z-score plot of Ace showing the Z value < 0.

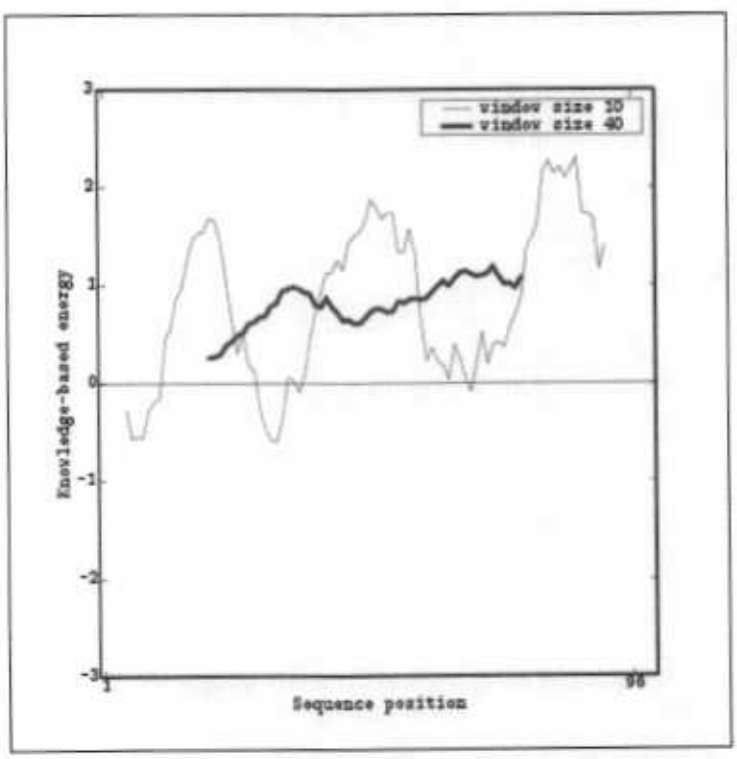


Fig.3.3.3b: Screenshot of ProSA plot of Ace showing the energy graph of residue scores of a predicted protein structure.

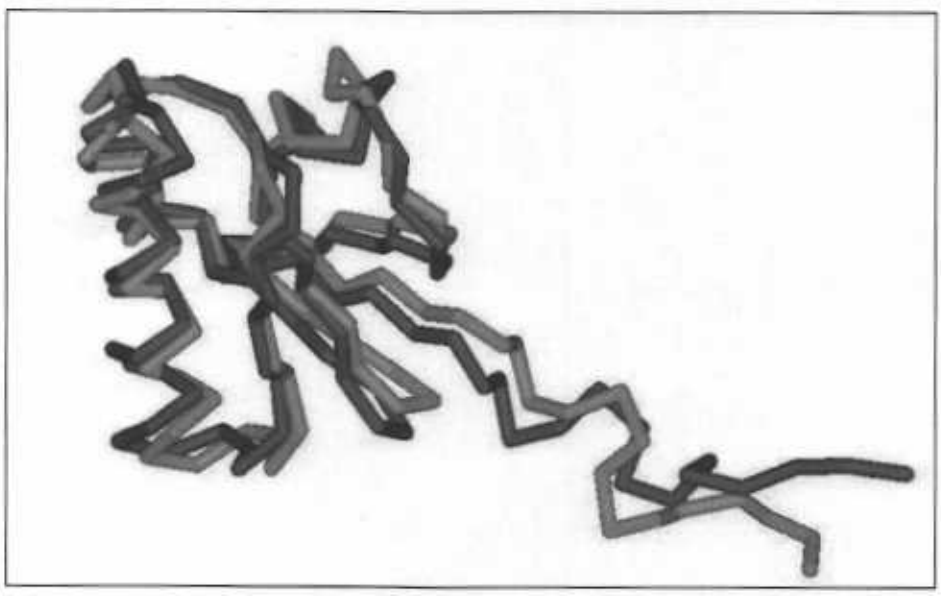


Fig.3.3.4. Structural superposition of Cα atoms of Ace (magenta) onto the crystal structure of 1CKV (blue).

CHAPTER

4.0

DISCUSSION

4.1. HEMAGGLUTININ/PROTEASE [HA/P (VIBRIOLYSIN)]

4.1.1. Biological Importance:

HA/P (vibriolysin) belongs to a family of metalloproteases (M4:M04.003) which is largely distributed among pathogenic bacteria (34,36). HA/P can proteolytically activate CT A subunit by nicking it into two (A_1 and A_2) peptide fragments (36,110) and increasing its biological activity (35). It can also activate the E1 Tor cytolysin and hemolysin (31, 36) and is capable of hydrolyzing several other physiologically important proteins such as mucin, fibronectin and lactoferin (111,32). It has been described previously that HA/P may play a vital role in intestinal colonization, and thus in virulence (2,112), though there is no clear evidence of direct involvement of HA/P in the pathogenesis of cholera (34). Both its hemagglutinative and protease functions are inhibited by a variety of chelating agents and inhibitors that are particularly designed to inhibit zinc metalloproteases (33).

The present study deals with the prediction of 3D model of hemagglutinin/Protease [HA/P (vibriolysin)] based on comparative homology modeling. The Enzyme-Inhibitor model was also constructed separately in order to identify important residues that contribute to the active site of the enzyme.

4.1.2. Sequence Analysis:

The M4 family is also known as Thermolysin like proteases (TLPs). All members of this family are synthesized as preproenzymes and processed after secretion of the inactive pro-enzyme to their respective active mature enzyme. It has been shown that the pro-sequence act as a molecular chaperone to support in proper folding and preventing unwanted proteolytic activity in the cytoplasm by acting as a specific inhibitor (113).

HA/P is synthesized as an inactive preproenzyme of 609 amino acid residues, 1-24 amino acid residues act as signal sequence whereas the prosequence comprises 25-196 residues. The mature protein has 413 amino acids after processing. Sequence similarity searches of HA/P (SwissProt AC: P24153) showed 62% identity with Pseudomonas Aeruginosa Elastase (PAE) (Swiss Prot AC: P14756) with two insertions at positions between Arg29 and Cys30 and between Thr60 and Asn61 in the template PAE, four gaps in each insertion and one insertion of two gaps in target sequence at position between Thr248 and Ser249. 190 amino acids (62%) out of 303 residues are identical and an additional 103 represent conservative amino acid substitutions.

ALIGN2D is based on a dynamic programming method, it is diverse from standard sequence-sequence alignment methods because it acquire into account structural information from the template when creating an alignment, as a result, the alignment errors are reduces. The align2d command of MODELLER realigned the target and template sequences, as a result reduced the gaps in vibriolysin sequence while shifted the gaps between Asn61 and Thr62 in the template (PAE) sequence as shown in Fig.3.1.1.

The first 20 amino acid residues of the N-terminal of two proteins shared 65% identity. The primary structures of both proteins show that the active site residues ($Glu147^{HA/P}$ & $His229^{HA/P}$) and the metal (Zn) binding amino acid residues ($His146^{HA/P}$, $His150^{HA/P}$, and $Glu170^{HA/P}$) are conserved. In these proteins 53% (1-140) of the residues are identical in the N-terminal domain. There is 80% sequence identity in the region surrounding the two active-site helices (141-185). In the remainder of the two proteins (P24153 & P14756), 69% of the residues (186-301) are identical. The active-site cleft is highly conserved; N-terminal domain is the least, while the C-terminal is intermediately conserved.

Comparison of multiple sequences shows considerable sequence similarity among primary structures of all known neutral proteases of M4 family (Appendix II). The active site and the metal binding residues are highly conserved. HEXXH is the putative signature of Zn-metalloproteases. This motif contains all of the residues that have been positively identified as part of the active site (114-115). HA/P also contains the typical HEXXH amino acid motif. It has been shown that the two conserved His and a Glu found in the unique signature (HEXXH) are crucial for the catalytic activity of neutral proteases (116-117). In addition, these enzymes also possess EXXXD motif. Both motifs are located in the helical region, which comprise the catalytic domain (118). It is clear from the multiple sequence alignment that the sequence similarity among the family extended beyond the HEXXH motif. Within a 10 amino acid stretch extending from three amino acids before the first His to two amino acids after the second His, there is no charged amino acid other than His and Glu. Two hydrophobic amino acids are located after the second His (114). Previously, the unique putative signature that has been redefined as (uncharged)-(uncharged)-H-E-(uncharged)-(uncharged)-H-(uncharged)-(hydrophobic) is common to Zn-dependent metalloproteases (114-115).Using this signature the similar regions among the neutral proteases are found to be conserved (115).

4.1.3. Topology of Predicted Homology Model of HA/P:

The overall tertiary structure of the HA/P strongly resembles that of its template PAE and other neutral metalloproteases, whose structures are known. Fig. 3.1.2 depicts the schematic representation of the homology model of vibriolysin showing arrangement of secondary structural elements. The tertiary structure comprises two major domains with $\alpha + \beta$ structural topology. The N-terminal domain contains mainly β-sheets, whereas the C-

terminal domain predominantly contains α-helices. The domains are connected by a central α-helix. The active site is located in the cleft between these two domains.

The catalytically essential Zn-ion and the central helix are located at the bottom of the active site cleft. Comparative analyses of sequence alignment and 3D structures of PAE and HA/P resulted in the identification of active site residues ($Glu147^{HA/P}$ and $His229^{HA/P}$), and the Zinc-binding residues ($His146^{HA/P}$, $His150^{HA/P}$ and $Glu170^{HA/P}$) that are located in the α-helical region (Fig.4.1). In both proteins the helices that cover the active site and Zn-binding residues are referred to as "Active-site helices" (119). These residues as well as the loops connecting those (comprising residues 145-230) are identical in length. Another structural component that is found to be similar to the known structures of bacterial neutral proteases is the presence of four-helix bundle within the residues of the C-terminal domain. Several of the β-strands, as well as the single α-helix in the amino-terminal domains of the proteins are similar (Fig.3.1.2).

4.1.4 Quality of the Model

Using different validation tools assessed quality of the model. The PROCHECK program was used to validate the model based on stereo chemical properties of main-chain and side chain residues and geometric considerations. The Ramachandran plot shows the main chain N-Cα and Cα-C bonds for each residue in protein known as phi and psi bond. The homology model of vibriolysin satisfies the stereo-chemical restraints and passed all criteria implemented in PROCHECK to the same degrees as the crystal structure refined at $2.5A^{0}$ resolution. Ramachandran plot indicates that 88.7% of the main-chain dihedral angles are found in the most favored regions, 10.5% in the additionally allowed, 0.8% in the generously allowed while no residue is found in the

disallowed regions (Appendix V). In accordance, the model show relatively good protein geometry, and most of the quality parameters are better than or in the range of tolerance. Therefore the models can be considered as structurally realistic.

The energetic architecture of protein folds was determined by using the program ProSA (Fig.3.1.3a & b). ProSA calculates an overall quality score for a specific input structure and displayed in a plot that shows the scores of all experimentally determined protein chains currently available in the Protein Data Bank (PDB). If this score is outside a range characteristic for native proteins the structure probably contains errors. In general, positive values correspond to problematic or erroneous parts of the input structure. Ideally the Z-score should be below the Zero point suggesting no significant stressed or strained folds with high energies (107). The energy graph of the predicted model was explored to observe the energy difference. Energy difference in the region between residues 10-100 might be due to substitution of amino acids in this region. Z-score was used as the measure of this energy, which indicated the quality of protein structures. The *z*-score (-7.89) indicates overall model quality (Fig.3.1.3a). Its value is displayed in a plot that contains the *z*-scores of all experimentally determined protein chains in current PDB. In this plot, groups of structures from different sources (X-ray, NMR) are distinguished by different colors. It can be used to check whether the *z*-score of the input structure is within the range of scores typically found for native proteins of similar size. Fig.3.1.3b shows local model quality by plotting energies as a function of amino acid sequence position. In general, positive values correspond to problematic or erroneous parts of the input structure. A plot of single residue energies usually contains large fluctuations and is of limited value for model evaluation. Hence the plot is smoothed by calculating the

average energy over each 40-residue fragment (Fig.3.1.3b, thick line). A second line with a smaller window size of 10 residues is shown in the background of the plot (Fig.3.1.3b, thin line). As is evident from the result, the energy graphs (Fig.3.1.3b) corresponded to highly stable structures and the values of energy, for the target molecule, well below Zero point, which was in consistence with the values of experimentally found Template molecule (Z-score: -8.63) (Appendix VI) . That is why, from the energy stabilizations point of view, there seemed to be no problems in the modeled structure and on the basis of this graph, the modeled 3-d structure can be acknowledged.

The α-carbon backbones of elastase and HA/P were aligned for comparison. The overall description of similarities and differences derived from backbone superposition is given in Fig.3.1.4. The RMSD value between template and the predicted model using all main-chain atoms was found to be $0.467A^0$ and $0.0916A^0$ in absence and presence of inhibitor HPI, respectively. Thus this value and the small variability between model and the template reflect the presence of strong restraints in most regions. As expected from the high sequence identity, the two structures are well superimposed and have similar folding topology but have different conformations at certain regions e.g. in region 29-34 and 70-72 because of the insertions of amino acids in these loop regions. The active site and metal binding residues are well superimposed indicating that the active sites of the two proteins are identical with similar geometry. The amino acid residues that constitute the sub-site region that might interact directly with the substrates and/or inhibitors are also identical and well superimposed. (Fig. 4.2)

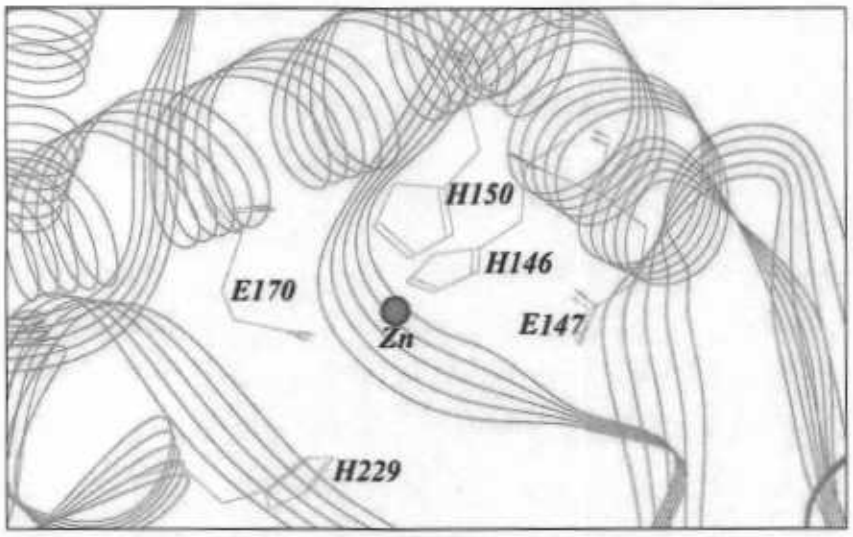


Fig.4.1. Zn-binding residues (line representation), His146, His150, and Glu170 and active site residues Glu147 and His229 are located in alpha helices (line ribbon).

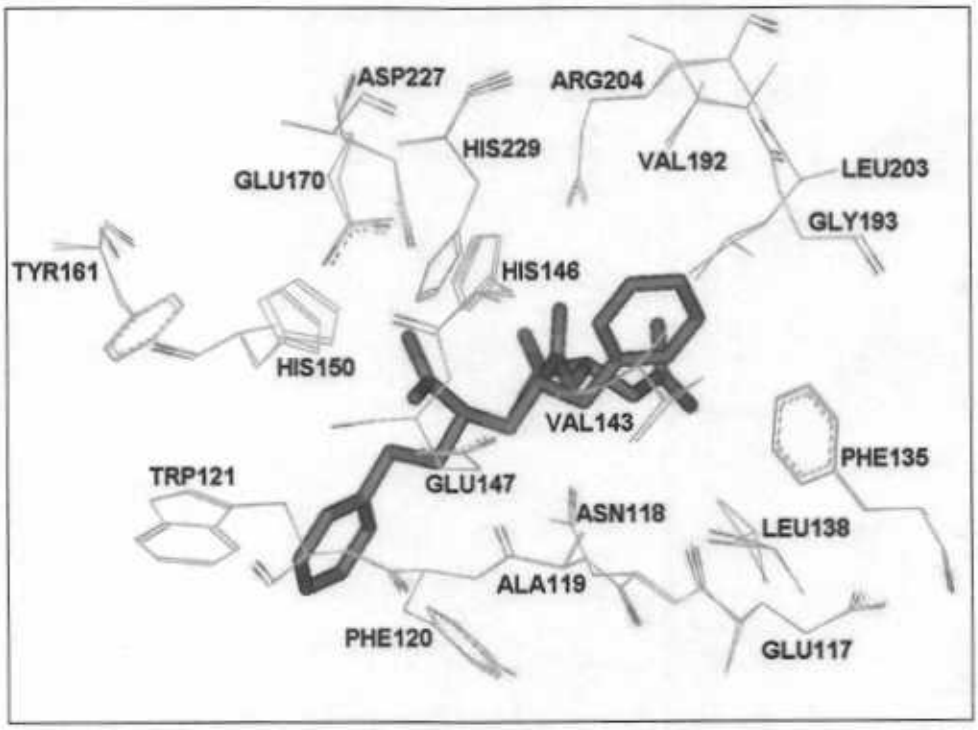


Fig. 4.2. The active site residues and metal binding residues (line representation) are well superimposed. The amino acid residues of enzyme segment that are involved in the hydrophilic and hydrophobic interactions (line representation) with the substrates and/or inhibitor (stick) are well superimposed.

4.1.5. Active Site and Protein-Ligand Interactions:

Both PAE and HA/P are metalloproteases, containing one zinc atom per molecule that is essential for their catalytic activity. The Zn atom is tetrahedrally co-coordinated with three ligands provided by HA/P His146 (NE2), His150 (NE2), Glu170 (OE2) and with a water molecule providing the fourth ligand (Fig.4.3). The crystal structure of thermolysin from Bacillus thermoproteolyticus, with various ligands, has been studied more comprehensively among the neutral metalloproteases (120). On the basis of crystallographic analyses Matthews and coworkers suggested a potential mechanism of catalysis for this enzyme (116-117, 121). On the basis of high sequence and structural homology with PAE and other neutral metalloproteases in the region of active site and Zn-binding environment it may be predicted that Glu147 and His229 are the critical residues for the catalytic activity of the HA/P as in the case with the analogous residues in PAE and thermolysin (122-123).

A mechanism of catalysis proposed by Matthews and coworkers for thermolysin activity is a general base type mechanism. According to this mechanism the carbonyl oxygen of the incoming substrate to be hydrolyzed is coordinated to the Zn-ion of protein. The water molecule, normally the fourth ligand to the zinc, is displaced towards Glu147 ($Glu143^{TLN}$) forming a pentacoordinated complex, but is not totally excluded from the zinc coordination sphere. There is a nucleopilic attack of this water molecule on the carbonyl carbon of the substrate to form a tetrahedral transition-state intermediate. The attacking water molecule is activated by the combined influence of the Zn-ion and Glu147 (Fig7a). The transition state intermediate is stabilized by interactions with the Zn-ion and by hydrogen bonds from His229 ($His231^{TLN}$) and Tyr161 ($Tyr157^{TLN}$).

According to the proposed model the active site glutamic acid ($Glu147^{HA/P}$) acts as a "Proton shuttle", transferring a proton from the attacking water molecule to the nitrogen of the scissile peptide bond leading to breakage of the bond and subsequent release of products (128, 133) (Fig.4.4a & 4.4b). In the enzyme inhibitor complex Zn-ion is penta-coordinated by three protein ligands and two atoms of the inhibitor HPI moiety, rather than three protein ligands plus a water molecule (Fig.4.4c).

The inhibitor (HPI) is associated with the enzyme segment by hydrogen bonding, hydrophobic and weak Van der Waal's interactions. In PAE, forming eleven hydrogen bonds with enzyme stabilizes HPI; similar hydrogen bonds were observed in HA/P. In the enzyme inhibitor complex, Asn118, Ala119, Glu147, Glu170, His229, and Arg204 (HA/P numbering) participate in a number of hydrogen bonds with the solvent and the inhibitor; two nitrogen atoms and three oxygen atoms of HPI moiety forming hydrogen bonds with the protein (Fig.4.5).

Residues Tyr114, His140, Val137, Ile186, Gly187, Leu197 of PAE participates in hydrophobic interactions with the inhibitor. In HA/P same residues (Val143, His146, Gly193, Leu203) are involved in hydrophobic interactions with the bound inhibitor except Tyr114 and Ile186, which are replaced by $Phe120^{HA/P}$ and $Val192^{HA/P}$ respectively (Fig.4.6). In addition to these interactions Glu117, Trp121, Glu147, Asp227, and His229 also interact with the ligand HPI by means of water molecules (Fig.4.7).

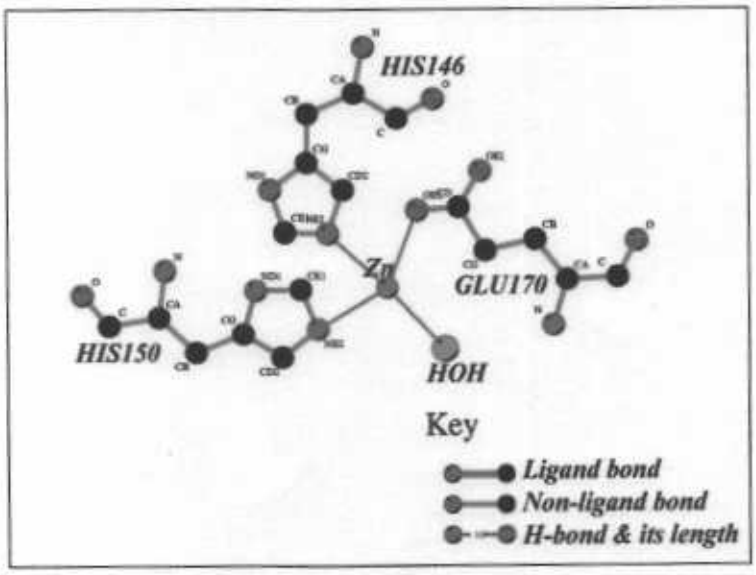


Fig.4.3. The ligplot result shows the interactions of metal ion (Zn) with the amino acid residues of the protein. An atom of Zn bound with a tetra-co-ordinate geometry to two Histidines (His146 & 150) and a Glutamate (Glu170), with water molecule providing the fourth ligand.

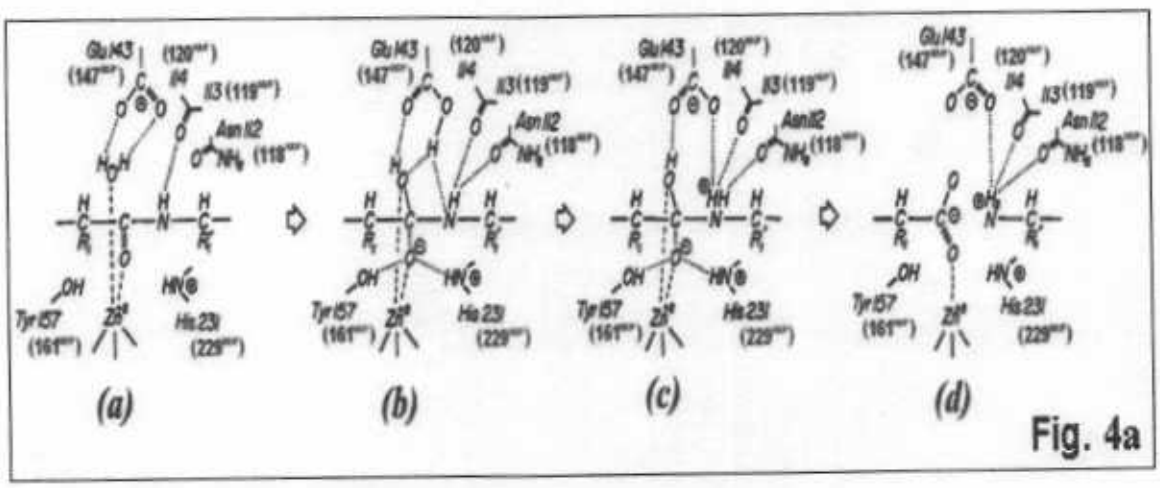


Fig.4.4 (a): Schematic representation of the key features of the presumed mechanism of action of thermolysin The catalysis proceeds by a general base-type mechanism as described by Matthew & coworkers [Modified from Matthews (116)]. Similar residues are involved in the catalytic activity of HA/P represented by the numbers with superscript HA/P.

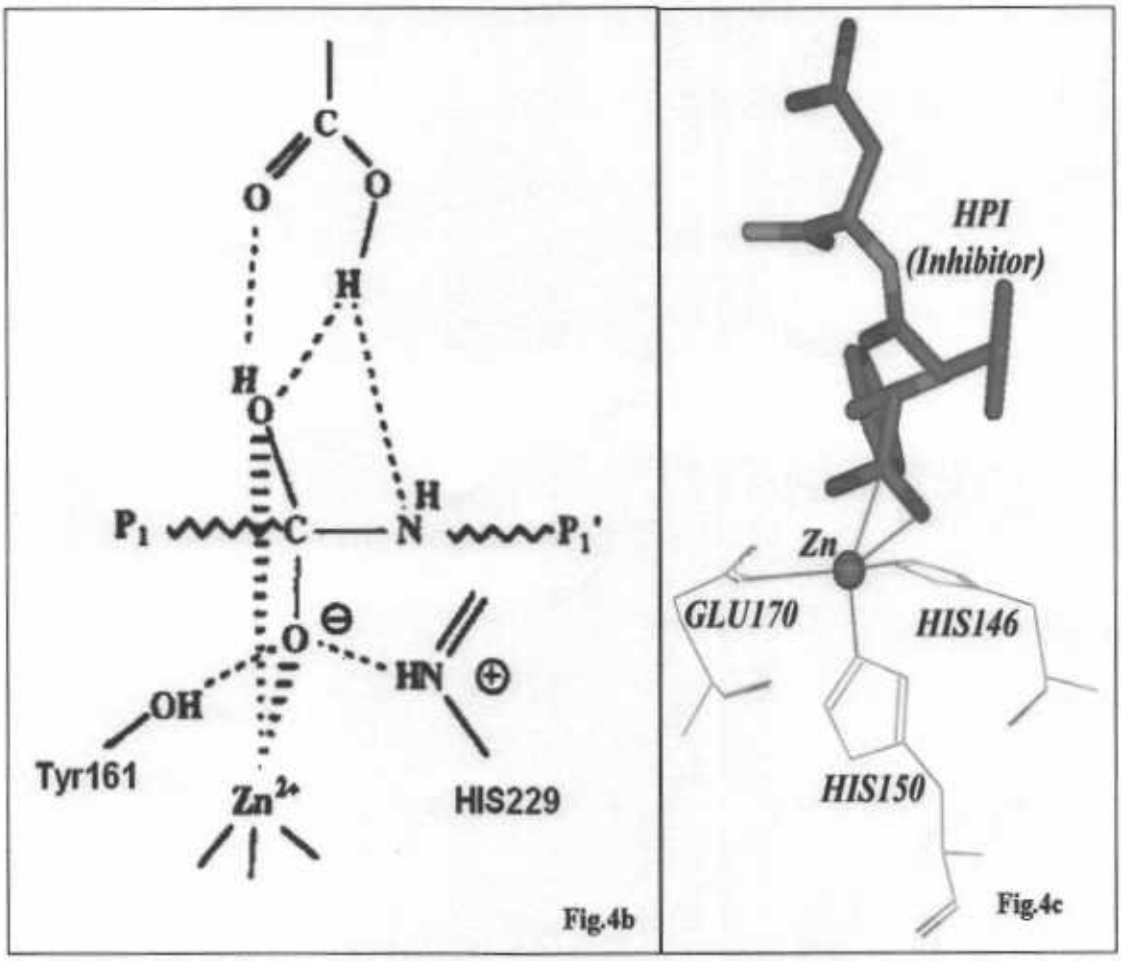


Fig.4.4 (b): An active site His-229 has been proposed to stabilize the transition state in catalysis by donating a hydrogen bond to the hydrated peptide [Adapted from Matthews (121)].**(c):** Zn-ion is penta-coordinated in the enzyme inhibitor complex.

In TLPs four substrate binding pockets (S_2, S_1, S_1', S_2') have been identified (118, 121). The S_1' sub-site is a deep hydrophobic pocket (112,123) that can accommodate side chains such as valine, leucine and phenylalanine (121). S_1' sub-site is mainly composed of Phe130, Leu/Phe133, Val139, and Leu202 (Thermolysin numbering) (112,123). In HA/P, Phe129, Leu132, Val143 and Leu203 constitute the S_1' sub-site. The hydrophobic S_1' sub-site is one of the major determinants of the substrate specificity of thermolysin and related M4 peptidase family (113, 124) Substrates with Phe residue at P_1' fit well in the S_1' sub-site (118,124). The S_1 and S_2' sub-sites are partially hydrophilic pockets. The substrates with benzyl group at R1 subsitituent oriented towards the S_1 binding sub-site that is composed of Asn112, Phe114, and Tyr157 (Thermolysin numbering) (113). In HA/P, the S_1 sub-site is composed of Asn118, Phe120 & Tyr161. In enzyme inhibitor complex of both proteins (HA/P & PAE) the inhibitor [N- {(1-carboxy-3-phenylpropyl)-phenylalanyl-alpha-aspargine} (HPI)] occupies the S_1-S_1' subsites.

4.1.6. Phylogenetic Analysis:

A phylogenetic tree was generated by using the multiple sequence alignment information of eighteen sequences with the help of PHYLIP program for exploring the evolutionary relationships among Thermolysin subfamilies. The tree clearly indicates the divergence of these eighteen operational taxonomic units (OTUs) from a common ancestral gene as indicated by the mid-point root. The phylogenetic tree for HA/P gene is presented in Appendix III.

Each branch in the dendrogram represents a point of divergence where HA/P gene has evolved to a new function. The distance between the branches corresponds to evolutionary distances as measured by how much the various sequences differ from one

another. It can observed from the dendrogram that the HA/P and PAE genes are clustered in a recently evolved branch. Another observation made from the dendrogram is that HA/P and PAE genes are more closely related to each other than either of the sequences in the tree. However, it can be seen clearly in the dendrogram that HA/P share a common ancestor with the PAE family as represented by the internal node.

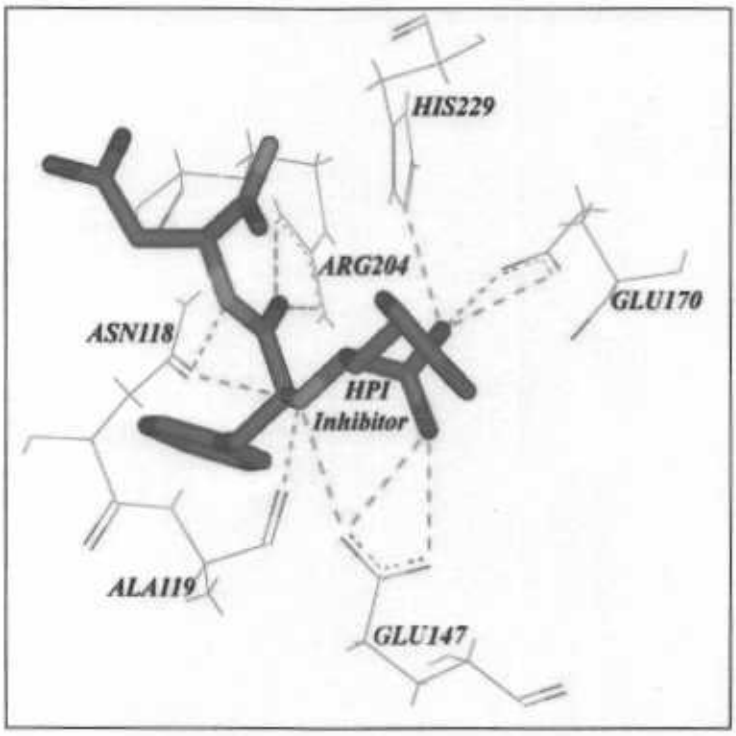


Fig.4.5. The binding of HPI inhibitor (Stick) with the enzyme segment (Line representation) by hydrogen bonding. Presumed Hydrogen bonds are represented as dotted line.

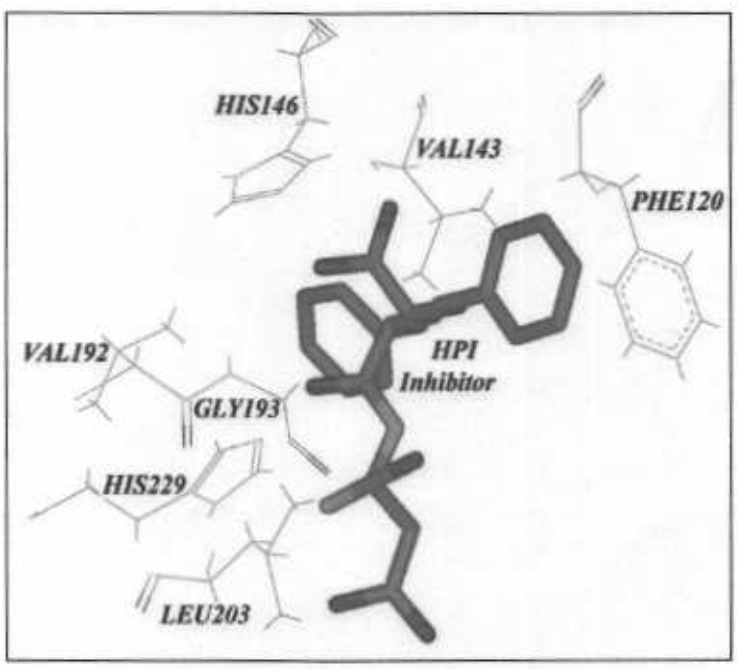


Fig.4.6. Amino acid residues (line representation) that are involved in hydrophobic interactions with inhibitor (Stick).

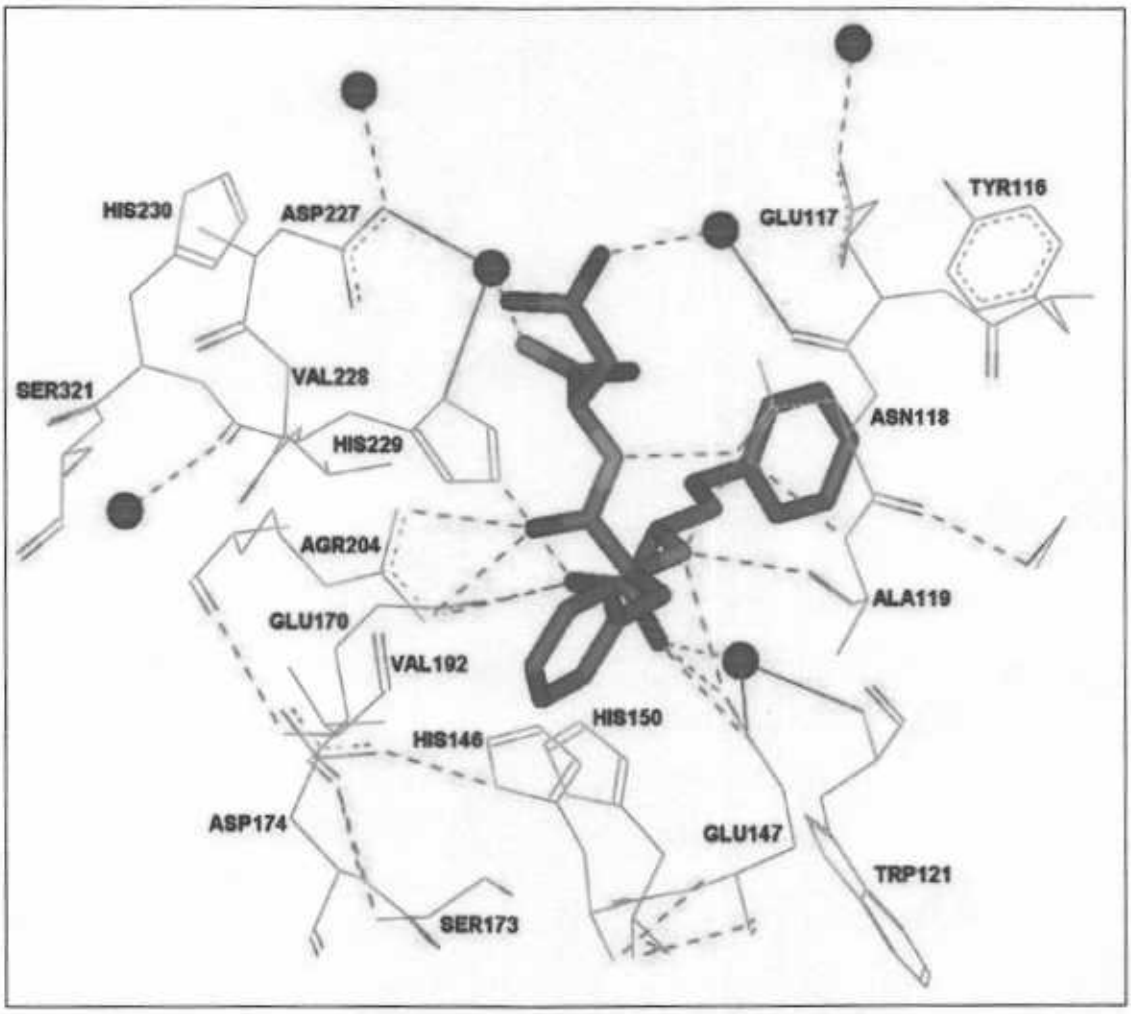


Fig.4.7. The binding of inhibitor (HPI) on the surface of active site of the HA/P model. Presumed protein-inhibitor, protein-protein and protein solvent hydrogen bonds are shown as dashed lines and inhibitor-protein interactions with water molecules are drawn as solid lines.

4.2. ZONA OCCLUDENS TOXIN (ZOt)

4.2.1. BIOLOGICAL IMPORTANCE:

Zonula occludens toxin (Zot) is an enterotoxin of a single polypeptide chain of 44.8 kDa (125-126).The gene encoding Zot is located on the CTX genetic element immediately upstream of *ctx* gene (127-128). In addition to Cholera Toxin, Zot has been reported a second toxin of a *V.cholerae* virulence cassette, that could contribute to reactogenicity suggested its role in the cause of acute dehydrating diarrhea typical of cholera (129-130). Zot affect the structure of epithelial tight junctions and increases the permeability of the small intestine. The mechanism of action of Zot on tight junctions involves a rearrangement of the epithelial cell cytoskeleton due to protein kinase C alphadependent F-actin polymerization (126-127).

4.2.2. Sequence Analysis:

Sequence similarity searches of Zot showed 26% identity with chain A of Peroxisomal 3-Ketoacyl-Coa Thiolase of *Saccharomyces Cerevisiae* with two insertions at positions between Arg274 and Tyr275 and between Val281 and Gly282 in the template 1PXT, seven gaps in each insertion and one insertion of eight gaps in target sequence at position between Val356 and Leu357 as shown in Appendix II. The two sequences showed no significant sequences similarity only 108 residues are found to align out of 399 residues. Only 25 residues are found to be conserved in both the sequences but these residues play no significant biological role (Fig.3.2.1). The amino acid sequence of Zot shoed no significant homology to sequences stored in SwissProt, as well as to the protein patterns found in the PROSITE and PRINTS databases.

Comparison of multiple sequences shows poor sequence similarity among primary structures of all known homologous proteins of proteobacteria Appendix II. As there is scant of knowledge about the primary sequence and function of Zot proteases so no enough information could be drawn from the sequence comparison of few sequences because each sequence belongs to the separate family. Only three residues are found to be least conserved among seven sequences of multiple alignment.

4.2.3. Topology of Predicted Homology Model of Zot:

The overall tertiary structure of the Zot strongly resembles that of its template 1PXT and other proteases belong to superfamily thiolase-like, whose structures are known. Fig. 3.2.2 depicts the schematic representation of the homology model Zot showing arrangement of secondary structural elements. The tertiary structure comprises with α/β structural topology, mainly antiparallel β-sheets (beta-alpha-beta units).The arrangement of secondary structural elements in Zot tertiary structure is thiolase-like consists of two similar domain and arrange in three layers of α/β/α.

4.2.4. Model Assessment:

Ramachandran plot indicates (Table 3.3.1a) that 85.9% of the main-chain dihedral angles are found in the most favored regions, 11.1% in the additionally allowed, 2.1% in the generously allowed while 0.9% (3 residues, Leu23, Leu97 & Thr300) was found in the disallowed regions (Appendix V). The structural conformation of the predicted Zot model is similar to X-ray structure of 1PXT. In accordance, the model show relatively good protein geometry, and most of the quality parameters are better than or in the range of tolerance. Therefore the models can be considered as structurally realistic.

The energetic architecture of protein folds was determined by using the program ProSA (Fig.3.2.3a & b). The energy graph of the predicted model was explored to observe the energy difference. Z-score was used as the measure of this energy, which indicated the quality of protein structures. The *z*-score (-0.09) indicates overall model quality (Fig.3.2.3a). Fig.3.2.3b shows local model quality by plotting energies as a function of amino acid sequence position. In general, positive values correspond to problematic or erroneous parts of the input structure. A plot of single residue energies usually contains large fluctuations and is of limited value for model evaluation. Hence the plot is smoothed by calculating the average energy over each 40-residue fragment (Fig.3.2.3b, thick line). A second line with a smaller window size of 10 residues is shown in the background of the plot (Fig.3.2.3b, thin line). As is evident from the result, the energy graphs (Fig.3.2.3b) corresponded to less stable structures and the values of energy, for the target molecule, is above Zero point, which is more than the values of experimentally found Template molecule (Z-score: -8.21) (Appendix VI). Thus from the energy stabilizations point of view, there seemed to be problems in the modeled structure which may be because of the low sequence identity between the target and template sequences but the overall Z-score is < 0 suggests that the predicted homology model fulfill the criteria of a good model.

The α-carbon backbones of 1PXT and Zot were aligned for comparison. The overall description of similarities and differences derived from backbone superposition is given in Fig. 3.2.4. The RMSD value between template and the predicted model using all main-chain atoms was found to be 0.5432A °. Thus this value and the small variability between model and the template reflect the presence of strong restraints in most regions. As

unexpected from the low sequence identity, the two structures are well superimposed and have similar folding topology but have different conformations at certain regions because of the difference in total number of residues between target (390AA) and template (443) alignment as well as because of the missing of few crystallized residues in the crystal structure of template (1PXT). In template 1PXT crystallized structure there are three regions (Asn161-Gln174, Asn192-Arg201 & Pro247-Gln275) of which the amino acids are not crystallized during the process of crystallization, so far these residues are missing from the 3D structure of template 1PXT, on the basis of which the homology model of Zot was constructed.

4.2.5. Phylogenetic Analysis:

A phylogenetic tree was drawn by using the multiple sequence alignment information of seven sequences for charting the evolutionary relationships. The tree supports the divergence of these seven operational taxonomic units (OUTs) from a common ancestral gene as indicated by the mid-point root. The phylogenetic tree for Zot has been presented in Appendix III. From the toplology of the Zot dendrogram the Zot gene is segregated in a clade that includes GIP_BPPHL and seemed to have separated in the evolutionary process. It can be observed from the dendrogram that THIK_YEAST (Template) seemed to have segregated in a slightly distinct branch, that could be attributed because of its taxonomic position as being the only fungi in the tree.

4.3.ACCESSORY CHOLERA ENTEROTOXIN (Ace)

4.3.1. BIOLOGICAL IMPORTANCE:

Accessory Cholera Enterotoxin (Ace) has been considered the third toxin of a *V.cholerae* virulence cassette. The gene encoding Ace is located on the CTX genetic element immediately upstream of the genes encoding Zot and cholera toxin. Across intestinal epithelial membrane, Ace increases the potential difference and alters ion transport may cause diarrhea. (127-128). The mechanism of action of Ace is still largely unknown, but its similarity with a family of ion-transporting ATPases may provide evidence. Ace is much smaller protein as compared to the proteins of this family but the greatest similarity of Ace and the ion-transporting ATPases is in the hydrophobic transmembrane domains. Ace is mainly a hydrophobic protein (128).

4.3.2. Sequence Analysis:

Sequence similarity searches of Ace showed 31% sequence identity with crystal structure of the Soluble Methane Monooxygenase Regulatory Protein B (1CKV.pdb) of *Escherichia coli* as shown in Appendix II. The two sequences showed significant sequences similarity but with low sequence identity as only 35 residues out of 96 residues showed sequence identity. The **align2d** command of MODELLER realigned the full length sequences of target (Ace) and template (1CKV) with improved sequence similarity. Only nine residues are found to be conserved in both sequences but these residues play no significant role in its function (Fig.3.3.1). The amino acid sequence of Zot shoed no significant homology to sequences stored in SwissProt, as well as to the protein patterns found in the PROSITE and PRINTS databases.

No appreciable numbers of sequences are found for constructing the multiple sequence alignment. Only sequences of four species are found to be homologous with Ace. Comparison of multiple sequences shows poor sequence similarity among primary structures of all known homologous proteins of proteobacteria (Appendix II).Only single residue (Gly73)is found to be conserved among the sequence of these four species, while only ten residues are found to be less conserved and only four residues are found to be least conserved. As there is no enough information is available about the primary sequence and function of Ace proteases so no significant information could be drawn from the sequence comparison of few sequences because each sequence belongs to the separate family.

4.2.3. Topology of Predicted Homology Model of ACE:

The overall tertiary structure of the ACE strongly resembles that of its template 1CKV and other proteases belong to superfamily hydroxylase regulatory proteins, whose structures are known. Fig.3.3.2 depicts the schematic representation of the homology model Ace showing arrangement of secondary structural elements. The tertiary structure is comprises with $\alpha + \beta$ structural topology, mainly antiparallel β-sheets (segregated alpha & beta regions).

4.3.4. Model Assessment:

Ramachandran plot indicates (Table 3.3.1a) that 75.9% of the main-chain dihedral angles are found in the most favored regions, 14.9% in the additionally allowed, 9.2% in the generously allowed while 0% is found in the disallowed regions (Appendix V). The structural conformation of the predicted Ace model is similar to X-ray structure of 1CKV. In accordance, the model show relatively good protein geometry, and most of the

quality parameters are better than or in the range of tolerance. Therefore the models can be considered as structurally realistic.

The energetic architecture of protein folds was determined by using the program ProSA (Fig.3.3.3a & b). The energy graph of the predicted model was explored to observe the energy difference. Z-score was used as the measure of this energy, which indicated the quality of protein structures. The *z*-score (-0.17) indicates overall model quality (Fig.3.3.3a). Fig.3.3.3b shows local model quality by plotting energies as a function of amino acid sequence position. In general, positive values correspond to problematic or erroneous parts of the input structure. A plot of single residue energies usually contains large fluctuations and is of limited value for model evaluation. Hence the plot is smoothed by calculating the average energy over each 40-residue fragment (Fig.3.3.3b, thick line). A second line with a smaller window size of 10 residues is shown in the background of the plot (Fig.3.3.3b, thin line). As is evident from the result, the energy graphs (Fig.3.3.3b) corresponded to less stable structures and the values of energy, for the target molecule, is above Zero point, which is more than the values of experimentally found Template molecule (Z-score: -4.28) (Appendix VI). Thus from the energy stabilizations point of view, there seemed to be problems in the modeled structure which may be because of the low sequence identity between the target and template sequences but the overall Z-score is < 0 suggests that the predicted homology model fulfill the criteria of a good model.

The α-carbon backbones of 1PXT and Zot were aligned for comparison. The overall description of similarities and differences derived from backbone superposition is given in Fig. 3.3.4. The RMSD value between template and the predicted model using all main-

chain atoms was found to be 0.5432A °. Thus this value and the small variability between model and the template reflect the presence of strong restraints in most regions. As unexpected from the low sequence identity, the two structures are well superimposed and have similar folding topology but have different conformations at certain regions because of the difference in total number of residues between target (390AA) and template (443) alignment as well as because of the missing of few crystallized residues in the crystal structure of template (1PXT). In template 1PXT crystallized structure there are three regions (Asn161-Gln174, Asn192-Arg201 & Pro247-Gln275) of which the amino acids are not crystallized during the process of crystallization, so far these residues are missing from the 3D structure of template 1PXT, on the basis of which the homology model of Zot was constructed.

4.2.5. Phylogenetic Analysis:

A phylogenetic tree was drawn by using the multiple sequence alignment information of four sequences for studying the evolutionary relationships. The tree supports the divergence of these four operational taxonomic units (OUTs) from a common ancestral gene as indicated by the mid-point root. The phylogenetic tree for Ace is presented in Appendix III. From the topology of the phylip dendrogram the Ace gene is segregated in a clade that includes SYFA_METCA and seemed to have separated in the evolutionary process. It can be observed from the dendrogram that MMOB_METCA (Template) seemed to have segregated in a slightly distinct branch, that could be attributed because it belongs to different subfamily in the tree.

Conclusion:

- The primary structure of both proteins shows strong amino acid homologies in the regions that are thought to be the active site and Zn-binding domains.
- Comparison of multiple sequences of all known neutral metalloproteases of M4 family shows significant sequence similarity and supports the reality of the putative signature of Zn-metalloproteases (HEXXH). In addition, the metalloprotease family also contains another conserved EXXXD motif. Both these conserved motifs contain crucial residues that might potentially play an important role in the mechanism of catalysis.
- The overall topology and secondary structural elements are quite conserved in the M4 family. They comprise a two-domain structure, the N-terminal domain comprises several β-strands and a single α-helix while C-terminal domain shows α-helix bundle and a β-sheet. The domains are connected by a central α-helix. The Glu141 and His223 are the key residues for the catalytic activity of the *P. aeruginosa* elastase.
- On the basis of high sequence and structural similarity with *P. aeruginosa* elastase and other known TLPs it might be proposed that the corresponding residues Glu147 and His229 in hemagglutinin/ protease are catalytically active and support the model for the mechanism of action of thermolysin as proposed by Matthews and coworkers.
- The hydrophobic S_1' binding sub-site is the major determinant of the substrate specificity of TLPs. The analysis of ligand interactions and assessment of

hydrophobic S_1' shows that the differences in specificity of M4 family towards substrates do not associate with sequence variation but depends on a small number of identifiable amino acids.

- It has been shown that M4 metalloproteases have almost identical substrate preferences as they have identical residues at positions 133^{TLN} and 202^{TLN}. In fact substitution of a single amino acid can convert catalytic characteristics of one family member into that of another. Therefore, on the basis of identifiable amino acids in TLPs, specific inhibitors to be used for blocking disease-related members of the metalloproteases family can be designed.

- It may be predicted that the hemagglutinin/protease may be inhibited by the same set of inhibitors that are specifically designed to inhibit Zn metalloproteases.

- Both elastase and hemagglutinin/protease differ markedly in their resistance to heat inactivation; a comparison of their structures may provide a chemical basis for explaining the differences in their thermal stability.

- The structural information obtained from the model may facilitate efforts at designing more selective inhibitors that can be used to further probe the distinct biological role of the members M4 family and to explore the potential of these macromolecular species as therapeutic targets.

- The overall tertiary structure of the Zot & ACE strongly resembles that of their template 1PXT & 1CKV respectively. The tertiary structure of Zot comprises with α/β structural topology, while the tertiary structure of ACE is comprises with $\alpha + \beta$ structural topology.

- The predicted 3D models of Zot & Ace fulfill the criteria of a good quality model.
- The function and biological role of Zot & Ace could not be predicted from its 3D predicted structure as there is no information available on their active site residues. Neither their residues show any conservation with the active site residues of their respective template.

CHAPTER

5.0

REFRENCES

1. Halpern, M., Landsberg, O., Raats, D., Rosenberg,E: **Culturable and VBNC Vibrio cholerae: Interactions with chironomid egg masses and their bacterial population.** *Microb Ecol.*, 53, 285-293, **(2007).**

2. Sarkar, M., Das, S., Bandyopadhaya, A., Ray, K., Chaudhuri, K: **Upregulation of human mitochondrial NADH dehydrogenase subunit 5 in intestinal epithelial cells is modulated by Vibrio cholerae pathogenesis.** *FEBS Lett.*, 579, 3449-3460, **(2005).**

3. Sack, D. A., Sack,R.B., Nair,G.B., Siddique,A.K.: **Cholera.** *The Lancet*, 363, 223 - 233, **(2004).**

4. Faruque, S.M., Albert, M.J., Mekalanos,J.J.: **Epidemiology, genetics, and ecology of toxigenic Vibrio cholerae.** *Microbiol Mol Biol Rev.*, 62, 1301-1314, **(1998).**

5. Vanden, B. D., Horvath, C., De Wolf, M.J.: **Vibrio cholerae: cholera toxin.** *Int J Biochem Cell Biol.*, 39,1771-1775, **(2007).**

6. Shimada,T., Arakawa,E., Itoh,K., Okitsu,T., Matsushima,A., Asai,Y., Yamai,S., Nakazato,T., Nair, G. B., Albert, M. J., and Takeda,Y.: **Extended Serotyping Scheme for *Vibrio cholerae.*** *Current Microbiology*, 28, 175-178, **(1994).**

7. Theophilo,G.N., Rodrigues,D.P., Leal, N.C., Hofer, E.: **Distribution of virulence markers in clinical and environmental Vibrio cholerae non-O1/non-O139 strains isolated in Brazil from 1991 to 2000.** *Rev Inst Med Trop Sao Paulo.*, 48, 65-70, **(2006)**

8. Liang,W., Wang,S., Yu,F., Zhang,L., Qi,G., Liu,Y., Gao,S., Kan,B.: **Construction and evaluation of a safe, live, oral Vibrio cholerae vaccine candidate, IEM108.** *Infect Immun.*, 71, 5498-5504 **(2003)**.

9. Li,M., Kotetishvili,M., Chen,Y., Sozhamannan,S.: **Comparative genomic analyses of the vibrio pathogenicity island and cholera toxin prophage regions in nonepidemic serogroup strains of Vibrio cholerae.** *Appl Environ Microbiol.*, 69, 1728-1738, **(2003)**.

10. Qadri,F., Ahmed,F., Karim,M.M., Wenneras,C., Begum,Y.A., Abdus Salam,M., Albert,M.J., McGhee,J.R.: **Lipopolysaccharide- and cholera toxin-specific subclass distribution of B-cell responses in cholera**. *Clin Diagn Lab Immunol.*, 6, 812-818, **(1999)**.

11. Stroeher,U.H., Karageorgos,L.E., Morona,R., and Manning,P.A.: **Serotype conversion in Vibrio cholerae O1.** *Proc Natl Acad Sci U S A.*, 89, 2566–2570, **(1992)**.

12. García,L., Jidy,M.D., García,H., Rodríguez,B.L., Fernández,R., Año,G., Cedré,B., Valmaseda,T., Suzarte,E., Ramírez,M., Pino,Y., Campos,J., Menéndez,J., Valera,R., González,D., González,I., Pérez,O., Serrano,T., Lastre,M., Miralles,F., Del Campo,J., Maestre,J.L., Pérez,J.L., Talavera,A., Pérez,A., Marrero,K., Ledón,T., Fando,R.: **The vaccine candidate Vibrio cholerae 638 is protective against cholera in healthy volunteers.** *Infect Immun.*, 73, 3018-24, **(2005)**.

13. Nair,G.B.: *Vibrio Cholerae*. Review. National Institute of Cholera & Enteric diseases, Calcutta, India.,

14. Zampini,M., Pruzzo,C., Bondre,V.P., Tarsi,R., Cosmo,M., Bacciaglia,A., Chhabra,A., Srivastava,R., Srivastava,B.S.: **Vibrio cholerae persistence in aquatic environments and colonization of intestinal cells: involvement of a common adhesion mechanism.** *FEMS Microbiol Lett.*, 244, 267-273, **(2005)**.

15. Roy, S., Dutta,B., Ghosh, A.R., Sugunan,A.P., Nandy,R.K., Bhattacharya,S. K., and Sehgal,S.C.: **Molecular tracking of the lineage of strains of *Vibrio cholerae* O1 biotype El Tor associated with a cholera outbreak in Andaman and Nicobar Islands, India.** *Trop Med Int Health.*, 10, 604-611, **(2005)**.

16. Iqbal,K.J., Anisia,J.S., and Jorge,A.B.: **Polyphosphate Stores Enhance the Ability of *Vibrio cholerae* To Overcome Environmental Stresses in a Low-Phosphate Environment.** *Appl Environ Microbiol.*, 72, 7043–7049, **(2006)**.

17. Singh,D.V., Matte,M.H., Matte,G.R., Jiang,S, Sabeena, F., Shukla,B.N., Sanyal,S. C., Huq,A., and Colwell,R.R.: **Molecular analysis of Vibrio cholerae O1, O139, non-O1, and non-O139 strains: clonal relationships between clinical and environmental isolates.** *Appl Environ Microbiol.*, 67, 910–921, **(2001)**.

18. Davis,B.M., Waldor,M.K.: **Filamentous phages linked to virulence of Vibrio cholerae.** *Curr Opin Microbiol.*, 6, 35-42, **(2003)**.

19. Sarkar,A., Nandy,R.K., Nair,G.B., Ghose,A.C.: **Vibrio pathogenicity island and cholera toxin genetic element-associated virulence genes and their expression in non-O1 non-O139 strains of Vibrio cholerae.** *Infect Immun.*, 70, 4735-4742, **(2002)**

20. Lencer,W.I.: **Microbes and Microbial Toxins: Paradigms for Microbial-Mucosal Interactions V. Cholera: invasion of the intestinal epithelial barrier**

by a stably folded protein toxin. *Am J Physiol Gastrointest Liver Physiol.*, 280,G781-G786, **(2001).**

21. Rivera,I.N., Chun.J., Huq,A., Sack,R.B., Colwell,R.R.: **Genotypes associated with virulence in environmental isolates of Vibrio cholerae.** *Appl Environ Microbiol.*, 67, 2421-2429, **(2001).**

22. Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L., Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson, D., Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I., Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D., Nierman,W.C, White, O., Salzberg,S.L, Smith,H.O, Colwell,R.R., Mekalanos,J.J., Venter,J.C., Fraser,C,M.: **DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.***Nature*,406, 477-483,**(2000).**

23. Edwards, K.A., March, J. C.: **GM (1)-functionalized liposomes in a microtiter plate assay for cholera toxin in Vibrio cholerae culture samples.** *Anal Biochem.*, 368, 39-48, **(2007).**

24. Fan,E., O'Neal,C.J., Mitchell,D.D., Robien,M.A., Zhang,Z., Pickens,J.C, Tan,X.J., Korotkov,K., Roach,C., Krumm,B., Verlinde,C.L., Merritt,E.A., Hol, W.G.: **Structural biology and structure-based inhibitor design of cholera toxin and heat-labile enterotoxin.** *Int J Med Microbiol.*, 294, 217-223, **(2004).**

25. Fujinaga, Y., Wolf, A.A., Rodighiero, C., Wheeler, H., Tsai, B., Allen, L., Jobling, M.G., Rapoport, T., Holmes, R.K., Lencer, W.I.: **Gangliosides that associate with lipid rafts mediate transport of cholera and related toxins**

from the plasma membrane to endoplasmic reticulm. *Mol Biol Cell.*, 14, 4783-4793, **(2003).**

26. Figueiredo,S.C., Neves-Borges,A.C., Coelho,A.: **The neuraminidase gene is present in the non-toxigenic Vibrio cholerae Amazonia strain: a different allele in comparison to the pandemic strains.** *Mem Inst Oswaldo Cruz.*, 100, 563-569, **(2005).**

27. Hinou,H., Kurogochi,M., Shimizu,H., Nishimura,S.: **Characterization of Vibrio cholerae neuraminidase by a novel mechanism-based fluorescent labeling reagent.** *Biochemistry*, 44(35), 11669-11675, **(2005).**

28. Galen,J.E., Ketley,J.M., Fasano,A., Richardson,S.H., Wasserman,S.S., Kaper, J.B.: **Role of Vibrio cholerae neuraminidase in the function of cholera toxin.** *Infect Immun.*, 60, 406-415, **(1992).**

29. Häse,C.C., Finkelsteinm,R.A.: **Comparison of the Vibrio cholerae hemagglutinin/protease and the Pseudomonas aeruginosa elastase.** *Infect Immun.*, 58, 4011-4015, **(1990).**

30. Silva,A.J., Leitch,G.J., Camilli,A., Benitez,J.A.: **Contribution of hemagglutinin/protease and motility to the pathogenesis of El Tor biotype cholera.** *Infect Immun.*, 74, 2072-2079, **(2006).**

31. Benitez,J.A., Silva,A.J., Finkelstein,R.A.: **Environmental signals controlling production of hemagglutinin/protease in Vibrio cholerae.** *Infect Immun.*, 69, 6549-6553, **(2001).**

32. Hazra,A., Silva, A.J., Benitez,J.A.: **Expression of foreign proteins in a Vibrio cholerae vaccine strain using the stationary phase hemagglutinin/protease promoter.** *Biotechnol Lett.*, 29, 1093-1097, **(2007).**

33. Halpern,M., Gancz,H., Broza,M., Kashi,Y.: **Vibrio cholerae hemagglutinin/protease degrades chironomid egg masses.** *Appl. And Env. Microbio.*,69, 4200-4204, **(2003).**

34. Wu,Z., Milton,D., Nybom,P., Sjö,A., Magnusson,K.E.: **Vibrio cholerae hemagglutinin/protease (HA/protease) causes morphological changes in cultured epithelial cells and perturbs their paracellular barrier function.** *Microb Pathog* ., 21, 111-123, **(1996).**

35. Booth,B.A., Boesman-Finkelstein,M., Finkelstein,R.A.: **Vibrio cholerae hemagglutinin/protease nicks cholera enterotoxin.** *Infect Immun*, 45, 558-560, **(1984).**

36. Silva,A.J., Pham,K., Benitez,J.A.: **Haemagglutinin/protease expression and mucin gel penetration in El Tor biotype Vibrio cholerae.** *Microbiology*, 149(Pt 7), 1883-1891, **(2003).**

37. Silva,A.J., Benitez,J.A.: **Transcriptional regulation of Vibrio cholerae hemagglutinin/protease by the cyclic AMP receptor protein and RpoS.** *J Bacteriol.*, 186, 6374-6382,**(2004).**

38. Ghosh,A., Saha,D.R., Hoque,K.M., Asakuna,M., Yamasaki,S., Koley,H., Das, S.S., Chakrabarti,M.K., Pal, A.: **Enterotoxigenicity of mature 45-kilodalton and processed 35-kilodalton forms of hemagglutinin protease purified from a**

cholera toxin gene-negative Vibrio cholerae non-O1, non-O139 strain. *Infect Immun.*, 74, 2937-2946, **(2006).**

39. Mukherjee,G., Biswas,A., Banerjee, K.K., Biswas, T.: **Vibrio cholerae hemolysin is apoptogenic to peritoneal B-1a cells but its oligomer shepherd the cells for IgA response.** *Mol Immunol.*, 45, 266-270, **(2008).**

40. Chattopadhyay,K., Banerjee,K.K.: **Unfolding of Vibrio cholerae hemolysin induces oligomerization of the toxin monomer.** *J Biol Chem.*, 278, 38470-38475, **(2003).**

41. Yamamoto,K., Ichinose,Y., Shinagawa,H., Makino,K., Nakata,A., Iwanaga,M., HondaT., Miwatani,T.: **Two-step processing for activation of the cytolysin/hemolysin of Vibrio cholerae O1 biotype El Tor: nucleotide sequence of the structural gene (hlyA) and characterization of the processed products.** *Infect Immun.*, 58, 4106-4116, **(1990).**

42. Wu, Z.: **Studies of Vibrio cholera (zonula occludens toxin and hemagglutinin/protease) and their effects on epithelial cells.** Department of Medical microbiology, Linköping University, Linköping, Sweden. **(1997).**

43. Mel, F.S., Fullner,J.K., Mackin,W.S., Lencer,I.W., and Mekalanos,J.J.: **Association of Protease Activity in *Vibrio cholerae* Vaccine Strains with Decreases in Transcellular Epithelial Resistance of Polarized T84 Intestinal Epithelial Cells.** *Infect Immun.*, 68, 6487–6492, **(2000).**

44. Bandyopadhaya, A., Sarkar, M., Chaudhuri, K.: **Human intestinal epithelial cell cytokine mRNA responses mediated by NF-kappaB are modulated by the**

motility and adhesion process of Vibrio cholerae. *Int J Biochem Cell Biol.*, 39, 1863-1876,**(2007).**

45. Sur, D., Sarkar, B.L., Manna, B., Deen, J., Datta , S., Niyogi, S.K., Ghosh, A.N., Deb,A., Kanungo,S., Palit,A., Bhatta,H.S.K: **Epidemiological, microbiological & electron microscopic study of a cholera outbreak in a Kolkata slum community.** *Indian J Med Res.*, 123, 31-36, **(2006)**

46. C.Flach., F.Qadri., T.Bhuiyan., N.Alam., E.Jennische., J.Holmgren., I.Lönnroth.: **Differential expression of intestinal membrane transporters in cholera patients.** *FEBS Letters.*, 581, 3183-3188, **(2007).**

47. Naumann M.: **Nuclear factor-kB activation and innate immune response in microbial pathogen infection.** *Biochem Pharmacol.*, 60, 1109-1114, **(2000).**

48. Giulietti,A., Overbergh,L., Valckx,D., Decallonne,B., Bouillon,R., and Mathieu,C.: **An Overview of Real-Time Quantitative PCR: Applications to Quantify Cytokine Gene Expression**. *Methods.* 25, 386-401, **(2001).**

49. Bandyopadhaya,A., Chaudhuri,K.: **Differential modulation of NF-kappaB-mediated pro-inflammatory response in human intestinal epithelial cells by cheY homologues of Vibrio cholerae.** *Innate Immun.*, 15, 131-42, **(2009).**

50. Bandyopadhaya,A., Chaudhuri,K.: **Involvement of intracellular signaling cascades in inflammatory responses in human intestinal epithelial cells following Vibrio cholerae infection.** *Mol Immunol.*, 46, 1129-39,**(2009).**

51. Zhou,X., Gao,D.Q., Michalski,J., Benitez,J.A., Kaper.J.B.: **Induction of interleukin-8 in T84 cells by Vibrio cholerae**. *Infect Immun.*, 72, 389-397, **(2004).**

52. Bandyopadhaya, A., Sarkar, M., Chaudhuri, K.: **IL-1beta expression in Int407 is induced by flagellin of Vibrio cholerae through TLR5 mediated pathway.** *Microb Pathog.*, 44, 524-36, **(2008)**.

53. Bandyopadhaya, A., Sarkar, M., Chaudhuri, K.: **Human intestinal epithelial cell cytokine mRNA responses mediated by NF-kappaB are modulated by the motility and adhesion process of Vibrio cholerae.** *Int J Biochem Cell Biol.*,39, 1863-1876, **(2007)**.

54. Bandyopadhaya, A., Sarkar, M., Chaudhuri, K.: **Transcriptional upregulation of inflammatory cytokines in human intestinal epithelial cells following Vibrio cholerae infection.***FEBS J.*,274, 4631-4642, **(2007)**.

55. Staats,H.F., Ennis,F.A.Jr.: **IL-1 is an effective adjuvant for mucosal and systemic immune responses when coadministered with protein immunogens.***J Immunol.*,162,6141-6147, **(1999)**.

56. Dube, P.H., Revell,P.A., Chaplin,D.D., Lorenz,R.G., Miller,V.L.: **A role for IL-1 alpha in inducing pathologic inflammation during bacterial infection.** *Proc Natl Acad Sci U S A.*, 98, 10880-10885, **(2001)**.

57. Denning,T.L., Campbell,N.A., Song,F., Garofalo,R.P., Klimpel,G.R., Reyes,V.E., Ernst,P.B., **Expression of IL-10 receptors on epithelial cells from the murine small and large intestine.** *Int Immunol.*, 12, 133-139, **(2000)**.

58. Jung,H. C., Eckmann,L., Yang,S. K., Panja,A., Fierer,J., Morzycka-Wroblewska, E., and Kagnoff,M.F.: **A distinct array of proinflammatory cytokines is expressed in human colon epithelial cells in response to bacterial invasion.** *J Clin Invest.*, 95, 55-65, **(1995)**.

59. Fichorova,R.N., Anderson,D.J.: **Differential expression of immunobiological mediators by immortalized human cervical and vaginal epithelial cells.** *Biol Reprod.*, 60,508-514,**(1999)**.

60. Fry, T.J., Mackall,C.L.: **Interleukin-7: from bench to clinic.** *Blood*, 99,3892-3904, **(2002)**.

61. Yan, Z., Yang, D.C., Jett, M.: **Cholera toxin induces tumor necrosis factor alpha production in human monocytes.** *Mol Cell Biol Res Commun.*,2, 124-30,**(1999)**.

62. Qadri, F., Bhuiyan, T.R., Dutta, K.K., Raqib,R., Alam,M.S., Alam,N.H., Svennerholm,A.M., Mathan,M.M.: **Acute dehydrating disease caused by Vibrio cholerae serogroups O1 and O139 induce increases in innate cells and inflammatory mediators at the mucosal surface of the gut.** *Gut*, 53, 62-69,**(2004)**.

63. Stokes,N.R., Zhou,X., Meltzer,S.J., Kaper,J.B.: **Transcriptional responses of intestinal epithelial cells to infection with Vibrio cholerae.***Infect Immun.*, 72, 4240-4248, **(2004).**

64. Bakhiet,M., Al-Salloom,F.S., Qareiballa,A., Bindayna,K., Farid,I., Botta,G.A.: **Induction of alpha and beta chemokines by intestinal epithelial cells stimulated with Campylobacter jejuni.** *J Infect.*, 48, 236-244, **(2004)**.

65. Haeberle,H.A., Dürrstein,C., Rosenberger,P., Hosakote,Y.M., Kuhlicke,J., Kempf,V.A., Garofalo,R.P., Eltzschig,H.K.: **Oxygen-independent stabilization of hypoxia inducible factor (HIF)-1 during RSV infection.** *PLoS One.*, 3, e3352, **(2008)**.

66. Royaee,A.R., Hammamieh,R., Mendis,C., Das,R., Jett,M.H., Yang, D.C.: **Induction of immunomodulator transcriptional responses by cholera toxin.** *Mol Immunol.*, 43, 1020-1028, **(2006)**.

67. Heinrich, P.C., Behrmann, I., Haan, S., Hermanns, H.M., Müller-Newen, G., Schaper, F.: **Principles of interleukin (IL)-6-type cytokine signalling and its regulation.** *Biochem J.*, 374, 1–20, **(2003)**.

68. Strober, W.: **Interactions between epithelial cells and immune cells in the intestine.** *Ann N Y Acad Sci.*, 859, 37-45, **(1998)**.

69. Zvelebil,M., and Baum,J.O.: **Understanding Bioinformatics**. Garland Science, Taylor & Francis Group, LLC, Newyork, **(2008)**.

70. Sippl, M.J. **Recognition of errors in three-dimensional structures of proteins**. *Proteins,* 17, 355-362,**(1993)**.

71. Eswar, N., John, B., Mirkovic, N., Fiser, A., Ilyin, V.A., Pieper, U., Stuart, A.C., Marti-Renom,M.A., Madhusudhan,M.S., Yerkovich,B., Sali,A.: **Tools for comparative protein structure modeling and analysis.** *Nucleic Acids Res.*,31, 3375-3380, **(2003)**.

72. Chivian, D., Baker, D.: **Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection.** *Nucleic Acids Res.*, 34, e112, **(2006)**.

73. Xu, D., Xu, Y., Uberbacher, E.C.: **Computational tools for protein modeling**. *Curr Protein Pept Sci.*, 1, 1-21, **(2000)**.

74. Bourne,P.E., and Weissig,H.: **Structural Bioinformatics**. John Wiley & Sons, Inc.,Hoboken, New Jersey, **(2003)**.

75. Huang, E.S., Samudrala, R., Ponder, J.W.: **Distance geometry generates native-like folds for small helical proteins using the consensus distances of predicted protein structures.** *Protein Sci.*,7, 1998-2003, **(1998).**

76. Berman, Helen. M., Westbrook, John., Feng, Zukang., Gilliland, Gary., Bhat, T. N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: **The Protein Data Bank.** *Nucleic Acids Res.*, 28, 235-242, **(2000).**

77. McGinnis, Scott., Madden, L.T.: **BLAST: at the core of a powerful and diverse set of sequence analysis tools.** *Nucleic Acids Res.*, 32, W20-25,**(2004).**

78. Silvanovich, A., Bannon, G., McClain, S.: **The use of E-scores to determine the quality of protein alignments.** *Regul Toxicol Pharmacol.*,54, S26-31, (2009).

79. Samudrala.R., Moult,J.: **Determinants of side chain conformational preferences in protein structures.** *Protein Eng.*, 11, 991-997, **(1998).**

80. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R., Thornton,J.M.: **AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.** *J Biomol NMR.*, 4, 477-486, **(1996)**

81. G.Vriend.: **WHAT IF: A molecular modeling and drug design program** *J. Mol. Graph.*, 8,52-56, **(1990)**

82. Bairoch,A., Apweiler,R., Wu,C.H., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H., Lopez,R., Magrane,M., Martin,M.J., Natale,D.A., O'Donovan,C., Redaschi,N., Yeh,L.S.: **The Universal Protein Resource (UniProt).** *Nucleic Acids Res.*, 33, D154-159, **(2005).**

83. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F., Higgins,D.G.: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence**

alignment aided by quality analysis tools *Nucleic Acids Res.*, 25, 4876-4882, **(1997).**

84. Edgar, R.C.: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res.*, 32, 1792-1797, **(2004)**.

85. Lukashin,A.V., Rosa,J.J.: **Local multiple sequence alignment using dead-end elimination.** *Bioinformatics,* 15, 947-953, **(1999)**.

86. Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D., Bairoch,A.: **ExPASy: The proteomics server for in-depth protein knowledge and analysis.** *Nucleic Acids Res.*, 31, 3784-3788, **(2003)**.

87. Sigrist, C.J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni,M., Bairoch, A., Bucher, P.: **PROSITE: a documented database using patterns and profiles as motif descriptors.** *Brief Bioinform.*, 3, 265-274, **(2002)**.

88. Felsenstein,J.: **Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates.** *Genet Res.*,59, 139-147, **(1992)**.

89. Timothy,V.Pyrkov., Anton,O.Chugunov., Nikolay,A.Krylov., Dmitry,E.Nolde., and Roman,G.Efremov., **PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes.** *Bioinformatics*, 25, 1201-2, **(2009)**.

90. Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H.: **Locating proteins in the cell using TargetP, SignalP and related tools.** *Nat Protoc.*, 2, 953-71, **(2007)**.

91. Shen,H.B., Chou,K.C.: **Signal-3L: A 3-layer approach for predicting signal peptides.** *Biochem Biophys Res Commun.*, 363, 297-303,**(2007)**.

92. Rost, B.: **PHD: predicting one-dimensional protein structure by profile-based neural networks.** *Methods Enzymol.*, 266, 525-39, **(1996)**.

93. Cuff,J.A., Clamp,M.E., Siddiqui,A.S., Finlay,M., Barton,G.J.: **JPred: a consensus secondary structure prediction server.** *Bioinformatics*, 14, 892-3, **(1998)**.

94. Guex,N., Peitsch,M.C.: **SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.** *Electrophoresis*, 18, 2714-23, **(1997)**.

95. Humphrey,W., Dalke,A., Schulten,K.: **VMD: visual molecular dynamics.** *J Mol Graph.*, 14, 33-8, 27-8, **(1996)**.

96. Nicholls,A., Sharp,K.A., Honig,B.: **Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.** *Proteins*, 11, 281-296, **(1991)**.

97. Andreeva, A., Howorth, D., Brenner, S.E., Hubbard, T.J.P., Chothia, C., Murzin, A.G.: **SCOP database in 2004: refinements integrate structure and sequence family data.** *Nucl. Acid Res.*, 32:D226-D229, **(2004)**.

98. Orengo,C.A., Michie,A.D., Jones,S., Jones,D.T., Swindells,M.B., Thornton,J.M.: **CATH--a hierarchic classification of protein domain structures.** *Structure*, 5, 1093-108, **(1997)**.

99. Jo,S., Kim,T., Iyer,V.G., Im,W.: **CHARMM-GUI: a web-based graphical user interface for CHARMM.** *J Comput Chem.*, 29,1859-65,**(2008)**.

100. Badger,J., Kumar,R.A., Yip,P., Szalma,S.: **New features and enhancements in the X-PLOR computer program.** *Proteins*, 35, 25-33, **(1999)**.

101. Case,D.A., Cheatham,T.E., Darden,T., Gohlke,H., Luo,R., Merz,K.M.Jr., Onufriev,A., Simmerling,C., Wang, B., Woods, R.J.: **The Amber biomolecular simulation programs.** *J. Computat. Chem.* **26,** 1668-1688 **(2005)**.

102. UniProt Consortium. **The universal protein resource (UniProt).** *Nucleic Acids Res.*, 36,D190-5, (2008).

103. Deléage,G., Blanchet,C., Geourjon,C.: **Protein structure prediction. Implications for the biologist.** *Biochimie.*, 79, 681-6, **(1997)**.

104. Pazel,D.P.: **DS-Viewer- interactive graphical data structure presentation facility.** *IBM Systems J.*, 28, 307–323, **(1989)**.

105. Jonathan,M. Goodman.: **WebLab ViewerPro**. *CTi Chemistry*, 17,34-35, **(1998)**.

106. Hartshorn, M.J.: **AstexViewer: A visualisation aid for structure-based drug design**. *J.Comput. Aided Mol.Des.*, 16,871-881, **(2002)**.

107. Wiederstein,M., Sippl,M.J.: **ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.** *Nucleic Acids Res.*, 35, W407-10,**(2007)**.

108. Ligand Explore Home Page. http://www.kukool.com/ligand

109. Wallace,A.C., Laskowski,R.A., Thornton,J.M.:. **LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions.** *Prot. Eng.*, **8,** 127-134, **(1995)**

110. Hase,C.C., Finkelstein,R.A.: **Cloning and nucleotide sequence of the Vibrio cholerae hemagglutinin/protease (HA/protease) gene and construction of an HA/protease-negative strain.** *J Bacteriol.*, 173, 3311–3317, **(1991).**

111. Booth,B.A., Boesman-Finkelstein,M., Finkelstein,R.A.: **Vibrio cholerae soluble hemagglutinin/protease is a metalloenzyme.** *Infect Immun.*, 42,639–644,**(1983).**

112. Benitez,J.A., Spelbrink,R.G., Silva,A., Phillips,T.E., Stanley,C.M., Boesman-Finkelstein,M., Finkelstein,R.A.: **Adherence of Vibrio cholerae to cultured differentiated human intestinal cells: an in vitro colonization model.** *Infect Immun.*, 65, 3474–3477, **(1997).**

113. De Kreij, A., Venema, G., vanden Burg, B.: **Substrate specificity in the highly heterogeneous M4 peptidase family is determined by a small subset of amino acids.** *J Biol Chem.*, 275, 31115-31120, **(2000).**

114. Jongeneel, C.V., Bouvier, J., Bairoch, A.: **A unique signature identifies a family of zinc-dependent metallopeptidases.** *FEBS Lett.*, 242, 211-214, **(1989).**

115. Domann,E., Leimeister-Wachter,M., Goebel,W., Chakraborty,T.: **Molecular cloning, sequencing, and identification of a metalloprotease gene from Listeria monocytogenes that is species specific and physically linked to the listeriolysin gene.** *Infect Immun.*, 59, 65-72, **(1991).**

116. Holden,H.M., Tronrud,D.E., Monzingo,A.F., Weaver,L.H., Matthews,B.W.: **Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic analysis of extended phosphonamidate transition-state analogues.** *Biochemistry,* 26, 8542-8553, **(1987).**

117. Beaumont, A., O. Donohue, M.J., Paredes, N., Rousselet, N., Assicot, M., Bohuon, C., Fournie-Zaluski, M.C., Roques,B.P.: **The role of histidine 231 in thermolysin-like enzymes. A site-directed mutagenesis study.** *J Biol Chem.*, 270, 16803-16808, **(1995)**.

118. Gaucher,J.F., Selkti, M., Tiraboschi, G., Prange, T., Roques, B.P., Tomas, A., Fournie-Zaluski, M.C.: **Crystal structures of alpha-mercaptoacyldipeptides in the thermolysin active site: structural parameters for a Zn monodentation or bidentation in metalloendopeptidases.** *Biochemistry*, 38, 12569-12576, **(1999)**.

119. Thayer,M.M., Flaherty,K.M., and McKay,D.B.: **Three-dimensional structure of the elastase of Pseudomonas aeruginosa at 1.5-A resolution.** *J Biol Chem.*, 266, 2864-2871, **(1991)**.

120. Kawamoto, S., Shibano, Y., Fukushima, J., Ishii, N., Morihara, K., Okuda, K.: *Infect Immun.*, 61,1400-1405, **(1993)**.

121. Holden,H.M., Matthews,B.W.: **The binding of L-valyl-L-tryptophan to crystalline thermolysin illustrates the mode of interaction of a product of peptide hydrolysis.** *J Biol Che.*, 263, 3256-3260, **(1988)**.

122. McIver,K.S., Olson,J.C., Ohman,D.E.: **Pseudomonas aeruginosa lasB1 mutants produce an elastase, substituted at active-site His-223, that is defective in activity, processing, and secretion.** *J Bacteriol.* 175, 4008-4015, **(1993)**.

123. McIver,K., Kessler,E., Ohman,D.E.: **Substitution of active-site His-223 in Pseudomonas aeruginosa elastase and expression of the mutated lasB alleles in Escherichia coli show evidence for autoproteolytic processing of proelastase.** *J Bacteriol.* 173, 7781-7789, **(1991)**.

124. De Kreij,A., VanDen Burg,B., Veltman,Or., Vriend,G., Venema,G., Eijsink,V.G.: **The effect of changing the hydrophobic S1' subsite of thermolysin-like proteases on substrate specificity.** *Eur J Biochem.*, 268, 4985-4991, **(2001).**

125. Baudry, B., Fasano, A., Ketley, J., Kaper, J.B.: **Cloning of a gene (zot) encoding a new toxin produced by Vibrio cholerae.** *Infect Immun.*, 60,428-34, **(1992).**

126. Marinaro, M., Fasano, A., De Magistris, M.T.: **Zonula occludens toxin acts as an adjuvant through different mucosal routes and induces protective immune responses.** *Infect Immun.*, 71, 1897-902, **(2003).**

127. Theophilo, G.N., Rodrigues, Ddos. P., Leal, N.C., Hofer, E.: **Distribution of virulence markers in clinical and environmental Vibrio cholerae non-O1/non-O139 strains isolated in Brazil from 1991 to 2000.** *Rev Inst Med Trop Sao Paulo.*, 48, 65-70, **(2006).**

128. Trucksis, M., Galen, J.E., Michalski, J., Fasano, A., Kaper, J.B.: **Accessory cholera enterotoxin (Ace), the third toxin of a Vibrio cholerae virulence cassette.***Proc Natl Acad Sci U S A.*, 90, 5267-71, **(1993).**

129. Fasano, A., Baudry, B., Pumplin, D.W., Wasserman, S.S., Tall, B.D., Ketley, J.M., Kaper, J.B.: **Vibrio cholerae produces a second enterotoxin, which affects intestinal tight junctions.** *Proc Natl Acad Sci U S A.*, 88, 5242-6, **(1991).**

130. Di Pierro, M., Lu, R., Uzzau, S., Wang, W., Margaretten, K., Pazzani, C., Maimone, F., Fasano, A.: **Zonula occludens toxin structure-function analysis. Identification of the fragment biologically active on tight junctions and of the zonulin receptor binding domain.** *J Biol Chem.*, 276, 19160-5, **(2001).**

APPENDIX

APPENDIX I

Reagents:

- **Cholera Toxins, A and B (Sigma, Saint Louis, Missouri USA):** CTA (1mg/mL) & CTB (0.5mg/mL) subunits from *V.cholerae* were reconstituted with 1mL of sterile water according to the manufacturer instructions and were kept at 4^0C for up to 6 months.
- **Neuraminidase (Sigma, Saint Louis, Missouri USA):** Neuraminidase from *V.cholerae* was in the form of sterile-filtered aqueous solution, pH 5.5 (1mg protein/mL; 5.0units/mg protein)
- **Phosphate Buffer Saline (PBS):** 1g tablet of PBS was dissolved in 800mL of deionized water. pH was adjusted up to 7.3-7.4 by using 1N HCl or 1N NaOH. The volume was adjusted to 1 litter with deionized water and the solution was autoclaved.
- **RNeasy Mini Kit (Qiagen)**
- **Ethanol,** 70%, 90%.
- **β-Mercaptoethanol (β-ME)**
- **DEPC Treated Water:** 1mL of diethylpyrocarbonate (DEPC) was added to 1 liter of deionized water, incubated overnight at 37^0C and autoclaved.
- **Isopropyl alcohol (Merck Marker Pvt. Ltd)**
- **Chloroform (Merck Marker Pvt. Ltd)**
- **Trizol Reagent (Invitrogen-Life Technologies, USA)**
- **Tris EDTA (TE) Buffer:** 0.1576g of Tris (10mM) and 18.61g of EDTA(1mM) were dissolved in 80mL of deionized water and pH was adjusted up to 7.5 by using 1N HCl or 1N NaOH. The volume was adjusted up to 100mL with deionized water and autoclaved.

- **Tris-Borate EDTA (TBE) Buffer (5X):** Tris Base (54g), Boric Acid (27.5g) and EDTA (2.92g) were dissolved in 800mL of deionized water and pH was adjusted to 8.0 by using 1N HCl or 1N NaOH. The volume was adjusted up to 1 litter with deionized water autoclaved.
- **Agarose (Invitrogen)**
- **Ethidium Bromide solution (Invitrogen)**

Appendix I

```
pdb|1EZM|  Elastase (E.C.3.4.24.26) (Zinc Metalloprotease) Length=301
pdb|1U4G|A Chain A, Elastase Of Pseudomonas Aeruginosa With An Inhibitor

 Score =  374 bits (961),  Expect = 2e-104, Method: Composition-based stats.
 Identities = 190/303 (62%), Positives = 230/303 (75%), Gaps = 10/303 (3%)

VIB    198  ATGTGPGGNQKTGRYEYGSNGLPGFTIDKTGTTCTMNNSAVKTVNLNGGT--SGSTAFSY  255
            A   GPGGNQK G+Y YGS+  P    D+    C M++  V TV++N  T  S +T F +
1EZM   1    AEAGGPGGNQKIGKYTYGSDYGPLIVNDR----CEMDDGNVITVDMNSSTDDSKTTPFRF  56

VIB    256  ACNNSTNYNSVKTVNGAYSPLNDAHFFGKVVFDMYQQWLNTSPLTFQLTMRVHYGNNYEN  315
            AC      N+ K VNGAYSPLNDAHFFG VVF +Y+ W  TSPLT +L M+VHYG + EN
1EZM   57   ACPT----NTYKQVNGAYSPLNDAHFFGGVVFKLYRDWFGTSPLTHKLYMKVHYGRSVEN  112

VIB    316  AFWDGRAMTFGDGYTRFYPLVDINVSAHEVSHGFTEQNSGLVYRDMSGGINEAFSDIAGE  375
            A+WDG AM FGDG T FYPLV ++V+AHEVSHGFTEQNSGL+YR  SGG+NEAFSD+AGE
1EZM   113  AYWDGTAMLFGDGATMFYPLVSLDVAAHEVSHGFTEQNSGLIYRGQSGGMNEAFSDMAGE  172

VIB    376  AAEYFMRGNVDWIVGADIFKSSGGLRYFDQPSRDGRSIDHASQYYSGIDVHHSSGVFNRA  435
            AAE++MRG  D+++G DI K SG LRY DQPSRDGRSID+ASQYY+GIDVHHSSGV+NRA
1EZM   173  AAEFYMRGKNDFLIGYDIKKGSGALRYMDQPSRDGRSIDNASQYYNGIDVHHSSGVYNRA  232

VIB    436  FYLLANKSGWNVRKGFEVFAVANQLYWTPNSTFDQGGCGVVKAAQDLNYNTADVVAAFNT  495
            FYLLAN  GW+ RK FEVF  AN+ YWT  S ++ G CGV+++AQ+ NY+ ADV  AF+T
1EZM   233  FYLLANSPGWDTRKAFEVFVDANRYYWTATSNYNSGACGVIRSAQNRNYSAADVTRAFST  292

VIB    496  VGV  498
```

Template Search result for HA/P (Vibriolysin) and Pairwise sequence analysis by BLAST and used to generate the alignment.ali file for align2d command of MODELLER.

```
pdb|1PXT|A    Chain A, The 2.8 Angstroms Structure Of Peroxisomal 3-Ketoacyl-
Coa
Thiolase Of Saccharomyces Cerevisiae: A Five Layered A-B-A-B-A Structure,
Constructed From Two Core Domains Of Identical Topology
Length=390

 Score = 30.8 bits (68),  Expect = 0.74, Method: Compositional matrix adjust.
 Identities = 29/108 (26%), Positives = 43/108 (39%), Gaps = 22/108 (20%)

ZOT    295  VQDGFVTVGDERYRLVDNLDIPYRGLWATGHHIYKDTLTVFFETESGSVPTELFASSYRY  354
            V DG   V   R  + + L++P  G        Y D  TV        VP E+       Y
1PXT   249  VSDGVAGVLLARRSVANQLNLPVLGR-------YIDFQTV-------GVPPEIMGVGPAY  294

ZOT    355  KV--------LPLPDFNHFVVFDTFAAQALWVEVKRGLPIKTENDKKG  394
             +        L + D + F + + FAAQAL+   K G+ +   N + G
1PXT   295  AIPKVLEATGLQVQDIDIFEINEAFAAQALYCIHKLGIDLNKVNPRGG  342
```

Template Search result for Zona Occludens Toxin (Zot) and Pairwise sequence analysis by BLAST and used to generate the alignment.ali file for align2d command of MODELLER.

```
pdb|1CKV|A Chain A, Structure Of The Soluble Methane Monooxygenase Regulatory
Protein B Length=141

 Score = 24.6 bits (52),  Expect = 7.1, Method: Compositional matrix adjust.
 Identities = 11/35 (31%), Positives = 22/35 (62%), Gaps = 0/35 (0%)

ACEV   23  GIMWIESKIFVIQFFWEMSQKVIDMFTIYPLIQQA  57
           GIM ++ K F  QFF + +Q V +  T+  +++++
1CKV   11  GIMGLKGKDFADQFFADENQVVHESDTVVLVLKKS  45
```

Template Search result for Accessory cholera enterotoxin (Ace) and Pairwise sequence analysis by BLAST and used to generate the alignment.ali file for align2d command of MODELLER.

```
C; A sample alignment in the PIR format; used in tutorial

>P1;1u4g
structureX:1u4g:1    :A:526  :A:elastase:Pseudomonas:-1.00:-1.00
AEAGGPGGNQKIGKYTYGSDYGPLIVNDR----CEMDDGNVITVDMNSSTDDSKTTPFRFACPTNTYKQ--VNGA
YSPLNDAHFFGGVVFKLYRDWFGTSPLTHKLYMKVHYGRSVENAYWDGTAMLFGDGATMFYPLVSLDVAAHEVSH
GFTEQNSGLIYRGQSGGMNEAFSDMAGEAAEFYMRGKNDFLIGYDIKKGSGALRYMDQPSRDGRSIDNASQYYNG
IDVHHSSGVYNRAFYLLANSPGWDTRKAFEVFVDANRYYWTATSNYNSGACGVIRSAQNRNYSAADVTRAFSTVG
V/.....
.....
.....
.....*

>P1;vib2
sequence:vib2:     : :     : :vibriolysin:Vibrio:-1.00:-1.00
ATGTGPGGNQKTGRYEYGSNGLPGFTIDKTGTTCTMNNSAVKTVNLNGGTSGSTAFSYACNNSTNYNSVKTVNGA
YSPLNDAHFFGKVVFDMYQQWLNTSPLTFQLTMRVHYGNNYENAFWDGRAMTFGDGYTRFYPLVDINVSAHEVSH
GFTEQNSGLVYRDMSGGINEAFSDIAGEAAEYFMRGNVDWIVGADIFKSSGGLRYFDQPSRDGRSIDHASQYYSG
IDVHHSSGVFNRAFYLLANKSGWNVRKGFEVFAVANQLYWTPNSTFDQGGCGVVKAAQDLNYNTADVVAAFNTVG
V/.....
.....
.....
.....*
```

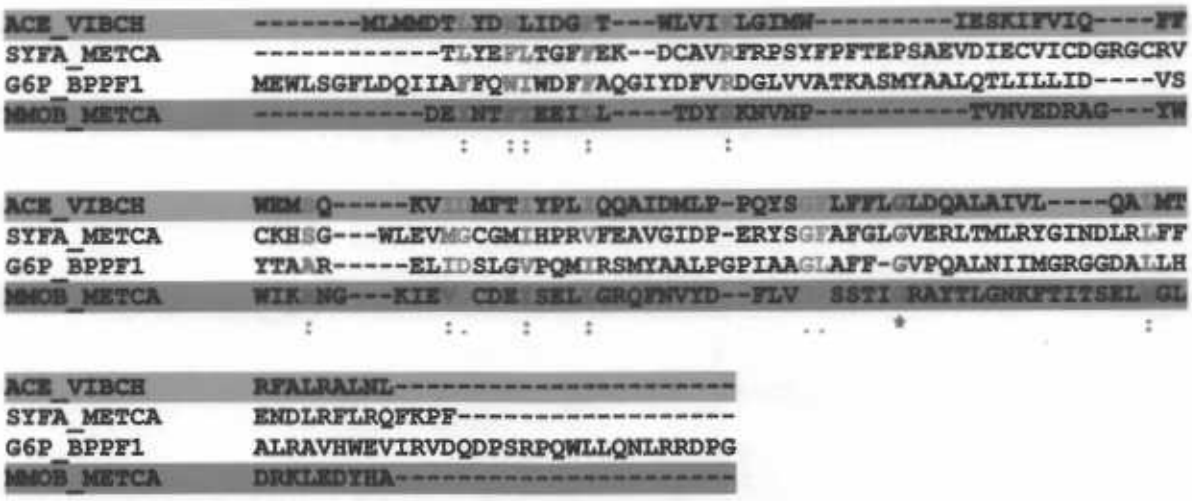
Alignment.ali **file used to contruct comparative model of Hemagglutinin/proteinase (Vibriolysin) from crystal structure of 1u4g.**

```
C; A sample alignment in the PIR format; used in tutorial

>P1;1PXT
structureX:1PXT:28:A:417:A:Keto acyl CoA:Saccharomyces cerevisiae:-1.00:-
1.00
LLEKRPEDVVIVAANRSAIGKGFKGAFKDVNTDYLLYNFLNEFIGRFPEPLRADLNLIEEVACGNVLNVGAGATE
HRAACLASGIPYSTPFVALNRQCSSGLTAVNDIANKIKVGQIDIGLALGVESMTNNYKNQKNREAKKCLIPMGIT
NENRKDQDEFAANSYQKAYKAKNEGLFEDEILPIKLPDGSICQSDEGPRPQVSDGVAGVLLARRSVANQLNLPVL
GRYIDFQTVGVPPEIMGVGPAYAIPKVLEATGLQVQDIDIFEINEAFAAQALYCIHKLGIDLNKVNPRGGAIALG
HPLGCTGARQVATILRELKKDQIGVVSMCIGTGMGAAAIFIKE*

>P1;ZOTV
sequence:ZOT:     : :     : :Zona occludens toxin:Vibrio cholerae:-1.00:-1.00
SIFIHHGAPGSYKTSGALWLRLLPAIKSGRHIITNVRGLNLERMAKYLKMDVSDISIEFIDTDHPDGRLTMARFW
HWARKDAFLFIDECGRIWPPRLTVTNLKALDTPPDLVAEDRPESFEVAFDMHRHHGWDICLTTPNIAKVHNMIRE
AAEIGYRHFNRATVGLGAKFTLTTHDAANSGQMDSHALTRQVKKIPSPIFKMYASTTTGKARDTMAGTALWKDRK
ILFLFGMVFLMFSYSFYGLHDNPIFTGGNDATIESEQSEPQSKATVGNAVGSKAVAPASFGFCIGRLCVQDGFVT
VGDERYRLVDNLDIPYRGLWATGHHIYKDTLTVFFETESGSVPTELFASSYRYKVLPLPDFNHFVVFDTFAAQAL
WVEVKRGLPIKTEND*
```

Alignment.ali **file used to contruct comparative model of Zona occludens toxin (Zot) from crystal structure of 1PXT.**

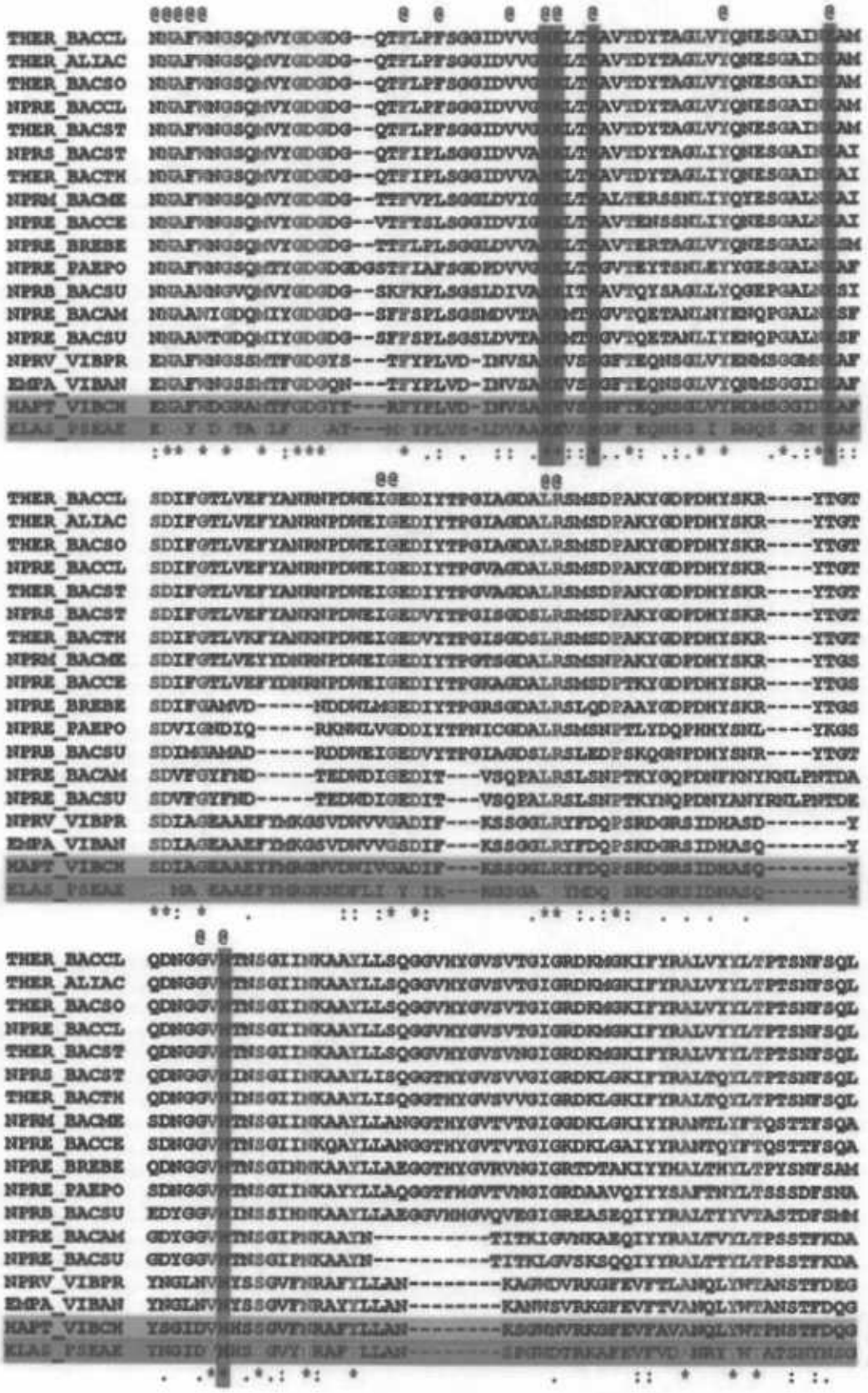
```
C; A sample alignment in the PIR format; used in tutorial

>P1;1CKV
structureX:1CKV:46:A:141:A:Methonol Momooxygenate:Methylococcus capsulatus:-1.00:-1.00
DEINTFIEEILLTDYKKNVNPTVNVEDRAGYWWIKANGKIEVDCDEISELLGRQFNVYDFLVDVSSTIGRAYTLG
NKFTITSELMGLDRKLEDYHA*

>P1;ACEV
sequence:ACEV: : : : :Accessory cholera enterotoxin:Vibrio cholerae:-1.00:-1.00
MLMMDTLYDWLIDGFTWLVIKLGIMWIESKIFVIQFFWEMSQKVIDMFTIYPLIQQAIDMLPPQYSGFLFFLGLD
QALAIVLQALMTRFALRALNL*
```

Alignment.ali **file used to contruct comparative model of Accessory cholera enterotoxin (Ace) from crystal structure of 1CKV.**

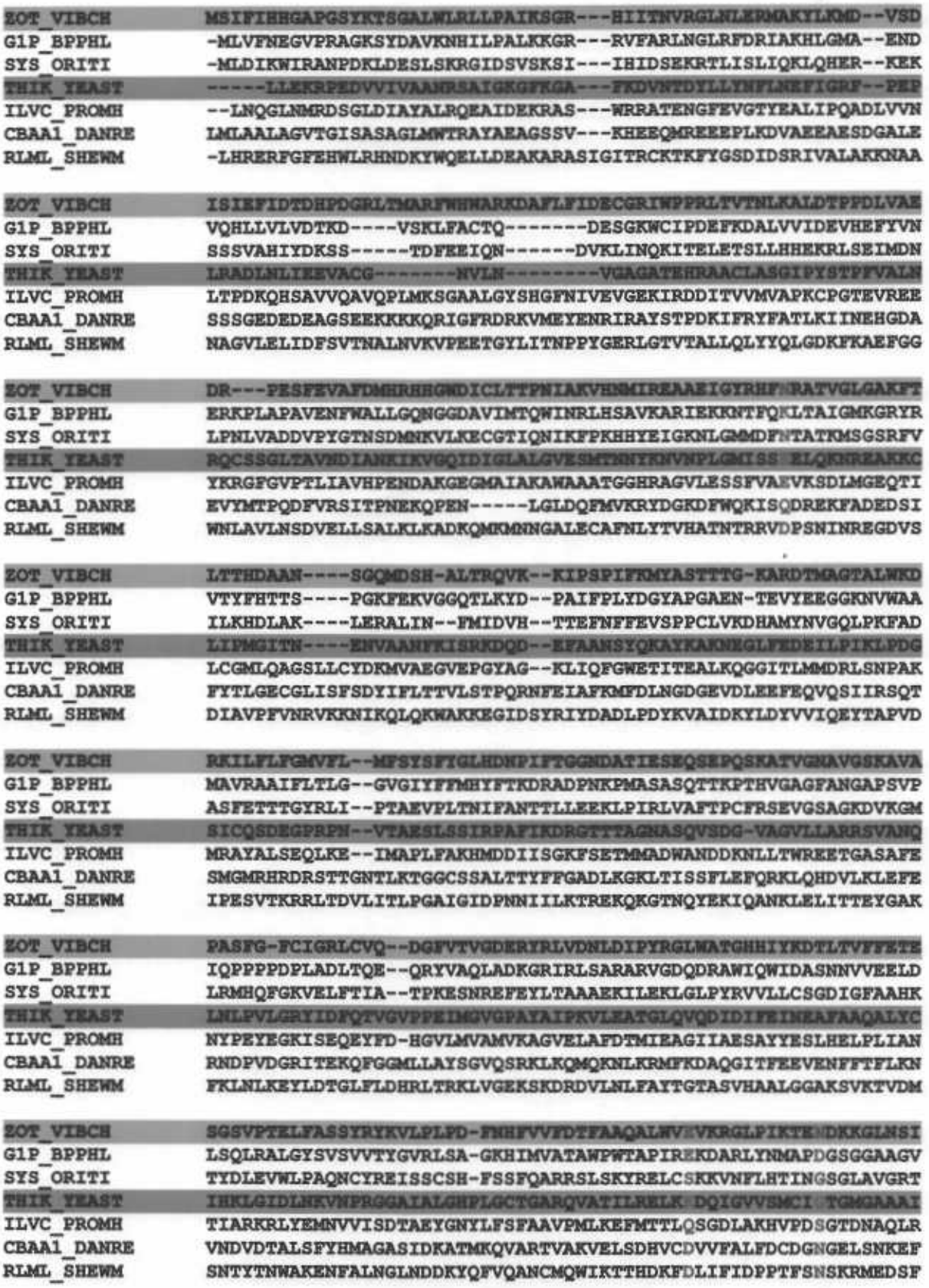


Multiple Sequence Alignment between Ace (*Vibrio cholerae)* and other species of protobacteria. Highly conserved residue is marked with * and highlighted with blue, less conserved are marked with : and highlighted with red and Least conserved residues are marked with · and highlighted with magneta.

CLUSTAL X (1.81) multiple sequence alignment

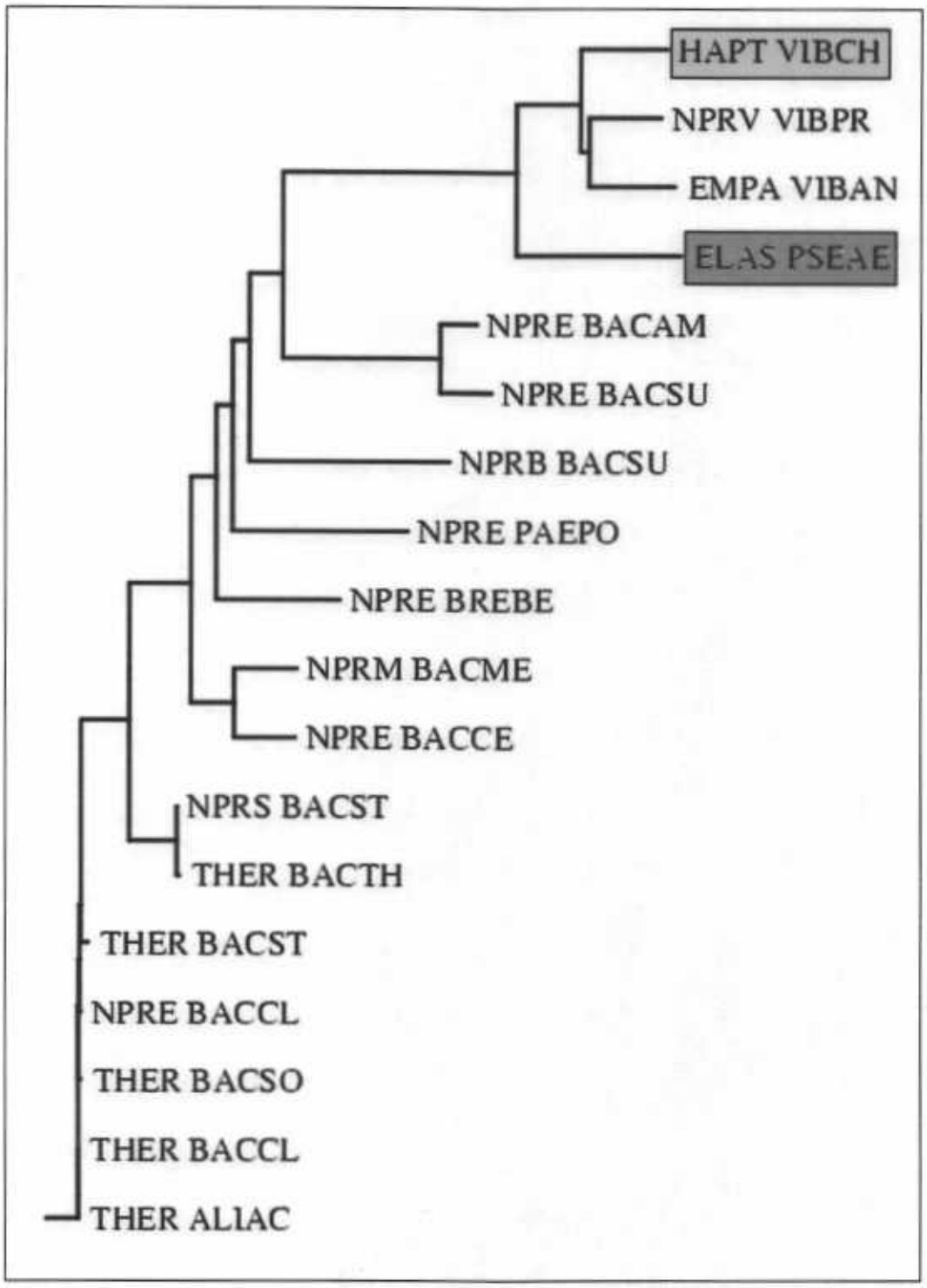


Multiple Sequence Alignment between Neutral Protease species & *Vibrio cholerae*. Conserved residues are highlighted with blue. Boxes designate residues that are important for catalytic activity (red) and metal binding (green). Residues that are involved in different interactions are designated as @. Sequences were first aligned using Clustal X and then manually edited for improvements.

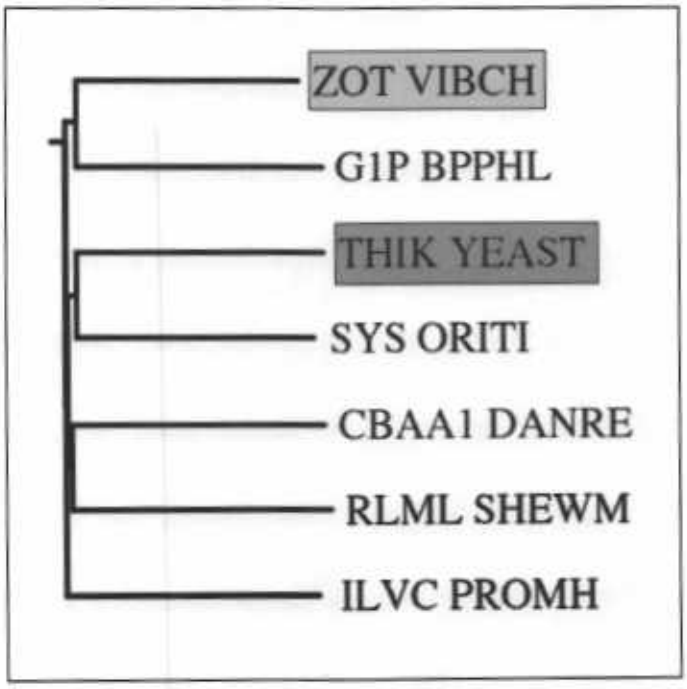


Multiple Sequence Alignment between Zot (*Vibrio cholerae)* and other species of protobacteria. Least conserved residues are highlighted with blue.

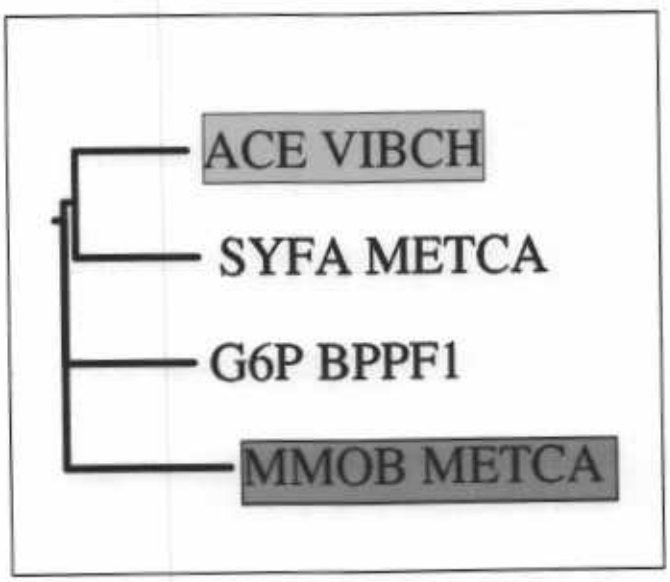
APPENDIX III



Dendogram derived for HAP_VIBCH protease from multiple sequence alignment using the neighbor joining distances method within Phylip package.

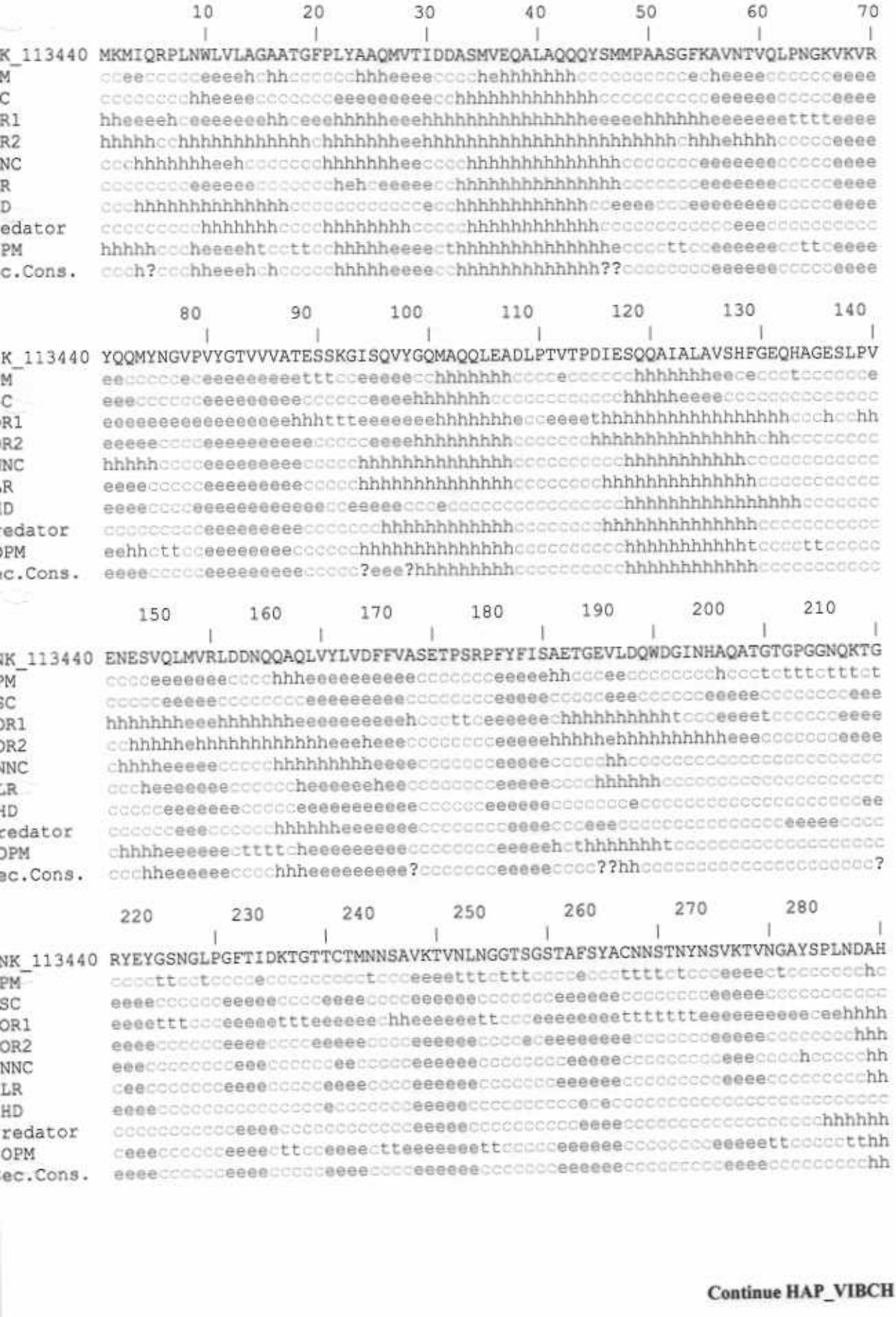


Dendogram derived for ZOT_VIBCH protease from multiple sequence alignment using the neighbor joining distances method within Phylip package.

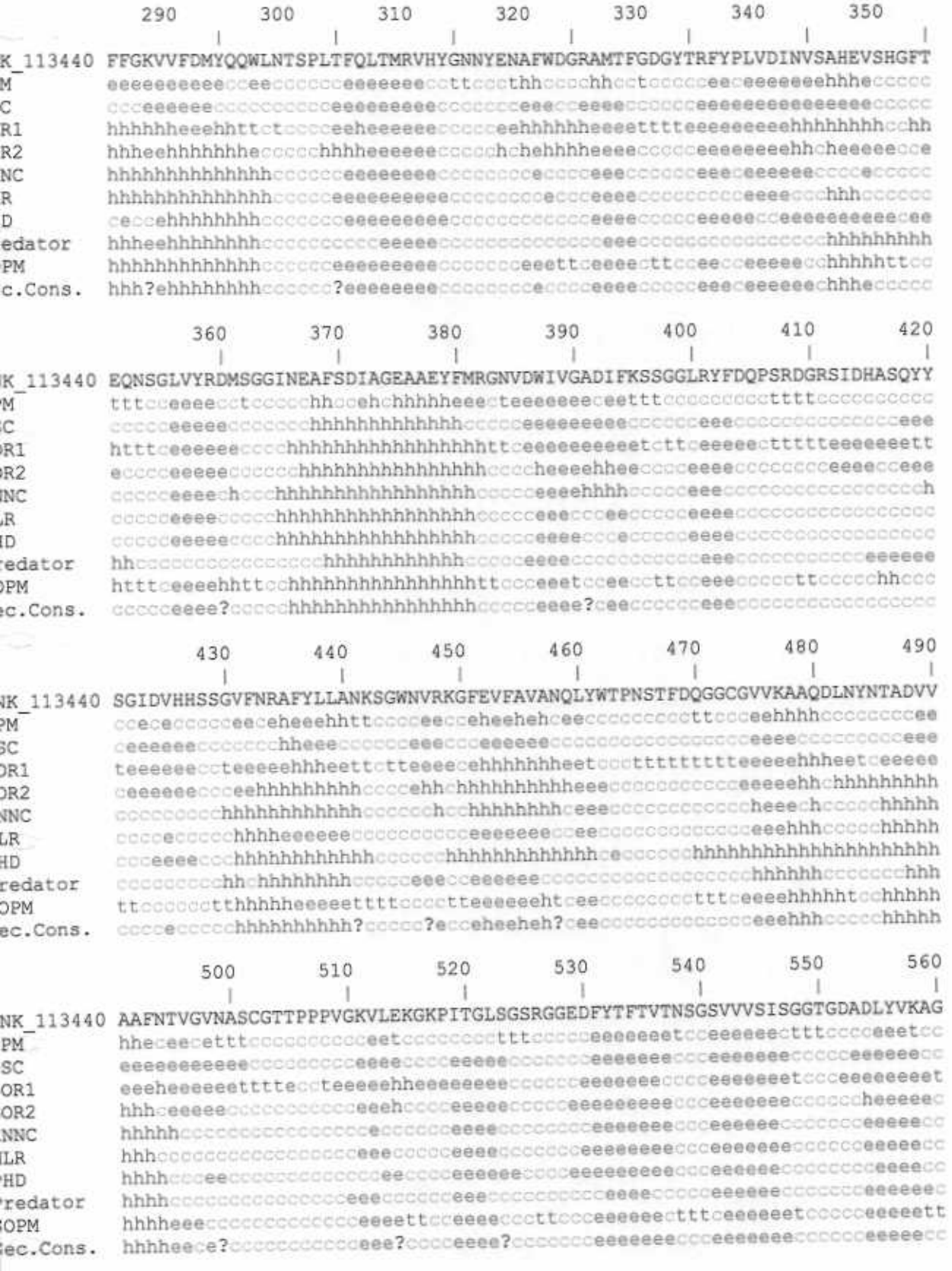


Dendogram derived for ACE_VIBCH protease from multiple sequence alignment using the neighbor joining distances method within Phylip package.

onsensus prediction result for : HAP_VIBCH



Continue HAP_VIBCH

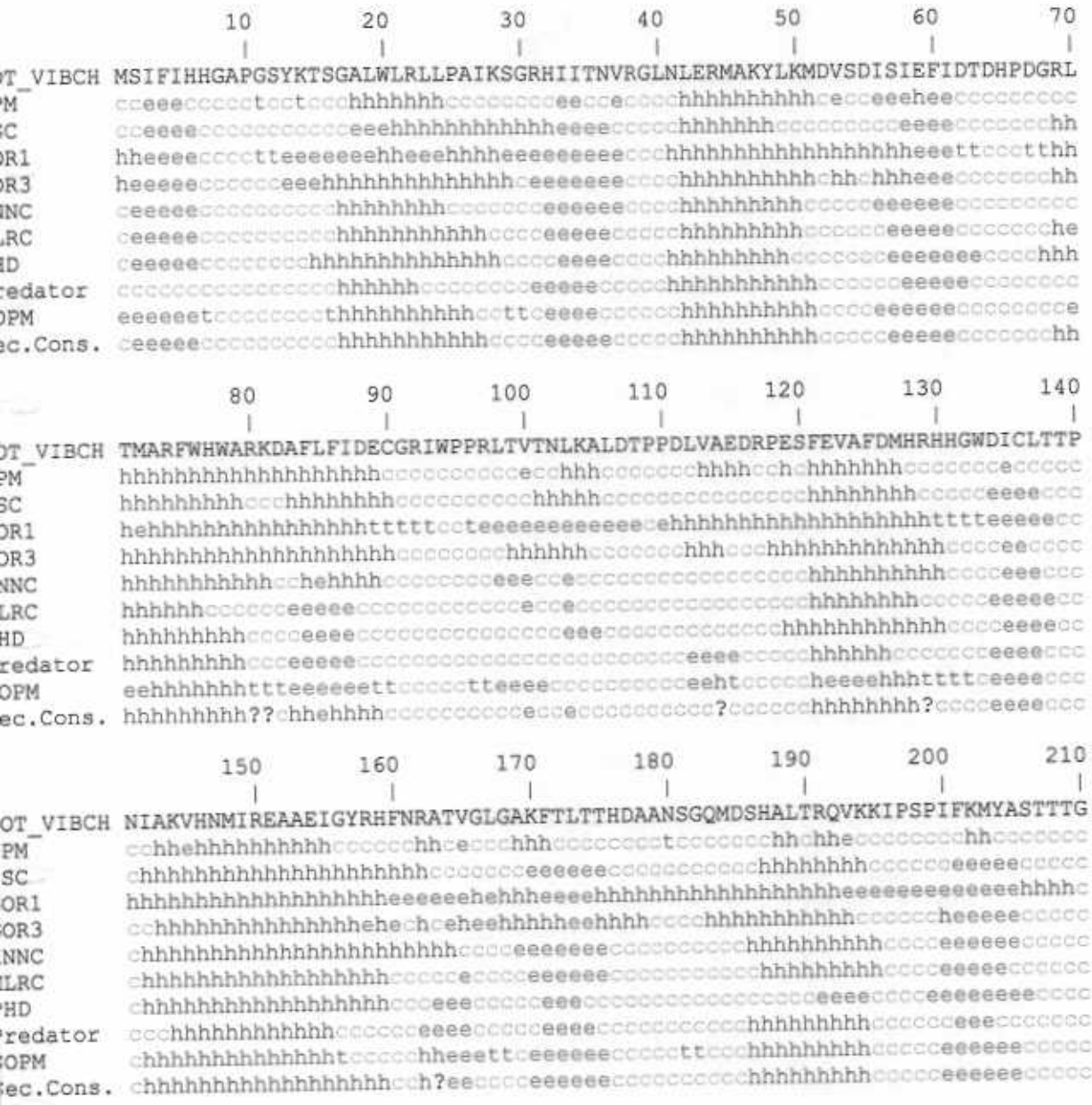


Continue HAP_VIBCH

quence length : 609

Continue HAP_VIBCH

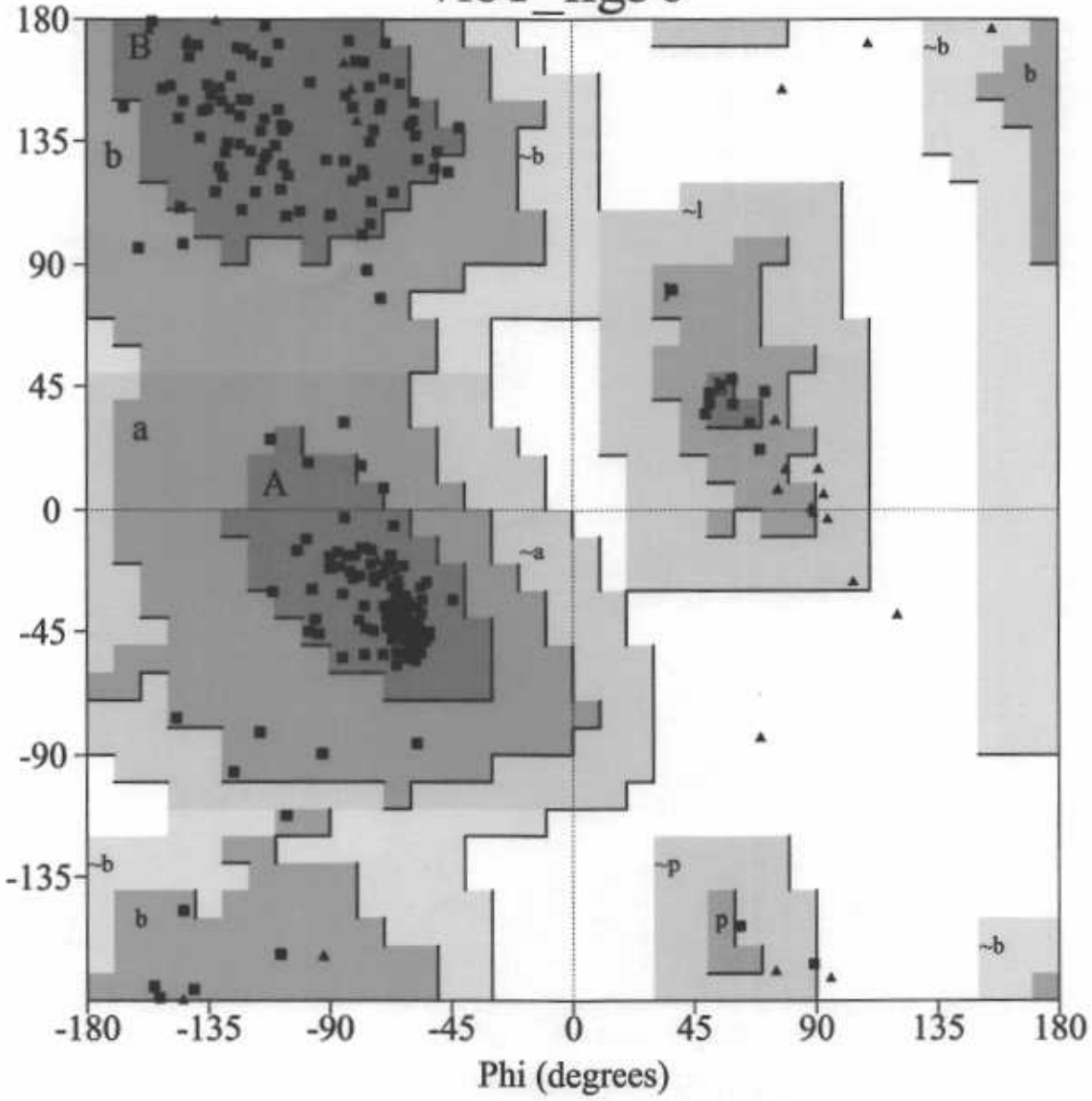
Consensus prediction result for : ZOT_VIBCH



onsensus prediction result for : ACE_VIBCH

```
                  10        20        30        40        50        60        70
                   |         |         |         |         |         |         |
E_VIBCH MLMMDTLYDWLIDGFTWLVIKLGIMWIESKIFVIQFFWEMSQKVIDMFTIYPLIQQAIDMLPPQYSGFLF
M       cchhheeeceeeeeeeeeeeeeeeeeeeeeeeeeeehhhheeeeeeeeceehhhehhccccccceee
C       ccchhhhhhhhhcchhhhhhhhhhhhhhhhhhheehhhhhhhhhhhhhhccchhhhhhhhccccccccchhh
R1      hhhhhhhhheeeethhhhhhhhhhhhhhhhhhhhehhhhhhhhhhhheeeeeehhheeeeeeeettteeeee
R3      hhhhhhhhhhhhhhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhehchhhhhhccccccccchhh
NC      cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhee
RC      chhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhee
D       ccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
edator  ccchhhhhhhhhhcceeeeeccccccccchhhhhhhhhhhhhhhhhheeehhhhhhhhhcccccccceee
PM      hhhhhhhhhhhhttthheeeeeeeeeecctteeeehhhhhhhhhhhhhhhhhhhhhhhhhhcccttcteee
c.Cons. cchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh?hhhhhhhhhh?ccccccchee

                  80        90
                   |         |
E_VIBCH FLGLDQALAIVLQALMTRFALRALNL
M       eechhhhhheeehhhhhhhhhhhcc
C       hhhhhhhhhhhhhhhhhhhhhhhccc
R1      ehhhhhhhhhhhhhhhhhhhhhhhh
R3      hhhhhhhhhhhhhhhhhhhhhhhhh
NC      hhhhhhhhhhhhhhhhhhhhhhhcc
RC      eehhhhhhhhhhhhhhhhhhhhhcc
D       hhhhhhhhhhhhhhhhhhhhhhhcc
edator  ecccchhhhhhhhhhhhhhhhhccc
PM      eechhhhhhhhhhhhhhhhhhhce
c.Cons. ehhhhhhhhhhhhhhhhhhhhhhcc

equence length :    96
```

Ramachandran Plot
vib1_lig36



Plot statistics

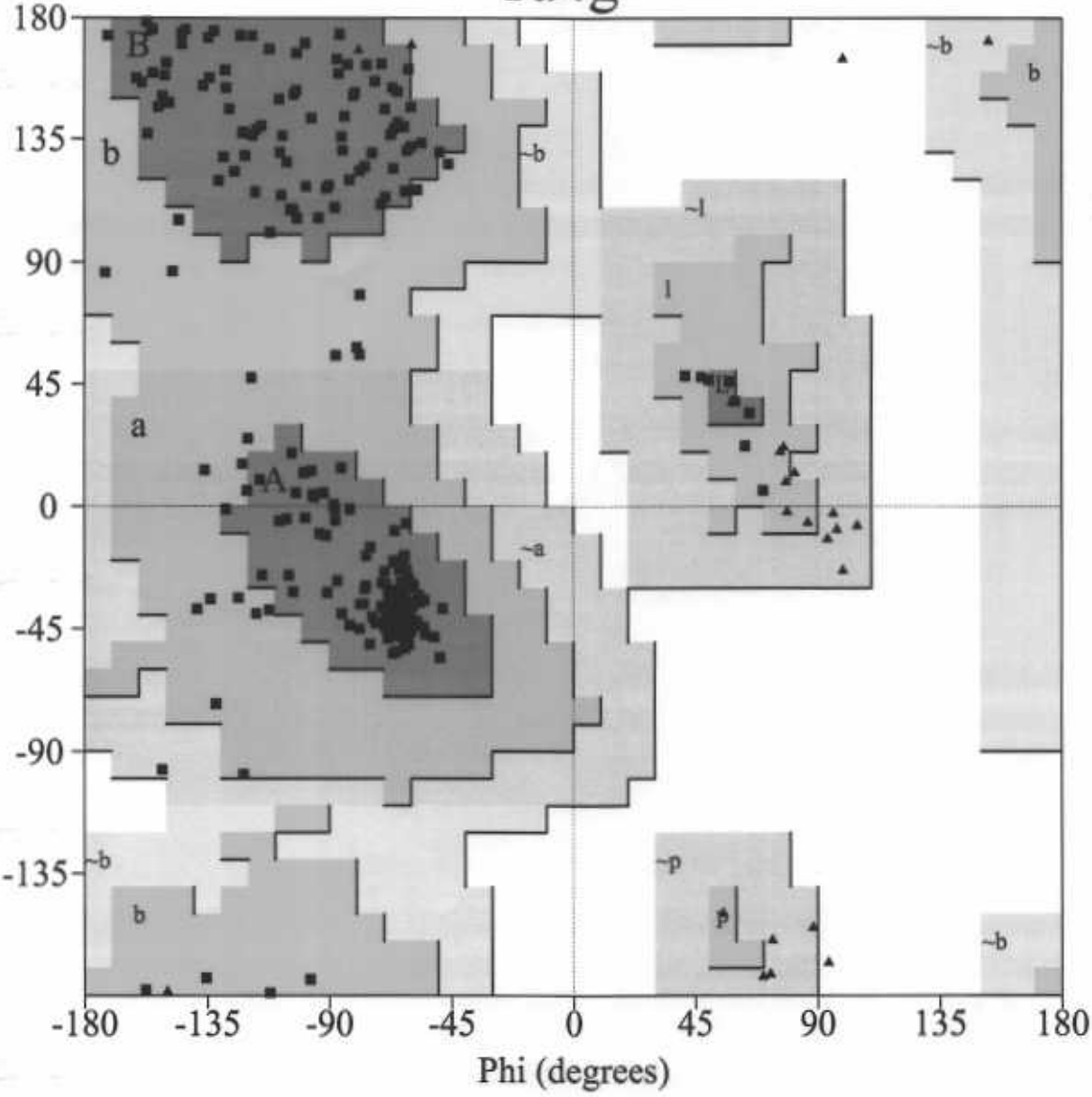
Residues in most favoured regions [A,B,L]	227	88.7%
Residues in additional allowed regions [a,b,l,p]	27	10.5%
Residues in generously allowed regions [~a,~b,~l,~p]	2	.8%
Residues in disallowed regions	0	.0%
Number of non-glycine and non-proline residues	256	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	36	
Number of proline residues	7	
Total number of residues	301	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

PROCHECK

Ramachandran Plot

1u4g

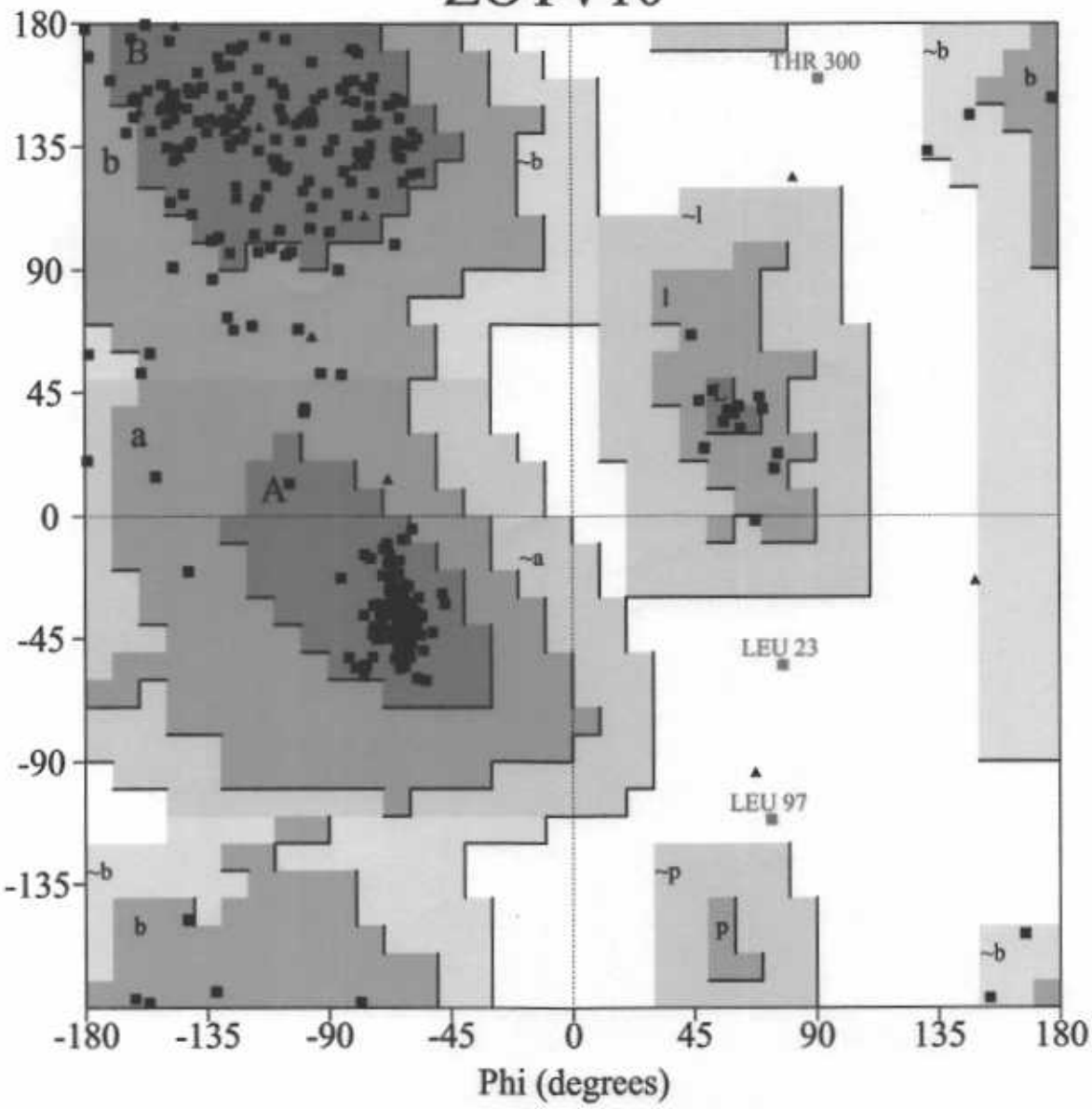


Plot statistics

Residues in most favoured regions [A,B,L]	220	87.6%
Residues in additional allowed regions [a,b,l,p]	30	12.0%
Residues in generously allowed regions [~a,~b,~l,~p]	1	.4%
Residues in disallowed regions	0	.0%
Number of non-glycine and non-proline residues	251	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	33	
Number of proline residues	9	
Total number of residues	295	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot
ZOTV10



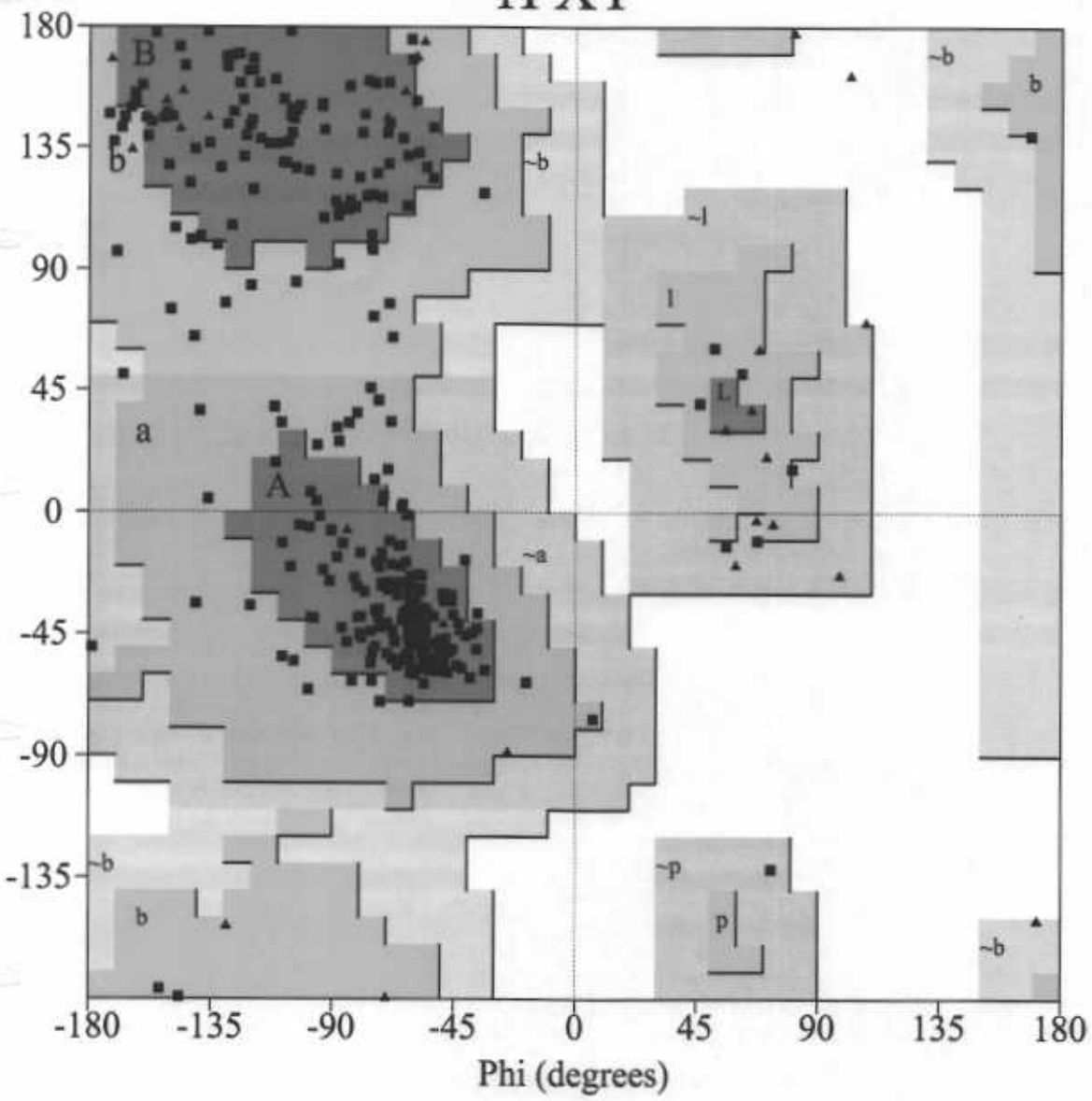
Plot statistics

Residues in most favoured regions [A,B,L]	293	85.9%
Residues in additional allowed regions [a,b,l,p]	38	11.1%
Residues in generously allowed regions [~a,~b,~l,~p]	7	2.1%
Residues in disallowed regions	3	.9%
Number of non-glycine and non-proline residues	341	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	28	
Number of proline residues	19	
Total number of residues	390	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot

1PXT



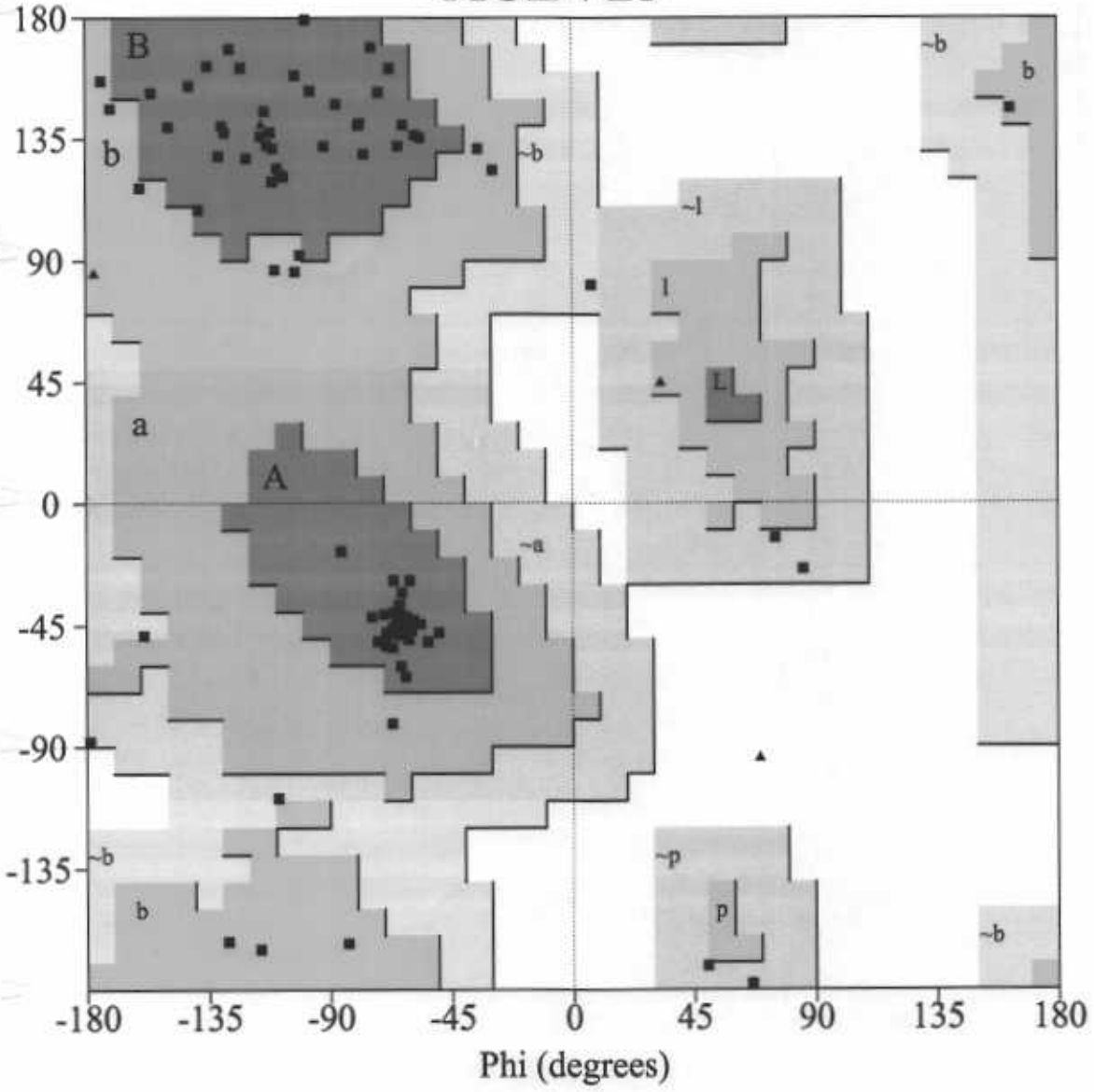
Plot statistics

Residues in most favoured regions [A,B,L]	226	79.0%
Residues in additional allowed regions [a,b,l,p]	53	18.5%
Residues in generously allowed regions [~a,~b,~l,~p]	7	2.4%
Residues in disallowed regions	0	.0%
Number of non-glycine and non-proline residues	286	100.0%
Number of end-residues (excl. Gly and Pro)	7	
Number of glycine residues (shown as triangles)	33	
Number of proline residues	17	
Total number of residues	343	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot

ACEV23



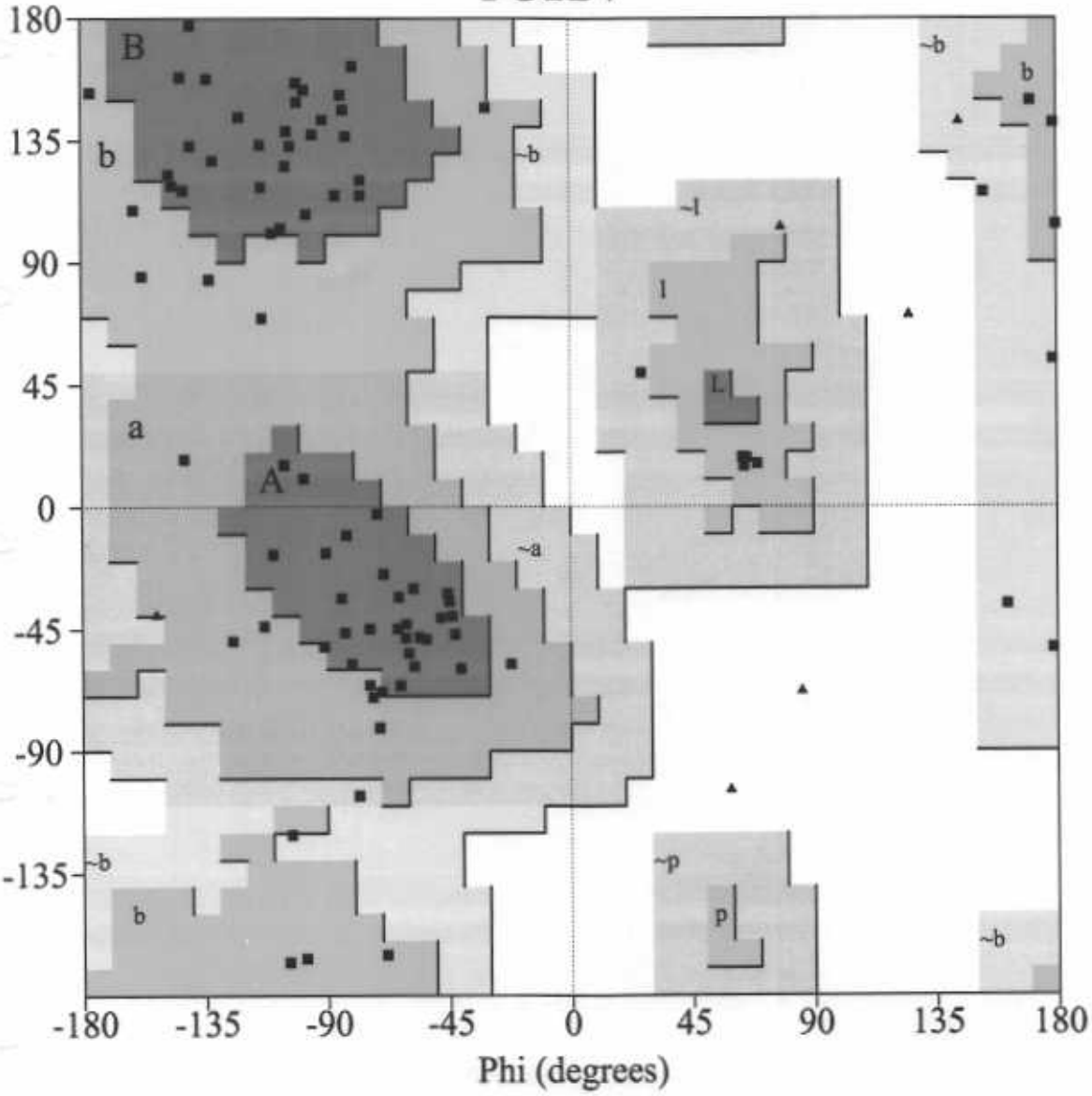
Plot statistics

Residues in most favoured regions [A,B,L]	66	75.9%
Residues in additional allowed regions [a,b,l,p]	13	14.9%
Residues in generously allowed regions [~a,~b,~l,~p]	8	9.2%
Residues in disallowed regions	0	.0%
Number of non-glycine and non-proline residues	87	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	3	
Total number of residues	96	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Ramachandran Plot

1CKV



Plot statistics

Residues in most favoured regions [A,B,L]	56	64.4%
Residues in additional allowed regions [a,b,l,p]	24	27.6%
Residues in generously allowed regions [~a,~b,~l,~p]	7	8.0%
Residues in disallowed regions	0	.0%
Number of non-glycine and non-proline residues	87	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	6	
Number of proline residues	1	
Total number of residues	96	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

APPENDIX VI

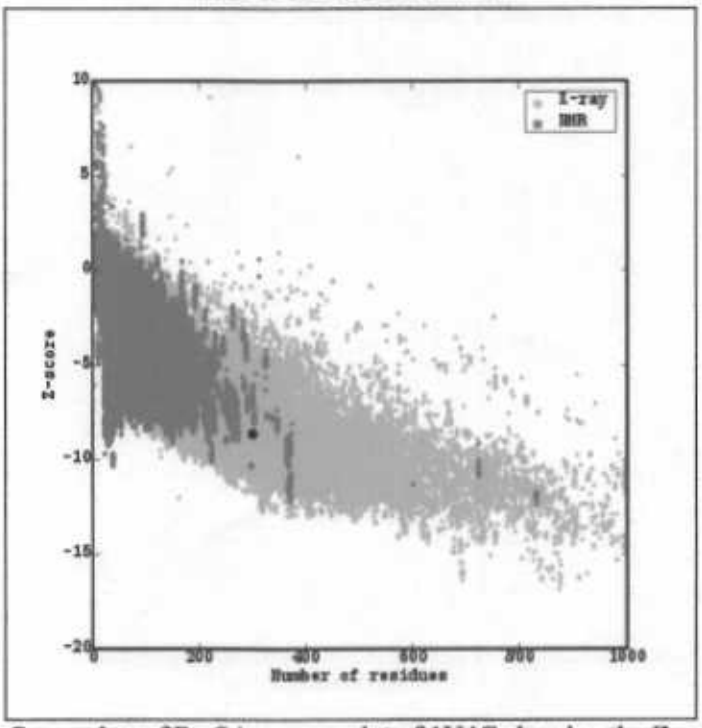


Fig.1 a: Screenshot of ProSA z-score plot of 1U4G showing the Z value < 0.

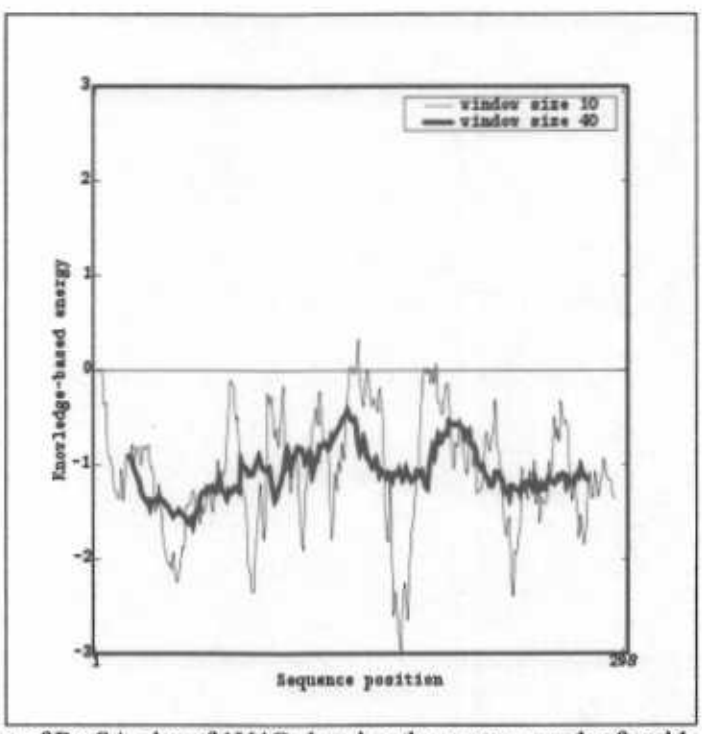


Fig.1b: Screenshot of ProSA plot of 1U4G showing the energy graph of residue scores of a native protein structure.

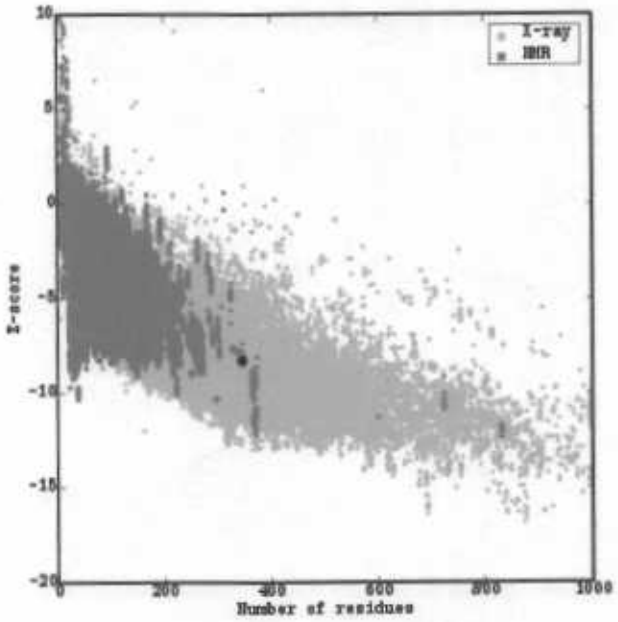


Fig.2a: Screenshot of ProSA z-score plot of 1PXT showing the Z value < 0.

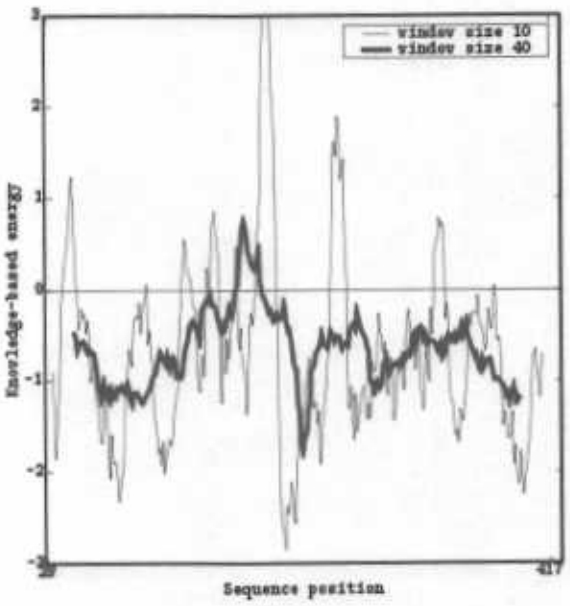


Fig.2b: Screenshot of ProSA plot of 1PXT showing the energy graph of residue scores of a native protein structure.

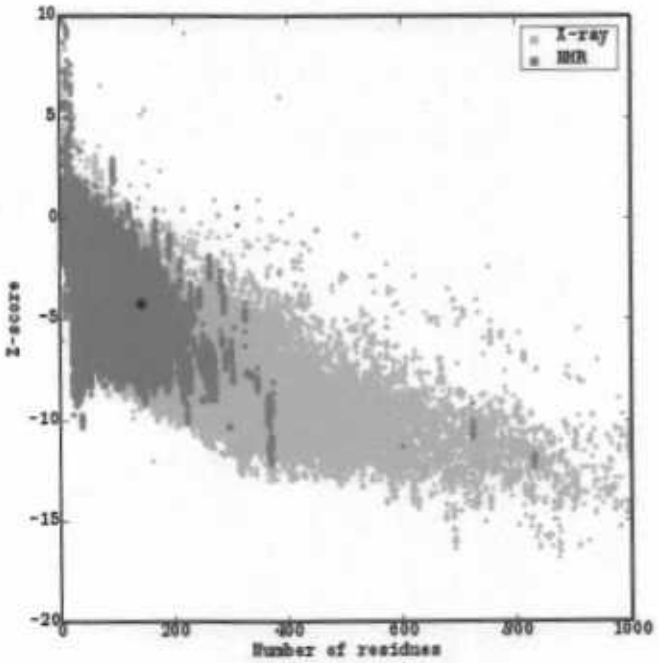


Fig.3a: Screenshot of ProSA z-score plot of 1CKV showing the Z value < 0.

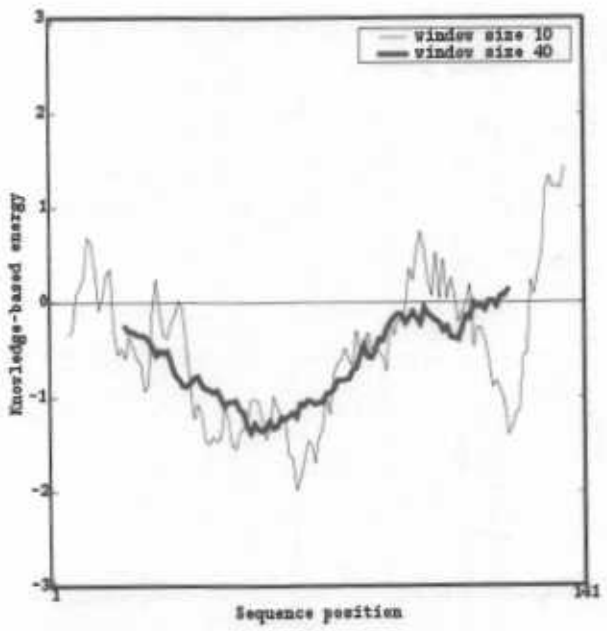


Fig.3b: Screenshot of ProSA plot of 1CKV showing the energy graph of residue scores of a native protein structure.