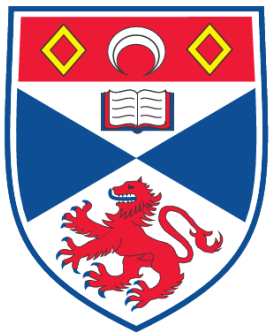
STRUCTURAL BIOLOGY OF *VIBRIO CHOLERAE* PATHOGENICITY FACTORS

Md. Arif Sheikh

A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2009

Structural biology of *Vibrio cholerae* pathogenicity factors

Md. Arif Sheikh

A thesis submitted for the degree of Doctor of Philosophy

University of St Andrews

United Kingdom

March 2009

Declaration

I, Md. Arif Sheikh, hereby certify that this thesis, which is approximately 39,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date ……...Signature of candidate …….....

I was admitted as a research student in April 2006 and as a candidate for the degree of PhD in April 2009; the higher study for which this is a record was carried out in the University of St Andrews between 2006 and 2009.

Date ……...Signature of candidate …….....

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ……...Signature of supervisor …….....

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis: Access to Printed copy and electronic publication of thesis through the University of St Andrews.

Date…….…Signature of candidate…..……...Signature of supervisor……

Dedicated to

The countless children who are killed by cholera every year

Acknowledgements

First of all, I would like to express my gratefulness to the Almighty for his divine help in this work and preparation of this.

I take profound privilege to express my sincere thanks and heartfelt gratitude to my supervisor Professor Garry L. Taylor, School of Biology, University of St Andrews for constant guidance throughout this research work.

I would like to express my great sense of gratitude to Dr. Jane Potter and Nasrin Jahan (Popy) for their constant guidance, inspiration and encouragement every day. I am also deeply grateful to Dr. Helen Connaris, Dr. Charlotte Ryan, Dr. Rupert Russell, Dr. Slava Zaitsev, Guogang Xu, mass spectrometry facilities and administrative support staff at the Centre for Biomolecular Sciences, University of St Andrews, for their valuable help, encouragement and co-operation throughout the research program.

I would also like to thank Professor E. Fidelma Boyd for supervising me on gene knockout studies (University of Delaware), Dr. Robert B. Sim for supervising me on C1q studies (University of Oxford), Professor Juergen Haas for supervising me on protein-protein interaction studies, Dr. Olav Schiemann for supervising me on protein domain movement studies and Professor Malcolm White for supervising me on protein-DNA complex studies.

Special thanks must goes to my family and friends; Azad, Thomas, Kannan, Melina, Christophe, Lester, Ma Qian, Lynn, Ana, Salvador, Pako, Mamun, Kazol, Babu, Nasim, Polash, Ifti, Samim, Arup, Polash (junior), Irfan, Sohana, Imtiaz and all my classmates at Dhaka University for their endless inspiration.

Contents

List of Figures

List of Tables

Abstract

The World Health Organization (WHO) states that 30,000 children under the age of five die each day worldwide. Around a quarter of these die from diarrheal disease caused by microbial infection. In addition to this high mortality rate, there are data emerging on the morbidity effects of diarrheal disease, for example a few episodes of diarrhea in the first two years of life can remove 10 IQ points and lead to growth deficiency. *Vibrio cholerae*, the causative agent of the diarrheal disease cholera, is a serious problem in third world countries, where sanitary and hygiene infrastructure is very poor, and claims several thousand lives every year. In order to better understand the pathogenicity regulation in *V. cholerae*, structural and functional investigations of a hypothetical protein family present in pathogenicity islands and a transcriptional regulator protein for DNA-binding were investigated.

Two adjacent genes, *vc1804* and *vc1805*, encode hypothetical proteins within the *Vibrio* pathogenicity island-2 (VPI-2) of *Vibrio cholerae*, and are part of a cluster of genes only present in pathogenic strains of the bacterium. Paralogous adjacent genes, *vc0508* and *vc0509*, are also present within a second pathogenicity island, the *Vibrio* seventh pandemic island-2 (VSP-2), of *V. cholerae* O1 El Tor and O139 serogroup isolates. Sequence similarity suggests that the VC0508, VC0509, VC1804 and VC1805 proteins will share a similar fold. The crystal structures of VC0508, VC0509 and VC1805 have been determined to a resolution of 1.9, 2.4 and 2.1 Å, respectively. Several recombinant constructs of *vc1804* were made, but no soluble proteins were expressed. This hypothetical protein family reveals structural homology to

human mitochondrial protein p32. Human p32 is a promiscuous protein known to bind to a variety of partners including the globular head component of C1q. We have shown that VC1805 binds to C1q. One possibility is that VC1805 is involved in adherence of the bacterium to membrane-bound C1q in the gut. To explore the roles of VC0508, VC0509, VC1804 and VC1805 *in vivo*, gene knockout and animal model studies of those proteins are underway.

The ferric uptake regulator (Fur), a metal-dependent DNA-binding protein, acts as both a repressor and activator of numerous genes involved in maintaining iron homeostasis in bacteria. It has also been demonstrated in *Vibrio cholerae* that Fur plays an additional role in pathogenesis, and this opens up the potential of Fur as a drug target for cholera. The first crystal structure of a Fur protein, from *Pseudomonas aeruginosa,* revealed a dimeric molecule with each monomer containing a dimerization domain, a helical DNA-binding domain and two metal binding sites: Zn1 is proposed to be a regulatory Fe-binding site, and Zn2 is proposed to be a structural Zn-binding site. Here we present the crystal structure of *V. cholerae* Fur (*Vc*Fur) that reveals a very different orientation of the DNA-binding domains. Accompanying these structural changes are alterations in the amino acids coordinating the zinc at the Zn2 site, and this lends support to this being the site regulated by iron. There is no evidence of metal binding to the cysteines that are conserved in many Fur homologues, including the much-studied *E. coli* Fur. An analysis of the metal binding properties shows that like other Fur proteins, *Vc*Fur can be activated by a range of divalent metals. EPR spectroscopy measurements of the movements of the DNA-binding domain, in the presence of DNA and different metals, are underway.

Abbreviations

aa	Amino acids
Arf6	ADP-ribosylation factor 6
bp	Base pair
C1q	A plasma complement component
cAMP	Cyclic adenosine monophosphate
CSS	Complexation significance score
CT	Cholera toxin
CTXΦ	Cholera toxin phage
D_2O	Deuterium oxide
DGBV++	Dextrose-gelatin-verenal buffer
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
DTT	Dithiothreitol
*Ec*Fur	*Escherichia coli* ferric uptake regulator
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
EPR	Electron paramagnetic resonance spectroscopy
ESI-TOF MS	Electrospray-time of flight mass spectrometry
FITC	Fluorescein isothiocyanate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMBP	His-tag maltose binding protein
HRP	Horseradish peroxidase
IS	Insertion sequence
LB	Luria–Bertani
LPS	Lipopolysaccharide
MALDI-TOF MS	Matrix-assisted laser desorption/ionization-time of flight mass spectrometry
MCS	Multiple cloning sites
MS/MS	Tandem mass spectrometry
MTSL	1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3 methyl-methanethiosulfonate
NAD	Nicotinamide adenine dinucleotide
NHS	Normal human serum
OD	Optical density
ORFs	Open reading frames
ORS	Oral rehydration solution
*Pa*Fur	*Pseudomonas aeruginosa* ferric uptake regulator
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PELDOR	Pulsed electron–electron double resonance
pI	Isoelectric point
RMSD	Root mean square deviation
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SOE-PCR	Splicing by overlapping extension-polymerase chain reaction
SSPF	Scottish Structural Proteomics Facility
TAE	Tris-Acetate-EDTA
TB	Tris-borate
TBE	Tris-borate-EDTA
TCA	Tricarboxylic acid
TCP	Toxin coregulated pilus
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TPB	Tryptose phosphate broth
*Vc*Fur	*Vibrio cholerae* ferric uptake regulator
VPI-1	*Vibrio* pathogenicity island-1
VPI-2	*Vibrio* pathogenicity island-2
VSP-1	*Vibrio* seventh pandemic island-1
VSP-2	*Vibrio* seventh pandemic island-2
WHO	World Health Organization
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
Y2H	Yeast-two-hybrid

Amino acids

Amino Acid	**3-Letter (1-Letter)**	**Side chain polarity**	**Side chain charge (pH7)**
Alanine	Ala (A)	Nonpolar	Neutral
Arginine	Arg (R)	Polar	Positive
Asparagine	Asn (N)	Polar	Neutral
Aspartic acid	Asp (D)	Polar	Negative
Cysteine	Cys (C)	Nonpolar	Neutral
Glutamic acid	Glu (E)	Polar	Negative
Glutamine	Gln (Q)	Polar	Neutral
Glycine	Gly (G)	Nonpolar	Neutral
Histidine	His (H)	Polar	Positive
Isoleucine	Ile (I)	Nonpolar	Neutral
Leucine	Leu (L)	Nonpolar	Neutral
Lysine	Lys (K)	Polar	Positive
Methionine	Met (M)	Nonpolar	Neutral
Phenylalanine	Phe (F)	Nonpolar	Neutral
Proline	Pro (P)	Nonpolar	Neutral
Serine	Ser (S)	Polar	Neutral
Threonine	Thr (T)	Polar	Neutral
Tryptophan	Trp (W)	Nonpolar	Neutral
Tyrosine	Tyr (Y)	Polar	Neutral
Valine	Val (V)	Nonpolar	Neutral

Chapter 1

Introduction

1. Introduction

1.1 Overview of cholera

The severe diarrhoeal disease cholera is caused by the Gram-negative bacterium *Vibrio cholerae* (Figure 1.1) and continues to be a major cause of morbidity and mortality. Annually, cholera causes the death of more than a hundred thousand people among millions of reported cases worldwide (Faruque *et al*., 1998a). *V. cholerae* was identified as the causative agent of cholera in 1884 during the fifth pandemic. The isolation of a bacteria capable of producing a toxin that stimulated massive loss of fluid from the intestine were reported (DiRita, 2001). But the cause of cholera associated with the contaminated water was determined in 1854 by an epidemiologist John Snow in London (Snow *et al*., 1936). *V. cholerae* has more than 200 known serotypes but not all the strains are pathogenic. Among them, only those strains with O1 and O139 antigens are highly pathogenic and acknowledged to cause epidemic and pandemic disease (Reidl and Klose, 2002). Strains of *V. cholerae* which do not agglutinate with O1 or O139 antisera are called non-O1 or non-O139 *V. cholerae* respectively and are not thought to be involved in epidemic cholera. However, these non-epidemic strains could be pathogenic (Morris *et al*., 1990) as they are occasionally associated with small outbreaks of diarrheal disease (Aldova *et al*., 1968). *V. cholerae* O1 and O139 typically cause intermittent outbreaks of diarrhoeal disease and rarely cause extra intestinal infection, but *V. cholerae* non-O1, non-O139 and other *Vibrio* species may cause frequent bacteremia in immunocompromised individuals (Ko *et al*., 1998; Safrin *et al*., 1988).

The main source of cholera is contaminated water and food (ICDDRB, 1993). The infective dose for *V. cholerae* is 10^6 cells. However the ultimate infection depends on some other factors, for instance, personal immunity, the gastric-acid barrier and overall sanitary conditions (Sack *et al*., 2004). Even more interestingly, people with blood group O are highly susceptible to *V. cholerae* El Tor strain infection compared to others (Barua and Paguio, 1977; Clemens *et al*., 1989; Glass *et al*., 1985).

The characteristic symptom of the acute form of cholera is rice water stools. Clinical manifestations of the disease range from mild symptoms such as abdominal cramps, nausea and vomiting, to more severe symptoms such as dehydration, shock and death (Weir and Haider, 2004). In case of severe cholera, the rate of fluid loss from the body could reach 500-1000 mL/h that can cause dangerous dehydration (Sack *et al*., 2004). Additionally, inappropriate treatment can lead to further complications such as acute renal failure due to insufficient fluid administration to the patient or severe hypoglycemia resulting from low blood glucose concentrations (Butler *et al*., 1989). If the infection is severe and the patient left untreated, an adult could die within 24 h (Schoolnik and Yildiz, 2000).

Over a century of research, knowledge about cholera has increased a lot, but the ultimate solution to the problem has not yet been found, as it is still devastating and can spread very rapidly in third-world countries where the degree of the disease is fatal.

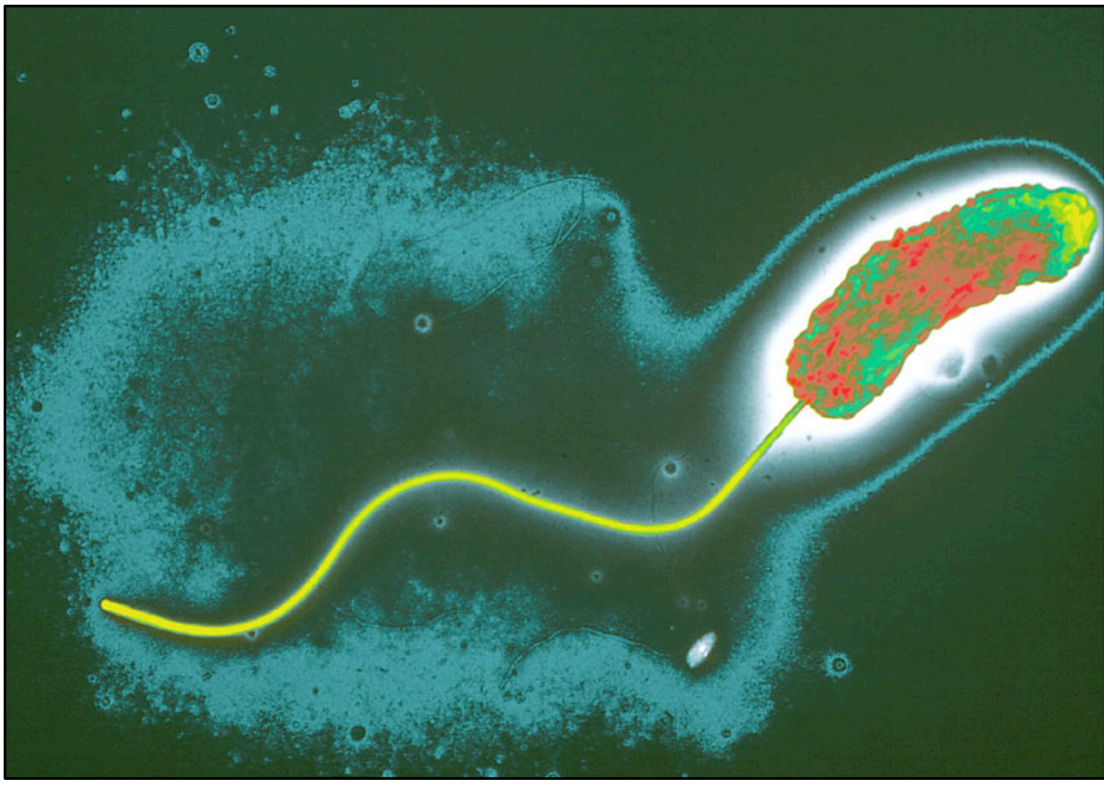


Figure 1.1 A 10,000 times magnification of *V. cholerae* bacterium (Waldor and RayChaudhuri, 2000).

1.2 Cholera: History of global pain

According to the recent reports in various international media in February 2009, the outbreak of cholera in Zimbabwe has already killed 3800 people, since the outbreak started in August 2008. Approximately 60,000 cases have been reported with a mortality rate of about 5%, according to United Nations and World Health Organization (WHO). Now this cholera outbreak is spreading to other neighbouring countries like South Africa where 36 people have already been killed among 5,696 reported cases and most of the affected regions are near Zimbabwean borders (Kapp, 2009).

The fearsome reputation of cholera as a disease has arisen due to its frequent occurrence in epidemic proportion coupled to its high mortality rates

(Sack *et al*., 2004). Every year cholera causes illness in millions of individuals including children who are highly susceptible (WHO, 1995). The first recorded cholera pandemic began in 1817. The infection rapidly spread from the Indian sub-continent along trade routes as far as southern Russia. A second pandemic, starting in 1826 leading to the third, fourth and fifth pandemics persisting until the 19th century, wrought havoc in Asia, Africa, Europe and the Americas (Pollitzer, 1954). The sixth pandemic, started in 1899 and persisted until 1923, spreading extensively in the Middle-East and Balkan regions (Pollitzer, 1959).

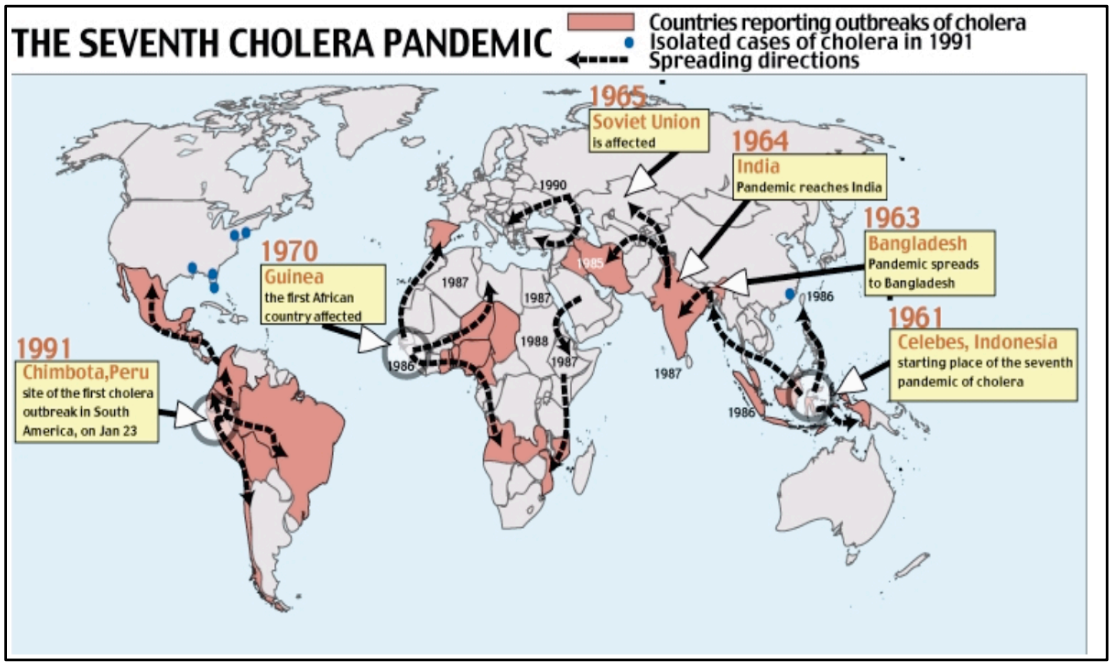


Figure 1.2 The seventh cholera pandemic (Institute of Tropical Medicine in Antwerp, Belgium).

In the Indian sub-continent, cholera epidemics emerge with alarming frequency and are usually initiated in the Ganges Delta region. These epidemics often spread to other continents over the course of many years claiming thousands of lives on their way (Pollitzer, 1954). Only the seventh

pandemic arose outside the Indian sub-continent, beginning in 1961 on the Island of Sulawesi in Indonesia before spreading to other countries. This pandemic is the most extensive with regard to geographic spread and duration, and the causative agent is *V. cholerae* O1 biotype E1 Tor (Cvjetanovic and Barua, 1972). The strain was named after the village El Tor in Egypt where it was first isolated in 1905 (Pollitzer, 1959). Later, it appeared in 1937 in Sulawesi, Indonesia (Tanamal, 1959). The strain began to spread all over the world in 1961 causing the seventh pandemic. The seventh pandemic continues to ravage parts of the world today, shown in Figure 1.2, with particularly devastating effect in developing countries such as India, Bangladesh, Rwanda and Mozambique (Sack *et al*., 2004).

V. cholerae serogroup O1 is classified into two biotypes, which are termed 'classical' and 'El Tor'. The current pandemic of cholera is caused by the El Tor biotype, but the fifth and sixth pandemics of cholera were caused by the classical biotype. The classical biotype of *V. cholerae* O1 is thought to have become extinct with the El Tor biotype becoming the dominant biotype. Indeed, the classical biotype has not featured in epidemic cholera outbreaks since 1983 (Samadi *et al*., 1983). Another serogroup called *V. cholerae* O139 was first discovered in 1992 when it caused severe outbreaks in India and Bangladesh (ICDDRB, 1993; Ramamurthy *et al*., 1993). There was another outbreak associated with O139 serotype where 30,000 people were infected in Dhaka, which surpassed the number of cases associated with seventh pandemic El Tor strain (Faruque *et al*., 2003a). It has been debated whether this serogroup may herald the beginning of an eighth cholera pandemic (Swerdlow and Ries, 1993).

1.3 Ecology of *Vibrio cholerae*

After the discovery of the *V. cholerae* bacteria in 1884 by Robert Koch, the first assumption was that an aquatic environment acted as the reservoir for vibrios. Koch isolated comma bacillus from a tank in Calcutta during the cholera epidemic in 1883 (DiRita, 2001). *V. cholerae* was isolated from areas free from fecal pollution and it was proposed that the bacterium belongs to the bacterial flora populating brackish water and estuarine systems and the phenomenon of aquatic environment as the natural reservoir of *V. cholerae* was established (Colwell *et al*., 1977). During an investigation to test the viability of *V. cholerae*, it was reported that the *V. cholerae* bacterium was able to survive in a number of different natural water sources over a period of time ranging from 1-6 days (Lahiri *et al*., 1939). Indeed, *V. cholerae* has the ability to survive up to 1 month at 4°C and 2 to 14 days at 20°C to 30°C in dechlorinated clean water (Feachem *et al*., 1981).

V. cholerae can survive in association with aquatic flora such as phytoplankton as well as zooplankton, particularly on copepods (Islam *et al.*, 1994; Samadi *et al.*, 1983). It has also been reported that *V. cholerae* can attach to the blue green algae *Anabaena variabilis* for up to 5 days within the algae's natural habitat. In addition, the bacterium does not lose its ability to produce the cholera toxin following its association with the algae. This study also suggested that *A. variabilis* has the ability to act as a reservoir of *V. cholerae* O1 in the aquatic environment (Islam *et al.*, 1990; Islam *et al.*, 1993, 1994). The association of *V. cholerae* strains with zooplankton and phytoplankton are well established and could also interfere with global climate

change by affecting the growth of planktons (Sack *et al*., 2004). In the absence of disease, the presence of pathogenic *V. cholerae* O1 in aquatic environments indicates both the tenacity of the organism and its ability to flourish within a natural environment (Colwell *et al*., 1981).

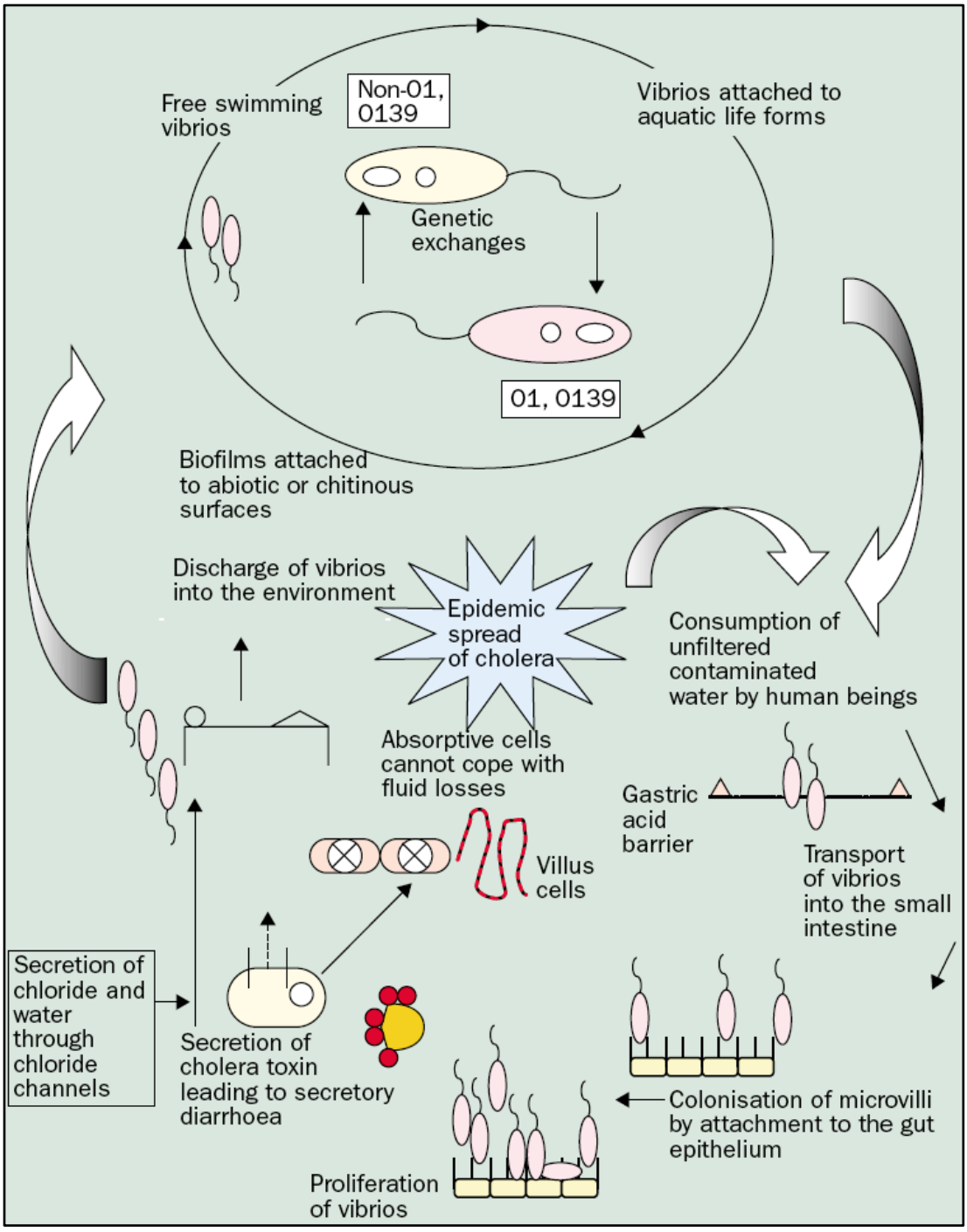


Figure 1.3 Life cycles of *V. cholerae* (Sack *et al*., 2004)

There are two distinctive life cycles for *V. cholerae* that occur inside the host (human) and outside the host (environment). These cycles, shown in Figure 1.3, involve different intermediate steps that sometimes intersect (Sack *et al*., 2004). Isolated pathogenic strains of *V. cholerae* from infected areas possess cholera toxin (CT) genes whereas the similar isolates from uninfected areas do not possess CT (Faruque *et al*., 1998b).

1.4 Evolution of *Vibrio cholerae*

V. cholerae had long been part of the environment before it became pathogenic. Most of the strains of *V. cholerae* are still part of the natural flora (Colwell and Spira, 1992). Pathogenic strains of *V. cholerae* contain a few chromosomal regions, acquired by horizontal gene transfer, which are absent in non-pathogenic strains, and are vital for the evolution of pathogenic strains (Faruque and Mekalanos, 2003). For example, *V. cholerae* have an extraordinary survival mechanism, as they can uptake new genetic material while they grow on chitin, an abundant natural habitat of *V. cholerae* (Meibom *et al*., 2005). Addition and deletion of some chromosomal regions between pathogenic strains assists the emergence of new classes of pathogenic clones. For example, the evolution of *V. cholerae* O139 epidemic strain occurred as a result of horizontal gene transfer events which deleted or replaced a few genes encoding the enzyme for the synthesis of lipopolysaccharide (LPS) O-antigen side chain (Bik *et al.*, 1995; Waldor *et al.*, 1994). The relationship between the evolution of pathogenic strains of *V. cholerae* and the ability of the organism to survive in the environment is not very clear, as some of the environmental strains are pathogenic but are not

involved in epidemic patterns of cholera (Boyd and Waldor, 2002; Faruque *et al.*, 2003b). The possible evolutionary linages of *V. cholerae* pathogenic strains are shown in Figure 1.4 (Faruque and Mekalanos, 2003).

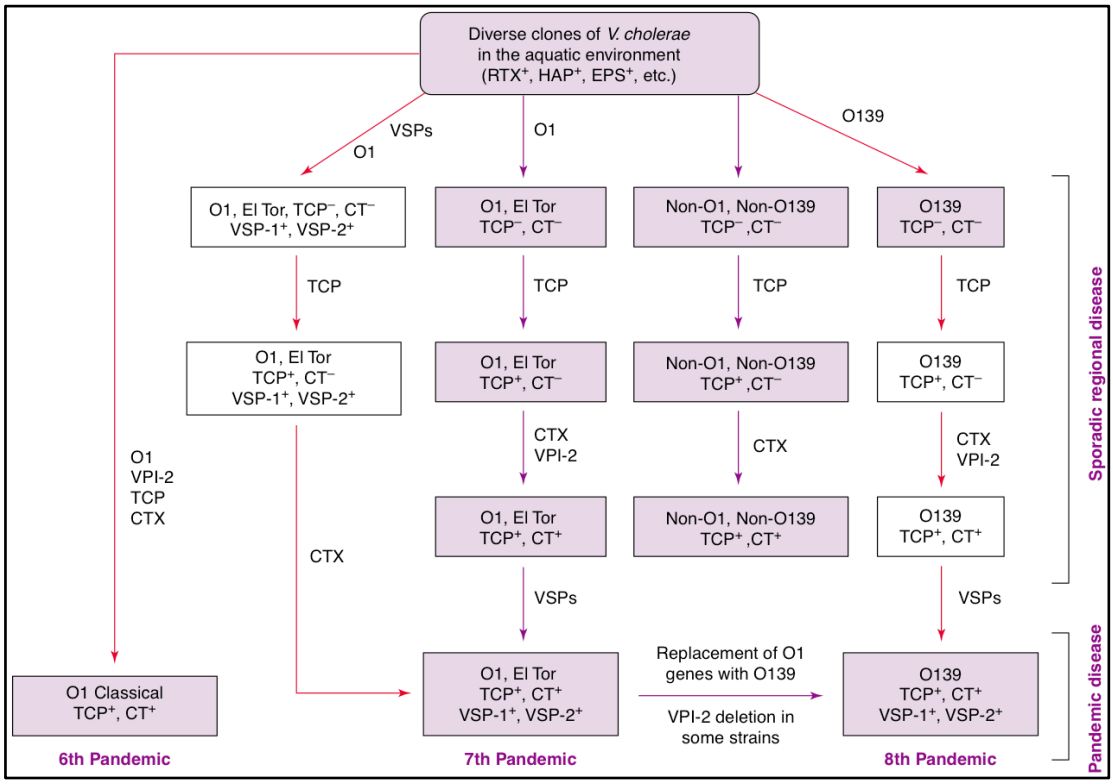


Figure 1.4 Evolution of the possible pathogenic strains of *V. cholerae* (Faruque and Mekalanos, 2003). The strains that have been reported are in shaded boxes and the white boxes represent possible strains yet to be discovered. Evolution of strains by horizontal gene transfer events are indicated by purple arrows whereas the red arrows indicate other transfer events that could be possible but the intermediate strains of these events have not been isolated yet. Abbreviations: EPS, extracellular protein secretion gene cluster; HAP, hemagglutinin protease gene.

1.5 Genome sequence of *Vibrio cholera*: inside of the killer

The *V. cholerae* strains consist of two separate chromosomes. The smaller chromosome is predicted to be a megaplasmid which became an essential part during the evolutionary process after gaining some essential genes (Egan

et al., 2005; Heidelberg *et al*., 2000). In the year 2000, the sequencing of the *V. cholerae* O1 El Tor genome started to reveal the significance of the pathogenicity of the organism. *V. cholerae* has two circular chromosomes of 2,961,151 bp and 1,072,914 bp, named as chromosome-I and chromosome-II and encode 2,770 and 1,115 open reading frames (ORFs), respectively (Figure 1.5). The larger chromosome encodes 19% hypothetical proteins and 17% conserved hypothetical proteins whereas the smaller chromosome encodes 38% hypothetical genes and 15% conserved hypothetical proteins. Most of the essential genes including pathogenicity island genes of *V. cholerae* are present in chromosome-I (Heidelberg *et al*., 2000).

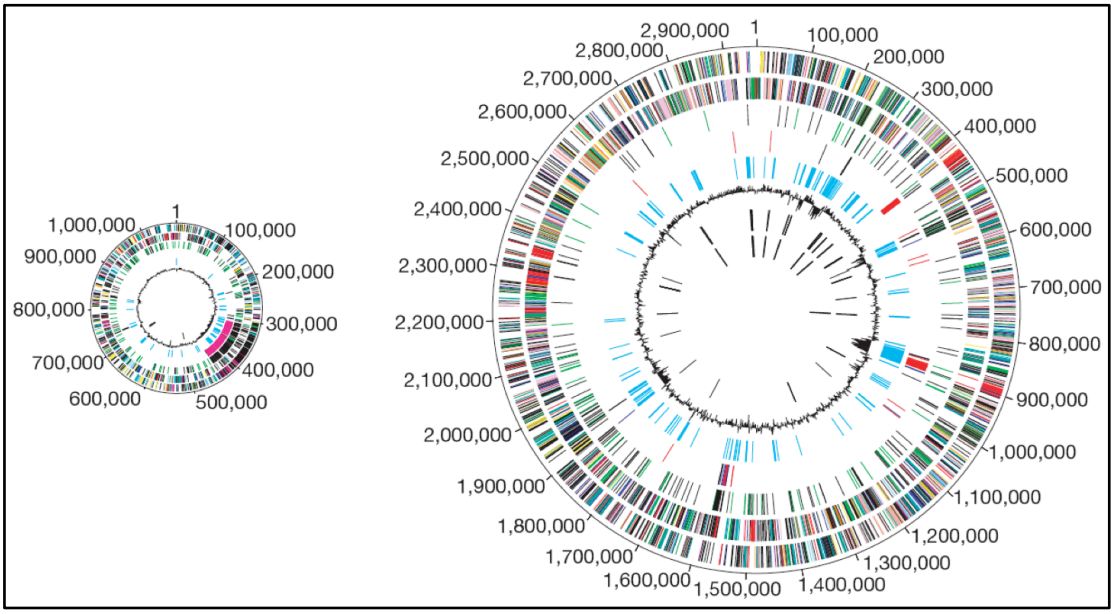


Figure 1.5 *V. cholerae* genome map. Circular representation of large (right) and small (left) chromosome. The outermost rings (circles 1 and 2) show predicted protein-coding regions (hypothetical proteins are in black). The third circle shows recently duplicated genes. The fourth circle shows transposon-related (black), phage-related (blue), *V. cholerae* repeat sequence (pink) and pathogenesis genes (red). The fifth circle shows χ^2 values for trinucleotide composition. The sixth circle shows G+C percentage. The seventh and eighth circles are tRNAs and rRNAs, respectively (Heidelberg *et al*., 2000).

1.6 Molecular basis of *Vibrio cholerae* pathogenesis

The molecular basis of *V. cholerae* pathogenesis is a complex process that has not been fully understood. It involves expression of numerous virulence genes which primarily helps the bacterium to pass the gastric acid barrier in order to colonize the small intestine, resulting in the secretion of cholera toxin (CT) (Kaper *et al*., 1995). Upon colonization of the mucosa of the small intestine, *V. cholerae* secretes CT as a major virulence factor. CT is a protein-protein complex made up of six separate subunits: a single copy of the globular A (29 kDa) subunit, and five copies of the B subunit. The five identical 11 kDa B subunits ally to form a pentameric ring-like structure that then assembles in a non-covalent fashion with the A subunit. The A subunit includes two sub-segments which are linked to each other via a disulphide bond (Sixma *et al*., 1991; Zhang *et al*., 1995).

Upon secretion of CT, the B subunit ring binds to GM1 gangliosides on the surface of host cells. After binding has occurred, the cell internalizes the entire CT complex and the CTA1 (23.5 kDa) chain is released via reduction of a disulfide bridge. CTA1 is then free to bind with a human partner protein called ADP-ribosylation factor 6 (Arf6). Binding to Arf6 induces a conformational change upon CTA1 exposing its active site and enabling catalytic activity. The CTA1 fragment catalyses ADP ribosylation from NAD to the regulatory component of adenylate cyclase resulting in increase of adenylate cyclase activity followed by increase of cyclic AMP (cAMP) synthesis. The presence of a higher concentration of cAMP pumps large amounts of chloride ion (Cl^-) from toxin damaged mucosal cells to the intestine. Consequently, water, Na^+

and other electrolytes are also pumped out due to the imbalanced osmotic and electrical gradients caused by the loss of Cl^-. The loss of water and electrolytes from the toxin damaged mucosal cells to intestine leads to characteristic symptoms of cholera (DiRita, 2001; O'Neal *et al.*, 2005; Zhang *et al.*, 1995). A schematic diagram of the cholera toxin activation mechanism is shown in Figure 1.6.

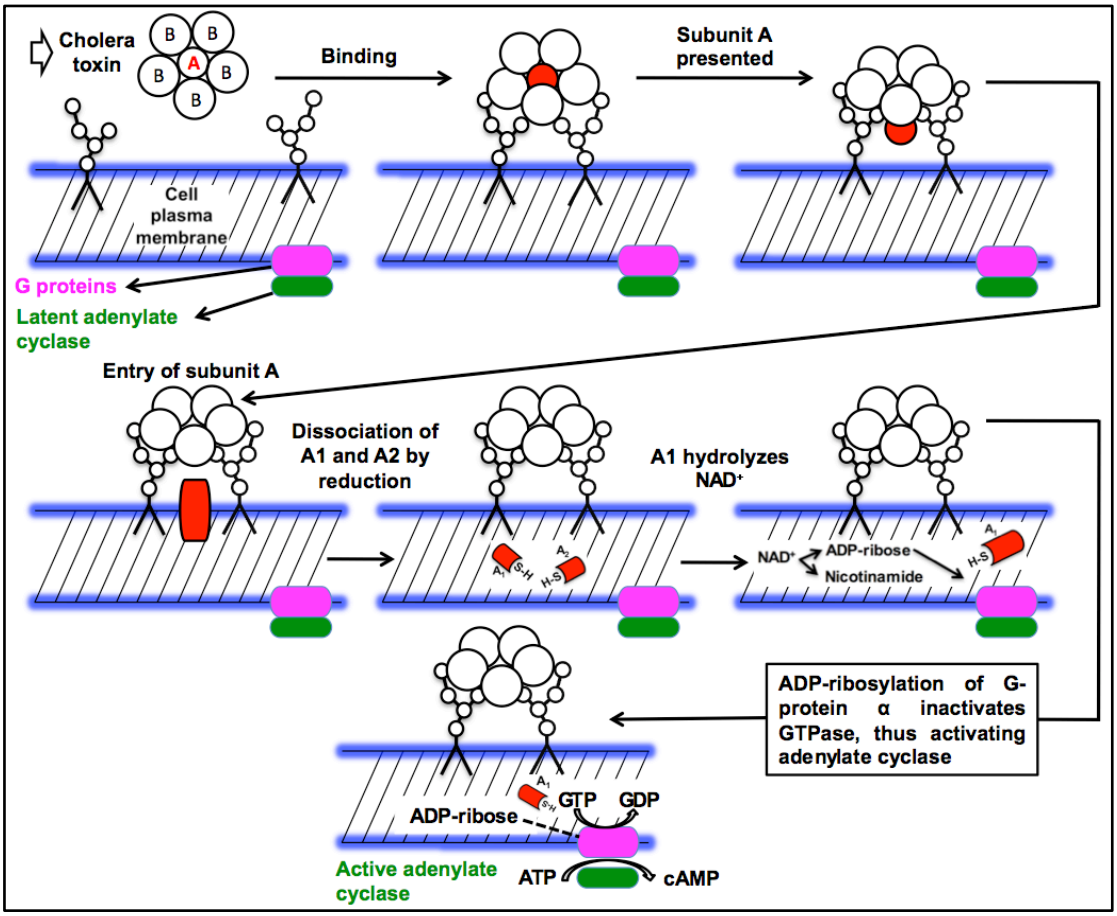


Figure 1.6 Mechanism of activation of cholera enterotoxin (http://gsbs.utmb.edu/microbook/ch024.htm).

1.7 Virulence factors of *Vibrio cholerae*

The pathogenesis of cholera depends on several factors that act in combination to assist the bacterium to colonize the gastrointestinal tract. During colonization in the epithelium of the small intestine, pathogenic strains

of *V. cholerae* produce CT that interrupts ion transport by epithelial cells and causes diarrhoea. Even though the manifestations of diarrhoea are associated with the production of CT, the pathogenesis of cholera depends on the synergistic action of several proteins (Kaper *et al*., 1995). The pathogenicity of *V. cholerae* involves a number of important genetic elements that act to distinguish them from non-pathogenic strains.

1.7.1 CTX (cholera toxin phage) genetic element

By horizontal gene transfer, the CTXΦ genome can integrate into the *V. cholerae* genome, or alternatively, the phage itself can stay as a plasmid inside *V. cholerae* due to the lack of an appropriate integration site, and can replicate during cell division. CTX genetic element is a chromosomally integrated CTXΦ genome that carries the gene *ctxAB* encoding CT. All epidemic *V. cholerae* isolates contain CTXΦ but it is rarely recovered from non-O1/non-O139 *V. cholerae* environmental isolates (Faruque *et al*., 1998a). The CTXΦ has a 6.9 kb genome consisting of two functionally diverse regions (Figure 1.7). The 'core region' that carries the *ctxAB* operon encodes CT as well as genes essential for phage morphogenesis which are thought to encode the major and minor phage coat proteins and a protein required for CTXΦ assembly (Table 1.1). The RS2 region consists of genes that encode products required for the integration, replication and regulatory functions of CTXΦ. The CTXΦ is frequently flanked by an additional region, named RS1 (Figure 1.7 and Table 1.1). RS1 is 2.7 kb in size and very similar to RS2 (Waldor and Mekalanos, 1996; Waldor *et al*., 1997). CTXΦ is present in the two available chromosomes in classical biotype strains of *V. cholerae*

serogroup O1 (Mekalanos, 1983; Trucksis *et al*., 1998), but only in the large chromosome of El Tor biotype strains of *V. cholerae* serogroup O1 (Mekalanos *et al*., 1983; Pearson *et al*., 1993). Non-pathogenic strains of *V. cholerae* may be converted to pathogenic strains by phage transduction of CTXΦ. This could conceivably take place in the gastrointestinal environment and would result in the production of a new pathogenic strain of *V. cholerae* (Reidl and Klose, 2002).

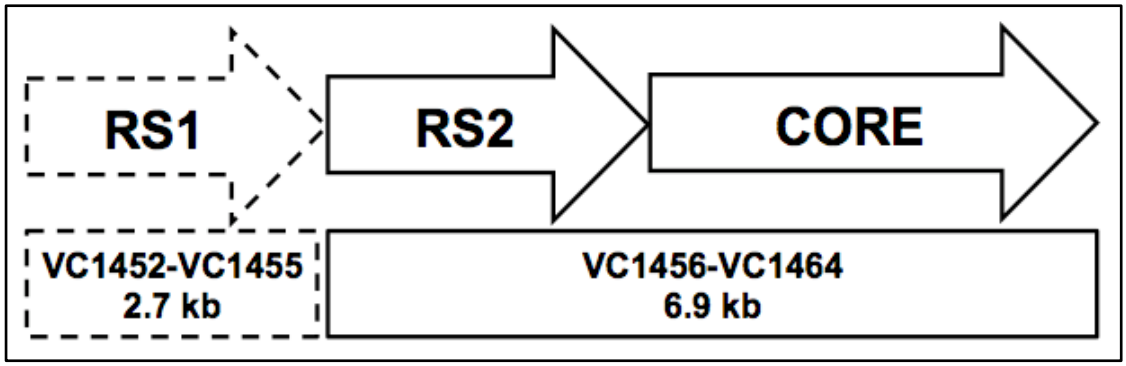


Figure 1.7 Schematic representation of CTXΦ

Table 1.1 ORFs within the *V. cholerae* CTXΦ

ORF	Length (aa)	Homologous protein*
VC1452	78	RstC protein
VC1453	124	RstB1 protein
VC1454	359	RstA1 protein (phage replication initiation protein)
VC1455	112	Transcriptional repressor RstR
VC1456	124	Cholera enterotoxin, B subunit
VC1457	258	Cholera enterotoxin, A subunit
VC1458	399	Zona occludens toxin
VC1459	97	Accessory cholera enterotoxin
VC1460	427	Hypothetical protein
VC1461	82	Colonization factor
VC1462	127	RstB2 protein
VC1463	359	RstA2 protein
VC1464	112	Transcriptional repressor RstR

*Homology based on BLASTP database analysis.

1.7.2 *Vibrio* pathogenicity island-1 (VPI-1) or TCP island

The term “pathogenicity island” requires that certain criteria are met, for example, it should be located within an unstable region of a chromosome, be of different GC content in comparison to the residual genome, be allied with a tRNA gene, contain an IS (insertion sequence) element near the integrase and finally the element should be absent in a non-pathogenic strain (Hacker *et al*., 1997). One of the vital components thought to be involved in colonization of *V. cholerae* in the small intestine is TCP (toxin coregulated pilus). TCP Island or VPI-1 (Figure 1.8) is the part of a region of pathogenic strains of *V. cholerae* which is about 40 kb in size with low GC content (35%) flanked on both sides by 20-bp *att*-like sequences, and consists of TCP-ACF (accessory colonization factor) gene clusters, the integrase and a transposase gene (Karaolis *et al*., 1998; Kovach *et al*., 1996). The gene *tcpA* (*vc0828*) encodes the major subunit of TCP, but the total formation and the function of TCP require proteins encoded by the *tcp* gene cluster (Ogierman *et al*., 1993). For example, the gene product of *tcpH and tcpI* influence TcpA synthesis, where *tcpI* acts as a negative regulator and *tcpH* is a positive regulator of pilin synthesis (Harkey *et al*., 1994). As part of *tcp* gene cluster, a total of 15 ORFs (VC0825-VC0839, Table 1.2) are located downstream of *tagD* gene (*vc0824*). The *acf* gene cluster is present downstream of the *tcp* gene cluster. The true functions of these genes have not been discovered with the exception of accessory colonization factor, *acfD* (*vc0845*), which encodes a lipoprotein (Kovach *et al*., 1996). TCP can also act as a receptor of CTXΦ for infecting new strains (Waldor and Mekalanos, 1996).

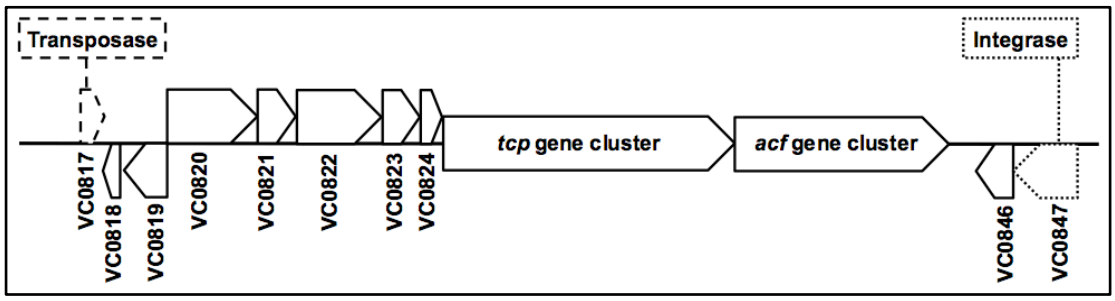


Figure 1.8 Schematic representation of TCP island or VPI-1 (VC0817-VC0847)

Table 1.2 ORFs within the TCP island or VPI-1 (VC0817-VC0847)

ORF	Length (aa)	Homologous protein*
VC0817	327	Transposase
VC0817	226	No significant match
VC0819	541	Aldehyde dehydrogenase
VC0820	1013	ToxR-activated gene A protein
VC0821	379	Hypothetical protein
VC0822	1112	Inner membrane protein, putative
VC0823	312	Hypothetical protein
VC0824	164	TagD protein ; thiol peroxidase, atypical 2-Cys peroxiredoxin
VC0825	620	Toxin co-regulated pilus biosynthesis protein I
VC0826	221	Toxin co-regulated pilus biosynthesis protein P
VC0827	136	Toxin co-regulated pilus biosynthesis protein H
VC0828	224	Toxin co-regulated pilin
VC0829	430	Toxin co-regulated pilus biosynthesis protein B
VC0830	150	Toxin co-regulated pilus biosynthesis protein Q
VC0831	489	Toxin co-regulated pilus biosynthesis outer membrane protein C
VC0832	151	Toxin co-regulated pilus biosynthesis protein R
VC0833	278	Toxin co-regulated pilus biosynthesis protein D
VC0834	152	Toxin co-regulated pilus biosynthesis protein S
VC0835	503	Toxin co-regulated pilus biosynthesis protein T
VC0836	340	Toxin co-regulated pilus biosynthesis protein E
VC0837	338	Toxin co-regulated pilus biosynthesis protein F
VC0838	276	TCP pilus virulence regulatory protein
VC0839	253	Toxin co-regulated pilus biosynthesis protein J
VC0840	626	Accessory colonization factor AcfB
VC0841	256	Accessory colonization factor AcfC
VC0842	157	Hypothetical protein
VC0843	302	TagE protein
VC0844	215	Accessory colonization factor AcfA
VC0845	1520	Accessory colonization factor AcfD
VC0846	229	No significant match
VC0847	422	Phage family integrase

*Homology based on BLASTP database analysis.

1.7.3 *Vibrio* seventh pandemic island-1 (VSP-1)

VSP-1 is a chromosomal island specific to the *V. cholerae* seventh pandemic strain of O1 El Tor and O139 serogroup isolates and is absent in more than 100 non-O1 and non-O139 and TCP negative O1 strains (Dziejman *et al*., 2002; Faruque and Mekalanos, 2003). VSP-1 is a 16 kb region, ranging from VC0175 to VC0185 (Table 1.3 and Figure 1.9), and has a low GC contents (40%) compared to the rest of the genome (47%). The low GC contents also reflect the recent acquisition of this island. Among 11 genes of VSP-1, seven of them were annotated as hypothetical proteins. The C-terminal domain of VC0175 shows similarity to deoxycytidylate deaminase-related proteins. The protein deaminase has various functions including nucleotide scavenging or DNA uptake during competence (Ribeiro *et al*., 1997; Richter *et al*., 1997). VC0176 might be a regulatory gene as it encodes a product which is paralogous to a family that includes RstR, a lysogeny repressor protein for CTXΦ (Dziejman *et al*., 2002). VC0178 is homologous to different phospholipases and VC0185 is a transposase which is similar to Tn554 in *Staphylococcus aureus* (Murphy *et al*., 1985).

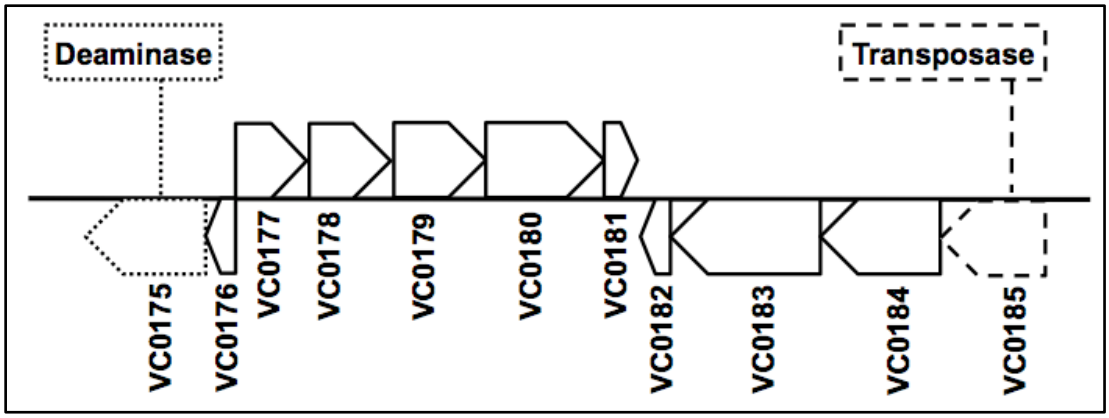


Figure 1.9 Schematic representation of VSP-1 (VC0175-VC0185)

Table 1.3 ORFs within VSP-1 (VC0175-VC0185)

ORF	Length (aa)	Homologous protein*
VC0175	532	Deoxycytidylate deaminase-related protein
VC0176	102	Transcriptional regulator, putative
VC0177	187	Hypothetical protein
VC0178	355	Patatin-related protein ; K06900
VC0179	436	Hypothetical protein
VC0180	584	Hypothetical protein
VC0181	156	Hypothetical protein
VC0182	143	Hypothetical protein
VC0183	703	Hypothetical protein
VC0184	562	Hypothetical protein
VC0185	404	Transposase, putative

*Homology based on BLASTP database analysis.

1.7.4 *Vibrio* seventh pandemic island-2 (VSP-2)

Like VSP-1, VSP-2 is a chromosomal island which is also specific to the *V. cholerae* seventh pandemic strain of O1 El Tor and O139 serogroup isolates and is absent in other non-pathogenic strains (Dziejman *et al*., 2002; Faruque and Mekalanos, 2003). In 2002, VSP-2 was initially identified as a 7.5 kb region consisting of 8 ORFs which includes locus *vc0490* to *vc0497* (Dziejman *et al*., 2002). In 2004, however, it was discovered that the island is much bigger than previously thought, about 27 kb and including the locus *vc0490* to *vc0516* (O'Shea *et al*., 2004a). VSP-2 consists of 24 functional ORFs and 3 non functional ORFs (Figure 1.10 and Table 1.4). The non-functional ORFs are VC0499, VC0500 and VC0501 which contain an authentic frame shift that is not the result of a sequencing artifact (Heidelberg *et al*., 2000). The overall GC content of VSP-2 is 43% whereas the overall GC contents of the *V. cholerae* genome is 47%. This difference reflects the typical characteristics of pathogenicity islands. VSP-2 ORFs also encodes

homologues of an RNase H1 protein (VC0498), a type IV pilus (VC0502), a lysostaphin peptidase (VC0503), a DNA repair protein (VC0510), two transcriptional regulators (VC0497 and VC0513), two methyl-accepting chemotaxis proteins (VC0512 and VC0514), a P4-like integrase (VC0516) adjacent to the tRNA-methionine locus (*vc0517.1*) and 15 hypothetical proteins (VC0490-VC0496, VC0504-VC0509, VC0511 and VC0515) (O'Shea *et al.*, 2004a). The pathogenicity island VSP-II shows homology to the 43.4 kb genomic island-I (VVI-I) of *Vibrio vulnificus*, a deadly human pathogen. *V. cholerae* ORFs VC0493–VC0498, VC0504–VC0510 and VC0516 are homologous to *Vibrio vulnificus* strain YJ016 ORFs VV0510–VV0516, VV0518–VV0525 and VV0560, respectively. The amino acid sequence identities of some ORFs in homologous regions of both islands are greater than 90% (O'Shea *et al.*, 2004a).

VSP-1 and VSP-2 were acquired by horizontal gene transfer, but the exact mechanism by which the current pandemic strains containing those chromosomal island evolved is still to be discovered. These islands may have been integrated into pre-seventh pandemic strain of *V. cholerae* El Tor O1 which already held TCP and CTXΦ. Another possibility could be the acquisition of VSP islands followed by gaining TCP and CTXΦ in non-pathogenic El Tor O1 strain (Figure 1.4). However, the success of the current pandemic strain over the last 48 years reflects its survival fitness in the host and the environment, as even after all these years it is still devastating and spreading faster than before. These islands could also be responsible for efficient infection in humans by allowing resistance to stomach acids, prolonged intestinal colonization etc (Dziejman *et al*., 2002).

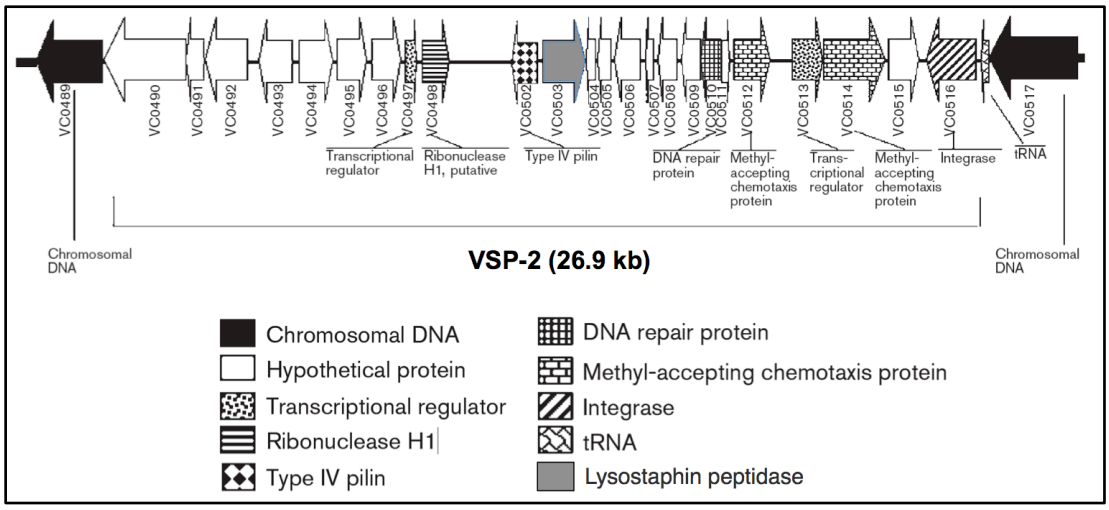


Figure 1.10 Schematic representation of VSP-2 (O'Shea *et al.*, 2004a).

Table 1.4 ORFs within VSP-2 (VC0490-VC0516)

ORF	Length (aa)	Homologous protein*
VC0490	653	Hypothetical protein
VC0491	178	Hypothetical protein
VC0492	388	Hypothetical protein
VC0493	291	Hypothetical protein
VC0494	214	Hypothetical protein
VC0495	224	Hypothetical protein
VC0496	199	Hypothetical protein
VC0497	66	Transcriptional regulator
VC0498	146	Ribonuclease H
VC0499	276	No significant match
VC0500	112	No significant match
VC0501	302	No significant match
VC0502	174	Type IV pilin, putative
VC0503	426	Lysostaphin peptidase
VC0504	75	Hypothetical protein
VC0505	122	Hypothetical protein
VC0506	244	Hypothetical protein
VC0507	58	Hypothetical protein
VC0508	147	Hypothetical protein
VC0509	147	Hypothetical protein
VC0510	157	DNA repair protein RadC-related protein
VC0511	39	Hypothetical protein
VC0512	529	Methyl-accepting chemotaxis protein
VC0513	271	Arac/Xyls family transcriptional regulator
VC0514	626	Methyl-accepting chemotaxis protein
VC0515	410	Hypothetical protein
VC0516	413	Phage integrase

*Homology based on BLASTP database analysis.

1.7.5 *Vibrio* pathogenicity island-2 (VPI-2)

VPI-2 is a 57.3 kb region which consists of 52 ORFs (VC1758 to VC1809) present in all pathogenic O1 and O139 serogroup isolates, but lacking in non-O1 and non-O139 non-pathogenic isolates (Figure 1.12), (Jermyn and Boyd, 2002). VPI-2 is 57 kb in length and the GC contents is 42% compared to 47% for the entire *V. cholerae* genome. Among 52 ORFs (Figure 1.11 and Table 1.5), 29 ORFs are homologous to genes of known function including hypothetical proteins, 13 ORFs are homologous to bacteriophage genes, and 10 ORFs do not show any significant match in the database. VPI-2 encodes an integrase (VC1758), which is similar to the integrase found in VPI-1 (Karaolis *et al*., 1998). In VPI-2, a different gene cluster was also observed, such as type 1 restriction modification system (VC1764-VC1769), which shows similarity to another restriction modification system found in *Xylella fastidiosa* (Simpson *et al*., 2000). Another region, named *nan-nag* region (VC1773-VC1784) next to the type 1, shows sequence similarity to another gene cluster present in the *Haemophilus influenzae* genome which encodes enzymes that are involved in utilization of amino sugars (Fleischmann *et al*., 1995; Jermyn and Boyd, 2002).

The *nanH* gene encodes neuraminidase (VC1784) which can act on higher gangliosides and convert them to GM1 gangliosides that results in the release of sialic acid. This sialic acid could be utilized by the proteins encoded by the *nan-nag* region and may provide alternative nutrient sources for pathogenic *V. cholerae* strains (Galen *et al*., 1992; Jermyn and Boyd, 2002; Vimr *et al*., 2004). VPI-2 has another phage like region (VC1791-VC1799) upstream of

two consecutive IS (insertion sequence) like elements (VC1789-VC1790). VPI-2 has an unstable region between VC1760 (helicase) and VC1789 (IS*911*-like element), which includes the restriction modification system and the *nan-nag* regions. This region is deleted from 13 of the 14 *V. cholerae* O139 serogroup strains examined (Figure 1.12) (Jermyn and Boyd, 2002).

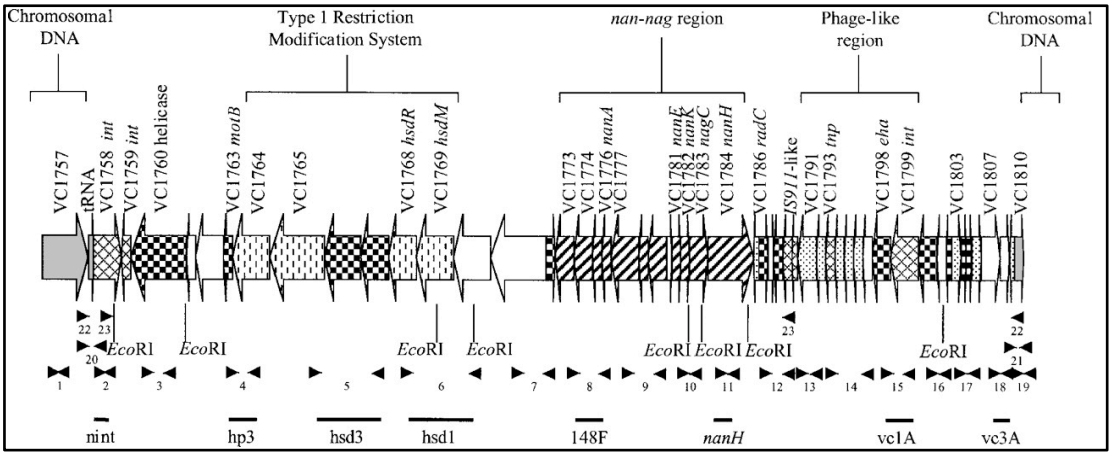


Figure 1.11 Schematic representation of the 57.3 kb VPI-2 (Jermyn and Boyd, 2002)

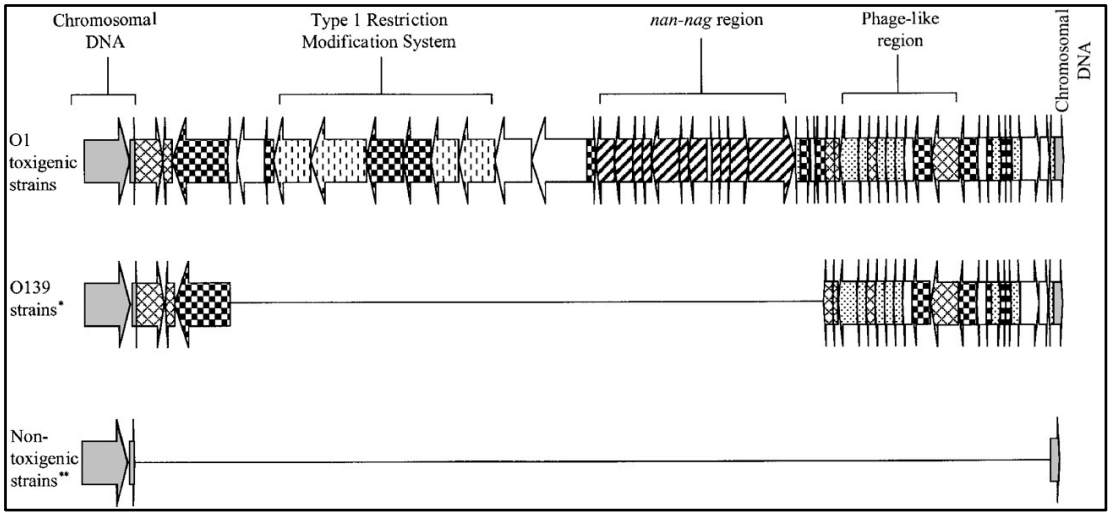


Figure 1.12 Structural organization of VPI-2 in *V. cholerae*. * A full VPI-2 could also be present in *V. cholerae* O139 strain. ** includes both *V. cholerae* O1 and non-O1/non-O139 strain (Jermyn and Boyd, 2002).

Table 1.5 ORFs within VPI-2 (Jermyn and Boyd, 2002)

ORF	Length (aa)	Homologous protein (organism)*
VC1758	411	Integrase (*V. cholerae*)
VC1759	159	Integrase (*V. cholerae*)
VC1760	940	Helicase (*Bacillus subtilis*)
VC1761	202	No significant match
VC1762	483	No significant match
VC1763	244	Chemotaxis protein MotB (*Aquifex aeolicus*)
VC1764	706	Methyl-accepting subunit (*Deinococcus radiodurans*)
VC1765	1022	Restriction enzyme helicase subunit (*Xylella fastidiosa*)
VC1766	663	Hypothetical protein (*P. multocida*)
VC1767	470	Hypothetical protein NMA2230 (*Neisseria meningitidis*)
VC1768	462	Type I restriction enzyme (*Xylella fastidiosa*)
VC1769	793	Type I restriction DNA methylase (*Xylella fastidiosa*)
VC1770	687	No significant match
VC1771	1220	No significant match
VC1772	286	Hypothetical protein NMA1157 (*Neisseria meningitidis*)
VC1773	356	Hypothetical protein HI0148 (*Haemophilus influenzae*)
VC1774	384	Hypothetical protein HI0148 (*H. influenzae*)
VC1775	278	Hypothetical protein HI0143 (*H. influenzae*)
VC1776	298	*N*-Acetylneuraminidase lyase HI0142 (*H. influenzae*)
VC1777	427	C4-Dicarboxylase transporter HI0147 (*H. influenzae*)
VC1778	173	C4-Dicarboxylase transporter HI0147 (*H. influenzae*)
VC1779	321	C4-Dicarboxylase binding protein HI0146 (*H. influenzae*)
VC1780	97	No significant match
VC1781	240	*N*-Acetylmannosamine-6-phosphate 2-epimerase (*H. influenzae*)
VC1782	287	*N*-Acetylmannosamine amide kinase (*H. influenzae*)
VC1783	378	*N*-Acetylamide glucosamide-6-phosphate deacetylase (*H. influenzae*)
VC1784	807	Neuraminidase (*Bacteroides fragilis*)
VC1785	68	DNA binding protein (P4 bacteriophage)
VC1786	158	RadC DNA repair (*Xylella fastidiosa*)
VC1787	45	No significant match
VC1788	231	Hypothetical protein chromosome II (*V. cholerae*)
VC1789	290	IS*911*
VC1790	114	IS*911*
VC1791	346	Mu-like gp42 (Mu prophage)
VC1792	119	Mu-like gp41 (Mu prophage)
VC1793	125	Transposase (*Caenorhabditis elegans*)
VC1794	192	gp12 protein (PSA bacteriophage)
VC1795	106	MOR protein (Mu phage)
VC1796	124	MOR protein (Mu phage)
VC1797	153	No significant match
VC1798	383	Eha protein (*Salmonella typhi*)
VC1799	585	Integrase (*Salmonella typhimurium* LT2)
VC1800	323	Plasmid replication protein C (*Clostridium butyricum*)
VC1801	120	No significant match
VC1802	78	Transcriptional regulator (*Salmonella typhi*)
VC1803	153	Repressor protein (CP-933O prophage)
VC1804	104	Hypothetical protein VC0509 (*V. cholerae*)
VC1805	148	Hypothetical protein VC0508 (*V. cholerae*)
VC1806	328	Replicase (Fr bacteriophage)
VC1807	214	No significant match
VC1808	281	No significant match
VC1809	76	Vis protein (P4 bacteriophage)

*Homology based on BLASTP database analysis.

1.7.6 Other pathogenicity factors

1.7.6.1 Integron islands

Integrons are genetic elements which have sophisticated gene capture systems which can, by site-specific recombination, recognize and capture mobile gene cassettes. Typically an integron contains an *integrase* gene cassette, captured gene cassettes and an integration site for the gene cassettes (Fluit and Schmitz, 2004; Hall and Collis, 1995). In *V. cholerae*, the integron island is a 125.3 kb gene cluster which is part of the gene capture and expression system. It is possible that integron Islands in *V. cholerae* express capture genes to enhance pathogenicity, in response to stress and the resistance against antimicrobial compounds (Faruque and Mekalanos, 2003).

1.7.6.2 RTX gene cluster

The RTX gene cluster (VC1447-VC1451) is a 10 kb region which encodes cytotoxin RtxA, an acyltransferase RtxC and an associated ABC transporter system RtxB/D. This gene cluster is physically related to the CTXΦ and is responsible for the cytotoxic activity in *V. cholerae* O1 El Tor and O139 strain. the RTX gene cluster may contribute to the residual adverse effects of certain live attenuated cholera vaccines (Lin *et al.*, 1999; O'Shea *et al.*, 2004b).

1.7.6.3 Type IV-A pilus gene cluster

The *V. cholerae* Type IV-A pilus gene cluster is 5.4 kb in size, and consists of five ORFs named PilA, PilB, PilC, PilD and YacE. This gene cluster also shows similarity to other type IV-A pilus gene clusters present in different

organisms. The *pilD* gene product is essential for the secretion of CT and is also important for the production of toxin-coregulated pili which is an important colonization factor in the host (Fullner and Mekalanos, 1999).

1.8 Role of iron in pathogenicity

In nearly all living organisms, iron is an essential element for survival as it is required for cellular metabolism, DNA metabolism, response to oxidative stress, the tricarboxylic acid (TCA) cycle, toxicity and pathogenesis. Regulation, transport and maintenance of iron in cells are vital (Andrews *et al*., 2003; Wandersman and Delepelaire, 2004). At biological pH, ferric iron is poorly soluble (10^{-18} M), and is not easily available. In mammalian hosts, free iron is maintained by the action of iron binding proteins such as the extracellular transferrin and lactoferrin; and the intracellular ferritin, hemosiderin and heme. Transferrin, a human iron acquisition protein, contains 30% saturated iron and exhibits high association with ferric iron (Fe^{3+}). Therefore, the free iron concentration in plasma is very low, at around 10^{-18} M (Sritharan, 2006). Although iron is important for microbes, a high concentration of iron in the cell is lethal for them as it leads to the production of reactive hydroxyl radicals via the Haber-Weiss and Fenton reactions (Haber and Weiss, 1932) .

A successful pathogen can obtain iron from the host cell by synthesized siderophores which are low molecular weight Fe^{3+}-specific ligand which are able to chelate iron from host iron-binding proteins. Alternatively, pathogens can uptake iron directly by producing receptors specific to host iron-containing molecules such as heme, hemoglobin, transferrin and lactoferrin (Braun *et al*.,

1998). *V. cholerae* has multiple systems for iron acquisition from the human host and environmental niches including siderophore synthesis and transport, receptors for heme and hemoglobin, TonB system, ABC transporters etc (Wyckoff *et al*., 2007). It is clear that the host contains very low levels of free iron and the pathogen requires a higher concentration of iron in order to multiply in host environments. In iron limiting situations in the host, the bacteria triggers a signal that activates virulence genes to kill host cells for nutrition including iron. Interestingly, huge numbers of virulence genes are regulated by ferric uptake regulator (Fur). In *Pseudomonas aeruginosa*, for example, expression of alternative sigma factor is regulated by Fur, which leads to the transcription of some virulence genes such as endoprotease (*PrpL*) and exotoxin A (Leoni *et al*., 1996; Ochsner *et al*., 1996; Wilderman *et al*., 2001).

1.9 Ferric uptake regulator of *V. cholerae*

The ferric uptake regulator of *V. cholerae* (*Vc*Fur) is a dimeric metalloprotein of monomeric size ~17 kDa with an N-terminal domain involved in DNA recognition domain and a C-terminal domain involved in dimerization (Coy and Neilands, 1991; Stojiljkovic and Hantke, 1995). Regulation of iron acquisition genes in *V. cholerae* is mediated primarily through the *Vc*Fur protein. The N-terminal domain contains a winged helix motif which helps to bind a 19-bp inverted repeat sequence known as a Fur-box (Escolar *et al*., 1999). As a classical repressor, the main function of Fur is iron homeostasis. It also helps organisms to survive acidic environments or stress (Hall and Foster, 1996). When iron levels are high, Fur binds to the Fur-box and

prevents access of RNA polymerase to the downstream iron acquisition genes whereas in the presence of low iron, derepression of the iron uptake genes occurs (Figure 1.13) (Andrews *et al*., 2003). A variety of *Vc*Fur homologues have been discovered in a wide range of Gram-negative human pathogens such as *Salmonella typhimurium* (Ernst *et al*., 1978), *Yersinia pestis* (Staggs and Perry, 1991), *Helicobacter pylori* (Bereswill *et al*., 1998), *Pseudomonas aeruginosa* (Prince *et al*., 1993), *Bordetella pertussis* (Brickman and Armstrong, 1995), *Neisseria gonorrhoeae* (Berish *et al*., 1993) and *Haemophilus ducreyi* (Carson *et al*., 1996). In Gram-positive bacteria such as *Bacillus subtilis* (Bsat *et al*., 1998) and *Staphylococcus epidermidis* (Heidrich *et al*., 1996), Fur-like proteins have been characterized.

Iron and Fur regulation in *V. cholerae* has been extensively studied and it has been shown that *Vc*Fur has both negative and positive regulatory functions and is able to repress almost all of the iron acquisition genes during iron-replete conditions (Mey *et al*., 2005). In addition, genes encoding the toxin-coregulated pilus (TCP) and the large integron are positively regulated by *Vc*Fur. As an activator, *Neisseria meningitidis* Fur protein can bind upstream of the target gene promoters that subsequently initiate positive regulation *in vivo* (Delany *et al*., 2004).

A strain of *V. cholerae* with the *fur* gene knocked out shows very weak autoagglutination that is possibly due to defective expression or assembly of the TCP, an important virulence factor. In infant mouse studies, it was found that the *fur* knockout strain colonized poorly compared to a wild type strain of *V. cholerae*. In practice, *Vc*Fur can control the expression of more than 100

genes in *V. cholerae* in an iron-dependent or iron-independent manner (Mey *et al*., 2005).

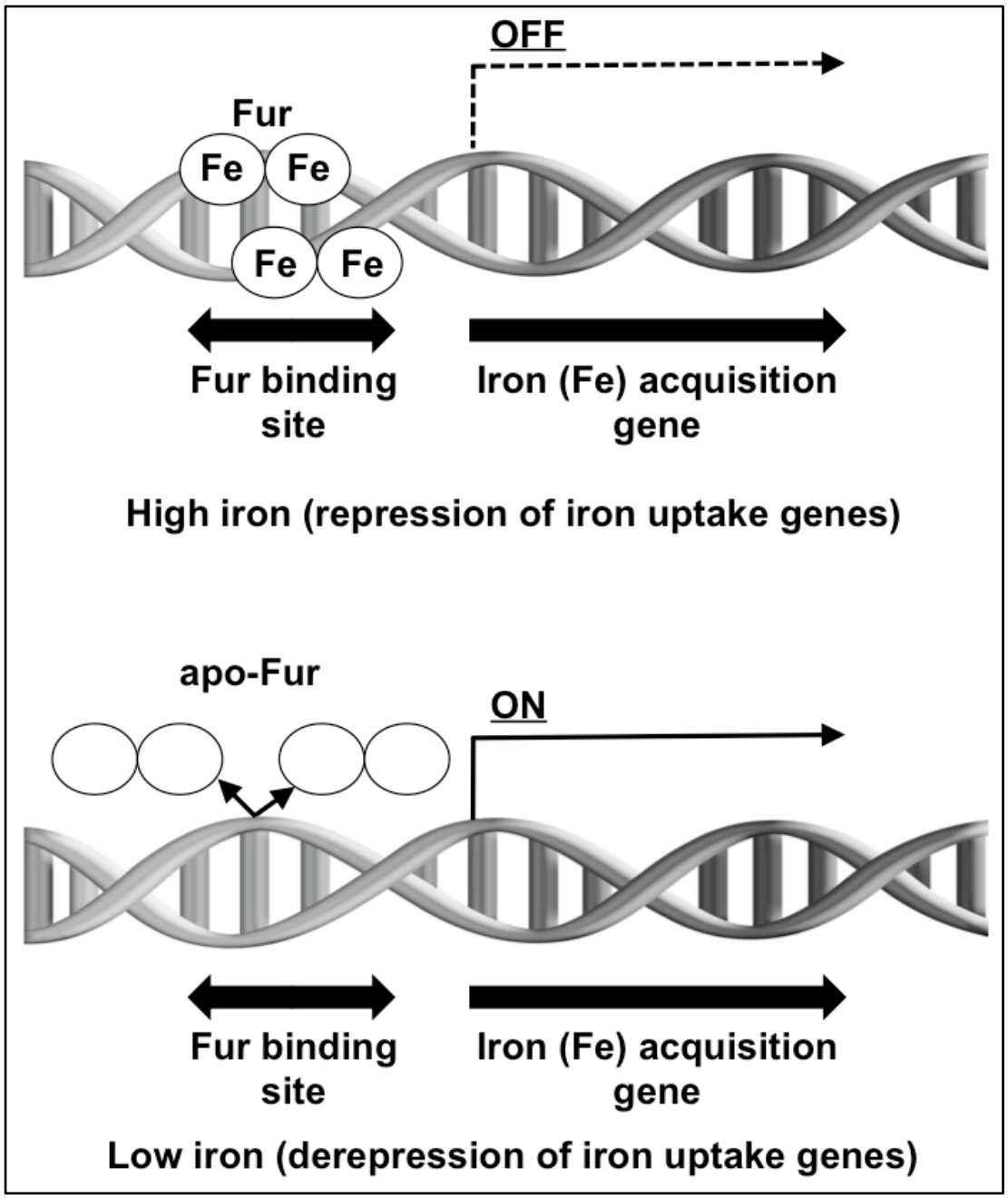


Figure 1.13 Schematic representation of Fur-mediated gene repression (Andrews *et al*., 2003).

1.10 Antibiotic resistance of *V. cholerae*

Antibiotic resistant pathogenic strains of *V. cholerae* are becoming a major challenge. Uptake of foreign genetic material by horizontal gene transfer is the primary cause of evolution of multi-drug resistant (MDR) strains of *V. cholerae* (Sack *et al*., 2001). Widespread antibiotic resistance to *V. cholerae* was reported after 1977 in Tanzania and Bangladesh (Glass *et al*., 1980; Mhalu *et al*., 1979). In the past two decades, reports generated from countries where cholera is endemic have shown *V. cholera* to be resistant to several commonly used antibiotics including tetracycline, ampicillin, kanamycin, streptomycin, sulphonamides, trimethoprim and gentamicin. Unlike *Shigella* spp, which show rising resistance to antibiotics, *V. cholerae* O1 and O139 show spatial and temporal fluctuations, with periods of fluctuating resistance or sensitivity indicating the possible misuse of antibiotics during treatment programs (Sack *et al*., 2001).

1.11 Prevention and treatment of cholera

The primary steps that could be useful in reducing the transmission of cholera includes the provision of a safe water supply, improvement of sanitation and hygiene, ensuring that food is fully cooked and health education. Unfortunately, in the developing world, where cholera is a particular problem, improvements in infrastructure for water and sanitation facilities has a long way to go (Sack *et al*., 2004). Replacing lost fluid and ions in order to prevent further dehydration is used as the major form of palliative care for patients infected with *V. cholerae*. In most cases, oral rehydration solution (ORS) is given to the patient. The therapy of the patient is guided by the degree of

dehydration. For severe cases, emergency intravenous polyelectrolyte solution for rehydration followed by ORS is required for maintaining hydration inside the body of the patient. In cases of milder dehydration, ORS is used for both rehydration and maintenance. The principles of rehydration therapy include a rapid replacement of fluid deficits; correction of the metabolic acidosis; correction of potassium deficiency; and replacement of continuing fluid losses. The case-fatality rate for severe cholera is about 50% without treatment (Sack *et al*., 2004). A course of antibiotics is usually given to patients with severe cholera, but the effectiveness of antimicrobial drugs is limited by antimicrobial resistance resulting in potential increased risk of *V. cholerae* (Sack *et al*., 2001).

In addition, soon after the discovery of *V. cholerae* as a cause of cholera, an anti-cholera vaccine was developed. From this point, the whole cell, injectable vaccine was used as a preventative method for disease control. However, the vaccine was shown to have only a very limited effectiveness and was withdrawn from use (Sack *et al*., 2004). Currently, different research groups are trying to develop a new vaccine for cholera which includes a few selected antigens derived from the pathogen, but no potential breakthrough has been made so far. It would be of great socioeconomic benefit to develop a safe vaccination agent against *V. cholerae* in order to protect from cholera and to alleviate the threat of the forthcoming eighth pandemic.

1.12 Overall project aims and background

The progress in the development of novel agents to fight against cholera has been slowed down during the past decade partly because of the success of ORS in reducing mortality and partly because of limited commercial interest in drugs for the developing world. The main objective of my study is to assist in the discovery of an effective drug against cholera and find a novel vaccination target using bioinformatics, molecular biology and structural proteomics to investigate the specific protein targets derived from pathogenic *V. cholerae*. It is hoped that the structural insight of these pathogenicity factors will improve our understanding of how these contribute to disease.

The first part of my PhD project was working on proteins of interest in VPI-2 which encodes several gene clusters and is likely to be important in pathogenesis. My particular interest is in determining the structures and ascribing functions to conserved hypothetical proteins and their homologues and paralogues in the reduced 20 kb VPI-2 region (VC1789 to VC1809), present in all pathogenic strain of *V. cholerae* O1 El Tor and O139 and missing in non-pathogenic strain (Jermyn and Boyd, 2002). One such conserved hypothetical protein is VC1805 and its four copies of homologous/ paralogous ORFs present in VPI-2 (VC1804 and VC1805) and VSP-2 (VC0508 and VC0509). In addition to the structure determination of these proteins, I want to answer some more questions: What is their function? How do they impart pathogenicity? Do they interact with host proteins? Are they potential therapeutic targets?

The second part of my PhD project was working on *Vc*Fur which is a well known metal-dependent DNA-binding protein that acts as both a repressor and activator of numerous genes involved in maintaining iron homeostasis in bacteria and virulence expression. Fur can be considered as one of the most promising novel drug targets against bacteria because in eukaryotic organisms, no homologues of Fur have been found so far. The *fur* deletion mutant shows reduced virulence in some pathogenic organisms like *V. cholerae* (Mey *et al*., 2005) and *Pseudomonas aeruginosa* (Vasil and Ochsner, 1999). The Fur protein from *Mycobacterium tuberculosis* has also been identified as a potential novel drug target (Lucarelli *et al*., 2008). Despite being a highly promising drug target, progress has not been advanced since the first discovery of the Fur family more than twenty years ago. This is because of the lack of studies of the functional diversity of this protein family in a variety of organisms and the lack of any structural details about Fur-DNA complex regulation. This project was aimed to solve the structures of *Vc*Fur and *Vc*Fur-DNA complex and to investigate its functional characterization including metal interaction which could provide important clues for future drug design.

Chapter 2

Materials and Methods

2. Materials and methods

This chapter describes the general methods used in this project.

2.1 Strains and preparation of genomic DNA

The *Vibrio cholerae* strain used for this study is *V. cholerae* O1 El Tor strain N16961 and was kindly provided by Dr. E. Fidelma Boyd (University of Delaware). All strains were grown in Luria–Bertani (LB) broth and stored as frozen stocks in LB broth with 20% (v/v) glycerol. Genomic DNA (gDNA) was prepared from this strain by using the Easy-DNA kit (Invitrogen) according to the manufacturers' instructions. The following *E. coli* strains were used as cloning and expression hosts:

Strain	Organism
BL21 (DE3)	*Escherichia coli*
Rosetta (DE3)	*E. coli*
DH5α	*E. coli*
DH5α λ*pir*	*E. coli*
K12 strain TB1	*E. coli*

2.2 Target proteins

A total of five proteins were targeted for crystallization, of which four are from pathogenicity islands and the fifth is the ferric uptake regulator (Table 2.1).

Table 2.1 Target proteins in *Vibrio cholerae*

ORF	Length (aa)	Function
VC1805	148	Hypothetical protein in VPI-2
VC1804	104	Hypothetical protein in VPI-2
VC0508	147	Hypothetical protein in VSP-2
VC0509	147	Hypothetical protein in VSP-2
VC2106	150	Transcriptional regulator, iron homeostasis

2.3 Vectors

2.3.1 pEHISTEV vector

The pEHISTEV vector, is an engineered variant of pET30 with an N-terminal 6X His-tag that is cleavable using tobacco etch virus (TEV) protease and leaves two additional residues (Gly-Ala) at the N-terminus (Liu and Naismith, 2009).

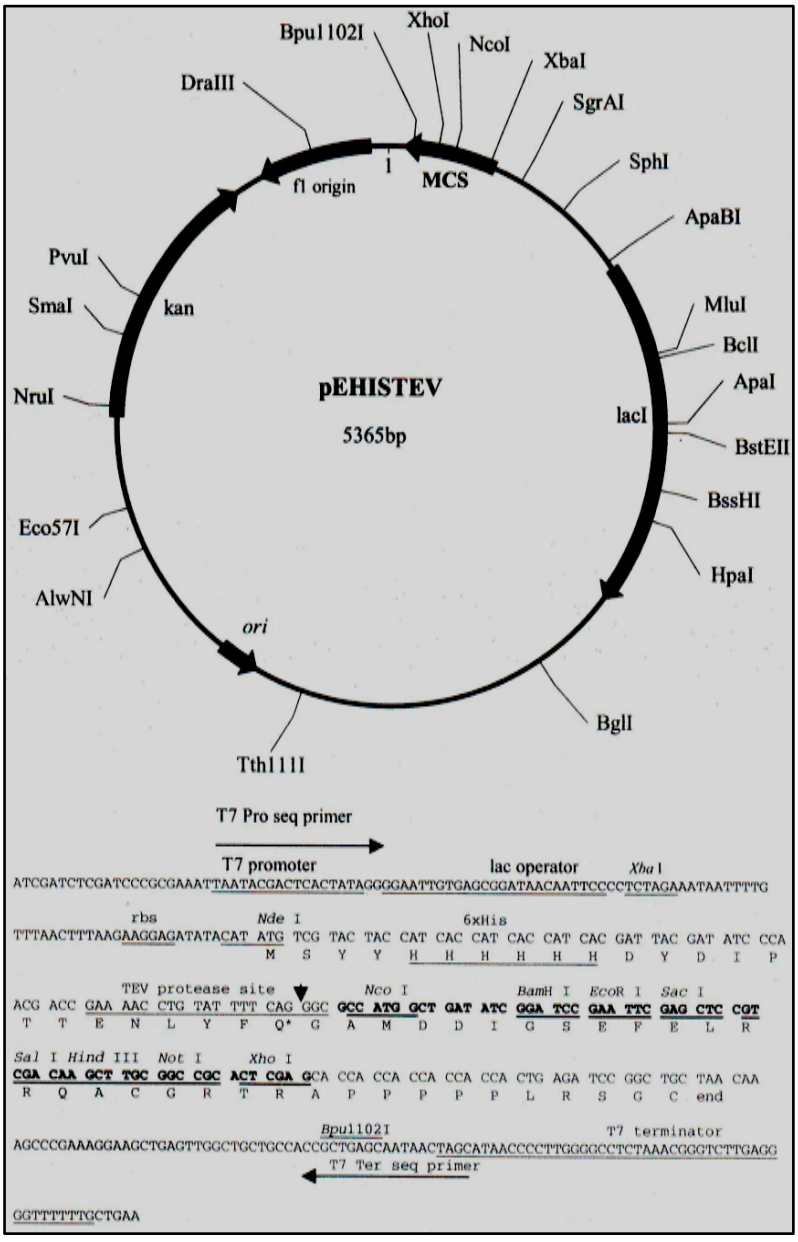


Figure 2.1 pEHISTEV vector

2.3.2 pLou3 vector

pLou3 is an engineered variant of NEB vector pMAL-c2X with an N-terminal 6X His-tag maltose binding protein (HMBP) that is cleavable using Tobacco Etch Virus (TEV) protease, and a lacZα complementation gene present after the MCS (multiple cloning sites) (Dr Louise Major, personal communication).

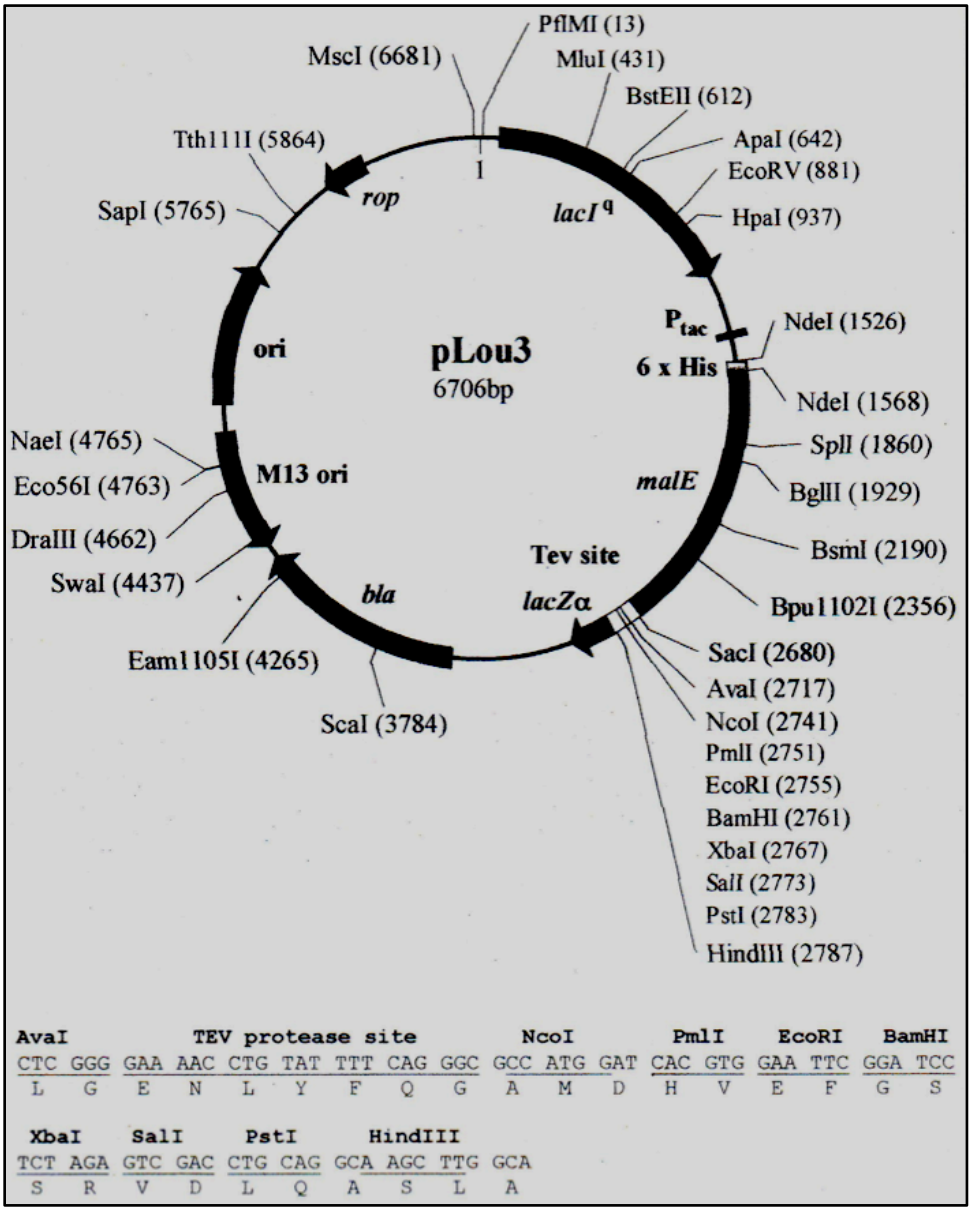


Figure 2.2 pLou3 vector

2.3.4 pCR2.1-TOPO vector

The pCR2.1-TOPO vector was purchased from Invitrogen, USA (www.invitogen.com). The vector is supplied linearized with single 3´-thymidine (T) overhangs for TA Cloning.

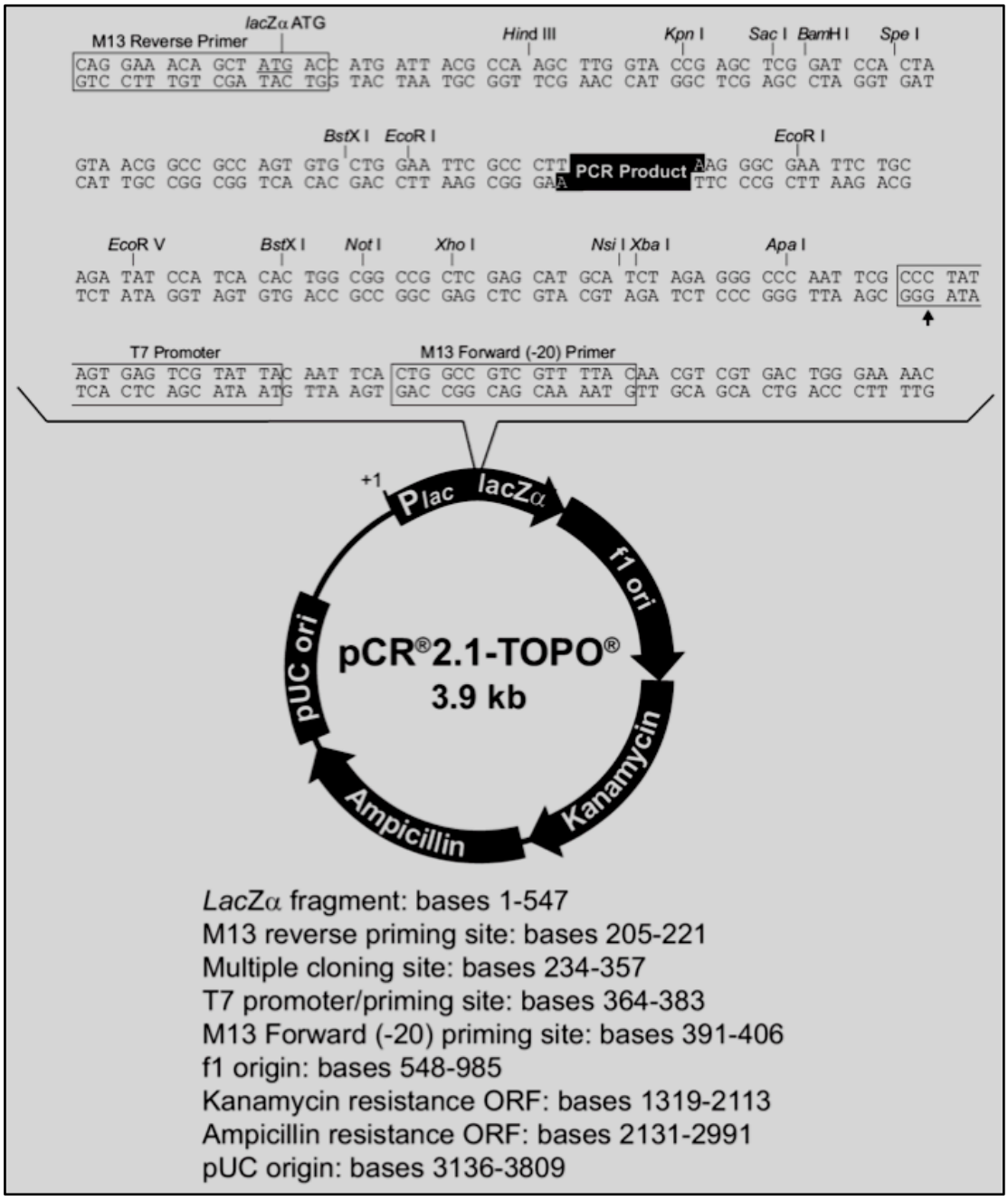


Figure 2.3 pCR2.1-TOPO cloning vector

2.3.5 pDS132 vector

pDS132 is an engineered variant of suicide vector pCVD442 (Donnenberg and Kaper, 1991) which contains *R6K ori,* the replication origin of plasmid R6K, the *mob RP4*, a plasmid mobilisation regions, *cat*, chloramphenicol resistance and *sacB* gene, a suicide marker .

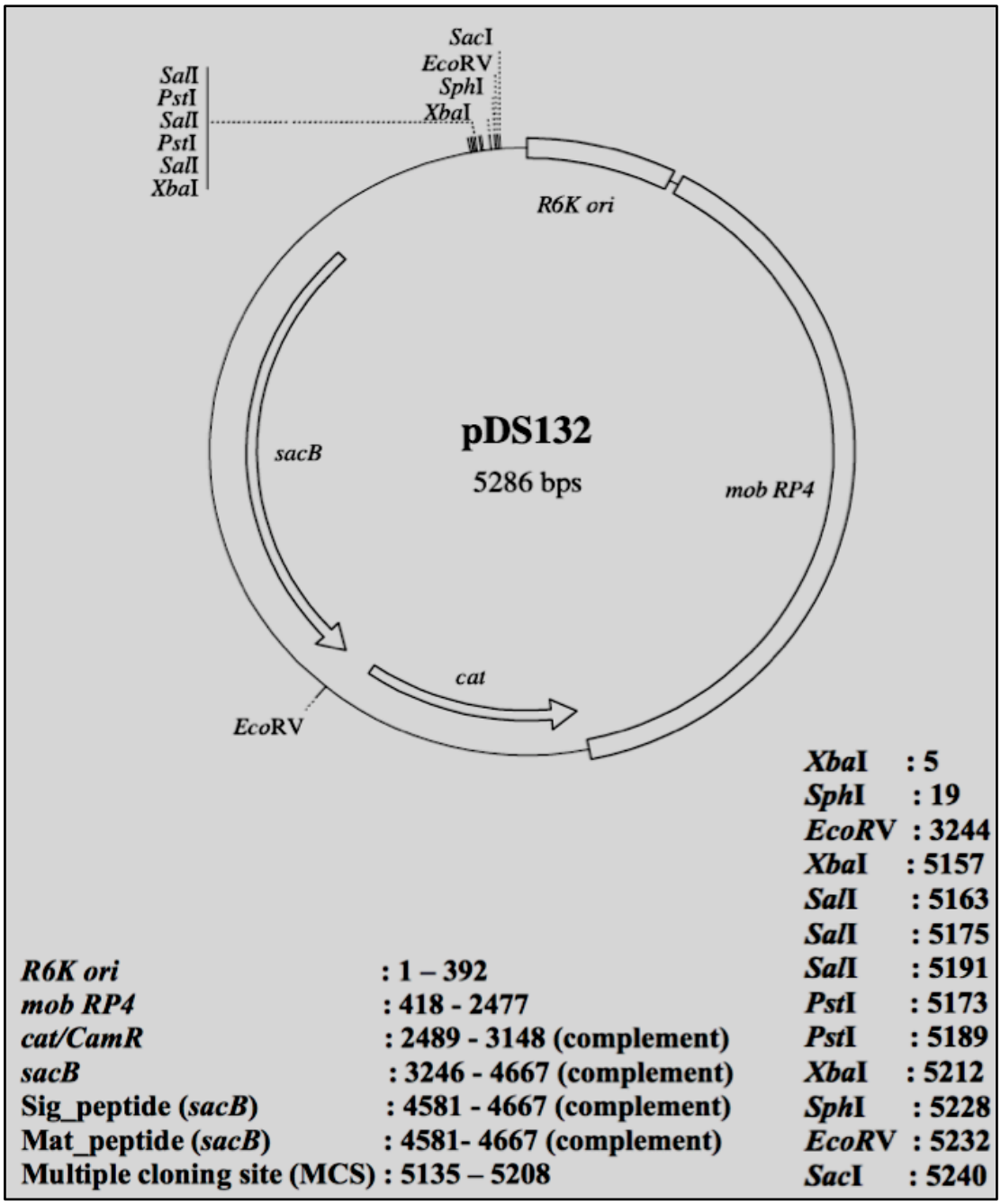


Figure 2.4 Suicide vector pADS132

2.4 Polymerase chain reaction (PCR)

The target genes were amplified by polymerase chain reaction (PCR) from genomic DNA of *V. cholerae* O1 El Tor strain N16961. Forward and reverse primers were designed corresponding to the 5′ and complementary 3′ ends of target genes with specific recognition sites such as *Nco*I, *EcoR*I, *Sal*I etc. PCRs were performed in a volume of 50 µL containing 1 µL of forward primer (50 µM), 1 µL of reverse primer (50 µM), 5 µL PCR buffer (10X), 100 ng DNA which was heated at 70 °C for 3 min, 1 µL of 10 mM deoxynucleotide triphosphates (dNTPs), 1 µL pfu DNA polymerase and water up to 50 µL. The amplification conditions were: pre-incubation at 94 °C for 5 min, followed by 30 cycles of 94 °C for 1 min, X °C for 1 min (X °C = 5 °C below primers' Tm) and 72 °C for Y min (Y min = 1 min per kb of nucleotide). Before storing at 4 °C samples were heated at 72 °C for 10 min for extension.

2.5 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out to visualize all the amplified PCR products for subsequent purification. For a 0.8% gel (which is normally used), 0.8 g of agarose was dissolved in 100 mL 1X Tris-Acetate-EDTA (TAE) buffer. Solutions were cooled slightly before addition of 5 µL of 10 mg/mL ethidium bromide and then poured into a mould. DNA samples were prepared by adding 6X gel loading solution and loaded into wells. DNA ladder (Promega), 1 kb is used depending on the product size. Gels were run at 120 V for 45 min.

2.6 Restriction digestion of PCR products and vector

The product band of target genes were excised from the gel and purified by QIAquick Gel Extraction Kit (QIAGEN) that results in 50 µL of eluted DNA. To open the specific recognition sites in the PCR products and vector (such as pEHISTEV), 200 ng to 1 µg of PCR products/vector were digested with 3 µL of 10X BSA (freshly prepared, can be stored in -20°C), 3 µL restriction digestion buffer (10 X), 1 to 2.5 µL of each corresponding restriction enzyme (Promega) and water up to 30 µL. The samples were mixed gently by using micropipette, spun down for 3 to 5 s, and incubated at 37 °C for 2-3 h followed by 65 to 80 °C for 10 min to inactivate enzymes. Alternatively, for rapid restriction digestion, a 20 µL reaction mixture was prepared in a PCR tube with 200 ng to 1 µg of PCR products/vector, 10X FastDigest buffer (Fermentas), 1 µL of each corresponding enzyme (Fermentas) and nuclease-free water up to 20 µL. The samples were mixed gently by using a micropipette, spun down for 3 to 5 s, and incubated for 5-15 min followed by 65 to 80 °C for 10 min to inactivate enzymes. To check digestion efficiency, 5 µL of restriction digestion product was run on an agarose gel electrophoresis.

2.7 Ligation

Different ratios of digested vector and insert have been used to optimize the ligation reaction. The use of 100 ng of DNA and a molar ratio of insert to vector of 3:1 is standard. The following formula is used to calculate the exact insert quantity to be needed for the ligation reaction.

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \frac{\text{insert}}{\text{Vector}} \text{ (molar ratio)} = \text{ng of insert}$$

Ligation of digested vector and insert was carried out in a microcentrifuge tube using 2 µL 10 X ligation Buffer (Promega), 1 µL T4 DNA ligase (Promega), digested vector (1 to 16 µL, depending on concentration) and corresponding insert (1 to 16 µL, depending on concentration) and water up to 20 µL. The samples were mixed gently using a micropipette, spun down for 3 to 5 s then incubated at 22 °C for 1 h or 16 °C overnight.

For rapid ligation, a 20 µL reaction mixture was prepared containing 4 µL 5X ligation Buffer (Fermentas), 1 µL T4 DNA Ligase (Fermentas), digested vector (1 to 14 µL, depending on concentration), corresponding insert (1 to 14 µL, depending on concentration) and nuclease-free water up to 20 µL. The samples were mixed gently using a micropipette, spun down for 3 to 5 s, and incubated for 5-15 min at 22 °C.

2.8 Transformation

5 µL of ligated DNA sample (5 to 50 ng) was transformed into 45 µL of competent cells such as *E. coli* DH5α, by cold shock (in ice 30-40 min), heat shock (in water bath 42 °C for 1 min) and cold shock (in ice 3-5 min). 950 µL of pre-warmed Luria-Bertani (LB) Media was added to the transformation reaction tube which was then placed horizontally in a shaking incubator (200 rpm) at 37 °C for 60-90 min. After incubation, the cells were spun down at 4000 rpm for 4 min and following removal of excess LB media, cells were

resuspended with the remaining LB supernatant. These transformed cells were plated on LB agar containing the appropriate antibiotic and incubated at 37 °C overnight. Alternatively, If the vector contained any antibiotic resistance gene other than kanamycin, a rapid transformation protocol was used in which 5 µL of ligated DNA sample (5 to 50 ng) was mixed with 45 µL of competent cells such as *E. coli* DH5α and incubated 5 min on ice and plated directly onto LB agar containing the appropriate antibiotic.

2.9 Blue/white colony screening

The substrate X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) is used to screen blue/white colonies on the plate. X-gal can be used by the bacterium if it is able to produce β-galactosidase. Cloning hosts such the *E. coli* DH5α strain do not contain the *lacZ*α gene which is one of the active parts of β-galactosidase, and can be complemented by vectors such as pLou3. In the presence of β-galactosidase, X-gal is cleaved into bromo-4-chloro-3-hydroxyindole followed by oxidization to 5,5'-dibromo-4,4'-dichloro-indigo, an insoluble blue product which gives characteristic blue coloured colonies on the culture plate. Some constructs of pLou3 cannot complement *lacZ*α gene if there is any disruption in the MCS due to the presence of our gene of interest. Transformed cells were spread on an LB plate containing 40 µL of 40 mg/mL X-gal, 40 µL of 100 mM IPTG and the appropriate antibiotic followed by incubation at 37 °C overnight. As X-gal is light sensitive, the stock was always wrapped with aluminium foil.

2.10 Confirmation of cloning

2.10.1 Colony PCR

Colonies containing recombinant vector DNA were identified using colony PCR screening. Several colonies were randomly picked by sterile toothpick and each resuspended in 10-15 µL water, then the wet toothpick was streaked on LB agar containing appropriate antibiotic and allowed to grow at 37 °C overnight. The remainder of each suspension was heated at 99 °C for 5 min before cell debris was removed by centrifugation. The cell free lysate was used in a PCR reaction containing 10 µL lysate, and 10 µL master mix which contained 0.2 µL 10 mM dNTPs, 0.2 µL 10X Taq DNA polymerase buffer, 0.2 µL DNA polymerase, 0.2 µL of 5 µM vector specific forward primer and 0.2 µL of 5 µM vector specific reverse primer and 9 µL of water. For the pEHISTEV vector, the primers used in colony PCR are T7 forward primer (5′-TAATACGACTCACTATAGGG- 3′) and T7 terminator (5′ -TATGCTAGTTATTGCTCAG- 3′). PCR reactions were carried out as described earlier (section 2.4). Colony PCR was analyzed by agarose gel electrophoresis.

2.10.2 Double digestion

As a further confirmation, double digestion of the cloned vector containing the gene of interest was carried out, using corresponding restriction enzymes according to the protocol described earlier (section 2.6). Then the sample was run on agarose gel to confirm the presence of the gene of interest.

2.10.3 DNA sequencing

The correct sequence of cloned constructs was confirmed by sequencing using vector specific primers. For example, T7 forward and reverse primers were used to sequence the pEHISTEV vector. 500 ng of DNA in 30 µL water was sent to the sequencing service at the School of Life Sciences, University of Dundee, Scotland (www.dnaseq.co.uk).

2.11 Purification of transformed vector DNA

A transformed colony containing the recombinant vector was inoculated in 10 mL LB containing the appropriate antibiotic followed by shaking incubation at 37 °C overnight. After incubation, cells were spun down at 3000 rpm for 10 min at 4 °C. The pellet was collected and vector DNA was purified by the QIAquick spin protocol (QIAGEN), which elutes vector DNA in 30 to 50 µL of water.

2.12 Transformation in expression host

3 µL of purified recombinant vector was transformed into 50 µL of BL21 (DE3) and/or Rosetta (DE3) and/or C43 strain of *E. coli* using the protocol described earlier (section 2.8) followed by plating out on LB agar containing the appropriate antibiotic and incubated at 37 °C overnight.

2.13 Protein expression trials and optimization of solubility

A transformed expression host colony was used to inoculate 10 mL LB containing the appropriate antibiotic followed by shaking incubation at 37 °C

overnight. After incubation, 200 µL culture was inoculated in 10 mL LB and 10 mL TPB (tryptose phosphate broth) containing the appropriate antibiotic, followed by shaking incubation at 37 °C for 2-3 h until an OD_{600} of 0.5-0.6 was reached. At this point, cells were exposed to various IPTG (isopropyl-beta-D-thiogalactopyranoside) concentrations and temperatures (Table 2.2) to induce the expression of cloned genes. 1 mL of culture of each sample was centrifuged and the pellet were lysed using 200 µL cell lysis buffer BugBuster (Novagen). The whole cell and soluble sample were analysed by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) to confirm soluble protein expression.

Table 2.2 Conditions used for the induction of expression of Targets

Expression (Shock)	**Expression Host (*E. coli*)**	**Incubation temperature**	**Incubation time**	**IPTG concentration**
Heat shock or Cold shock or No shock	Rosetta (DE3) or BL21 (DE3) or C43 (DE3)	37 °C or 25 °C or 16 °C	3 h or 24 h or 36 h	1.0 mM or 0.5 mM

2.14 SDS-PAGE

Pre-cast 4- 12% Nupage gels were set up according to manufacturers' instruction (Invitrogen). Rigs were filled with 800 mL 1 X MES buffer and 1 mL of antioxidant was added to the upper chamber. Loading buffer, 4X LDS was added to the protein samples followed by denaturation at 99 °C for 5 min and samples were loaded into wells alongside an appropriate unstained marker such as Mark12™ standard (Invitrogen). Electrophoresis was carried out at

200 volts (V) and a current of 70-115 mA for 35 min. On completion, gels were stained with Coomassie Blue to allow protein detection.

2.15 Mass spectrometry

Mass spectrometry is an important method for protein molecule identification. The protein, either in solution or as a band excised from an SDS-PAGE gel, was submitted to the mass spectrometry and proteomics facility in the University of St Andrews. Tryptic fragments of the protein were analyzed by MALDI-TOF MS (matrix-assisted laser desorption/ionization-time of flight mass spectrometry) and MS/MS (tandem mass spectrometry) using a 4800 MALDI-TOF/TOF analyzer (Applied Biosystems).

2.16 Bulk-scale purification of target protein

2.16.1 Nickel column purification

The recombinant protein expressed from the pEHISTEV/ pLou3 vector, contained an N-terminal 6X His-tag. The cell pellet was resuspended in an appropriate solubilizing buffer (such as 50 mM HEPES, 500 mM NaCl, 10 mM imidazole and 10% (v/v) glycerol), which is usually made up with high salt concentration, containing 5 µg/mL DNase I and EDTA-free protease inhibitor cocktail (Roche Applied Science) and sonicated. After centrifugation, the soluble fragment was syringe filtered through a 0.22 µm filter unit (Millipore) and the target protein was purified using a HisTrap™ HP column (Ni^{2+}-affinity column, GE Healthcare), according to the manufacturers' instructions. Purification of target proteins was also confirmed by SDS-PAGE.

2.16.2 6X His-tag cleavage

His-tag cleavage was done in dialysis buffer (appropriate solublizing buffer) in the presence of DTT (Dithiothreitol) and/or EDTA (ethylenediaminetetraacetic acid). TEV (Tobacco Etch Virus) protease was added into the dialysis tube containing His-tagged protein (1 µg protease for 1 mg at 25 °C overnight). The digested protein sample was analysed by SDS-PAGE to check the efficiency of TEV protease digestion.

To remove EDTA and/or DTT, the digested sample was dialysed against the same dialysis (for TEV cleavage) buffer without EDTA and/or DTT. Protein samples were syringe filtered and loaded into the nickel column as before and the flow-through was collected in fractions and analysed by SDS-PAGE. As the 6X His-tag is cleaved from the protein, it should bind to the nickel column and the proteins will be eluted in the flow-through. The corresponding fractions were pooled, dialysed and concentrated using an appropriate centrifugal concentrator (Sigma- Aldrich) to prepare the samples for gel filtration column chromatography for further purification.

2.16.3 Gel filtration chromatography

In the final stage of purification, up to 4.5 mL of sample was loaded onto a Hiprep™ 16/60 Sephacryl™ S-200 gel filtration column (GE Healthcare) according to the manufacturers' protocol. Samples from the observed peaks were analysed separately by SDS-PAGE.

2.17 Pre crystallization test

The PCT™ (Pre-Crystallization Test) is used to determine the appropriate protein concentration for crystallization screening. Sample concentration is a significant crystallization variable. If the samples are too concentrated they can result in amorphous precipitate, while too dilute samples can result in clear drops. Precipitate and clear drops are typical crystallization screen results for reagent conditions which do not promote crystallization and are part of every crystallization screen. However, by optimizing protein concentration for the screen, the number of clear and precipitated results can often be reduced, which in turn results in more efficient sample utilization while at the same time enhancing the chances for crystallization. The ideal concentration to crystallize the target protein can be determined by using PCT kit (Hampton research) according to the manufacturers' protocol.

2.18 Protein crystallization and optimization

Different commercially available crystallization screens were used to screen initial conditions of crystallization, such as Index (Hampton Research), JCSG+ (Qiagen) and Wizard I+II (Emerald Biosciences). Manually, this experiment was carried out using the sitting drop diffusion method in 96-well plates following standard procedures. Alternatively, a nano-drop crystallization robot (Cartesian HoneyBee, Genomic Solutions), as part of the Hamilton-Thermo Rhombix system, was used to set-up crystal plates. The main advantage of using a robot is the smaller amount of protein needed for plates (150 or 300 nl per drop). All the crystallization plates were kept at room temperature. The initial crystallization hits needed to be optimized to get better quality crystals.

Optimization is based on varying the concentration of salt, protein and precipitant in the screen.

2.19 X-ray data collection of crystals

Using 10-30% (v/v) glycerol in the crystallization buffer, the crystals of target proteins were cryoprotected. A suitable protein crystal was picked out from the drops with a crystal loop and transferred to 5 µl cryoprotectant drop. Then the crystal was re-looped and mounted onto X-ray at 100K in-house using a Rigaku/MSC MicroMax-007HF rotating anode equipped with a Saturn 944+ CCD detector at wavelength 1.54178 Å. For higher resolution data, crystals were taken to the European Synchrotron Radiation Facility (ESRF), France.

2.20 Heavy metal soaking of VC1805 crystal

In order to obtain heavy-metal derivatives, crystals were soaked at different metal solution concentrations (1 mM/ 5 mM/ 10 mM/ 15mM/ 20mM) for different time intervals (10 min/ 30min/ 16 h). Heavy metals such as platinum potassium cyanide, K_2PtCN_4; potassium aurocyanide, $KAu(CN)_2$; ytterbium sulphate, $Yb_2(SO_4)_3$; samarium acetate hydrate, $Sm(C_2H_3O_2)_3.H_2O$; mercury nitrate, $Hg(NO_3)_2$; p-chloromercuri benzene sulphonate (PCMBS), $C_6H_5O_3HgS$; were used for soaking. Next, the crystals were cryoprotected in preparation for data collection.

2.21 C1q binding assay

Using a 96-well microtitre plate (Maxisorb, NUNC, Denmark), wells were coated with 50 µL of a two-fold serial dilution of VC1805 in PBS (phosphate buffered saline), which was in range of 4 µg/mL to 0.125 µg/mL. For negative controls, wells were also coated with 50 µL of a two-fold and five-fold dilution of milk in PBS buffer, and with 50 µL of the archaeal DNA-binding protein hssb1 at 4 and 2 µg/mL, diluted in PBS buffer. Coating was done overnight at 4 °C. The protein Hsbb1 (which, like C1q is highly positively charged) and milk (which is a mixture of different proteins) were used as negative controls in this experiment. The wells were then washed four times with PBS + 0.1 % Tween 20 (Buffer A). The remaining protein binding sites on the wells were saturated with 360 µL of blocking buffer (10% milk in PBS) and incubated for 1 h at room temperature, then wells were washed four times with buffer A. 50 µL of blocking buffer containing 5 µg C1q was added to the wells and incubated at 37°C for 1 h followed by washing four times with buffer A. 100 µl biotinylated anti- human C1q IgG in blocking buffer which contained 5 µg of antibody were added to the wells and incubated at 37 °C for 1 h. After incubation, the wells were washed four times with Buffer A. Streptavidin HRP (horseradish peroxidase, KPL Inc.) was added to the wells and incubated at 37 °C for 1 h, followed by four washes with buffer A. 100 µL of TMB peroxide substrate (prepared as per manufacturer's instructions, KPL) was added to the wells and kept at room temperature. As a positive control for C1q binding, wells were coated with hen ovalbumin, and then reacted with rabbit-anti-hen ovalbumin antiserum, to form immune complexes in the wells. C1q was purified from Human plasma (Scipac Ltd.) (Salvador-Morales *et al.*, 2006).

Rabbit anti-C1q IgG was biotinylated with biotin N-hydroxysuccinimide.

2.22 Complement consumption assay

NUNC Maxisorp 96-well microtitre plate wells were coated with 100 µL (5 µg) of VC1805 in PBS. As a positive control, wells on the same plate were coated with 100 µL per well of 100 µg/mL hen ovalbumin (Sigma) in 0.1 M sodium carbonate buffer, pH 9.6. The plate was incubated overnight at 4°C, then washed 4 times with buffer A. The remaining sites for protein binding on the plate were saturated by blocking buffer and incubated for 1 h at room temperature before again washing 4 times with buffer A. Rabbit anti-ovalbumin antiserum, prepared by MRC Immunochemistry Unit in Oxford, was heat-treated for 30 min at 56 °C to inactivate rabbit C1q, followed by ten-fold dilution in PBS and 100 µL was added to the ovalbumin-coated wells and incubated for 1 h at room temperature. Wells were washed 4 times with buffer A. The plates were now ready for the complement consumption assay. Reagent and buffer details were as previously described (Salvador-Morales *et al*., 2006). NHS (normal human serum) was diluted with 1 volume of DGBV++ buffer (dextrose-gelatin-verenal buffer). The diluted serum (300 µL) was added in triplicate to wells coated with VC1805, ovalbumin anti-ovalbumin and non-coated wells followed by incubation at 37 °C for 30 min. Serum was removed from wells and kept on ice until ready to assay for total complement haemolytic activity (Salvador-Morales *et al*., 2006). Briefly, serial 2-fold dilutions of serum were incubated with antibody-sensitised sheep erythrocytes and the degree of lysis determined after 1 h incubation at 37°C. The dilution of serum required to give 50% lysis was calculated.

2.23 Gene knockout

2.23.1 Preparation of electrocompetent DH5α λ*pir* strain

E. coli DH5α λ*pir* strain from -70˚C stock was streaked on LB agar plate. A single colony of DH5α λ*pir* was inoculated into a flask containing 50 mL LB medium followed by the incubation of the culture at 37˚C overnight with shaking at 250 rpm. Next day, 500 mL of prewarmed LB medium in separate 2 L flasks was inoculated with 25 mL of the overnight bacterial culture. The flasks were incubated at 37˚C with shaking at 250 rpm until the OD_{600} was about 0.35-0.4. This density is usually achieved after around 2.5 h. The culture was then rapidly transferred into an ice-water bath for 15-30 min. The culture was swirled occasionally to make sure that cooling occurred evenly. For maximum efficiency of transformation, the temperature was not allowed to exceed 4˚C at any stage in this procedure. The culture was transferred to pre-chilled centrifuged bottles (kept at -20˚C) and harvested by centrifugation at 2500 rpm for 20 min at 4˚C. The supernatant was discarded and the pellet was resuspended in 500 mL sterilized ice-cold water. Then, the cells were harvested again by centrifugation at 2500 rpm for 20 min at 4˚C. The supernatant was discarded and the pellet was resuspended in 250 mL sterilized ice-cold water containing 10% (v/v) glycerol. Again, the cells were harvested by centrifugation at 2500 rpm for 20 min at 4˚C. The supernatant was carefully discarded to remove any remaining drops of buffer. Finally, the pellet was resuspended in 1 mL sterilized ice-cold GYT medium (10% (v/v) glycerol, 0.125% (w/v) yeast extract, 0.25% (w/v) tryptone) by gentle swirling rather than vortexing and pipetting. The cell suspension was diluted to a concentration of 1 X 10^{10} cells/mL (1.0 OD_{600} = ~2.5 X 10^{8} cells/mL) with ice-

cold GYT medium. 50 µL of cells were used for each transformation. For storage, 50 µL aliquots of the cell suspension was dispensed into sterile (ice-cold) 0.5 mL eppendorfs and flash-frozen with liquid nitrogen before transferral to - 70˚C.

2.23.2 Preparation of electrocompetent *Vibrio cholerae*

V. cholerae electrocompetent cells were prepared and electroporated as described by the published procedure with minor some modifications (Marcus *et al*., 1990). *V. cholerae* O1 El Tor strain N16961 (Streptomycin resistant) from a -70˚C stock were streaked onto LB agar media containing 200 µg/mL streptomycin followed by overnight culture of *V. cholerae* in 50 mL LB broth with 200 µg/mL streptomycin. Using a 2% inoculums of overnight culture in 300 mL LB broth containing 200 µg/mL streptomycin, cells were grown until the OD_{600} reached about 0.4. After incubation, cultures were split into 50 mL falcon tubes and pelleted by centrifugation at 4000 rpm for 15 min at 4˚C. The supernatant was discarded and the pellet was washed three times with 250 mL of 10% (v/v) glycerol containing 272 mM sucrose. Sucrose/ glucose is not suitable for autoclaving so it was sterilized by using 0.22 µm filter. Finally, the cells were resuspended with 10% (v/v) glycerol containing 272 mM sucrose to make 1 X 10^{10} cells/mL. This electrocompetent *V. cholerae* was stored at -70˚C in 50 µL aliquots for future use after flash-freezing with liquid nitrogen.

2.23.3 SOE-PCR (Splicing by Overlapping Extension PCR)

PCR has been carried out to amplify portions at the beginning and the end of gene of the interest to be knocked out (AB and CD) (Figure 2.5). Primers were commercially synthesized (Table 2.3) and purchased from IDT, USA. PCRs were performed in a volume of 50 µL containing 1.5 µL of forward primer (25µM), 1.5 µL of reverse primer (25 µM), 5 µL PCR buffer (10X), 1 µL template genomic DNA of *V. cholerae* O1 El Tor strain N16961 (80 ng), 2 µL dNTPs (10 mM), 1 µL Taq DNA polymerase and 38 µL water. The amplification conditions were the same as the standard PCR described earlier (Section 2.4). Use of Taq DNA polymerase in PCR gives 3´ A-overhangs which is suitable for further cloning into pTOPO vector. The PCR products were run on agarose gel. The product bands was excised from the gel and purified by the protocol given in the QIAquick Gel Extraction Kit (QIAGEN).

2.23.4 PCR to ligate beginning and end of gene

1 µL of each construct (40 ng each) was used as template per reaction. The same SOE-PCR reaction as described above was carried out until the 4th cycle without primer (Figure 2.5). The primers (A and D) were added after the 4th cycle (allowing overhang in constructs to anneal) and the PCR was continued for another 26 cycle according to the standard protocol described earlier. Agarose gel electrophoresis and gel extraction of PCR products was done (Section 2.5).

Table 2.3 Primers used in Gene knockout

Target protein	Primer name	Primer sequence (5′- 3′)
VC050809	M050809F_A	ctgagctcagcaatggctg
	M050809R_B	ggctggctaccaatgcaag
	M050809F_C	cttgcattggtagccagcccggcagttcaacacgagtc
	M050809R_D	gaagtgaatgccgcagcag
	M050809F_E	ccacaatgacttcatcgagg
	M050809R_F	cctaccattgattccgacag
VC050809	M180405F_A	caatctaagcgggatcgatc
	M180405R_B	ctggtaacaccaacggttag
	M180405F_C	ctaaccgttggtgttaccagcggatggaatcctatcgtag
	M180405R_D	gttgaatctaacgagttgcga
	M180405F_E	gcggtgatattgccacaatc
	M180405R_F	gatgagtctcaatggtgttcc

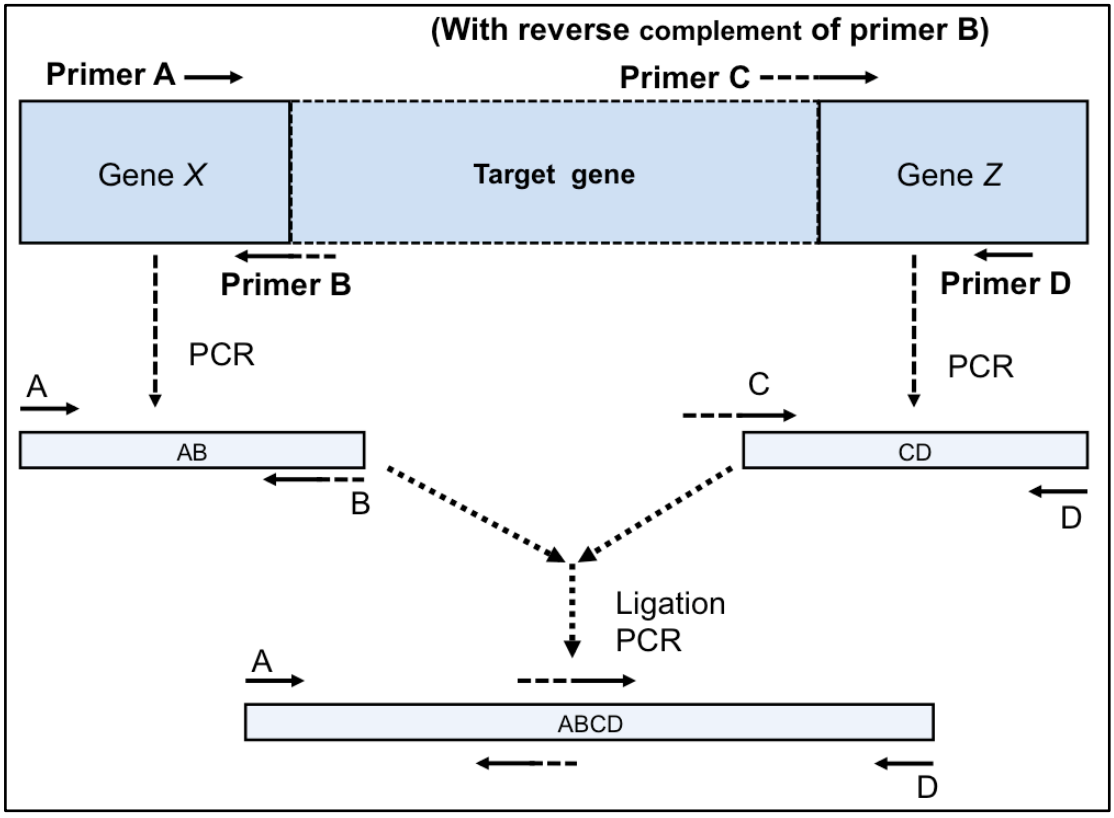


Figure 2.5 SOE-PCR (Splicing by Overlapping Extension PCR)

2.23.5 Cloning into pCR 2.1 –TOPO vector

TOPO cloning reactions were setup in 6 µL of reaction mixture consisting of 50 ng of insert (ABCD in Figure 2.5), 1 µL supplied salt solution (Invitrogen), 1 µL TOPO vector and water to a total volume of 5 µL. The reaction was mixed gently and incubated for 15 to 20 min at room temperature (22 °C) and placed on ice. 2 µL of the TOPO cloning reaction was transformed into chemically competent One Shot TOP10F´ *E. coli* cells using the standard protocol supplied by Invitrogen. Finally, positive clones were selected by blue/white colony screen followed by colony PCR (Section 2.10.1) using A and D primers.

2.23.6 Construction of pADS132 vector with multiple cloning sites

To get suitable cloning sites in the pDS132 suicide vector (Philippe *et al.*, 2004) for our gene of interest, new MCS were introduced (*Xba*I- *Nde*I- *Nsi*I- *Xho*I- *Nhe*I- *Pst*I- *Spe*I- *Sac*I) and the modified vector was named pADS132. A forward (5′- gctctagacatatgatgcatctcgaggctagcctgcagactagtgagctcag -3′) and a reverse (5′- ctgagctcactagtctgcaggctagcctcgagatgcatcatatgtctagagc -3′) primer were commercially synthesized and purchased (IDT, USA). The synthesized fragments were annealed (80 °C for 10 min followed cooling it down to room temperature), digested with *Xba*I and *Sac*I and ligated with similarly digested pADS132 according to the protocol described earlier (Section 2.6 and 2.7). The new suicide vector pADS132 contains unique *Xho*I and *Nco*I sites which was confirmed by double digestion (Figure 2.6). DNA sequencing also confirmed the presence of the MCS in pADS132.

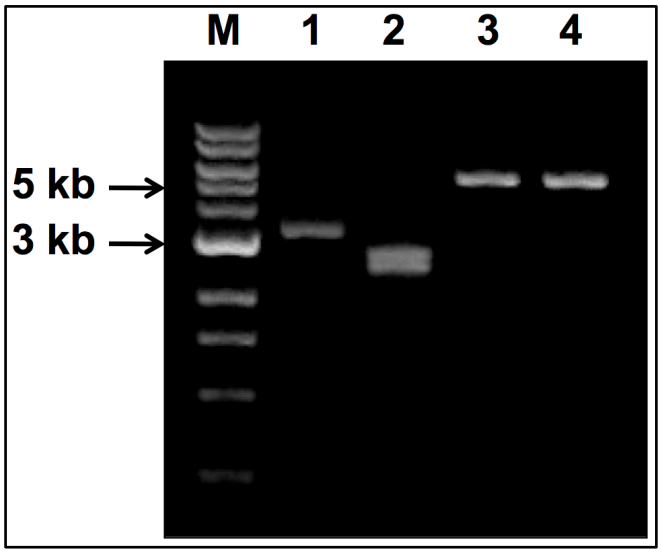


Figure 2.6 Construction of pADS132 vector

M, 1 kb DNA ladder (NEB, USA); lane 1, circular pADS132 vector; lane 2, double digest (*Xho*I and *Nco*I) of pADS132 results two linear bands; lane 3, single digest of pADS132 by *Xho*I; lane 4, single digest of pADS132 by *Nco*I.

2.23.7 cloning of genes of interest in pADS132

SOE-PCR product cloned into pCR 2.1 –TOPO vector was digested with *Xho*I and *Spe*I and ligated to similarly digested pADS132 vector. The ligation reactions were transformed by electroporation into *E. coli* DH5α λ*pir* strains to allow replication of the vector. Briefly, 5 µl of ligation reaction (50 ng) was electroporated in 50 µl cells of DH5α λ*pir* which contain 5 X 10^8 cells. Electroporation was carried out with an Eppendorf 2510 electroporator with the pulse controller set at 10 µF, 900 V and 600 Ω. After electroporation, 900 µL LB medium was added to the suspension and the tube was placed horizontally in a shaking incubator for 1 h at 37 °C and cells were plated onto LB agar containing 34 µg/mL chloramphenicol. Positive clones were identified by colony PCR using A and D primers.

2.23.8 Electroporation of vector in *V. cholerae*

50 ng of vector DNA of pADS132 containing the gene of interest was electroporated into 50 µL of electrocompetent *V. cholerae* O1 El Tor strain N16961 containing 5 X 10^8 cells. Electroporation was carried out with a Eppendorf Electroporator 2510 using cuvettes with a 0.2 cm gap distance and settings of 10 µF, 830 V and 600 Ω. After electroporation, the suspension was diluted with 900 µL LB medium, placed horizontally in a shaking incubator for 1 h at 37 °C and plated on LB agar containing 34 µg/mL chloramphenicol and 200 µg/mL streptomycin.

2.23.9 Colony PCR to screen single crossover

Colonies of *V. cholerae* O1 El Tor strain N16961 containing the correct insert were identified using colony PCR screening. Colonies were randomly picked by sterile tips and incubated overnight at 37 °C in 5 mL LB with 34 µg/mL chloramphenicol and 200 µg/mL streptomycin. Colony PCR was performed with corresponding internal primers A and D and flanking primers E and F. Glycerol stocks were made for positive single crossover colonies.

2.23.10 Select for double crossover mutant by sucrose sensitivity

Single crossover positive colonies were incubated overnight in 5 mL LB broth with 200 µg/mL streptomycin and without chloramphenicol. Next day, 100 µL of 10^{-2} to 10^{-7} serially diluted cultures were plated on LB containing 10% sucrose and incubated at 30°C overnight. After incubation, single colonies were picked and streaked on a plate of LB containing 10% sucrose and

200 µg/mL streptomycin; LB containing 34 µg/mL chloramphenicol; and 5 mL LB with 200 µg/mL streptomycin; followed by incubation at 30 °C overnight. Plates were divided into 8 sections with one section used per single colony.

2.24 Characterization of *Vc*Fur

2.24.1 Electrophoretic mobility shift assay (EMSA)

The *fur* promoter of gene locus *vc2694* was chosen for the gel-shift assays, based on a study of *fur* promoters in *V. cholerae* (Mey *et al*., 2005), with forward and reverse oligonucleotides purchased from VHBIO, with one oligo labeled with FITC (fluorescein isothiocyanate). The oligonucleotide was 5′-ATTGATAATGATTATTATTAAC, where the 19-bp Fur-box is underlined. Working solutions containing 2 µM of forward and reverse oligos were prepared in buffer (20 mM MES pH 6.5, 1 mM DTT). Eppendorfs containing the oligo mix were placed into a water bath at 80-85 °C then allowed to cool down at room temperature for 2 h. The resultant double stranded DNA duplex was stored at -20 °C. To prepare acrylamide gels, 5 mL of Accuflow acrylamide, 2 mL of 10X TBE (Tris-borate-EDTA) buffer or 10X TB (Tris-borate) buffer, 18 mL of deionised water, 20 µL of TEMED (tetramethylethylenediamine), 100 µL of 10% ammonium persulphate were mixed together and poured between the glass plates with comb inserted. After drying, the gels were always pre-run for 10 min at 200 V with appropriate running buffer before loading samples.

The DNA mobility shift assay was performed using the published method with minor modifications (Ochsner *et al.*, 1995). A final concentration of 50 nM FITC-labeled double stranded DNA fragment was mixed into 20 µL binding

buffer which containing 10 mM bis-Tris borate (pH 7.5), 40 mM KCl, 0.1 mg/mL of bovine serum albumin (BSA) (Sigma fraction V) and 10% (v/v) glycerol. Different concentrations of native *Vc*Fur were added and the mixture was incubated at 37 °C for 15 min using a static incubator. For the negative control, 50 nM FITC-labeled double stranded DNA (poly A and poly T) was used. 2 µL of 10X Ficoll loading dye was added to the samples and 10 µL of each samples was loaded into gel wells running at 60 V constant voltage. The gel was run at 10 mA constant current for 80 min in running buffer (20 mM bis-Tris borate and 0.1 mM $MnSO_4$ or 20 mM bis-Tris borate only). After the run, the gel was scanned at 473 nm using Fujifilm FLA-5000 series scanner.

2.24.2 Metal binding studies of *Vc*Fur

Two batches of *Vc*Fur were purified into a final buffer containing 20 mM HEPES, 200 mM NaCl and 10% (v/v) glycerol at pH 7.5. Batch 1, was dialyzed in 20 mM HEPES, 200 mM NaCl, 1 mM $ZnCl_2$ and 2 mM DTT and 10% (v/v) glycerol, followed by removal of excess metal ion by dialysis of protein samples in the same buffer without $ZnCl_2$. Batch 2 was not dialysed against any zinc-containing buffer. Finally both samples were dialysed overnight at 4 °C against 20 mM HEPES, 20 mM NaCl and 5% (v/v) glycerol at pH 7.5. The concentration of protein was measured using a NanoDrop spectrophotometer (Thermo Scientific).

2.24.3 Removal of metals from *Vc*Fur

To remove metal ions (mainly Zn) from *Vc*Fur, the experiment was carried out with minor modification as previously described (Althaus *et al*., 1999). 50 µM

of both batches of *Vc*Fur were dialysed overnight against 200 mM EDTA, 20 mM HEPES, 20 mM NaCl and 5% (v/v) glycerol at pH 8.0. To remove extra EDTA, the samples were dialysed in the same buffer without EDTA for 3 h followed by dialysis into 20 mM HEPES, 20 mM NaCl and 5% (v/v) glycerol at pH 7.5 with 3 buffer changes. All the procedures were carried out at 4 °C.

2.24.4 Addition of metals into EDTA-treated *Vc*Fur

Different metals Zn(II) (as $ZnCl_2$), Mn(II) (as $MnSO_4$) and Fe(III) (as $FeCl_3$) were added into EDTA-treated *Vc*Fur. The procedure for addition of metal into *Vc*Fur was followed with minor modifications (Mills and Marletta, 2005). To incorporate Zn(II), 50 μM of *Vc*Fur was dialysed overnight with 3 equivalent of $ZnCl_2$ (150 μM at pH 6.2) in 20 mM HEPES, 50 mM NaCl and 5% (v/v) glycerol at pH 6.2 in the presence of a 20-fold excess of DTT. For Mn(II), 50 μM of *Vc*Fur was dialysed overnight with 4 equivalent of $MnSO_4$ (200 μM) in 20 mM HEPES, 50 mM NaCl and 5% (v/v) glycerol at pH 7.5. For the incorporation of the Fe (III), ferric chloride ($FeCl_3$) was first dissolved with a 10-fold excess of citric acid which was adjusted to pH 7.0 and 50 μM of *Vc*Fur was dialysed overnight with 3 equivalent of $FeCl_3$ (150 μM) in 20 mM HEPES, 50 mM NaCl and 5% (v/v) glycerol at pH 7.0. To remove the excess metal, all samples were dialysed in 20 mM HEPES, 20 mM NaCl and 5% (v/v) glycerol at pH 7.5, without adding metal salts for 4 h followed by 3 buffer changes.

2.24.5 Analysis of metals

The metal contents of the proteins were analyzed using inductively coupled plasma mass spectrometry (ICP-MS) at the University of Edinburgh. All the

samples which were subject to metal analysis were finally dialysed against a buffer containing 20 mM HEPES, 20 mM NaCl and 5% (v/v) glycerol at pH 7.5 and the machine was equilibrated with the same buffer to minimize any metal contamination from the buffer itself. Standard metal solutions were measured between every sample measurement to maintain higher sensitivity. Measurements were carried out in triplicate.

2.24.6 Spin labeling of *Vc*Fur

Spin labeling of *Vc*Fur was carried out in 2.5 mL of reaction mixtures containing 20 μM of protein, 100 μM of spin labeling reagent, MTSL (1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3 methyl-methanethiosulfonate), and a buffer X (20 mM HEPES, 200 mM NaCl at pH 7.5) to a final volume of 2.5 mL followed by incubation at 4°C for 1 h. Next, the unbound spin label (MTSL) from the sample was removed by gel filtration using Sephadex G-25 minicolumns PD-10 (Amersham Biosciences). Briefly, the column was equilibrated with 25 mL elution buffer (buffer X) and the flow-through was discarded. Next, 2.5 mL of reaction mixture was loaded onto the column and the flow-through was discarded. The proteins in the column eluted with 8 mL of buffer and the first 8 fractions of 0.5 mL each and the next 4 fractions of 1 mL each were collected. The appropriate fractions, confirmed by Bradford protein assay, were pooled and concentrated to 200 μL. Then, 1 mL of buffer X prepared in D_2O (deuterium oxide) was added into the concentrator and the sample was concentrated again to 200 μL. This process was repeated three times and finally 40-50% deuterated ethylene glycol was added to the sample to obtain a final concentration of 100 μM protein sample in 100 μL.

Chapter 3

Structural determination of VC0508, VC0509, VC1804 and VC1805

3. Structural determination of VC0508, VC0509, VC1804 and VC1805

This chapter describes the results of cloning, expression, purification and structural determination of VC0508, VC0509, VC1804 and VC1805. All the structures were determined by X-ray crystallography except VC1804 which was not soluble. The first helix of VC1804 is missing compared to VC0508, VC0509 and VC1805 and may be required for solubility in *E. coli*. All the structures show similar folds, as predicted from the sequence analyses. However, the only structural homologue of these proteins is human p32 which is a versatile protein and can bind to several other proteins.

3.1 Overview of VC0508, VC0509, VC1804 and VC1805

VC0508, VC0509, VC1804 and VC1805 are hypothetical proteins of *V. cholerae* O1 El Tor strain N16961 that contain 147, 147, 104, 148 amino acid residues respectively. The molecular weights of VC0508, VC0509, VC1804 and VC1805 proteins are 17, 17.1 12.3 and 16.8 kDa and their calculated isoelectric points (pI) are 5.0, 5.2, 5.4 and 4.7, respectively. Sequence analyses of these proteins did not suggest any structural homologues in the existing protein structure database.

3.2 Cloning

3.2.1 *vc0508* gene

Amplification of the *vc0508* gene was carried out by PCR from genomic DNA of *V. cholerae* O1 El Tor strain N16961 using standard protocols (Figure 3.1.A) (Section 2.4). Primers (VHBIO, UK) for PCR amplification from genomic DNA contained specific recognition sites *Nco*I in the forward primer

(5′- GGACTCACCATGGCCAAATTAAC -3′) and *Sal*I site in the reverse primer (5′- GAAGGGAAGTCGACAGTTATTGC -3′). The forward primer introduced a Ser2Ala mutation after the start codon of the VC0508 protein. After amplification, the PCR product was digested with *Nco*I and *Sal*I enzymes and cloned into the similarly digested pLou3 expression vector (Section 2.6 and 2.7). After propagation in *E. coli* DH5α strain, the positive clones were picked up by blue/white screening due to the presence of *lacZα* complementation gene in pLou3 vector. Positive clones were verified by double digestion (Figure 3.1.B) (Section 2.10.2) and DNA sequencing in the University of Dundee (Section 2.10.3).

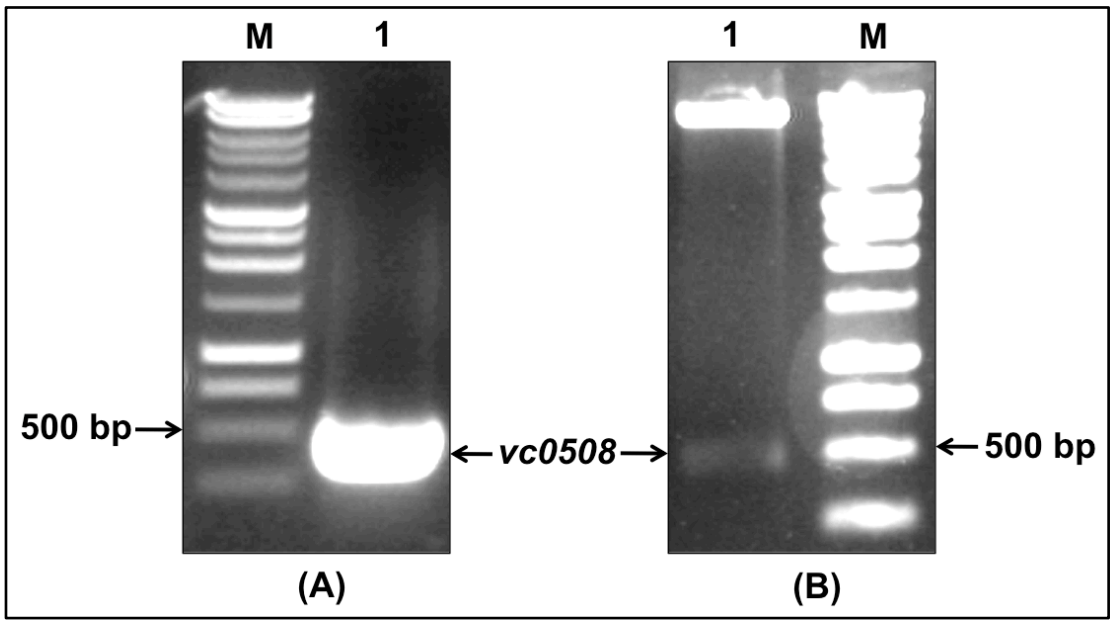


Figure 3.1 Agarose gel electrophoresis of *vc0508* gene products

(A) Amplification of *vc0508* gene from genomic DNA of *V. cholerae* O1 El Tor strain N16961. M, 1 kb DNA ladder (Promega); lane 1, Amplification of *vc0508* gene products. **(B)** Double digestion of cloning vector pLou3 containing *vc0508* gene. Lane 1 (bottom band), insert *vc0508* gene fragment; lane 1 (top band), linear pLou3 vector without insert gene after double digestion.

3.2.2 *vc0509* gene

V. cholerae O1 El Tor strain N16961 strain was used in the amplification of the *vc0509* gene by PCR from genomic DNA using standard protocols (Figure 3.2.A). Primers (VHBIO, UK) for PCR amplification from genomic DNA contained specific recognition sites *Nco*I in the forward primer (5′- CTGAAAGGCCATGGTTATGATTG -3′) and *Sal*I site in the reverse primer (5′- GAGAGAGTCGACTTTAGTTGAAC -3′). After amplification, the PCR product was digested with *Nco*I and *Sal*I enzymes and cloned into the similarly digested pLou3 expression vector. Positive clones were identified by blue/white screening followed by double digestion Figure 3.2.B) and DNA sequencing.

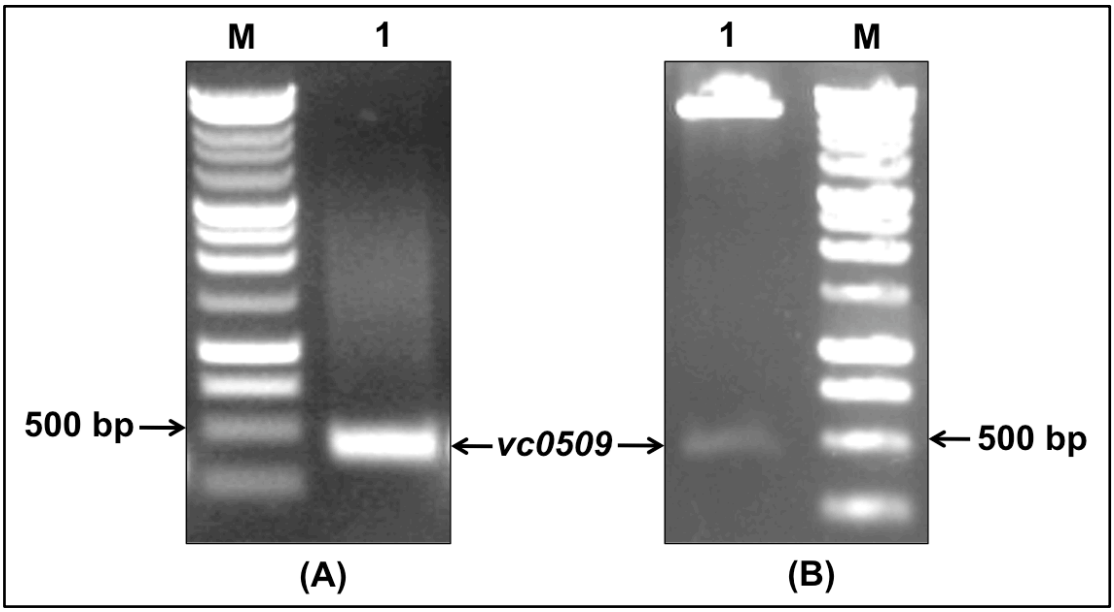


Figure 3.2 Agarose gel electrophoresis of *vc0509* gene products

(A) Amplification of *vc0509* gene from genomic DNA of *V. cholerae* O1 El Tor strain N16961. M, 1 kb DNA ladder (Promega); lane 1, Amplification of *vc0509* gene products. **(B)** Double digestion of cloning vector pLou3 containing *vc0509* gene. Lane 1 (bottom band), insert *vc0509* gene fragment;

lane 1 (top band), linear pLou3 vector without insert gene after double digestion.

3.2.3 *vc1804* gene

Standard PCR protocol was used to amplify *vc1804* gene by PCR from genomic DNA of *V. cholerae* O1 El Tor strain N16961 (Figure 3.3.A). Primers (VHBIO, UK) for PCR amplification from genomic DNA contained specific recognition sites *Nco*I in the forward primer (5′- CCAATCCATTATTACCATGGTTCT -3′) and *Sal*I site in the reverse primer (5′- GAATCGCGTCGACTTAGCGAAT -3′). The amplified PCR product was digested with *Nco*I and *Sal*I enzymes and cloned into the similarly digested pLou3 expression vector After propagation in *E. coli* DH5α, Positive clones were identified by blue/white screening followed by double digestion (Figure 3.3.B) and DNA sequencing.

Another construct of *vc1804* gene was made using the pEHISTEV vector (Liu and Naismith, 2009). The *vc1804* gene was amplified by PCR from genomic DNA of *V. cholerae* O1 El Tor strain N16961. Primers (VHBIO, UK) for PCR amplification contained specific recognition sites *Nco*I in the forward primer (5′- GCCAATCCATTATTACCATGGTTCTATAC -3′) and *EcoR*I site in the reverse primer (5′- GAAGAATCGAATTCGGTTAGCGAATCATG -3′). After amplification, the PCR product was digested with *Nco*I and *EcoR*I enzymes and cloned into the similarly digested pEHISTEV vector. After propagation in *E. coli* DH5α, the positive clones were verified by colony PCR (Section 3.10.1) followed by DNA sequencing.

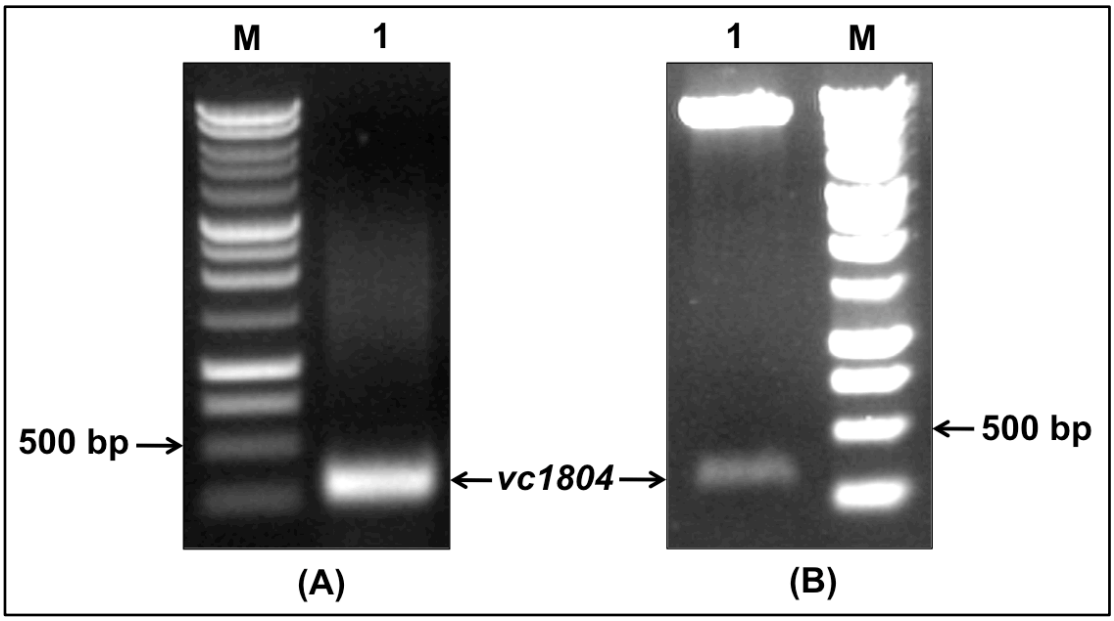


Figure 3.3 Agarose gel electrophoresis of *vc1804* gene products

(A) Amplification of *vc1804* gene from genomic DNA of *V. cholerae* O1 El Tor strain N16961. M, 1 kb DNA ladder (Promega); lane 1, Amplification of *vc1804* gene products. **(B)** Double digestion of cloning vector pLou3 containing *vc1804* gene. Lane 1 (bottom band), insert *vc1804* gene fragment; lane 1 (top band), linear pLou3 vector without insert gene after double digestion.

3.2.4 *vc1805* gene

The *vc1805* gene was amplified by PCR from genomic DNA of *V. cholerae* O1 El Tor strain N16961 (Figure 3.4.A). The forward primer containing the specific recognition site *Nco*I (5′-GGAGGCATAACCATGGCTAACATTTTG -3′) and reverse primer containing the specific recognition site *EcoR*I (5′-CTTGATGGAATTCACTAAGGAGCCGAG -3′), were purchased from VHBIO, UK. After amplification, the PCR product was digested with *Nco*I and *EcoR*I enzymes and cloned into the similarly digested pEHISTEV vector. After propagation in *E. coli* DH5α, the positive clones were verified by colony PCR (Figure 3.4.B) and DNA sequencing.

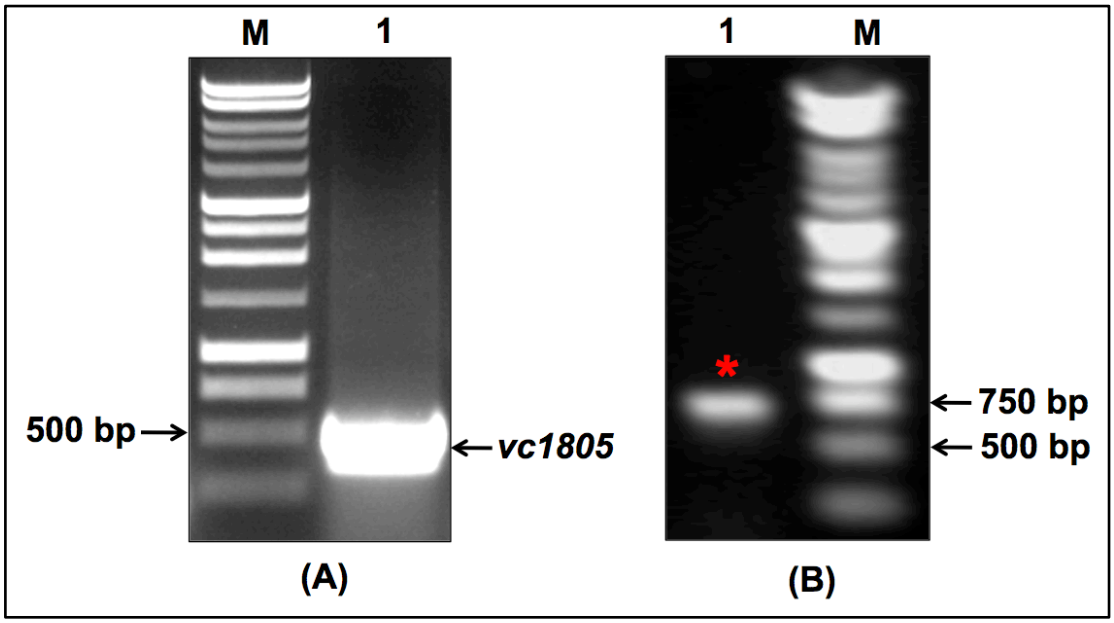


Figure 3.4 Agarose gel electrophoresis of *vc1805* gene products

(A) Amplification of *vc1805* gene from genomic DNA of *V. cholerae* O1 El Tor strain N16961. M, 1 kb DNA ladder (Promega); lane 1, Amplification of *vc1805* gene products. **(B)** Colony PCR of cloning vector pEHISTEV containing *vc1805* gene. Lane 1, colony PCR products of pEHISTEV vector containing *vc1805* gene (indicated by red asterisk).

3.3 Protein expression, optimization of solubility and purification

3.3.1 VC0508 protein

The *E. coli* K12 strain TB1 (DE3), was used to transform the expression vector pLou3 containing the *vc0508* gene using standard methods (Section 2.8). The best soluble protein expressions were achieved in TPB medium and the condition was 25 °C overnight incubation for 24 h with the addition of 0.5 mM IPTG (Section 2.13). These conditions were also used to scale up recombinant VC0508 protein expression in a 5 L culture.

The cell pellet was resuspended in buffer A (50 mM HEPES, 500 mM NaCl, 10 mM imidazole and 10% (v/v) glycerol) containing DNase and EDTA free

protease inhibitor cocktail and sonicated. After centrifugation, the soluble fragment was syringe filtered through a 0.22 µm filter unit and loaded on to a HisTrap™ HP column (Ni^{2+}-affinity column, GE Healthcare), to purify the target proteins. His-tagged proteins bound to the HisTrap™ HP column were eluted by a higher imidazole concentration buffer. Fractions were analysed by SDS-PAGE and pooled (Figure 3.5.A).

To remove the extra imidazole, pooled samples were dialysed for several h against 4 L of buffer A and 1 mM DTT at 4 °C, with two buffer changes. TEV protease was added into the dialysis tube containing 6X his-tagged protein (1 µg protease for 1 mg at 25 °C overnight). Digestion of protein sample was analyzed by SDS-PAGE to check the efficiency of TEV protease digestion (Figure 3.5.B). After TEV digestion, protein samples were syringe filtered through a 0.22 µm filter unit and loaded into the HisTrap™ HP column as before and the flow-through was collected in fractions and analyzed by SDS-PAGE (Figure 3.5.C). Corresponding fractions containing HMBP- cleaved target protein were pooled and dialysed against 4 L of buffer containing buffer A and 1 mM DTT overnight at 4 °C. After dialysis, the samples were then concentrated to 4.5 mL by using a 10,000-Da MWCO centrifugal concentrator (Sigma- Aldrich). Up to 4.5 mL of concentrated samples were loaded onto a Hiprep™ 16/60 Sephacryl™ S-200 gel filtration column (GE Healthcare) and eluted in buffer A and 1 mM DTT. Corresponding fractions of VC0508 were identified by chomatogram (Figure 3.5.E) and confirmed by SDS-PAGE (Figure 3.5.D). For crystallization trials, fractions containing VC0508 were pooled and dialysed overnight at 4 °C against 4 L of buffer containing 10 mM

HEPES, 50 mM NaCl and 5% (v/v) glycerol. Purified proteins of VC0508 were analyzed by mass spectrometry which confirmed the expressions of the target proteins (Figure 3.5.F).

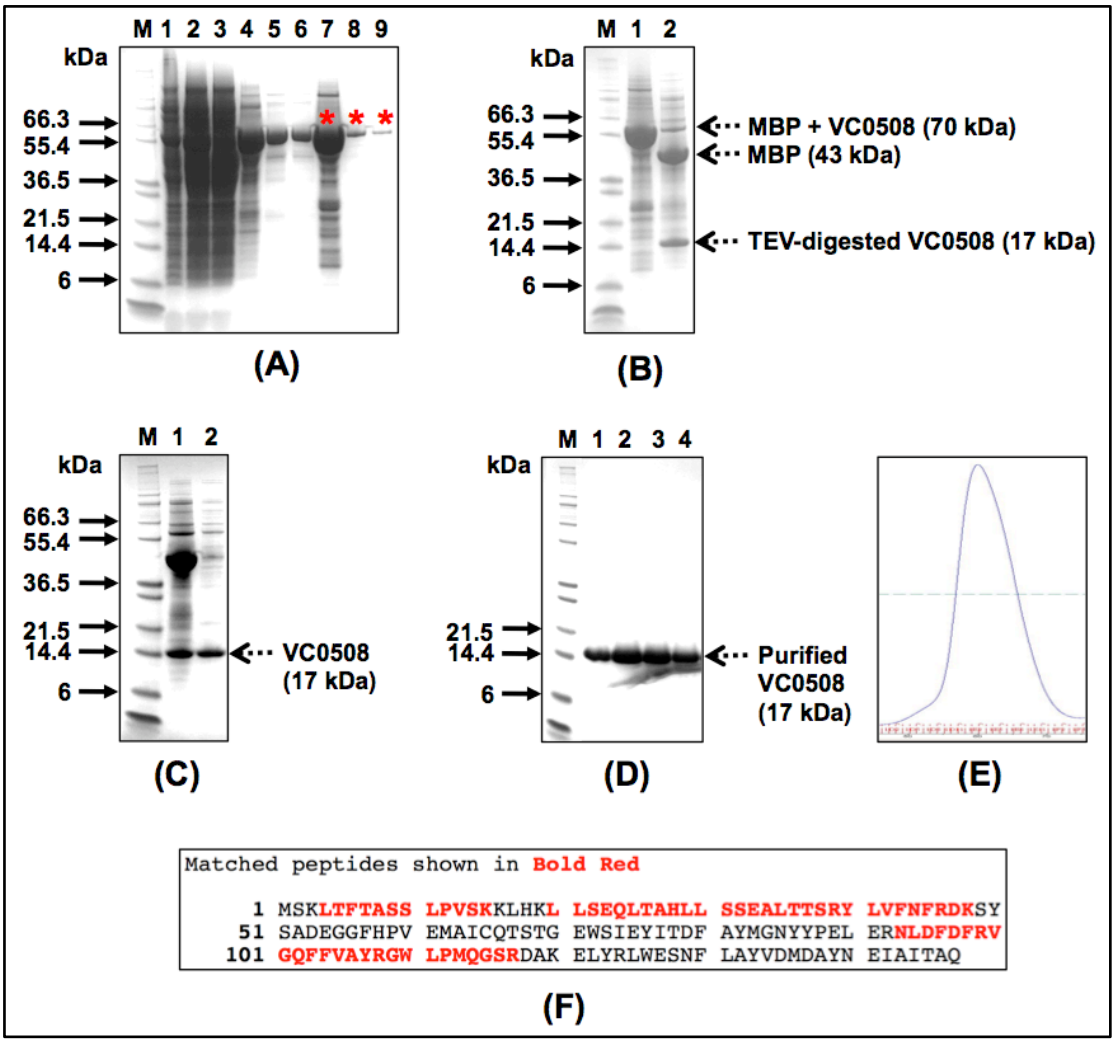


Figure 3.5 Expression and purification of VC0508 protein

(A) First Ni^{2+}-affinity column pull down. M, protein marker (Mark12™, Invitrogen); lane 1, cell lysate after IPTG induction; lane 2, supernatant after sonication; lane 3, flow-through (unbound protein) after loading sample on to the column; lane 4-6, column wash with the buffer containing 10 mM imidazole; lane 7-9, elution in buffer containing 350 mM imidazole. Bands indicated by the red asterisks are MBP-fused VC0508. **(B)** TEV cleavage of his-tagged VC0508. Lane 1, dialysed eluted protein from first Ni^{2+}-affinity column; lane 2, TEV digested sample of VC0508. **(C)** Second Ni^{2+}-affinity

column. Lane 1, filtered TEV digested sample; lane 2, VC0508 without HMBP **(D)** Eluted VC0508 after gel filtration chromatography. Lane 1-4, fractions containing VC0508 protein. **(E)** Gel filtration peak corresponding to purified VC0508. **(F)** Matched peptides of VC0508 identified by mass spectrometry.

3.3.2 VC0509 protein

The expression vector pLou3 containing the *vc0509* gene was transformed in *E. coli* K12 strain TB1 using standard methods (Section 2.8). The best soluble protein expressions were achieved in TPB medium and the condition was 25 °C overnight incubation for 24 h with the addition of 0.5 mM IPTG (Section 2.13). Using these conditions, expression, optimization of solubility and purification of the protein VC0509 was done according to the procedure described earlier (Section 3.3.1) (Figure 3.6). For crystallization trials, purified VC0509 was dialysed in buffer containing 10 mM HEPES, 50 mM NaCl and 5% (v/v) glycerol. Purified VC0509 protein was analyzed by mass spectrometry which confirmed the expressions of the target protein (Figure 3.6.F).

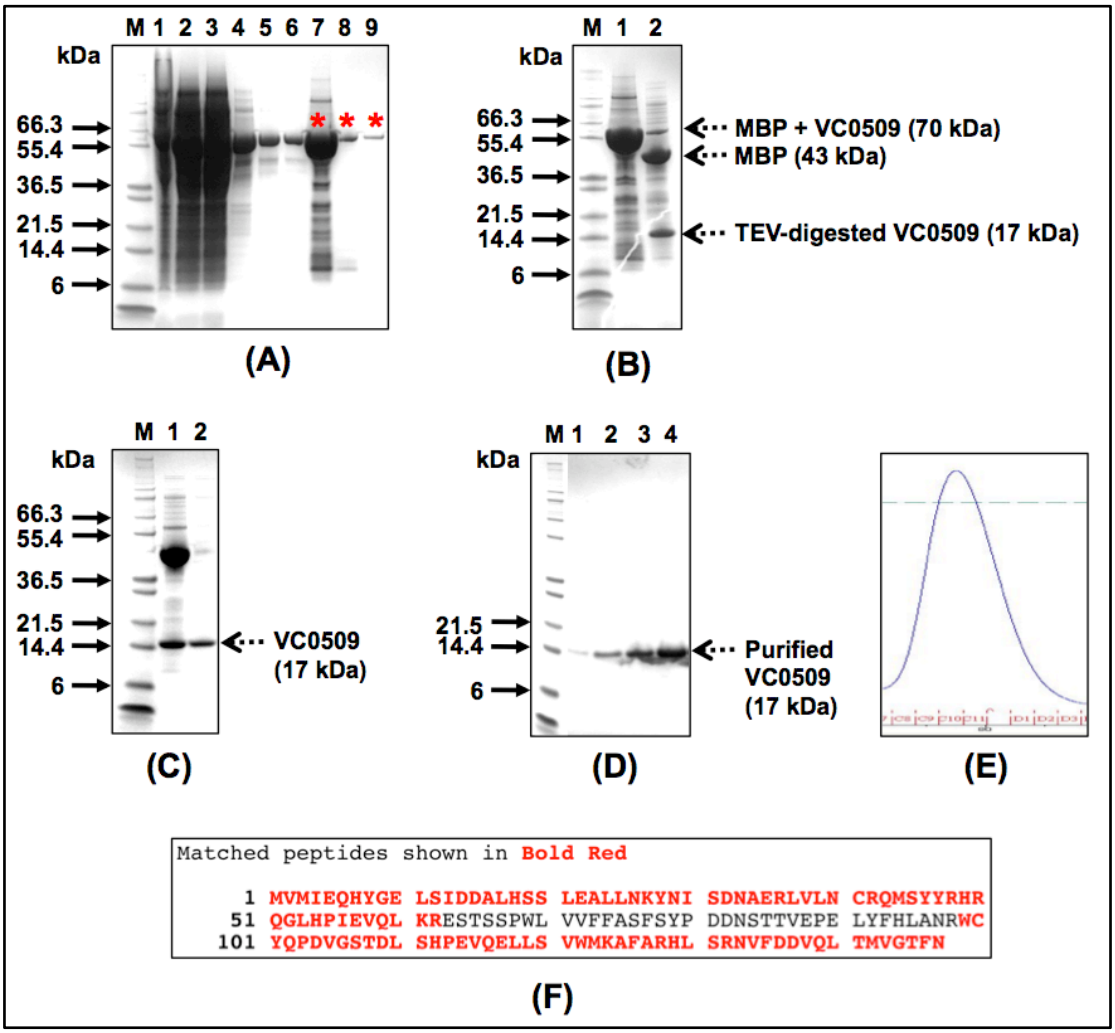


Figure 3.6 Expression and purification of VC0509 protein

(A) First Ni^{2+}-affinity column pull down. M, protein marker (Mark12™, Invitrogen); lane 1, cell lysate after IPTG induction; lane 2, supernatant after sonication; lane 3, flow-through (unbound protein) after loading sample on to the column; lane 4-6, column wash with the buffer containing 10 mM imidazole; lane 7-9, elution in buffer containing 350 mM imidazole. Bands indicated by the red asterisks are MBP-fused VC0509. **(B)** TEV cleavage of his-tagged VC0509. Lane 1, dialysed eluted protein from first Ni^{2+}-affinity column; lane 2, TEV digested sample of VC0509. **(C)** Second Ni^{2+}-affinity column. Lane 1, filtered TEV digested sample; lane 2, VC0509 without HMBP **(D)** Eluted VC0509 after gel filtration chromatography. Lane 1-4, fractions containing VC0509 protein. **(E)** Gel filtration peak corresponding to purified VC0509. **(F)** Matched peptides of VC0509 identified by mass spectrometry.

3.3.3 VC1804 protein

The vectors pLou3 and pEHISTEV containing the *vc1804* gene were transformed into *E. coli* K12 strain TB1 and *E. coli* strain BL21 (DE3), respectively. The expression, optimization of solubility and purification was done according to the procedure described earlier (Section 3.3.1) (Figure 3.7.A-B). Over-expression of VC1804 protein was confirmed by mass spectrometry (Figure 3.7.C), but no soluble proteins were obtained (Figure 3.7.A-B). Cells were exposed to various IPTG concentrations (1 mM/ 0.5 mM/ 0.1 mM) and temperatures (such as 37 °C/ 25 °C/ 16 °C), shock (heat shock/ cold shock), but unfortunately no soluble expression was obtained.

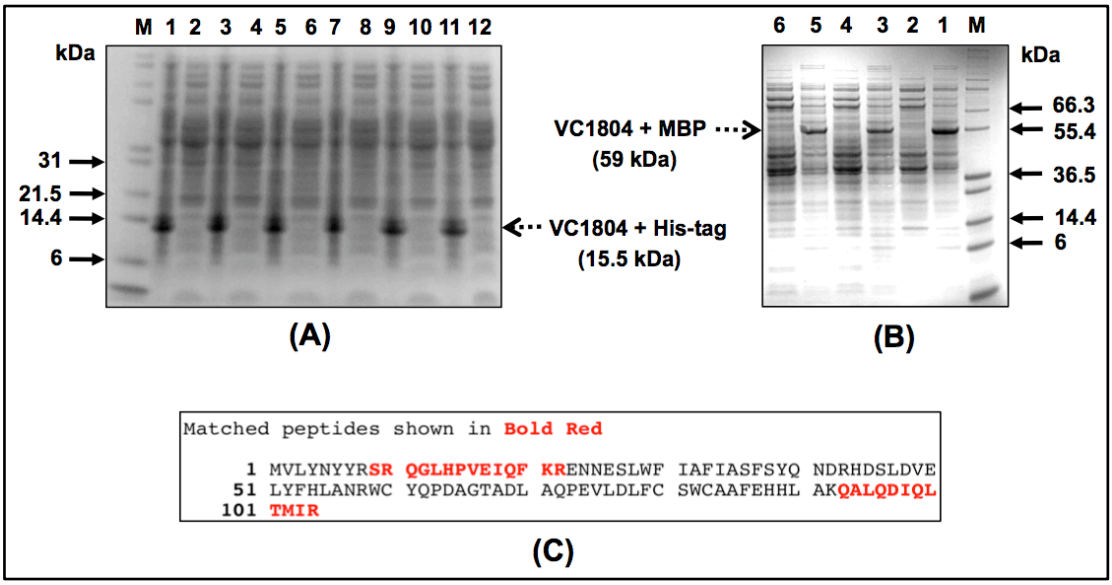


Figure 3.7 Expression and identification of VC1804 protein

(A) Over-expression of VC1804 in pEHISTEV vector. M, protein marker (Mark12™, Invitrogen); lane 1-2, cell lysate and soluble fraction at 37 °C induction; lane 3-4, cell lysate and soluble fraction at 37 °C induction with heat shock; lane 5-6, cell lysate and soluble fraction at 37 °C induction with cold shock; lane 7-8, cell lysate and soluble fraction at 25 °C induction; lane 9-10,

cell lysate and soluble fraction at 25 °C induction with heat shock; lane 11-12, cell lysate and soluble fraction at 25 °C induction with cold shock. **(B)** Over-expression of VC1804 in pLou3 vector. Lane 1-2, cell lysate and soluble fraction at 37 °C induction; lane 3-4, cell lysate and soluble fraction at 25 °C induction, **(C)** Matched peptides of VC1804 identified by mass spectrometry.

3.3.4 VC1805 protein

The plasmid DNA containing the *vc1805* gene was transformed into *E. coli* strain BL21 (DE3). The best soluble protein expression was achieved in LB medium and the condition was 25 °C overnight incubation for 24 h with the addition of 0.5 mM IPTG. These conditions were also used to scale up recombinant VC0508 protein to 5 L culture according to the procedure described earlier (Section 3.3.1)

The cell pellet was resuspended in PBS buffer containing DNase and EDTA-free protease inhibitor cocktail and sonicated. Cell lysate was syringe filtered through a 0.22 µm filter unit and VC1805 was purified using a HisTrap™ HP column, according to the manufacturer's instruction manual. Higher concentrations of imidazole were used to elute His-tagged proteins bound to the HisTrap™ HP column. Fractions were analyzed by SDS-PAGE and pooled (Figure 3.8.A). Protein samples were dialysed against 3 L of PBS buffer containing 50 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 mM DTT, 20 mM imidazole and 0.3 M NaCl for 4 h at 4 °C, with two buffer changes, to remove excess imidazole from protein sample. Next, TEV protease was added into the dialysis tube containing his-tagged protein at 25 °C overnight. Digestion of sample by TEV was analyzed by SDS-PAGE (Figure 3.8.B) and corresponding fractions were pooled. The digested sample was dialysed

against the same dialysis buffer without EDTA and DTT for 2 h at 4°C and the dialysis buffer was changed twice. Protein samples were syringe filtered and loaded onto the nickel column and the flow-through was collected in fractions and analysed by SDS-PAGE (Figure 3.8.C). The flow-through containing VC1805 was dialysed against 3 L of PBS buffer containing 50 mM Tris-HCl pH 7.5, 20 mM imidazole and 20 mM NaCl overnight at 4 °C. After dialysis, the samples were concentrated to 4.5 mL and loaded onto a gel filtration column. Fractions were collected based on the observed peak of gel filtration (Figure 3.8.E), also analysed by SDS-PAGE (Figure 3.8.D) and the correct identity confirmed by mass spectrometry (Figure 3.8.F).

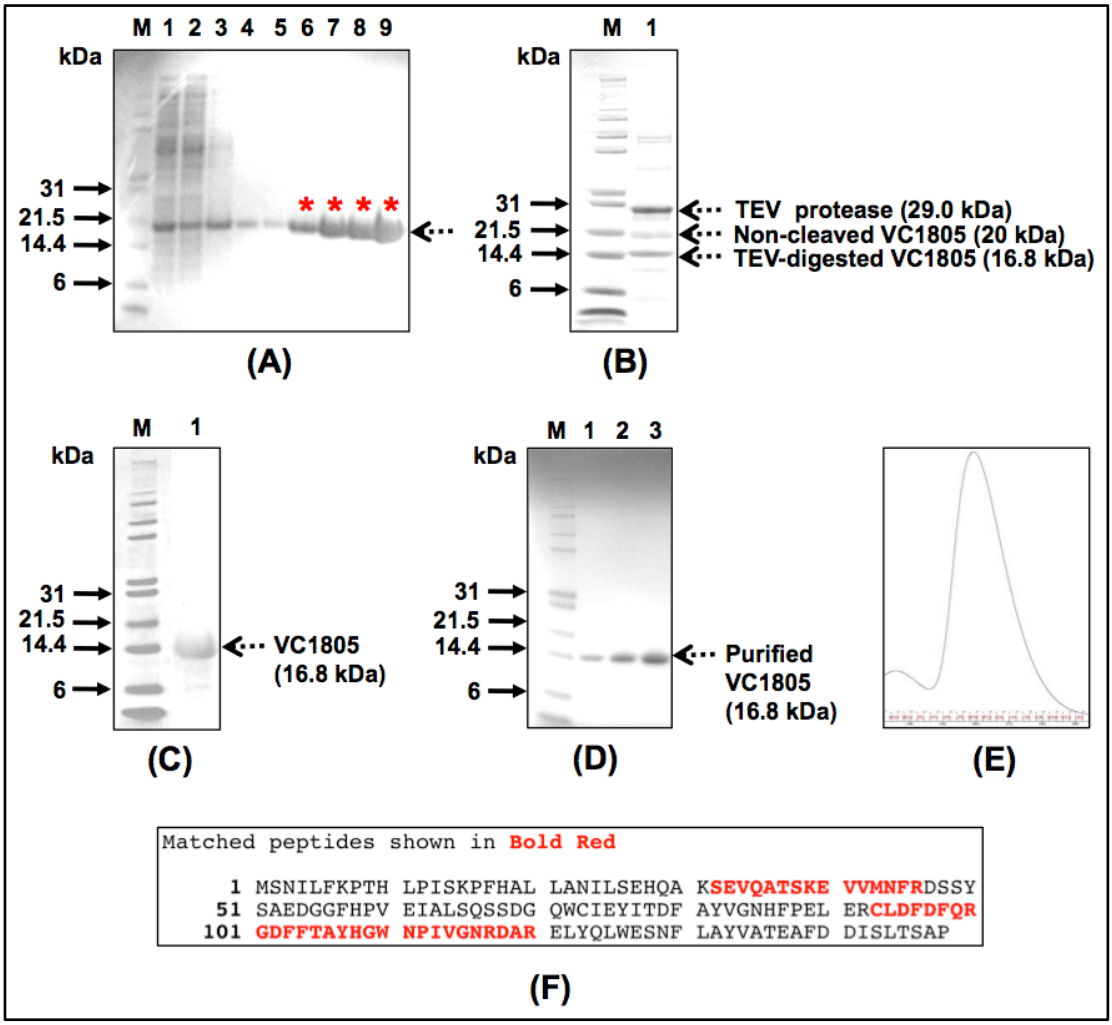


Figure 3.8 Expression and purification of VC1805 protein

(A) First Ni^{2+}-affinity column pull down. M, protein marker (Mark12™, Invitrogen); lane 1, cell lysate after IPTG induction; lane 2, supernatant after sonication; lane 3, flow-through (unbound protein) after loading sample on to the column; lane 4-5, column wash with the buffer containing 10 mM imidazole; lane 6-9, elution in buffer containing 350 mM imidazole. Bands indicated by the red asterisks are His-tagged VC1805. **(B)** TEV cleavage of his-tagged VC1805. Lane 1, TEV digested sample of VC1805. **(C)** Second Ni^{2+}-affinity column. Lane 1, VC1805 without his-tag **(D)** Eluted VC1805 after gel filtration chromatography. Lane 1-3, fractions containing VC1805 protein. **(E)** Gel filtration peak corresponding to purified VC1805. **(F)** Matched peptides of VC1805 identified by mass spectrometry.

3.4 Crystallization, optimization and X-ray data collection of targets

3.4.1 VC0508 protein

The initial concentration of VC0508 for crystallization was 12 mg/mL, which was determined by the PCT kit. A total of six crystal screens were set up with two different concentrations which were 12 mg/mL and 6 mg/mL. The screens were: The PEGs, Ammonium salt, pH clear, JCSG+, Classics (Qiagen) and Wizard I and II (Emerald Biosciences). Crystal plates were initially set up using a nano-drop crystallization robot, with 300 nl of buffer and 300 nl of protein per drop. Initial hits were obtained in the JCSG+ screen (Well number 91; 0.2 M ammonium sulfate, 0.1 M bis-TRIS pH 5.5 and 25 % w/v PEG 3350) (Figure 3.9.A). Further optimizations gave the best quality crystals in the original condition using 8 mg/mL of protein (Figure 3.9.B). Using 20% (v/v) glycerol in crystallization buffer, the crystal of VC0508 was cryoprotected and diffracted X-rays at 100K in-house using a Rigaku/MSC MicroMax-007HF rotating anode equipped with a Saturn 944+ CCD detector at wavelength 1.54178 Å (Figure 3.9.C).

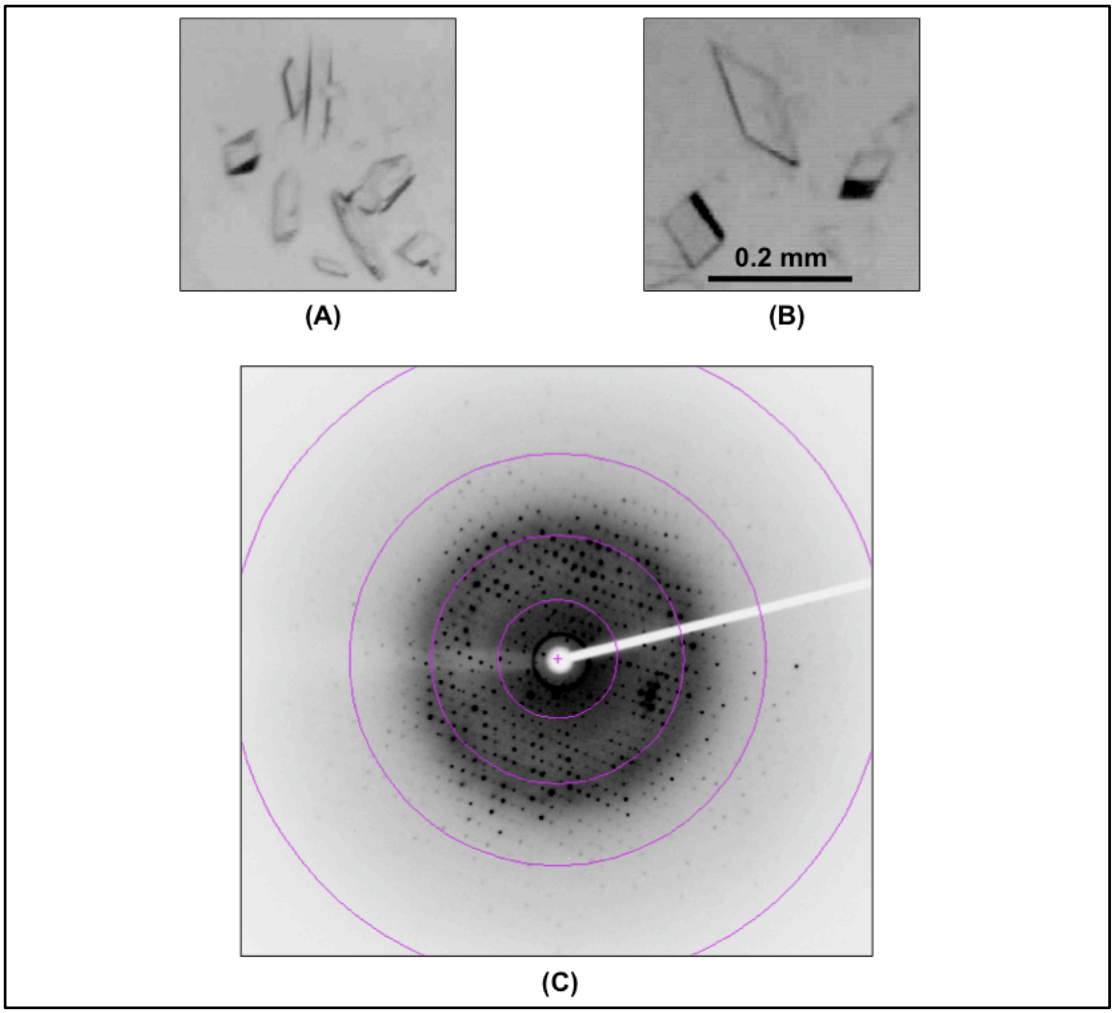


Figure 3.9 VC0508 protein crystal and diffraction pattern

(A) Initial hits in JCSG+ screen (well number 91; 0.2 M ammonium sulfate, 0.1 M bis-TRIS pH 5.5 and 25 % w/v PEG 3350). **(B)** Optimized crystal, using the initial hit condition, containing 8 mg/mL of protein. **(C)** X-ray diffraction image of VC0508 protein crystal. The resolution of the outer circle is 1.9 Å.

3.4.2 VC0509 protein

After PCT, crystallization plates were set up using 12 mg/mL and 6 mg/mL of VC0509 protein by robot as described earlier (Section 3.4.1). The best initial hits were obtained in PEGs screen (well number 12; 0.2 M ammonium sulfate, 0.1 M MES pH 6.5 and 25 % w/v PEG 2000 MME) (Figure 3.10.A). After optimization, good diffraction quality crystals were obtained using 7 mg/mL of

protein in the original condition (Figure 3.10.B). The crystals of VC0509 were cryoprotected using 20% (v/v) glycerol in crystallization buffer and diffracted X-rays at 100K in-house at wavelength of 1.54178 Å (Figure 3.10.B).

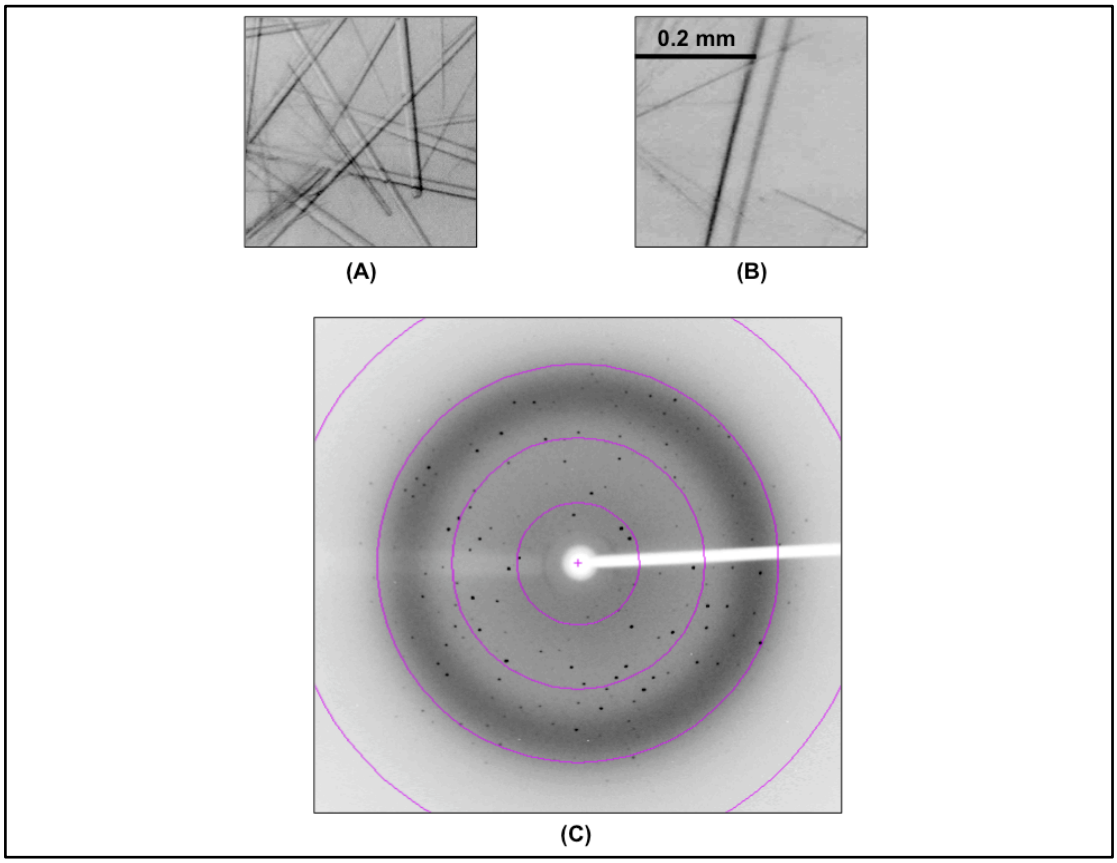


Figure 3.10 VC0509 protein crystal and diffraction pattern

(A) Initial hits PEGs screen (well number 12; 0.2 M ammonium sulfate, 0.1 M MES pH 6.5 and 25 % w/v PEG 2000 MME). **(B)** Optimized crystal, using the initial hit condition, containing 7 mg/mL of protein. **(C)** X-ray diffraction image VC0509 protein crystal. The resolution of the outer circle is 2.4 Å.

3.4.3 VC1805 protein

A pre-crystallization test, PCT, indicated the appropriate protein concentration (8 mg/mL) for crystallization screening. Three screens from Hampton research (Index, Crystal Screen and Crystal Screen 2) were used in initial crystallization trials of native VC1805. This experiment was carried out using

the sitting drop diffusion method following standard procedure supplied by the manufacturers. Crystals were observed in Index (33) (Figure 3.11.A) and Crystal Screen (41) (Figure 3.11.B). After optimization, the best quality crystals were obtained in 0.1 M HEPES-Na pH 7.5, 10% v/v 2-propanol, 24% w/v PEG 4,000 which is an optimization of condition 41 of Crystal Screen (Figure 3.11.C). Crystals were cryoprotected in crystallization buffer with 20% (v/v) glycerol added before being flash-frozen in a nitrogen gas stream at 100 K. X-ray data were collected on station ID14-1 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France (Figure 3.11.D).

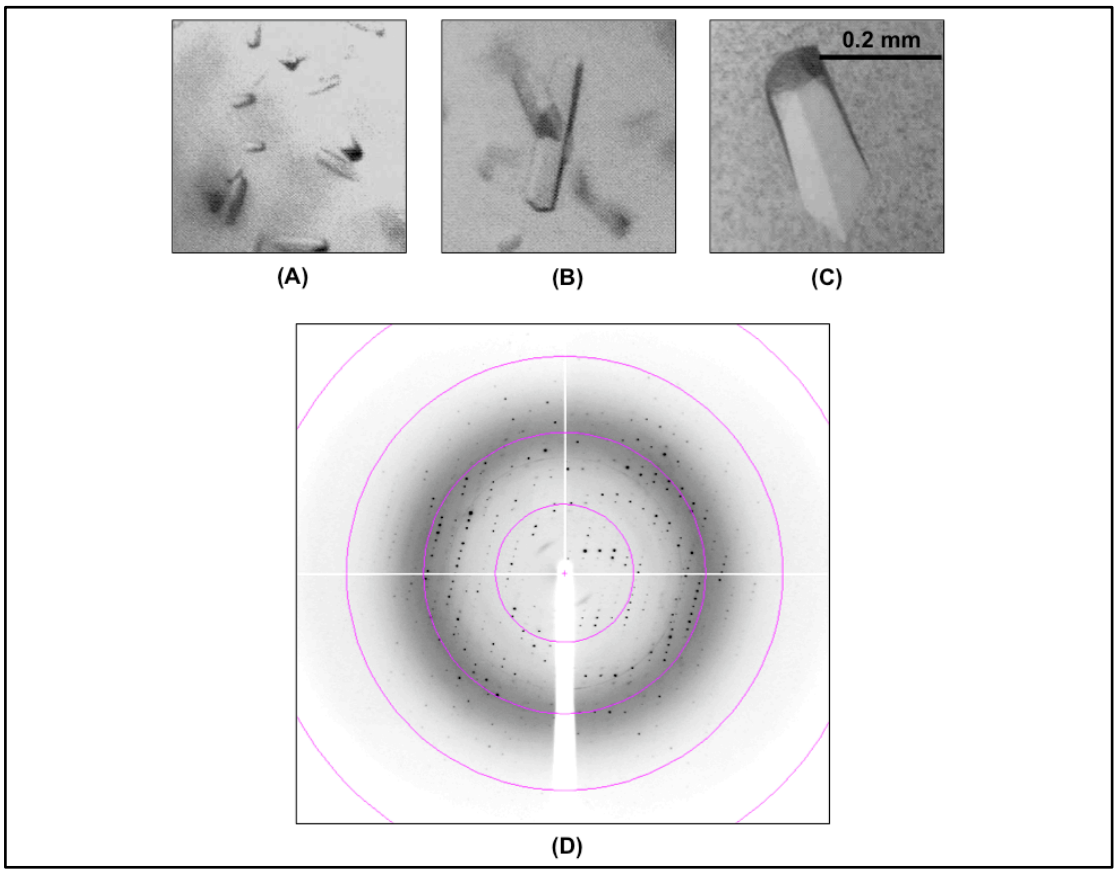


Figure 3.11 VC1805 protein crystal and diffraction pattern

(A) Initial hits in Index 33 (1.1 M sodium malonate pH 7.0, 0.1 M HEPES pH 7.0, 0.5% v/v jeffamine ED-2001 ® pH 7.0). **(B)** Initial hits in Crystal screen 41 (0.1 M HEPES-Na, pH 7.5, 10% v/v 2-propanol, 20% w/v PEG 4,000).

(C) Optimized crystal in 0.1 M HEPES-Na pH 7.5, 10% v/v 2-propanol, 24% w/v PEG 4000. **(D)** X-ray diffraction image of VC1805 optimized protein crystal. The resolution of the outer circle is 1.9 Å.

3.5 Structure solution and refinement

3.5.1 VC1805 crystal

The native cryoprotected crystals diffracted to a maximum resolution of 2.13 Å. The structure was solved by multiple heavy atom isomorphous replacement using data from a crystal soaked for 30 min in 5 mM mercury nitrate or 5 mM PCMBS and data were collected on station ID14-1 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data were indexed and integrated using the program MOSFLM (Leslie, 1992). The integrated data were merged and scaled using the programs of the CCP4 (CCP4., 1994) suite. The initial phases were obtained using the program SOLVE (Terwilliger and Berendzen, 1999) with density modification using the program RESOLVE (Terwilliger, 2000, 2003a, b), confirming the space group as $P3_121$. Manual model rebuilding and refinements were conducted using the programs O (Jones TA, 1991), COOT (Emsley and Cowtan, 2004) and REFMAC (Murshudov *et al.*, 1997). The final model had an R-factor of 0.235 and an R_{free} of 0.316. Parts of the protein have missing electron density as they are highly disordered, possibly contributing to the higher R_{free}. Part of the final $2F_o$-F_c electron density map is shown in Figure 3.12. The data collection and refinement statistics for native and heavy atom derivatives are summarized in Table 3.1. Atomic coordinates and structure factors of VC1805 have been deposited in the Protein Data Bank (PDB) with code 2v1l (www.rcsb.org).

Table 3.1 Data collection and refinement statistics of VC1805 crystal. Numbers in parentheses refer to highest resolution shell.

Data collection	**Crystal (data collected at synchrotron)**		
	Native	Mercury Nitrate	PCMBS
Unit cell parameters (Å) Space group $P3_121$	a = b = 78.3 c = 42.3	a = b = 78.4 c = 41.3	a = b = 78.7 c = 42.3
Maximum resolution (Å)	2.13 (2.19-2.13)	2.9 (3.06-2.9)	2.3 (2.42-2.3)
R_{merge}[a]	0.057 (0.375)	0.079 (0.359)	0.038 (0.278)
Observed reflections	75,489	17,207	35,617
Unique reflections	8,549	3,223	6,496
Completeness (%)	99.4 (99.4)	94.8 (94.8)	94.6 (96.7)
$\langle I/\sigma I \rangle$	23.2 (6.4)	21.8 (5.3)	29.2 (4.6)
Multiplicity	8.8 (9.0)	5.3 (5.3)	5.5 (4.7)
No. of heavy atom sites		2	1
<figure of merit > after SOLVE	0.43		
Refinement			
Resolution limits	39.2-2.13		
Number of reflections	8131		
R factor[b]	0.235		
Free R factor[c]	0.316		
RMSD ideal bond length (Å)	0.013		
RMSD ideal bond angles (°)	1.39		
No. of protein atoms	1054		
No. of water atoms	35		
Average B-factor for protein ($Å^2$)	39.2		
Average B-factor for water ($Å^2$)	40.4		

[a]*R*merge = $\sum_h \sum_j | I_{hj} - \langle I_h \rangle | / \sum_h \sum_j$, where I_{hj} is the intensity of the *j*th observation of unique reflection *h.*

[b]Overall R factor $\sum h || F_{oh}| - |F_{ch} || / \sum h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes for reflection *h.*

[c]Free R factor is equivalent to overall R factor , but is calculated using 5% of reflections excluded from the maximum-likelihood refinement stages.

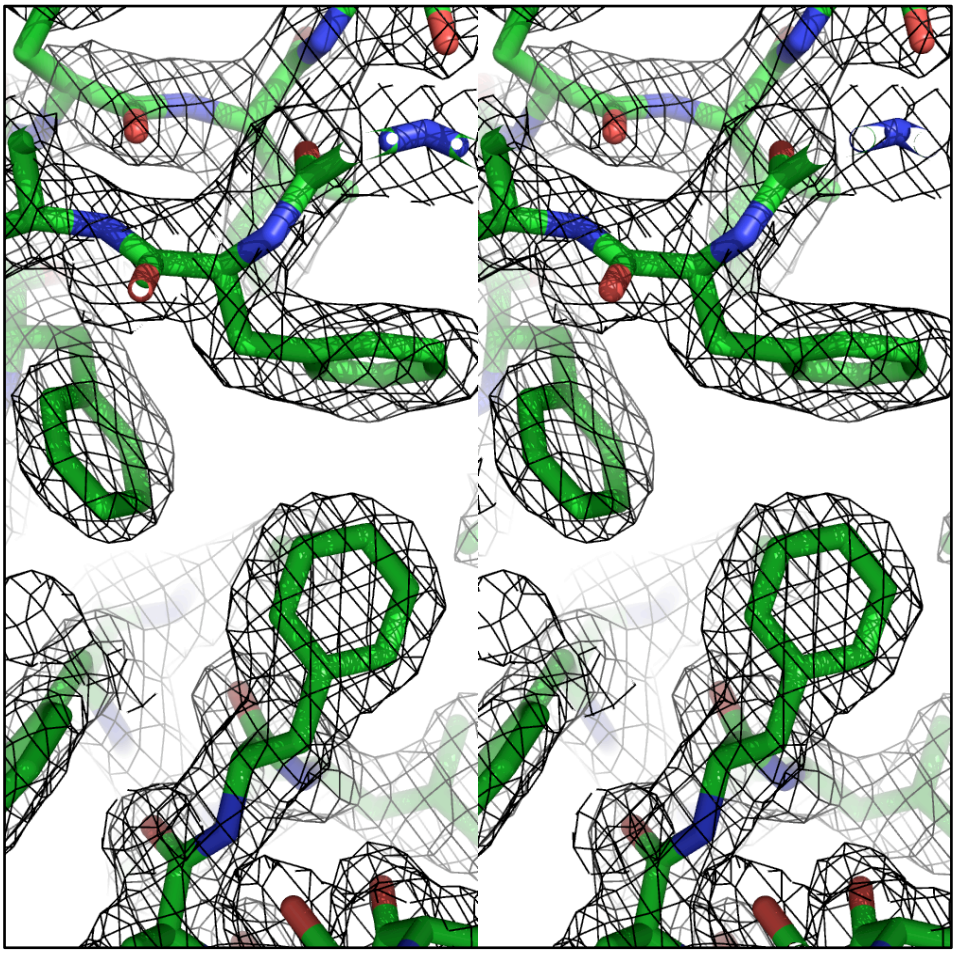


Figure 3.12 Stereo view of $2F_o$-F_c electron density map of VC1805. A region of the map is shown, contoured at 1σ. $2F_o$-F_c refers to the difference between twice the observed structure factor amplitudes and the calculated amplitudes. This and other figures were created with PyMOL (DeLano, 2007).

3.5.2 VC0508 crystal

A cryoprotected crystal of VC0508 diffracted to 1.9 Å and the data were integrated and scaled by using the program HKL2000 (Otwinowski and Minor, 1997) and the CCP4 (CCP4., 1994) suite. The spacegroup of the crystal was $P2_1$. Analysis of solvent content of VC0508 crystal, suggested two chains per asymmetric unit which gave 41% solvent content and a Matthews coefficient of 2.05 $Å^3$/Da (Kantardjieff and Rupp, 2003; Matthews, 1968). An automated molecular replacement server BALBES (Long *et al.*, 2008) was used to solve

the VC0508 structure using the VC1805 structure (PDB code 2v1l) as a search model confirming the spacegroup as $P2_1$. The search model has 59% sequence similarity to VC0508. The initial R-factor and R_{free} were 0.337 and 0.384, respectively. The partially built output model was completed manually using the program COOT (Emsley and Cowtan, 2004) followed by refinement using REFMAC (Murshudov *et al.*, 1997) and PHENIX (Adams *et al.*, 2002). The final R-factor was 0.227 and R_{free} was 0.275. Part of the final $2F_o$-F_c electron density map is shown in Figure 3.13. The statistics of data collection and refinement statistics are summarized in Table 3.2. Atomic coordinates and structure factors of VC0508 have been deposited in the PDB (2w56).

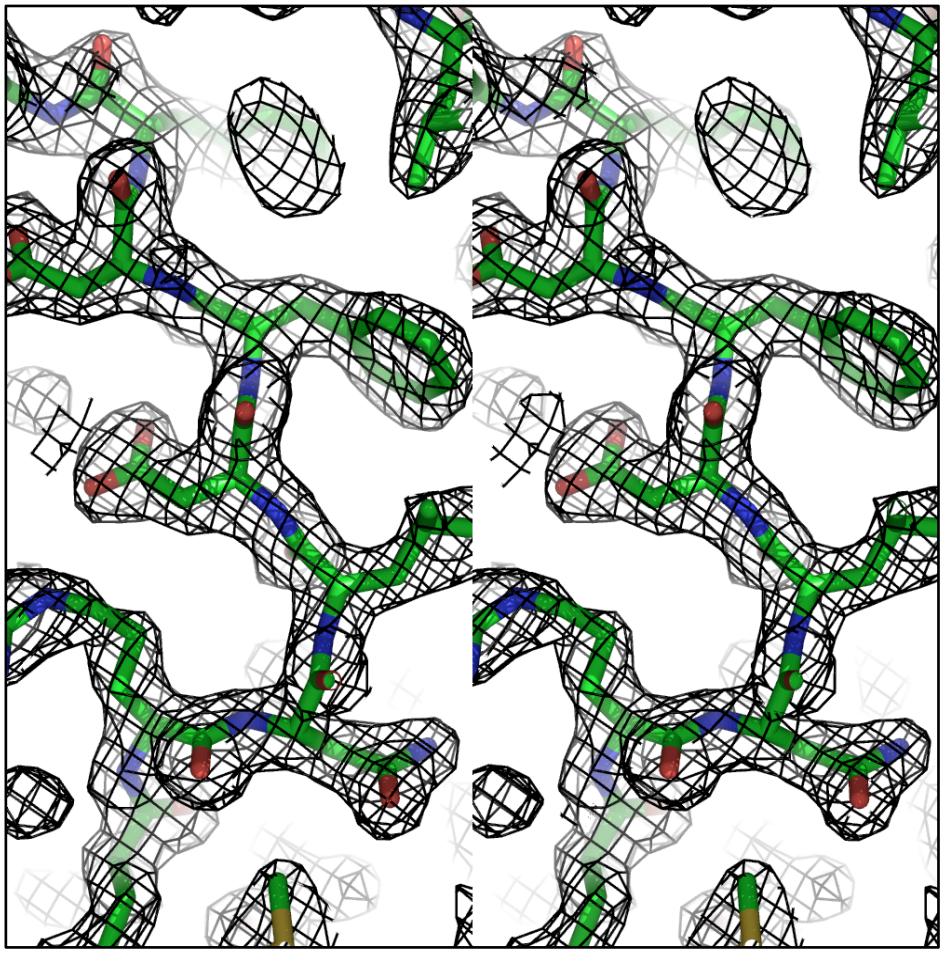


Figure 3.13 Stereo view of $2F_o$-F_c electron density map of VC0508. A region of the map is shown, contoured at 1σ.

Table 3.2 Data collection and refinement statistics of VC0508 crystal. Numbers in parentheses refer to highest resolution shell.

Data collection	**Crystal (data collected at in-house)**
Space group	$P2_1$
Unit cell parameters (Å)	a = 41.10, b = 70.60, c = 51.38, β = 110.64
Resolution (Å)	48.06 – 1.9 (1.97 – 1.9)
R_{merge}[a]	0.033 (0.304)
Observed reflections	135,957
Unique reflections	21,026
Completeness (%)	97 (96)
< I / σ(I) >	62 (5.6)
Multiplicity	6.5 (6.2)
Refinement	
No. of protein atoms	2441
No. of water atoms	194
Resolution limits (Å)	48.06 – 1.9
Number of used reflections	19,870
R factor[b]	0.227
Free R factor[c]	0.275
RMSD bond lengths (Å)	0.010
RMSD bond angles (°)	1.339
Average B-factor (Å^2) for protein	36.22

[a]*R*merge = $\sum_h \sum_j | I_{hj} - \langle I_h \rangle | / \sum_h \sum_j$, where I_{hj} is the intensity of the *j*th observation of unique reflection *h.*

[b]Overall R factor $\sum h || F_{oh}| - |F_{ch} || / \sum h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes for reflection *h.*

[c]Free R factor is equivalent to overall R factor , but is calculated using 5% of reflections excluded from the maximum-likelihood refinement stages.

3.5.3 VC0509 crystal

The crystals of VC0509 were cryoprotected and diffracted to 2.36 Å and the data were integrated and scaled by using the program HKL2000 (Otwinowski and Minor, 1997) and the CCP4 (CCP4., 1994) suite. The spacegroup of the crystal was H3. Two monomers were present per asymmetric unit of VC0508 crystal which gave 56% solvent content and a Matthews coefficient of 2.83 $Å^3$/Da as revealed by the analysis of solvent contents in the crystal (Kantardjieff and Rupp, 2003; Matthews, 1968). To obtain phases for VC0509, a automated molecular replacement server BALBES (Long *et al.*, 2008) was used. The search model was the VC1805 structure (PDB code 2v1l) which has 26% sequence identity to VC0509. The initial model had an R-factor of 0.428 and an R_{free} of 0.496. The partially built output model obtained from the server was built up manually using the program COOT (Emsley and Cowtan, 2004) followed by refinement using REFMAC (Murshudov *et al.*, 1997) and PHENIX (Adams *et al.*, 2002). This process was repeated several times and the final R-factor and R_{free} were 0.202 and 0.249, respectively. The statistics of data collection and refinement for VC0509 are summarized in Table 3.3. Part of the final $2F_o$-F_c electron density map is shown in Figure 3.14.

Table 3.3 Data collection and refinement statistics of VC0509 crystal. Numbers in parentheses refer to highest resolution shell.

Data collection	**Crystal (data collected at in-house)**
Space group	H3
Unit cell parameters (Å)	a = b = 153.77, c = 42.21
Resolution (Å)	76.92 – 2.36 (2.44 – 2.36)
R_{merge}[a]	0.048 (0.302)
Observed reflections	42,268
Unique reflections	15194
Completeness (%)	99 (93)
< I / σ(I) >	15.7 (2)
Multiplicity	2.8 (2.4)
Refinement	
No. of protein atoms	2520
No. of water atoms	124
Resolution limits (Å)	76.92 – 2.36
Number of used reflections	14412
R factor[b]	0.218
Free R factor[c]	0.259
RMSD bond lengths (Å)	0.010
RMSD bond angles (°)	1.209
Average B-factor (Å^2) for protein	50.65

[a]*R*merge = $\sum_h \sum_j | I_{hj} - \langle I_h \rangle | / \sum_h \sum_j$, where I_{hj} is the intensity of the *j*th observation of unique reflection *h.*

[b]Overall R factor $\sum h || F_{oh}| - |F_{ch} || / \sum h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes for reflection *h.*

[c]Free R factor is equivalent to overall R factor , but is calculated using 5% of reflections excluded from the maximum-likelihood refinement stages.

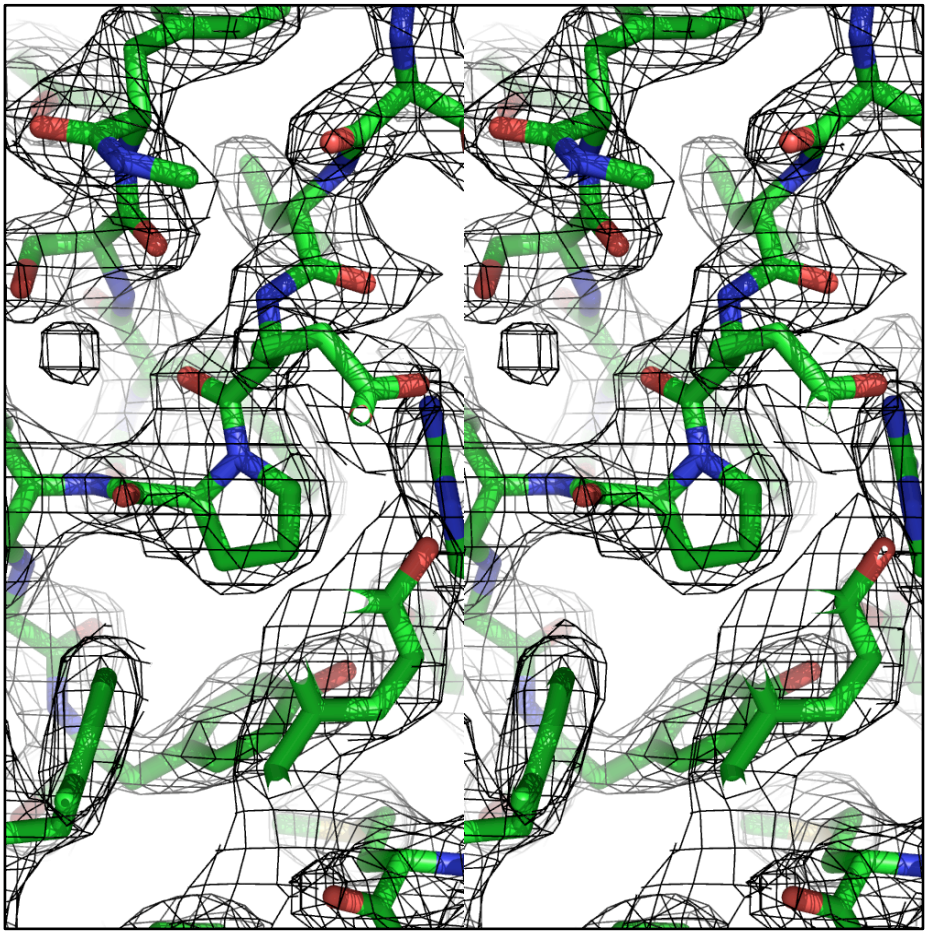


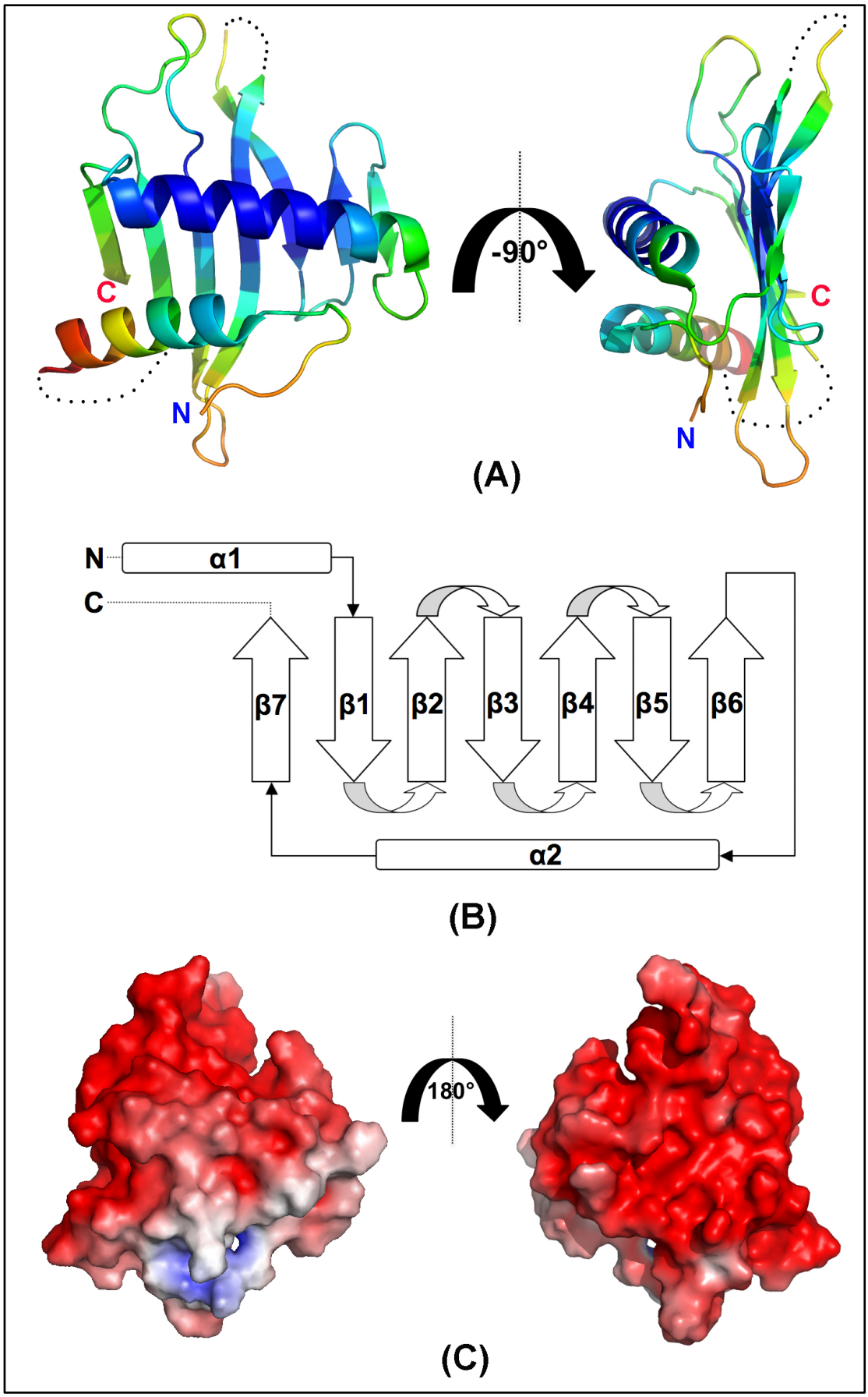
Figure 3.14 Stereo view of $2F_o$-F_c electron density map of VC0509. A region of the map is shown, contoured at 1σ.

3.6 Overall structure

3.6.1 Structure of VC1805

Using multiple heavy atom isomorphous replacement, the crystal structure of VC1805 was solved at a resolution of 2.13 Å. The crystals belong to the $P3_121$ spacegroup and contain one molecule per asymmetric unit. The crystal structure of VC1805 reveals a flat seven-stranded anti-parallel β-sheet with two helices sitting on the same face of the sheet. The first helix (α1) is at the N-terminus, followed by six β-strands (β1 to β6) of varying length that form a simple consecutive up-down topology. The second helix (α2) then stretches

back across these six β-strands, leading to a seventh strand (β7) at the C-terminus that completes the sheet. There are four disordered regions where no electron density was observed: the first four residues (1-4), the connection between α1 and β1 (31-39), the loop connecting β3 and β4 (84-86) and the final three residues (146-148). A cartoon representation based on crystallographic B-factor and surface topology of VC1805, is shown in Figure 3.15.A-B. The theoretical pI of the protein is 4.99 which is reflected by the negatively charged surface calculated by APBS (Baker *et al.*, 2001) shown in Figure 3.15.C.The structure of VC1805 was validated by the Molprobity server which shows that the structure is in the top 37% for the corresponding resolution with a Molprobity score of 2.22 (64th percentile). In the Ramachandran plot 96.8% of the residues are present in the favored regions and the rotamer outliers are 2.7% (Davis *et al.*, 2007).



3.15 Structure of VC1805

(A) Cartoon representation of the VC1805 structure which is colored by crystallographic B-factor (dark blue, B~19 $Å^2$ to red, B~76 $Å^2$). The black dotted lines are the disordered regions, N denotes N-terminus and C denotes C-terminus of the protein. **(B)** Surface topology of chain A of VC1805. There are a total of seven anti-parallel β-strands and two α-helices. **(C)** Electrostatic surface of VC1805 from -6 kT/e to +6 kT/e, calculated using APBS (Baker *et al.*, 2001).

3.6.2 Structure of VC0508

The crystal structure of VC0508 was solved to a resolution of 1.9 Å using molecular replacement. The crystal belongs to the $P2_1$ spacegroup and contains two molecules per asymmetric unit. Residues 3-83 and 86-147 of monomer A, and residues 5-28 and 40-146 monomer B are well ordered. The RMSD (root mean square deviation) fit between chain A and B is 0.98 Å, revealed by SSM (Krissinel and Henrick, 2004) analysis. The buried surface area of VC0508 in the asymmetric unit is 720 $Å^2$ with the complexation significance score (CSS) of 0.0 which was calculated by the PISA server (Krissinel and Henrick, 2007). The CSS score may depend on the interface area, residue/ atom composition and contacts, hydropathy index, charge distribution and topological complementarity (Krissinel and Henrick, 2007). The CSS score ranges from 0 to 1 as interface relevance to complexation increases. The results therefore indicate that the VC0508 interface might be an artifact of crystal packing. A cartoon representation of VC0508 (chain A) is shown in Figure 3.16.A, which is based on crystallographic B-factor. The structure of VC0508 has the same topology as VC1805. The surface topology and negatively charged surface of chain A of VC0508 is shown in Figure 3.16.B-C. Molecules in the asymmetric and unit cell of the crystal of

VC0508 is shown in Figure 3.17 and Figure 3.18, respectively. The final coordinates of the protein were analysed by the Molprobity server which shows that the structure is in the top 20% for the corresponding resolution with a Molprobity score of 1.55 (94th percentile). In the Ramachandran plot 97.7% of the residues are present in the favored regions and the rotamer outliers are only 0.4% (Davis *et al.*, 2007).

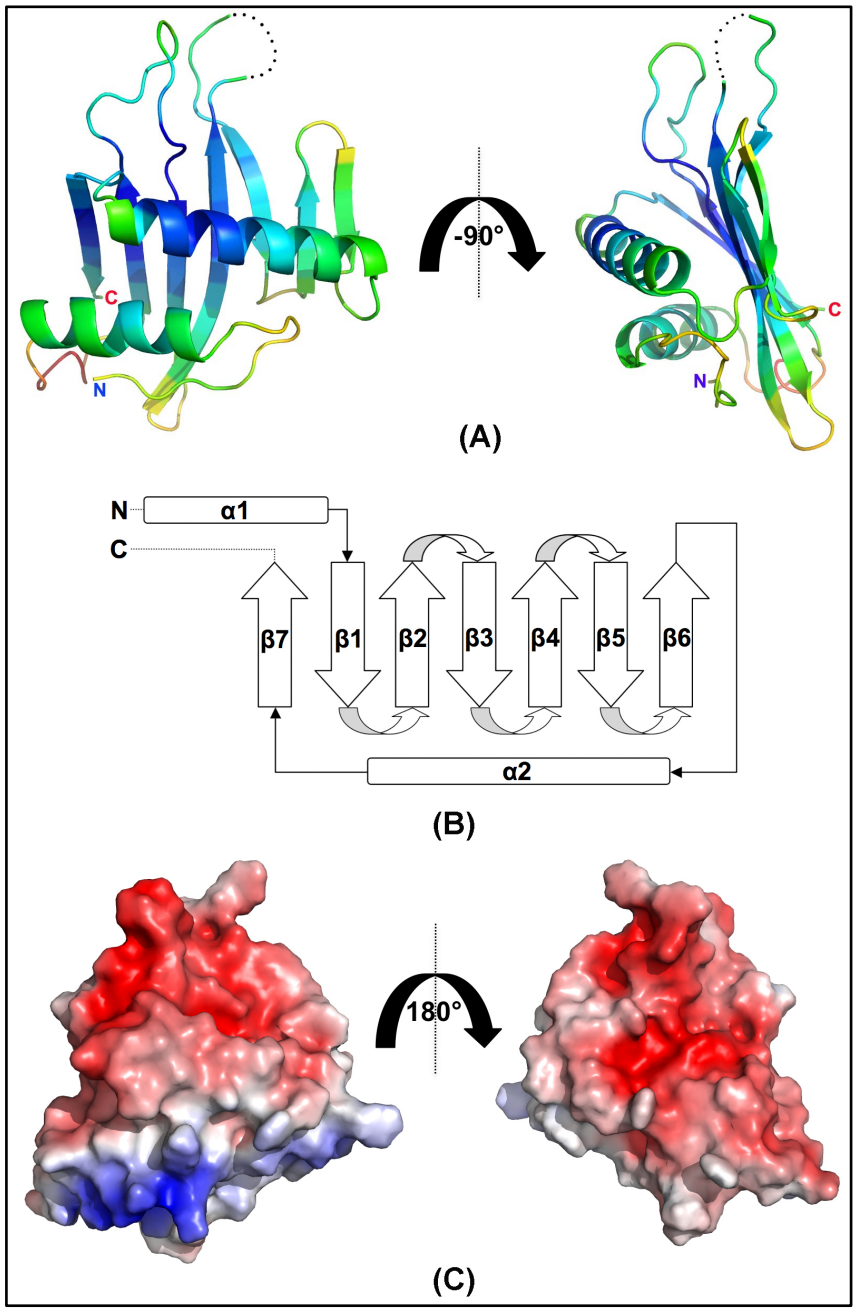


Figure 3.16 Structure of VC0508

(A) Cartoon representation of the VC0508 structure (chain A) which is colored by crystallographic B-factor (dark blue, B~23Å^2 to red, B~52Å^2). The black dotted lines are the disordered region, N denotes N-terminus and C denotes C-terminus of the protein. **(B)** Surface topology of chain A of VC0508. A total of seven anti-parallel β-strands lie on top of two α-helices. **(C)** Electrostatic surface of chain A of VC0508 from -6 kT/e to +6 kT/e, calculated using APBS (Baker *et al.*, 2001).

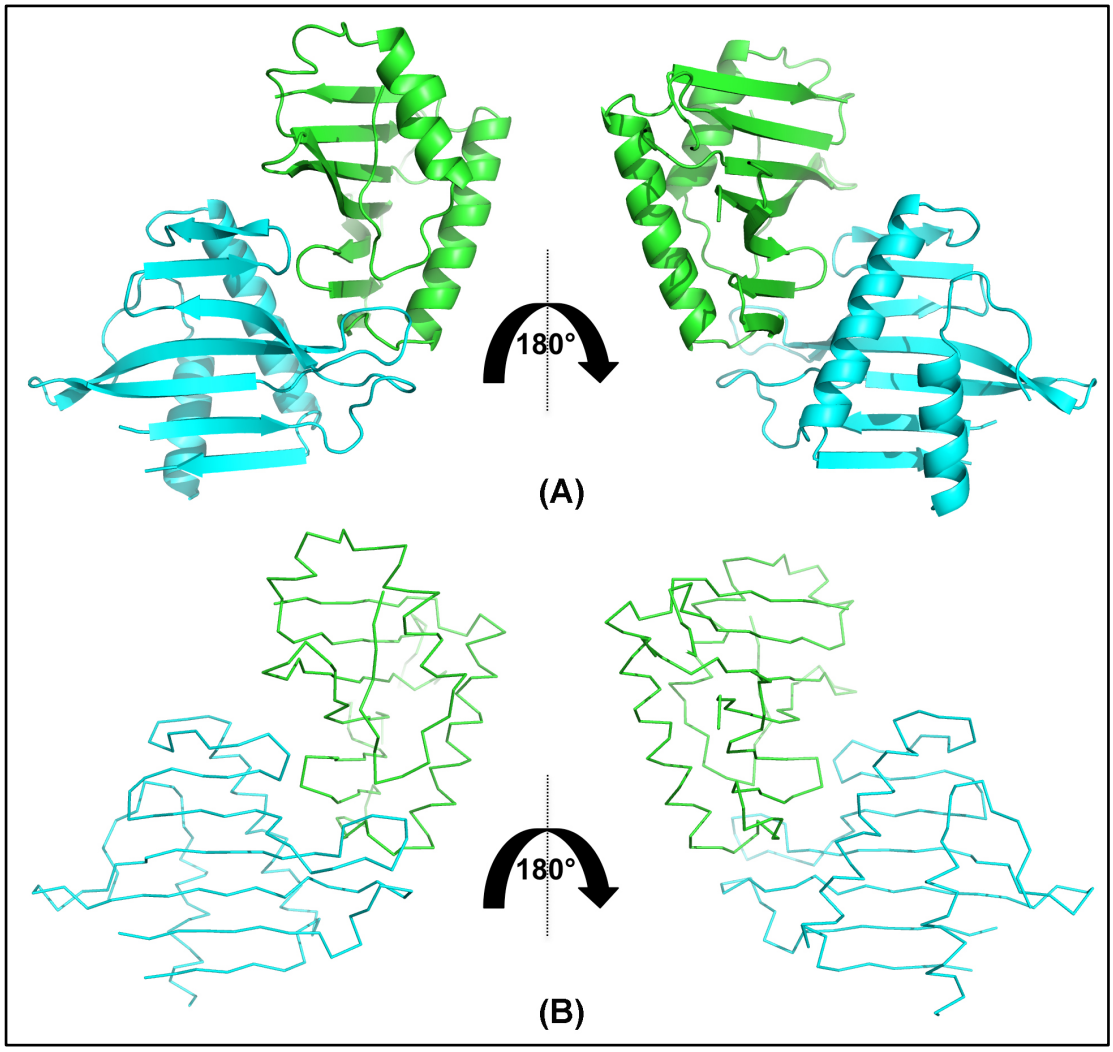


Figure 3.17 Molecules in the asymmetric unit of VC0508 crystal

(A) The crystallographic dimeric form of VC0508. Chain A is green in colour whereas chain B is cyan in colour. **(B)** Ribbon representation of VC0508 dimer.

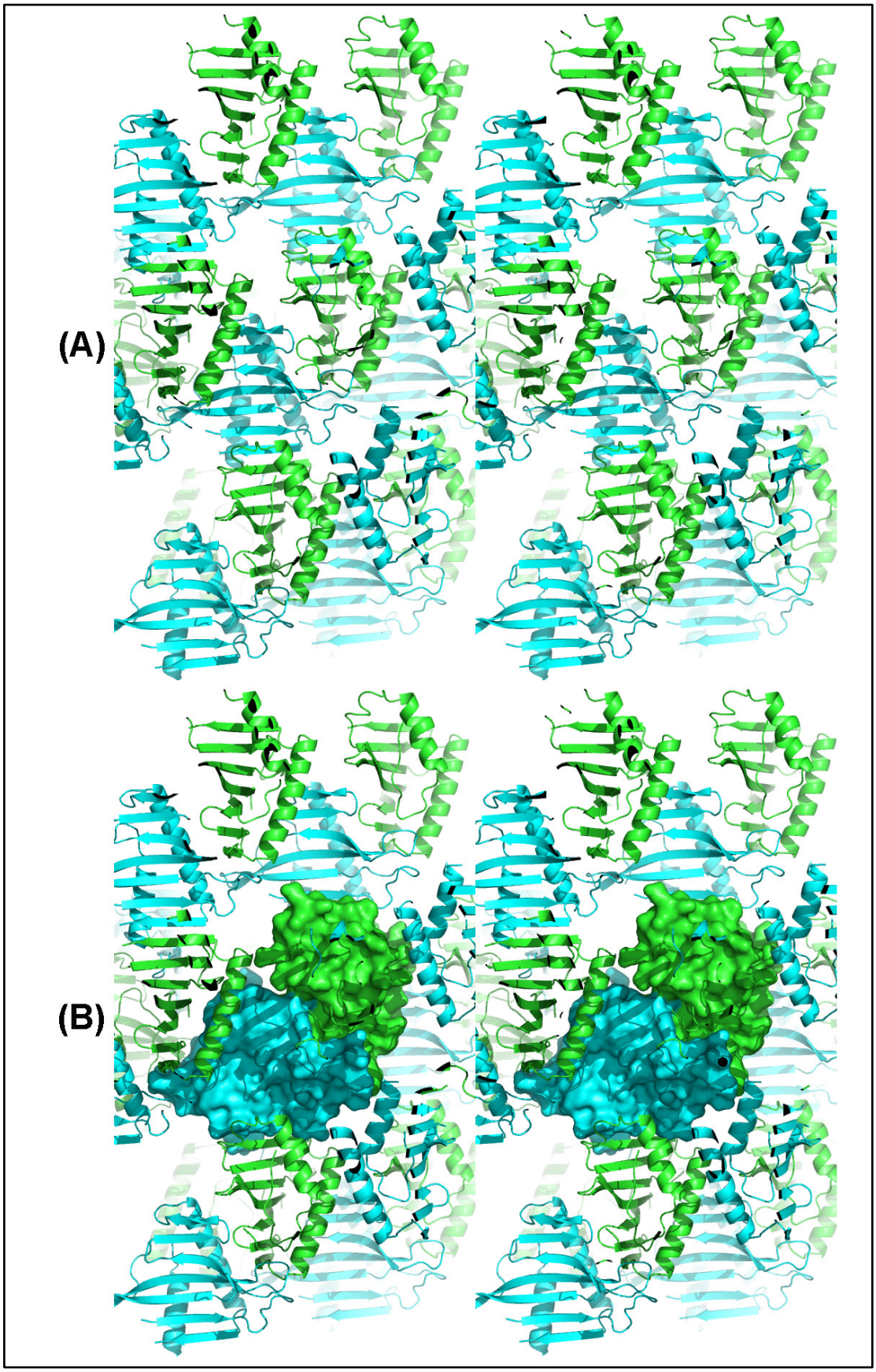


Figure 3.18 Molecules in the unit cell of VC0508 crystal. In the crystal, the protein is packed in such a way that β-strands of adjacent molecules form a

continuous β-sheet. This is likely to be an artifact of crystallization. **(A)** Stereo view of molecules in $P2_1$ unit cell of VC0508. Chain A is green in colour whereas chain B is cyan in colour **(B)** Stereo view of molecules in $P2_1$ unit cell of VC0508 with the surface of chain A and B in the asymmetric unit.

3.6.2 Structure of VC0509

The VC0509 crystal diffracted to a resolution of 2.36 Å, and the structure was solved by molecular replacement. The asymmetric unit of the VC0509 crystal contains two molecules of protein and the crystal belongs to the spacegroup H3. The last residue (Asn147) is missing in both chains A and B. The RMSD fit between chain A and B of VC0509 structure is 0.85 Å, revealed by SSM (Krissinel and Henrick, 2004) analysis. Analysis by the PISA server (Krissinel and Henrick, 2007) revealed that the buried surface area of VC0509 in the asymmetric unit, which contains two molecules, is 1012 $Å^2$ with the CSS score of 0.1, indicating that the interface might be the result of crystal packing. The overall structure chain A of VC0509, coloured by crystallographic B-factor is shown in Figure 3.19.A. The N-terminus of the protein starts with the first helix (α1) followed by first anti-parallel β-sheet (β1), a small α-helix (α2), five consecutive anti-parallel β-sheets (β2 to β6) connected to the third helix (α3) and the seventh anti-parallel β-sheet (β7). The surface topology and negatively charged surface of chain A of VC0509 is shown in Figure 3.19.B-C. Molecules in the asymmetric and unit cell of the crystal of VC0508 are shown in Figure 3.20 and Figure 3.21, respectively. The overall structural quality was analyzed by the Molprobity server which shows the structure is in the top 13% for the corresponding resolution with a molprobity score of 1.6 (99^{th} percentile). In the Ramachandran plot 98.7% of the residues are present in

the favored regions and the rotamer outliers are only 0.7% (Davis *et al.*, 2007).

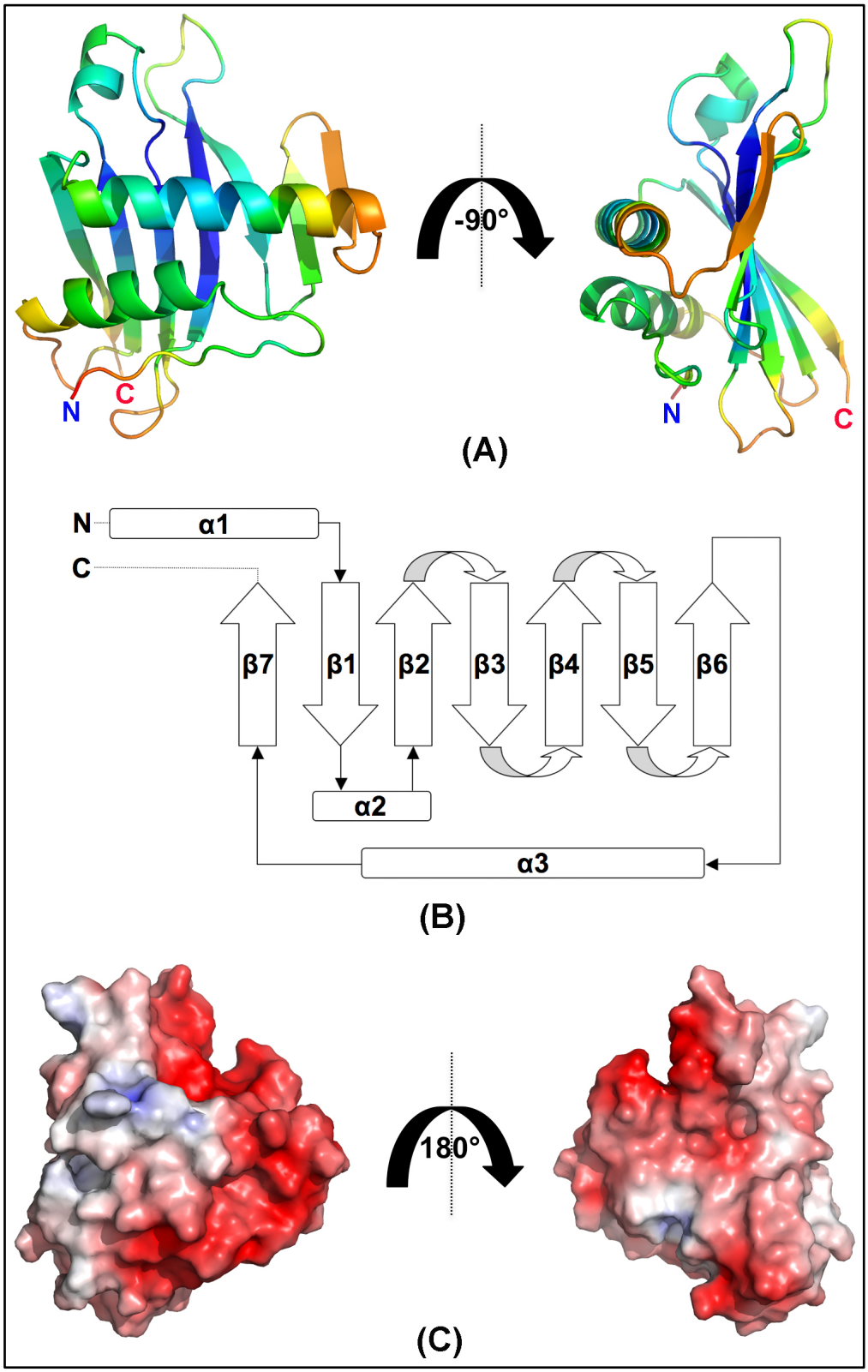


Figure 3.19 Structure of VC0509

(A) Cartoon representation of the VC0509 structure (chain A) which is colored by crystallographic B-factor (dark blue, B~37 Å^2 to red, B~69 Å^2). N denotes N-terminus and C denotes C-terminus of the protein. **(B)** Surface topology of chain A of VC0509. There are a total of seven anti-parallel β-strands and three α-helices. **(C)** Electrostatic surface of chain A of VC0509 from -6 kT/e to +6 kT/e, calculated using APBS (Baker *et al.*, 2001).

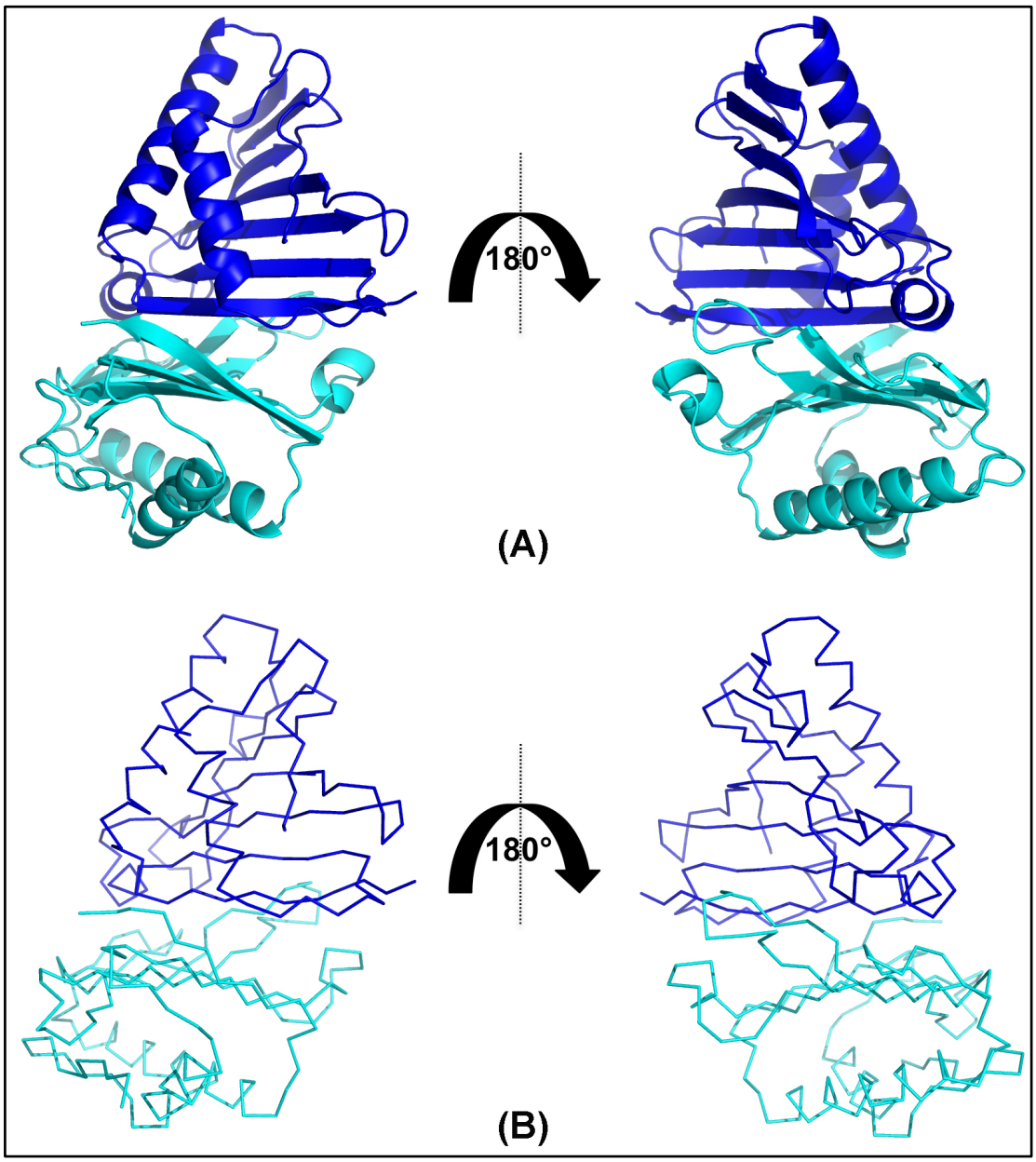


Figure 3.20 Molecules in the asymmetric unit of VC0509 crystal

(A) The crystallographic dimeric form of VC0509. Chain A is blue in colour whereas chain B is cyan in colour. **(B)** Ribbon representation of VC0509 dimer.

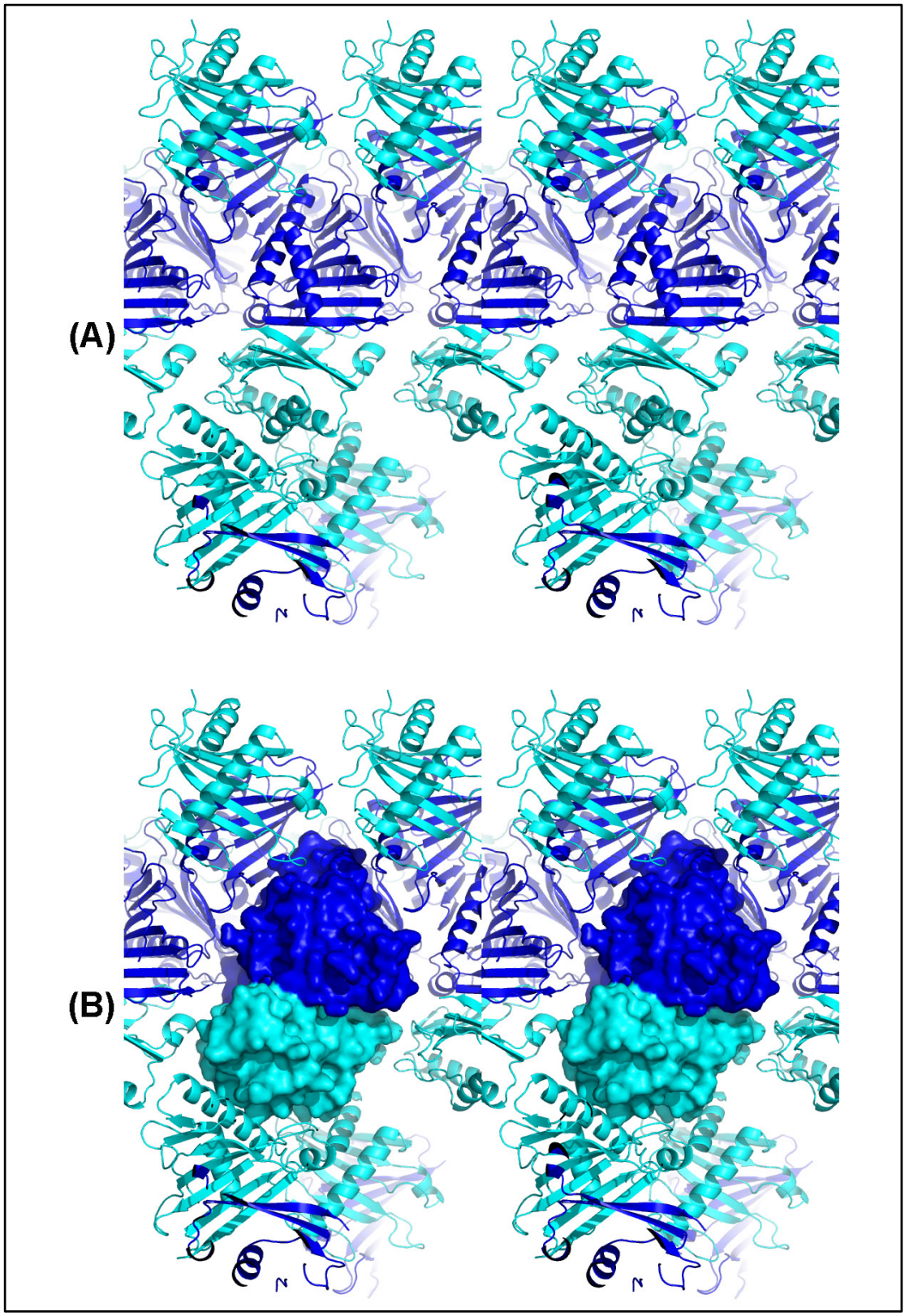


Figure 3.21 Molecules in the unit cell of VC0509 crystal. **(A)** Stereo view of molecules in H3 unit cell of VC0509. Chain A is blue in colour whereas chain B is cyan in colour **(B)** Stereo view of molecules in H3 unit cell of VC0509 with the surface of chain A and B in the asymmetric unit.

3.7 Multiple sequence alignments and conserved amino acids

A sequence similarity search with BLAST (Altschul *et al.*, 1997) and multiple sequence alignment with CLUSTAL-W (Chenna *et al.*, 2003) indicated that VC0508 is a conserved hypothetical protein with high sequence identity among different *Vibrio* spp. as well as other pathogenic bacteria. The sequence identities of VC0509, VC1804, VC1805, *V. cholerae* non-O1, *V. harveyi*, *V. splendidus*, *V. vulnificus*, *V. alginolyticus*, *V. fischeri*, *Providencia stuartii*, *Yersinia intermedia* and *Aeromonas hydrophila* to VC0508 are 26, 20, 59, 99, 84, 81, 94, 50, 48, 38, 45 and 37%, respectively (Figure 3.22). All of the proteins listed in Figure 3.18 carry a negative charge having isoelectric points between 4.44 and 5.36. There are 9 amino acids that are invariant among this family and these are mapped onto the surface of VC0508 (chain A) and coloured blue in Figure 3.23.A. If we analyze the sequence identities among VC0509, VC1804, VC1805 and VC1805, then a total of 16 amino acids residues are conserved, indicated by the green arrows in Figure 3.22, which are also mapped onto the surface of VC0508 in red and shown in Figure 3.24.A. There appear to be two conserved patches, both carrying a negative potential, on opposite sides of the protein. One patch forms a groove formed between α2 and the loop connecting β1 and β2 with the invariant Tyr50 and His58 coming close in space at one end of the groove (Figure 3.23.B, left hand image). The other patch is centred around another groove on the opposite face that includes a deep hole lined with invariant residues Pro59, Glu61, Phe80 (Figure 3.23.B, right hand image). It is interesting to note that VC1804 would be predicted to be missing the first helix, a region that does not contain any conserved residues. Among the

conserved residues, Pro59, Gly56 and Gly109 might have a structural role as they interrupt the regularity of the α-helical backbone conformation. Hydrophobic residues Tyr50, Trp72, Leu94, Phe96, Leu122, Trp126 and Phe130 are buried in the hydrophobic core and are likely to be structurally important (Figure 3.24.B). Surface exposed conserved residues are Glu61, Val42 and Thr145. Being hydrophobic, surface exposed Val42 might be involved in binding and recognition of hydrophobic ligands (Figure 3.24.B).

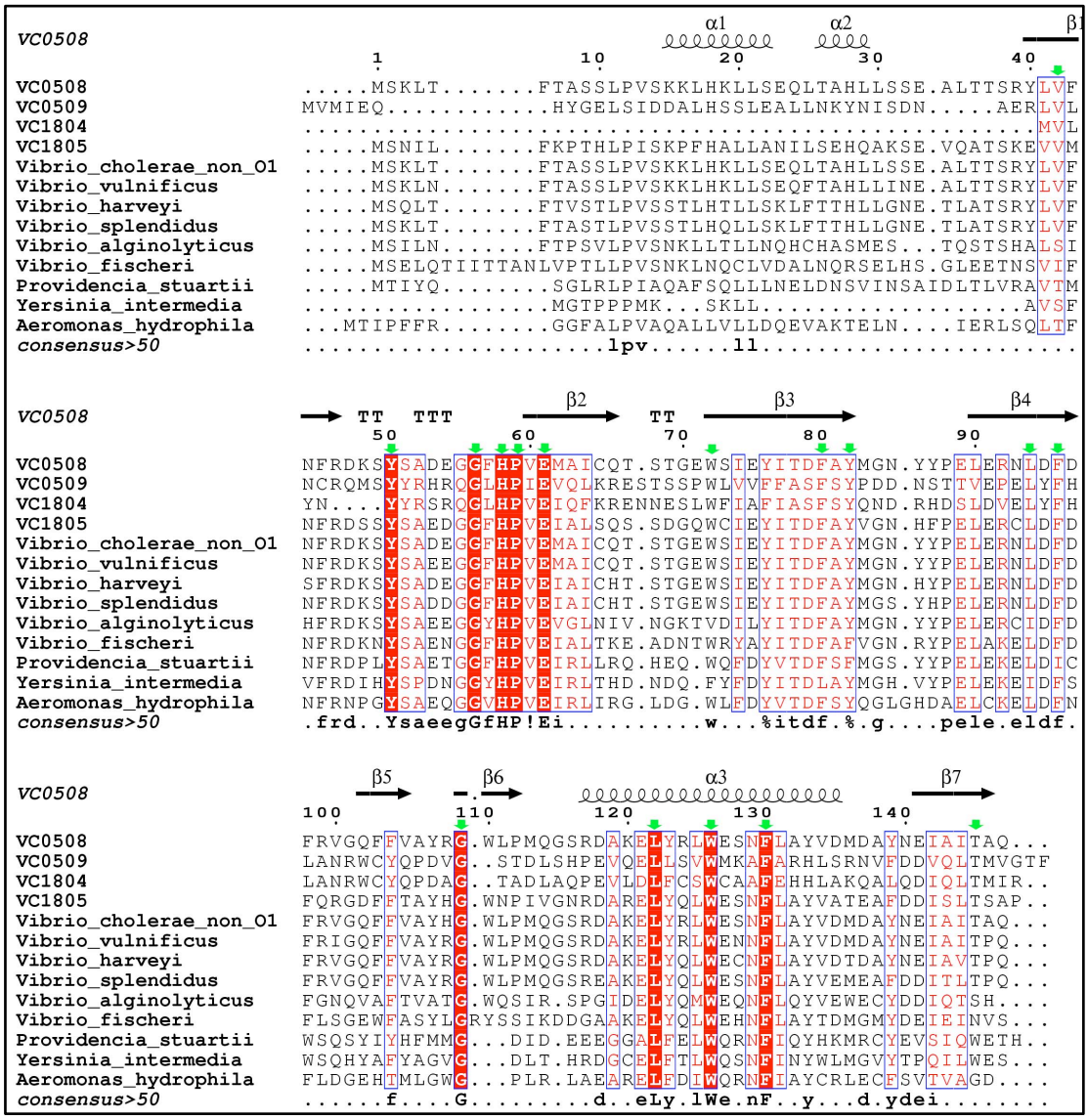


Figure 3.22 Multiple sequence alignment of VC0508 homologues. Green arrows indicate the conserved residues among VC0508, VC0509, VC1804 and VC1805. The secondary structure of VC0508 is shown on the top line.

NCBI entry name and genome annotation name for the sequences are *V. cholerae* VC0508 (NP_230159, VC0508), *V. cholerae* VC0509 (NP_230160, VC0509), *V. cholerae* VC1804 (NP_231439, VC1804), *V. cholerae* VC1805 (NP_231440, VC1805), *V. cholerae* non O1 (ZP_01956054, A51_B0486), *V. harveyi* (ZP_01986167, AIQ_2013), *V. splendidus* (ZP_00989426, V12B01_02775), *V. vulnificus* (NP_933316, VV0523), *V. cholerae* VC1805 (NP_231440, VC1805), *V. alginolyticus* (ZP_01262456, V12G01_09492), *V. fischeri* (YP_206478, VF_A0520), *Providencia stuartii* (ZP_02961693, PROSTU_03744), *Yersinia intermedia* (ZP_00835059, YintA_01000511) and *Aeromonas hydrophila* (YP_855622, AHA_1080). Alignment was created by with CLUSTAL-W (Chenna *et al.*, 2003) and the figure was produced by ESPRIPT (Gouet *et al.*, 1999).

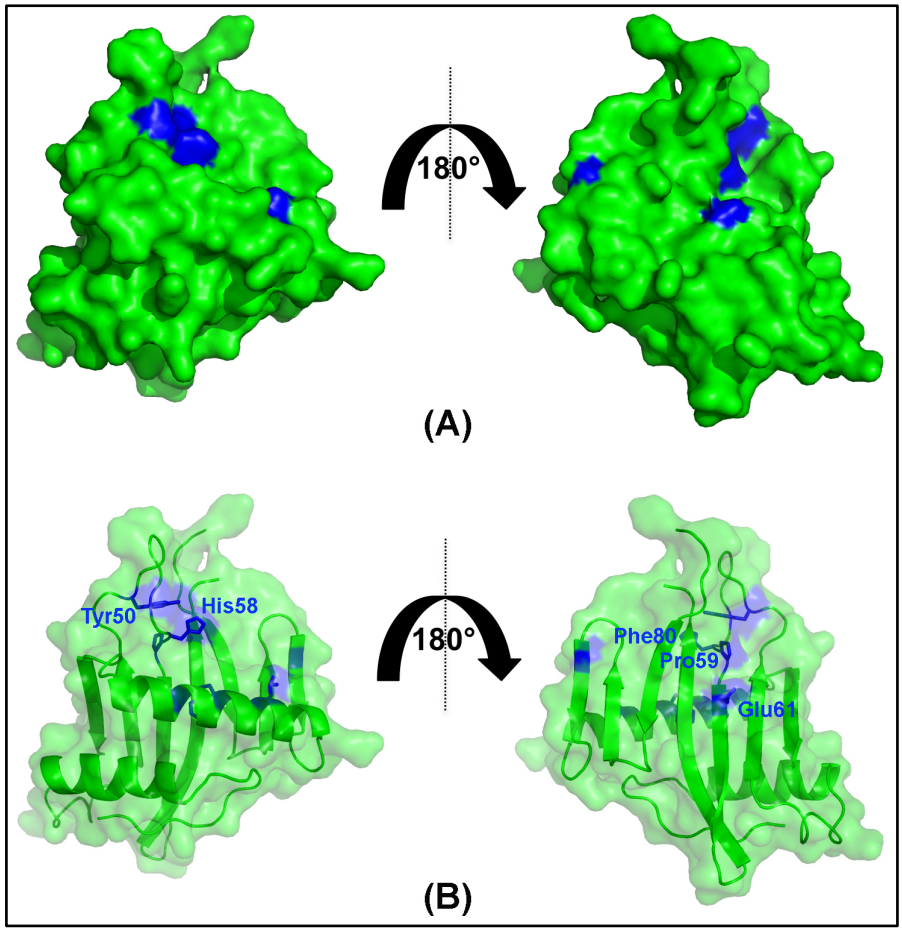


Figure 3.23 Location of the 9 conserved residues of VC0508. **(A)** Residues conserved across VC0508 homologues are mapped on the surface of VC0508 (chain A) and coloured in blue. **(B)** The positions of these conserved residues in the structure of VC0508 (chain A) are shown coloured in blue.

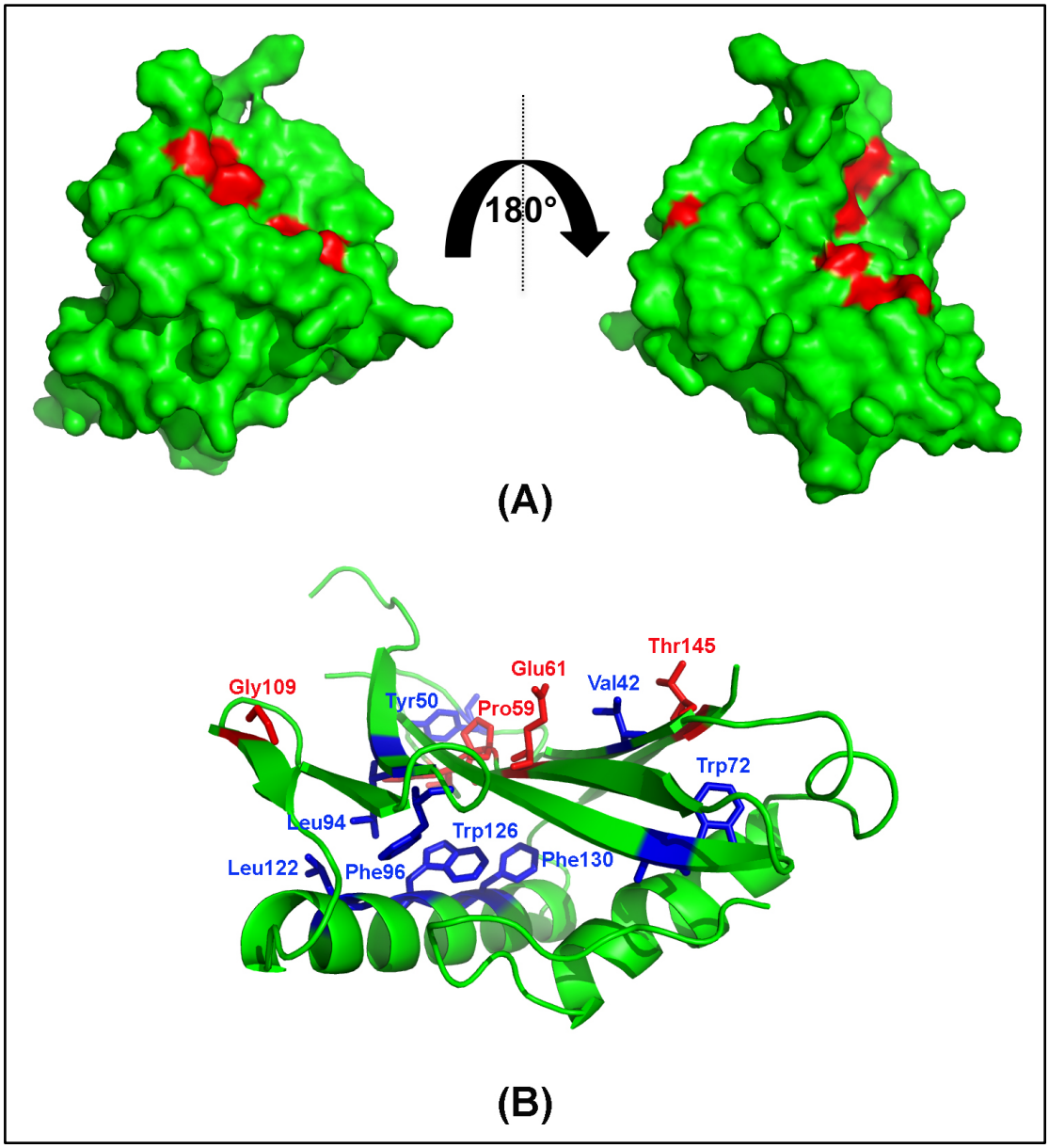


Figure 3.24 Location of the 16 conserved residues of VC0508

(A) Residues that are conserved among VC0508, VC0509, VC1804 and VC1805 are mapped onto the surface of VC0508 (chain A) and coloured in red. **(B)** The positions of these conserved residues in the structure of VC0508 (chain A). Hydrophobic conserved residues are coloured in blue and other conserved residues are coloured red.

Chapter 4

Functional studies of VC0508, VC0509, and VC1805 and discussion

4. Functional studies of VC0508, VC0509, and VC1805 and discussion

One of the proteins encoded by the 20 kb truncated version of VPI-2, named VC1805, is a hypothetical protein. Interestingly, a total of four proteins, VC0508, VC0509, VC1804 and VC1805 that are homologous/ paralogous to each other are present in two different pathogenicity islands, VSP-2 and VPI-2. In this chapter, the structural comparisons of VC0508, VC0509, and VC1805 with one another and with their structural homologue p32 will be discussed. One of the binding partners of p32 is a plasma complement component C1q, which has been found to bind to VC1805. Single crossovers for the construction of VC0508/ VC0509 and VC1804/ VC1805 double knockouts have also been produced.

4.1 Overall structural comparison

The topology of VC0508, VC0509 and VC1805 is remarkably similar. Superposition of these structures is shown in Figure 4.1. Using SSM (Krissinel and Henrick, 2004) analysis, chain A of VC0508 shows an RMSD fit of 2.27 Å over 136 Cα atoms (sequence identity 18%) with chain A of VC0509 and RMSD fit of 1.33 Å over 129 Cα atoms (sequence identity 60%) with VC1805, respectively. Chain A of VC0509 shows an RMSD fit of 1.74 Å over 121 Cα atoms (sequence identity 26%) with VC1805. Sequence comparisons had failed to suggest a fold or function for VC0508, VC0509 and VC1805. The coordinates of VC0508, VC0509 and VC1805 were submitted to the ProFunc server (Laskowski *et al.*, 2005), that revealed a structural homology with human p32 (PDB code 1p32) an acidic mitochondrial matrix protein. Human p32 is a homotrimer with each chain containing three helices and a seven-

stranded anti-parallel β-sheets, and has a pI of 4 (Jiang *et al.*, 1999). SSM (Krissinel and Henrick, 2004) analysis revealed that, respectively, chain A of VC0508, chain A of VC0509 and VC1805 have RMSD fits of 2.59 Å over 72 Cα atoms (sequence identity 4%); 2.38 Å over 73 Cα atoms (sequence identity 8%) and 2.51 Å over 70 Cα atoms (sequence identity 4%) with chain A of human p32 protein, respectively.

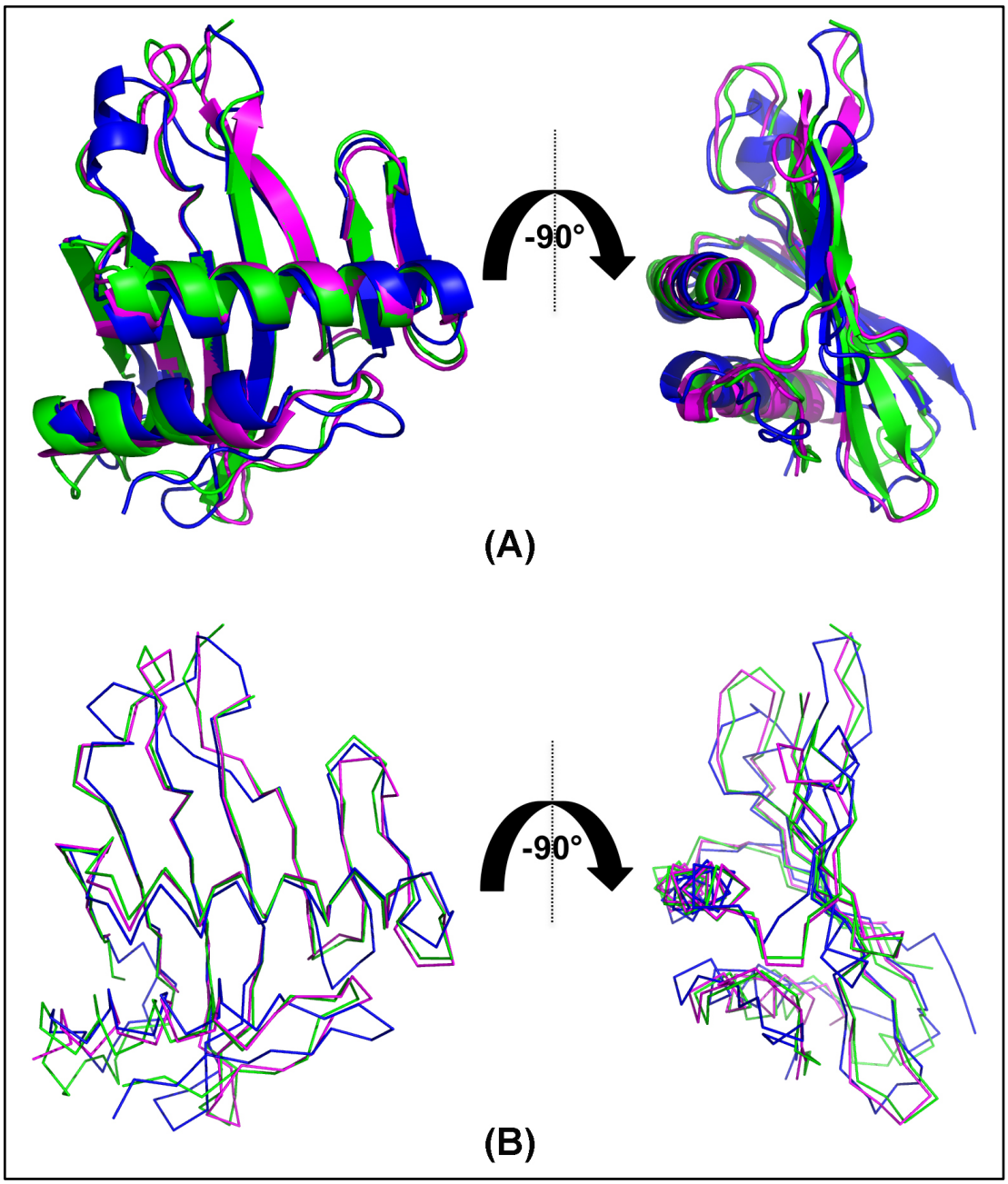


Figure 4.1 Superposition of VC0508, VC0509 and VC1805

(A) Cartoon representation of VC0508 (green), VC0509 (blue) and VC1805 (red). **(B)** Ribbon representation of VC0508 (green), VC0509 (blue) and VC1805 (magenta).

4.2 Structural comparison of VC1805 and human p32

The structures of VC0508, VC0509 and VC1805 show significant structural homology to human p32 protein but the sequence identities are only 11, 11 and 6%, respectively. Although p32 has a low pI, the charge distribution is asymmetric with one face of the p32 trimer carrying a predominantly negative charge, and the other face far less negative (Figure 4.2). There is little sequence similarity between VC1805 and p32, and the residues forming the two patches conserved within the VC1805 family are not conserved in p32. Conversely, the two regions conserved among the eukaryotic homologues of p32 are not conserved in VC1805 (Jiang *et al.*, 1999). The structural comparison of VC1805 and p32 reveals that the conserved fold between the two proteins involves the seven-stranded β-sheet and the helix (α2) that straddles the sheet, although the connectivity is slightly different (Figure 4.3).

Human p32 has two long helices at its N- and C-termini that are involved in the association with other monomers in the p32 trimer. The first β-strand after the N-terminal helix of p32 is in the same location as β7 in VC1805, but from then on the topology is the same for both proteins until the structurally conserved helix (α2) which in p32 is slightly shorter and turns 90° to form the C-terminal helix that associates with the long N-terminal helix. As noted above, in VC1805 the final strand of the sheet (β7) follows helix α2. The only homologous structure to p32 in the protein databank is that of a hypothetical

protein LMAJ011689 from *Leishmania major* (PDB code 1yqf) that shares only 16% sequence identity with human p32 and 15% sequence identity with VC1805. Like p32, LMAJ011689 has a low pI of 4.5, and possesses the same trimeric structure with monomers having an RMSD fit of 1.64 Å over 155 Cα atoms using SSM (Krissinel and Henrick, 2004). The p32 fold is therefore an unusual fold in the current protein databank, so the fact that VC1805 shares the same core fold of a helix straddling a seven-stranded anti-parallel β-sheet, and also has a dominant negative charge potential, suggests that the proteins may have a related function.

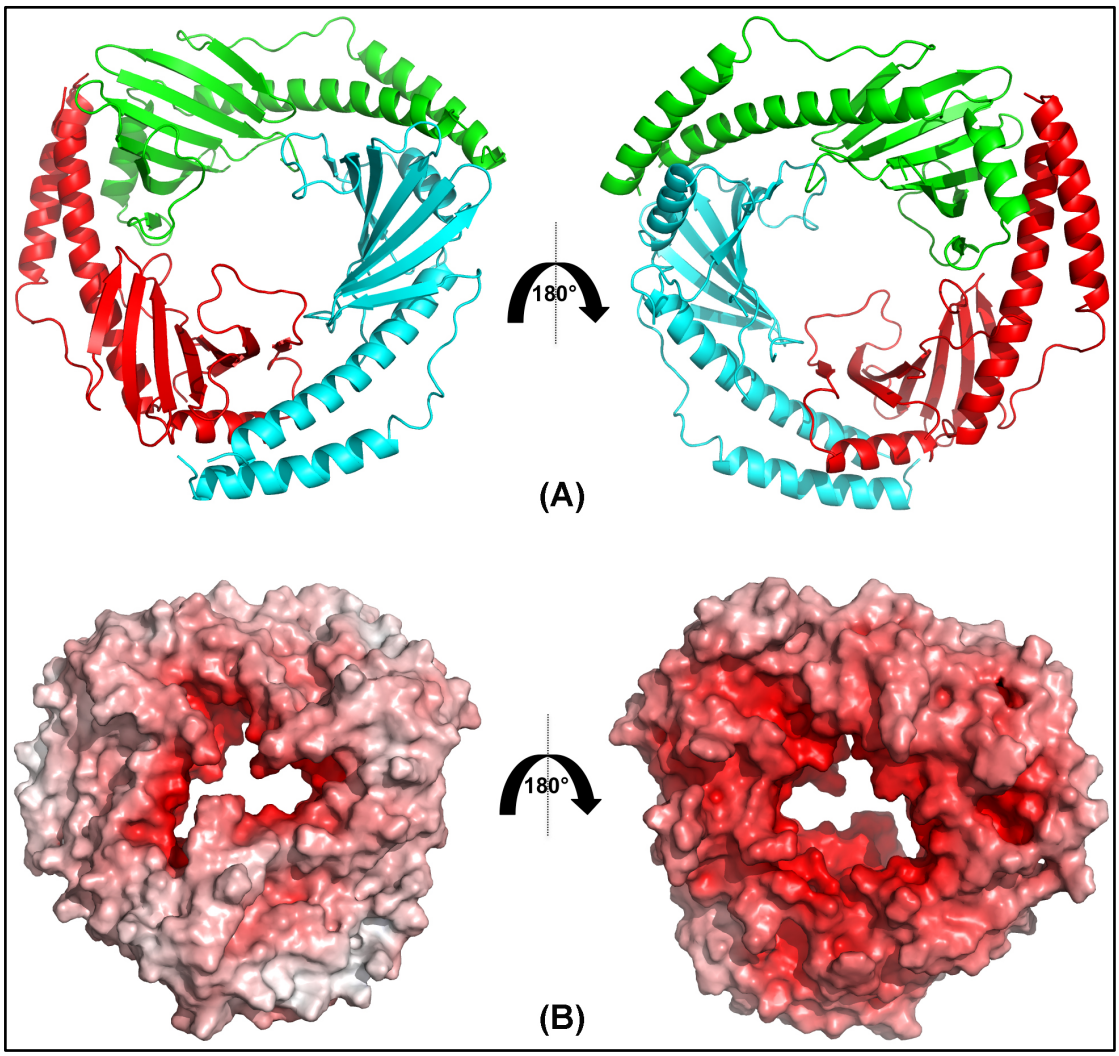


Figure 4.2 Structure of human mitochondrial protein p32

(A) Cartoon representation of p32 protein which is coloured by chain where chain A is cyan, chain B is green and chain C is red. **(B)** Electrostatic surface of p32 protein from -26 kT/e to +26 kT/e, calculated using APBS (Baker *et al.*, 2001).

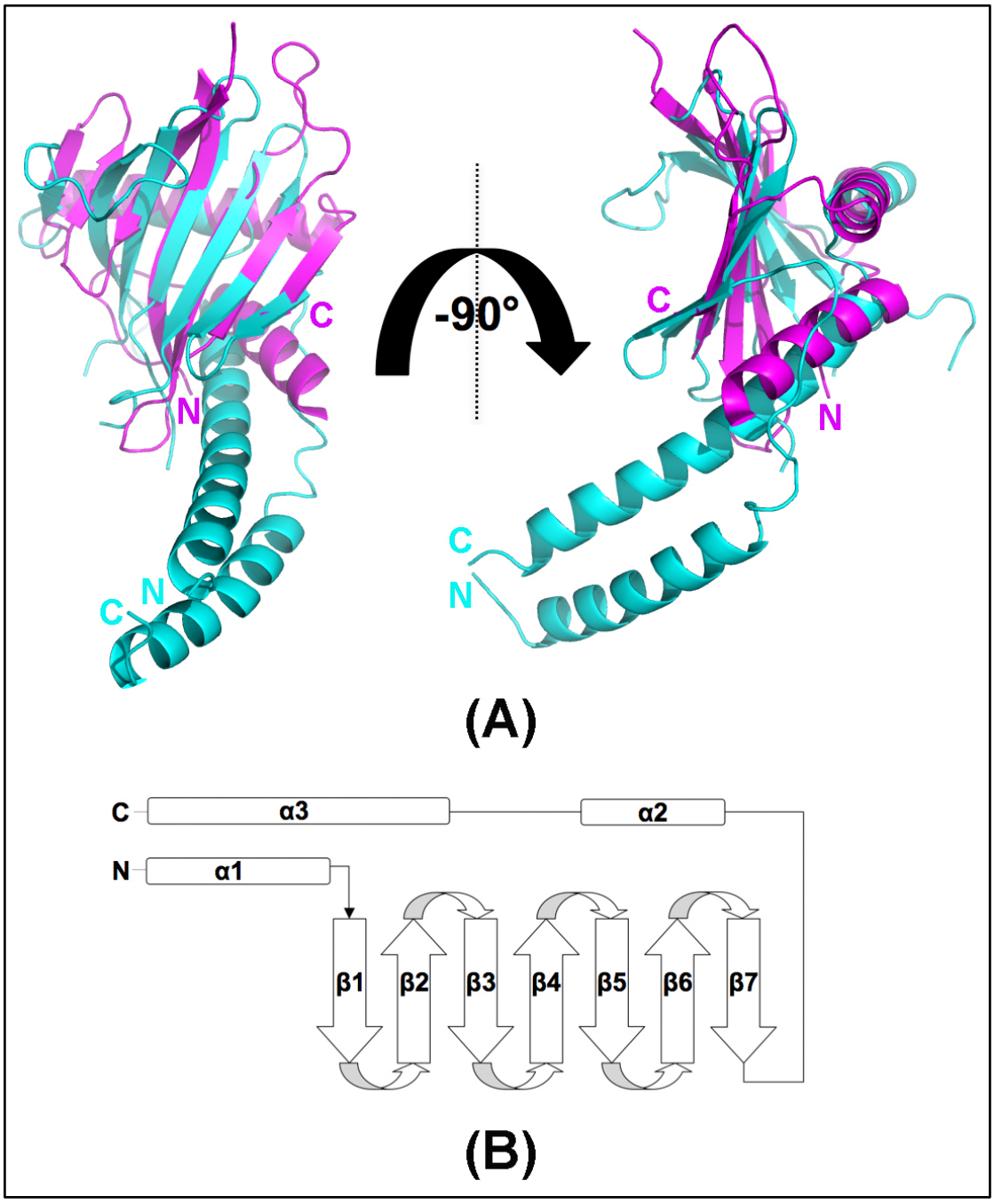


Figure 4.3 Superposition of VC1805 with human p32**. (A)** Cartoon representation of VC1805 (magenta) and chain A of human p32 protein (cyan). N denotes N-terminus and C denotes C-terminus of the protein. **(B)** Surface topology of chain A of p32.

4.3 Functional aspects

4.3.1 C1q studies

Human p32 protein, a structural homologue of VC0508, VC0509 and VC1805, is an evolutionarily conserved eukaryotic protein originally discovered in association with the nuclear pre-mRNA splicing factor SF2/ ASF, although no evidence for a splicing function has been reported (Krainer *et al.*, 1991). Human p32 is a promiscuous protein, reported to bind to many viral proteins (Bruni and Roizman, 1996; Desai *et al.*, 1991; Matthews and Russell, 1998; Tange *et al.*, 1996; Wang *et al.*, 1997; Yu *et al.*, 1995), transcription factor IIB (Yu *et al.*, 1995), the lamin B receptor (Simos and Georgatos, 1994), high molecular weight kininogen and factor XII (Herwald *et al.*, 1996), vitronectin (Lim *et al.*, 1996), hyaluronic acid (Deb and Datta, 1996) and the globular head domain of the plasma complement component C1q (Ghebrehiwet *et al.*, 1994).

We were interested to explore the functional aspects of VC1805 with C1q as a quest to reveal any functional clues for this hypothetical protein family. C1q was found to bind to VC1805 immobilized on microtitre plate wells (Section 2.21 and Figure 4.4.A). Immobilized VC1805 also caused complement consumption when incubated with human serum (Section 2.22 and Figure 4.4.B), indicating that VC1805 binds C1 and activates the complement classical pathway. Complement consumption by VC1805 was at a similar level (25–30%) to consumption by immune complexes. C1, a complex of C1q with the serine protease proenzymes C1r and C1s, is the form in which C1q circulates in blood (Figure 4.4.B).

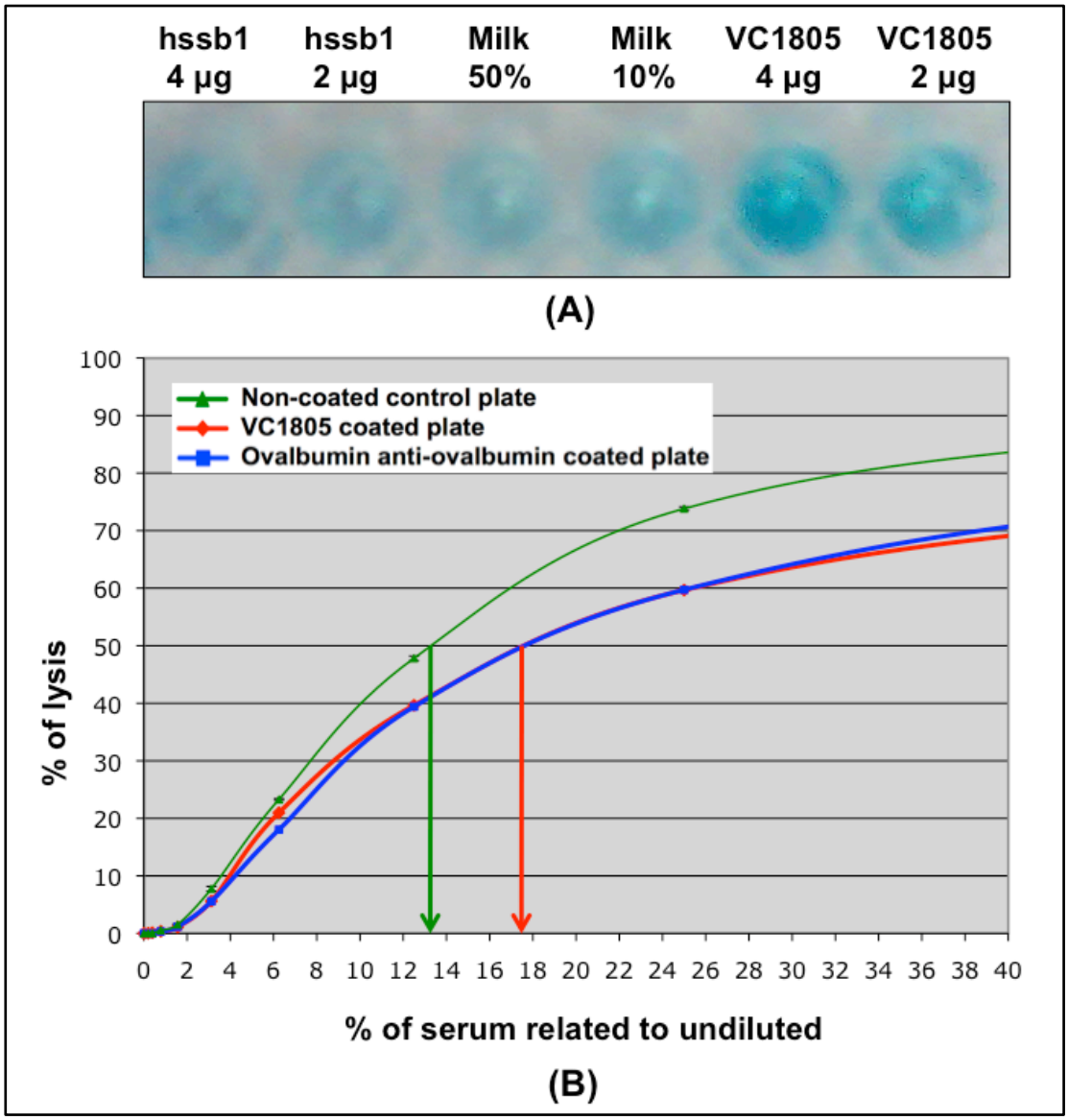


Figure 4.4 Binding of VC1805 to C1q. **(A)** ELISA assay showing C1q binding VC1805 with negative controls. **(B)** Complement consumption assay. Error bars, calculated using triplicate results, are included in the graph but are too small to visualize. The non-coated control plate gives 50% lysis with 13.2% serum at a 1/10 dilution (indicated by green arrow). The 1/10 diluted serum has 100/13.2 = 7.58 units per 100 µL, which corresponds to 758 units per mL of undiluted serum. VC1805 and ovalbumin anti-ovalbumin coated plate give 50% lysis with 17.5% serum at a 1/10 dilution (indicated by green arrow). The 1/10 diluted serum has 100/17.5 = 5.71 units per 100 µL which corresponds to 571 units per mL of undiluted serum. A and B have 75.33% of the activity of C which reflects 24.67% complement consumption by both A and B.

There was also no consistent dose-dependent complement inhibitory effect when soluble VC1805 was added to serum. This is consistent with the requirement for multivalent binding of C1q to a target array to initiate complement activation. Thus soluble VC1805, which is a monomer in solution, may bind weakly to sites on the C1q globular heads, but engagement of two or more heads of C1q, and summation of multiple weak interactions, as can occur with an array of immobilized VC1805, is required for high avidity C1q binding and complement activation.

4.4 Gene knockout

4.4.1 Overview

To find out the exact role of VC0508, VC0509, VC1804 and VC1805 proteins in *V. cholerae* pathogenesis, knockouts of those genes were made by SOE-PCR. The ultimate target was to test knockout stains of *V. cholerae* in an animal model to find out whether there are any significant differences in pathogenesis compared to wild type strain. A series of experiments was designed to produce three different mutant strains which are termed M050809 (double mutant of VC0508 and VC0509 of *V. cholerae* N16961), M180405 (double mutant of VC1804 and VC1805 of *Vibrio cholerae* N16961) and M04050809 (quadruple mutant of VC0508, VC0509, VC1804 and VC1805 of *V. cholerae* N16961).

4.4.2 Cloning SOE-PCR products in pCR 2.1 –TOPO vector

A total of six primers for each construct were designed (Table 2.3) which contained a portion of the beginning and the end of the mutant gene of

interest (M050809 and M180405). The genes were successfully amplified and ligated using the procedures described earlier (section 2.23.3-4 and Figure 4.5.A-B). The ABCD products (Figure 2.5) of M050809 and M180405 were cloned into pCR 2.1 –TOPO and the positive clones were confirmed by blue/white screening and colony PCR (Figure 4.5.C-D).

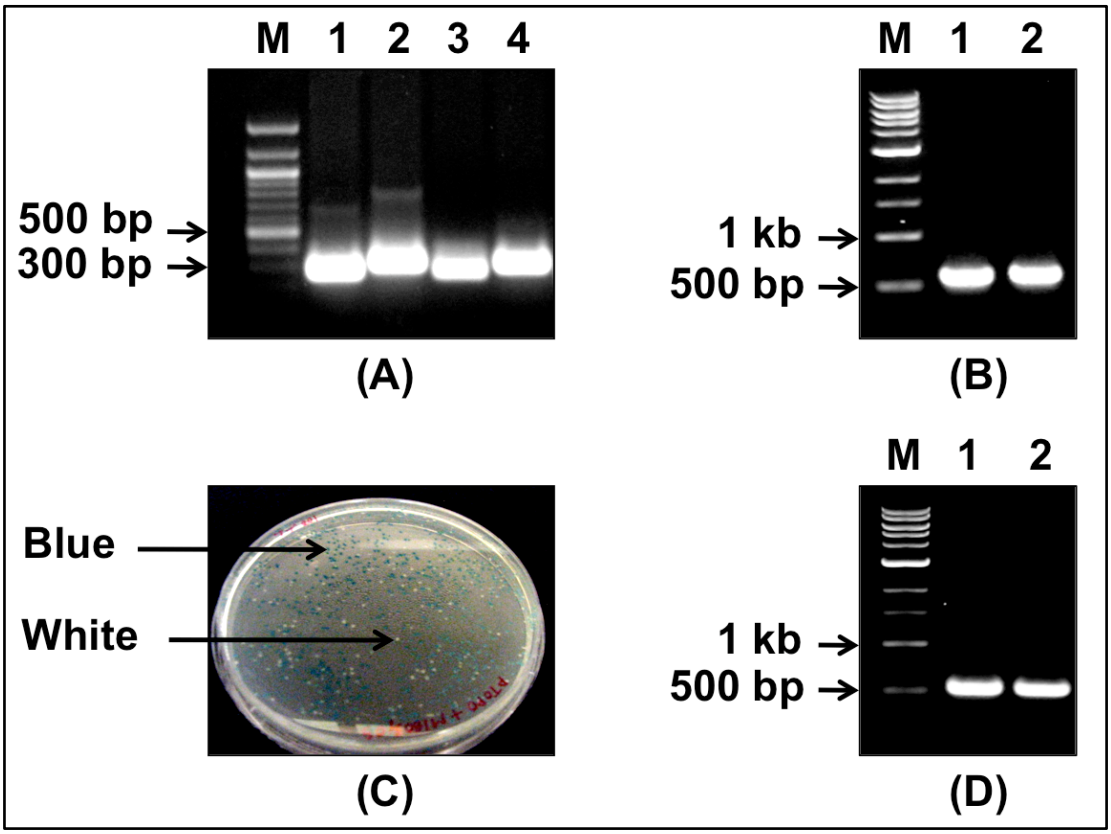


Figure 4.5 Cloning SOE-PCR products in pCR 2.1 –TOPO vector

(A) SOE-PCR of target genes. M, 100 bp DNA ladder (NEB, USA); lane 1, AB products of M050809 (278 bp); lane 2, CD products of M050809 (301 bp); lane 3, AB products of M180405 (270 bp); lane 4, CD products of M180405 (295 bp). **(B)** PCR reaction to ligate beginning and end of gene. M, 1 kb DNA ladder (NEB, USA); lane 1, ABCD products of M050809 (579 bp); lane 2, ABCD products of M180405 (565 bp). **(C)** Transformed colonies on LB-Amp plate. Blue colonies are galactose fermentor whereas white colonies are non-galactose fermentor and are possibly the right constructs. **(D)** Colony PCR to screen for the correct insert in pCR2.1-TOPO vector. M, 1 kb DNA ladder

(NEB, USA); lane 1, correct insert of M050809 of 579 bp; lane 2, correct insert of M180405 of 565 bp.

4.4.3 Single crossover mutants

ABCD products of M050809 and M180405 in pCR 2.1–TOPO, were successfully cloned into pADS132 plasmid using the procedure described earlier (section 2.23.6). The recombinant plasmids were then purified and transformed into *V. cholerae* O1 El Tor (section 2.23.7). The aim of single crossover was selection of plasmid integration into the recipient chromosome (which does not carry the *pir* gene) through allelic exchange by plating cells on chloramphenicol-LB plates. As *V. cholerae* O1 El Tor is not resistant to chloramphenicol, the observed growth of the strains in chloramphenicol-LB plate indicates the presence of recombinant pADS132 plasmid on the chromosome of the strain. This event is likely to produce single cross-over recombination due to the presence of a small piece of homologous DNA to *V. cholerae* O1 El Tor in pADS132 plasmid. The positive single crossover was validated by colony PCR using internal (AD) and flanking primers (EF) (Table 2.3). For positive single crossover, an internal primer should give a product smaller (ABCD products, Figure 2.5) than wild type, whereas the flanking primer would not give any product, as the whole plasmid of pADS132 sits in an amplification region which needs a higher extension time (about 6 min) in PCR. The absence of flanking products and the presence of internal products confirmed the construction of single crossover mutant (Figure 4.6).

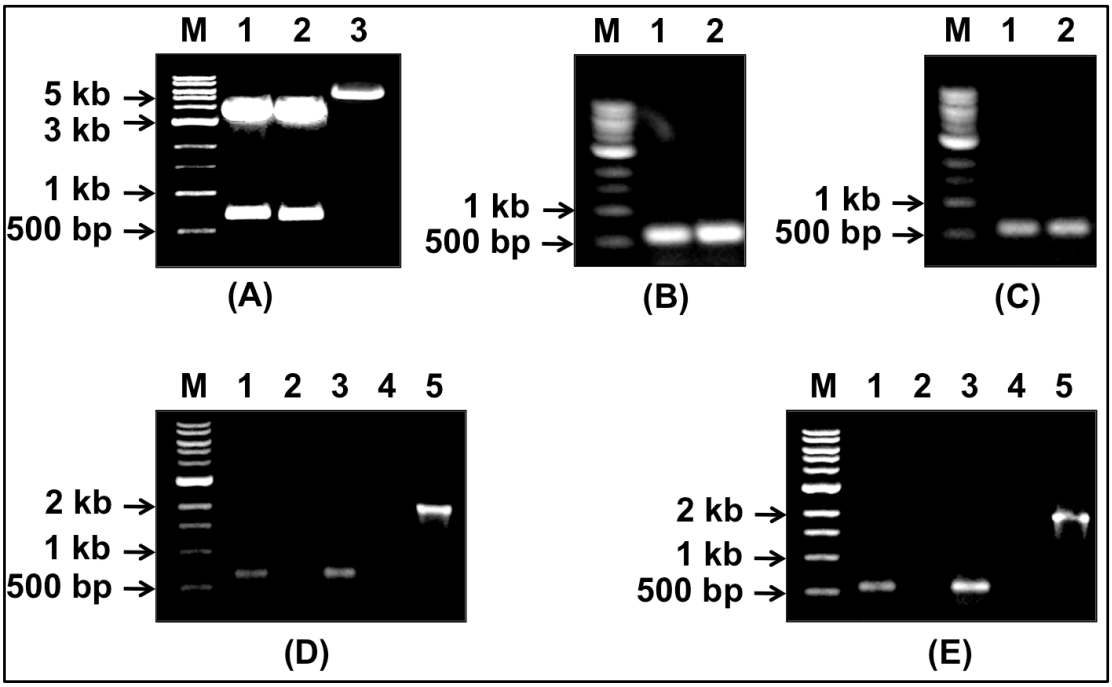


Figure 4.6 Construction of single crossover mutants

(A) Double digest of pCR2.1-TOPO vector containing insert gene of interest and pADS132. M, 1 kp DNA ladder (NEB, USA); lane 1 (upper band) digested pTOPO of M050809 construct of 3.85 kb; lane 1 (lower band), digested M050809 inserts of 0.66 kb (ABCD products with restriction site); lane 2 (upper band), digested pTOPO of M180405 construct of 3.85 kb; lane 2 (lower band), digested M180405 inserts of 0.64 kb (ABCD products with restriction site); lane 3, digested linerized pADS132 suicide vector of 5.2 kb. **(B)** Colony PCR to screen colonies contain M050809 insert with vector pADS132. Lane 1-2, ABCD products of M050809 of 579 bp. **(C)** Colony PCR to screen colonies contain M180405 insert with vector pADS132. Lane 1-2, ABCD products of M180405 of 565 bp. **(D)** Single crossover in *V. cholerae* by pADS132 containing homologous region of M050809. Lane 1 and 3, Internal product of M050809, AD (579 bp); lane 2 and 4, No flanking products of M050809, EF (1.9 kb); lane 5, positive control of 1.9 kb flanking products of wild type *V. cholerae*. **(E)** Single crossover in *V. cholerae* by pADS132 containing homologous region of M180405. Lane 1 and 3, Internal product of M180405, AD (565 bp); lane 2 and 4, No flanking products of M180405, EF (1.8 kb); lane 5, positive control of 1.8 kb flanking products of wild type *V. cholerae*.

4.4.4 Double crossover mutants

The growth of single crossover mutants *V. cholerae* O1 El Tor in the presence of sucrose will assist the removal of the pADS132 plasmid from its chromosome. The plasmid pADS132 contains the *sacB* gene which encodes the secreted enzyme levansucrase. This enzyme catalyzes the hydrolysis of sucrose and the synthesis of levans (high-molecular-weight fructose polymers) which are lethal for Gram-negative bacteria (Gay *et al.*, 1985; Rapoport and Dedonder, 1966). During excision of plasmid from host chromosome, some of the plasmid will participate in homologous recombination with the host chromosome (*V. cholerae* O1 El Tor) due to the presence of host chromosome specific homologous regions in the pADS132 plasmid. Excision of plasmid was confirmed from both mutants of M050809 and M180405, as there was no growth in LB plates containing chloramphenicol but there was growth in sucrose containing plates (Figure 4.7). Screening of double crossover mutants and animal model studies are underway.

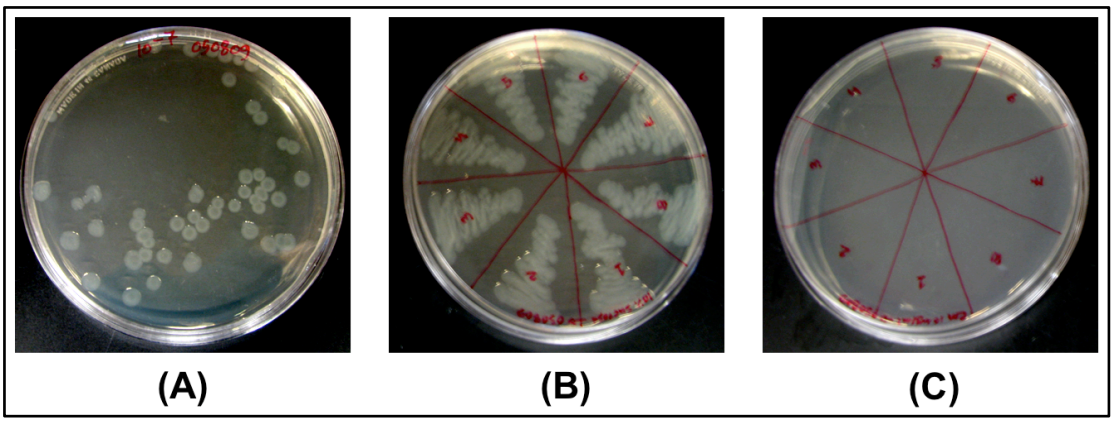


Figure 4.7 Double crossover mutant screening by sucrose sensitivity plating

(A) LB plate with 10% sucrose and Streptomycin at 10^{-7} dilution of M050809 mutant. **(B)** Colonies of M050809 mutant on LB 10% sucrose. **(C)** No colonies of M050809 mutant on chloramphenicol containing LB plate.

4.5 Discussion

Genome sequencing of microorganisms started a new era for scientists working in different areas of life sciences. More than 420 bacterial genomes have been sequenced to date (February 2009) (http://cmr.jcvi.org/tigr-scripts/CMR/shared/Genomes.cgi). In a newly sequenced genome, 30 to 40% of genes are annotated having no function (Bork, 2000), or named as hypothetical protein if the gene is unique to the genome, or named as conserved hypothetical proteins if the gene is found in more than one genome (Roberts, 2004). The complete understanding of a microbial organism requires the proper understanding of its non-annotated genes. Determination of the three-dimensional structure of a protein is very important in understanding its function at the atomic level, but achieving this goal is often difficult and expensive. Generally, determination of a structure would cost about $66,000 in the present day, but back in 2002 it was four times more costly (Lattman, 2004; Service, 2008). Structural genomics consortia around the world are putting a great deal of effort and money into solving the three-dimensional structures of novel proteins including hypothetical or conserved hypothetical proteins which often provide excellent clues about the function of hypothetical proteins (Zarembinski *et al.*, 1998).

Genes which sit on pathogenicity islands are of great interest as they are likely related to the pathogenesis of the organism. In a quest to find out the

role of hypothetical genes in the pathogenicity island, I designed some experiments to investigate structural, biochemical and microbiological analysis of four hypothetical proteins present in VPI-2 and VSP-2 of *V. cholerae* El Tor O1 pathogenic strain. These proteins, VC0508, VC0509, VC1804 and VC1805, are homologues and paralogues to each other, and share the same folds as revealed by PSI-BLAST (Altschul and Koonin, 1998) and PHYRE (Bennett-Lovsey *et al.*, 2008). I have cloned all four genes of *vc0508*, *vc0509*, *vc1804* and *vc1805* as part of the structural investigation. Proteins expressed from these genes were soluble with the exception of VC1804 even after extensive optimization. About 40 amino acid residues at the N-terminal of the protein are missing in VC1804 (104 aa) compared to VC0508 (147 aa), VC0509 (147 aa) and VC1805 (148 aa) which could be essential for the solubility of this protein.

VC0508, VC0509, VC1804 and VC1805 all carry a predominantly negative charge. The presence of those proteins in all pathogenic strains of *V. cholerae*, and other pathogenic strains, suggests that these proteins must have a role that gives an advantage to those strains of *V. cholerae* possessing them. There are a number of other *Vibrio* spp. and γ-proteobacteria that have homologues to those proteins: *V. harveyi* which is a pathogen of fish and invertebrates, including sharks, seahorses, lobster and shrimp; *V. splendidus* is associated with mortality in a variety of marine animals including fish, oysters, mussels, and scallops; *V. vulnificus* may cause gastroenteritis and can enter into the blood stream, resulting in septic shock with a mortality rate of about 50%; *V. alginolyticus* is an opportunistic pathogen that may cause disease in marine cultured animals, including fish,

shellfish and shrimp; *P. stuartii* is also an opportunistic pathogen which may cause invasive diarrhea; *Y. intermedia* may cause enterocolitis in humans; *A. hydrophila* may cause cholera like gastroenteritis and dysenteric gastroenteritis in humans and may cause disease in fish and amphibians. So the complete understanding of the functional aspects of this protein could answer some questions regarding pathogenicity in different organisms.

The crystal structures VC0508, VC0509 and VC1805 have been solved to 1.9 Å, 2.4 Å and 2.1 Å resolution, respectively. The structures of these proteins reveal the same folds, in agreement with the prediction made by sequence analysis. The structures of VC0508, VC0509 and VC1805 have an unusual topology which includes a flat seven-stranded, anti-parallel β-sheet and α-helices and shows structural homology to human p32 protein. Of the many reported binding partners of p32, the binding to the globular head domain of the complement component C1q appeared worthy of exploration. In our experiment, we found that VC1805 does bind to the complement protein C1q. In fact, C1q can bind a remarkable variety of immune and non-immune ligands through charge pattern recognition mediated mainly through its globular head domain (Ghai *et al.*, 2007). The crystal structure of the globular head domain of human C1q revealed the heterotrimeric organization of the three chains each with similar jelly-roll topology (Gaboriaud *et al.*, 2003). The three modules within the C1q head domain show clear differences in their electrostatic surface potentials, with one module carrying a predominantly positive charge that has been implicated in the C1q-IgG interaction (Gaboriaud *et al.*, 2003). However, the details of the interactions between C1q and its various ligands remain to be elucidated (Malhotra *et al.*, 1993).

V. cholerae is generally accepted to be a non-invasive gut pathogen, and is thought to be relatively non-inflammatory (Satchell, 2003; Silva *et al.*, 1996), although a more recent report described an inflammatory response in both O1 and O139 cholera infections (Qadri *et al.*, 2004). Could VC1805 therefore be engaging C1q as part of a strategy to modify complement attack, or to mediate cell adhesion? It would not be advantageous for *V. cholerae* to have a protein that would activate complement, but this would be consistent with *V. cholerae* being non-invasive. It is not known if VC1805 is secreted, and if so whether it remains attached to the bacterial surface. What therefore could be the relevance of binding to C1q? It is known that some other complement-related components (DAF-CD55 and MCP-CD46) serve as anchors for gut microbes, but DAF and MCP are cell-surface proteins, not soluble like C1q. C1q is synthesized in gut epithelial cells (Loos *et al.*, 1989; Morris *et al.*, 1978) and may be present in the gut lumen, although this has been shown only in inflammatory bowel disease (Halstensen *et al.*, 1992). C1q can associate with membranes via C1q receptors which bind the collagenous region of C1q. The identification of C1q receptors has been controversial, but it is now widely accepted that calreticulin, bound to cell surfaces via CD91, HLA class 1 heavy chain, or CD59, acts as a receptor for C1q and the related "defence collagens" SPA, SPD and MBL (Ghiran *et al.*, 2003; Ogden *et al.*, 2001; Sim *et al.*, 1998). C1q receptor activity is present on lung epithelial cells (Malhotra *et al.*, 1993), but it does not appear to have been investigated in gut epithelium.

It may be fortuitous that VC1805, being negatively charged, happens to be able to bind to positively charged C1q *in vitro*, and this may be completely

unrelated to its biological function *in vivo*. Perhaps a role for VC1805 is to aid in cell adhesion and to act as a bridge between membrane immobilized C1q and some ligand on the bacterial surface and this may be consistent with the two conserved surface patches on opposite sides of VC1805.

Human p32 also binds to hyaluronic acid, a non-sulfated glycosaminoglycan found in the extracellular matrix of epithelial tissues. Perhaps hyaluronic acid is the ligand recognized by VC1805. An intriguing scientific question is why *V. cholerae* keeps four copies of structurally similar proteins in its genome. The true biological role of these proteins remains to be discovered.

4.6 Ongoing and future work

This structural study of VC0508, VC0509 and VC1805 has stimulated further investigations that will focus on quantifying the interaction with C1q and carrying out pull-down experiments to identify other binding partner(s) in *V. cholerae*. To better understand the interactions of the VC0508, VC0509, VC1804 and VC1805 proteins with both human proteins and other *V. cholerae* pathogenicity island proteins, Y2H (yeast-two-hybrid) analyses are underway. These proteins are already cloned in prey vector pGADT7 and bait vector pGBKT7 in preparation for further Y2H experiments in the department of Systems Biology of Infectious Disease at the University of Edinburgh. In order to explore the potential role of VC0508, VC0509, VC1804 and VC1805 *in vivo*, gene knockouts and animal model are currently underway.

Chapter 5

Structural determination of *Vibrio cholerae* ferric uptake regulator

5. Structural determination of *Vibrio cholerae* ferric uptake regulator

This chapter describes the results of cloning, expression, purification and structural determination of *V. cholerae* ferric uptake regulator (*Vc*Fur). The structure *Vc*Fur was determined to 2.6 Å resolution, and revealed a dimeric molecule with each monomer containing a dimerization domain, a helical DNA-binding domain and two metal binding sites occupied by zinc. Like Fur from *Pseudomonas aeruginosa* (*Pa*Fur), *Vc*Fur has two metal binding sites. However, the Zn2 site which is proposed to be a structural Zn-binding site in *Pa*Fur, has different ligands in *Vc*Fur.

5.1 Overview of *Vibrio cholerae* ferric uptake regulator (*Vc*Fur)

The *Vibrio cholerae* ferric uptake regulator (*Vc*Fur) is a metalloprotein which is relatively small in size (about 17000 Da) (Litwin *et al*., 1992). Fur is a classical repressor protein and its main function is iron homeostasis but it can also regulate the genes involve in pathogenesis. A wide variety of Fur homologues are present widely in Gram-positive and Gram-negative human pathogens. The locus *vc2106*, containing the *fur* gene, encodes a protein of 150 amino acid residues and a calculated isoelectric point (pI) of 5.81. It has two domains involved in DNA binding and dimerization. *In vitro*, a number of other divalent cations can also activate Fur to bind DNA (Bagg and Neilands, 1987; de Lorenzo *et al*., 1987), and their binding affinities for Fur have been studied in detail: in the case of *E. coil* Fur (*Ec*Fur), Fe(II), Fe(III), Co(II), Mn(II) and Zn(II) all activate Fur binding to DNA (Mills and Marletta, 2005).

5.2 Multiple sequence similarity and conserved amino acids

A sequence similarity search with BLAST (Altschul *et al*., 1997) and a multiple sequence alignment with CLUSTAL-W (Chenna *et al*., 2003) indicated that *Vc*Fur is a conserved member of the Fur family. There is a high sequence identity of *Vc*Fur homologues in different *Vibrio* sp. The sequence identities of *Vibrio mimicus*, *Vibrio harveyi*, *Vibrio alginolyticus*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, *Vibrio splendidus* and *Vibrio fischeri* to *Vc*Fur are 98, 96, 94, 92, 93, 90 and 86% respectively. Among other pathogenic species such as *Aeromonas hydrophila*, *Escherichia coli*, *Salmonella typhimurium*, *Yersinia enterocolitica*, *Klebsiella pneumoniae*, *Shewanella baltica*, *Yersinia pestis*, *Haemophilus influenzae*, *Legionella pneumophila*, *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*, *Bordetella pertussis* and *Neisseria meningitidis* the sequence identities to *Vc*Fur are 82, 76, 77, 76, 76, 78, 73, 58, 59, 58, 52, 53 and 52%, respectively. A multiple sequence alignment of *Vc*Fur homologues is shown in Figure 5.1. Metal binding residues of *Vc*Fur; His33, Glu81, His87, His88 Asp89, His90, Glu108 and His125, are conserved in all homologues and shown by green arrows in Figure 5.1. Using RONN analysis (Yang *et al.*, 2005), it was found that residues 1, 72-86, and 141-150 in *Vc*Fur are predicted to be disordered (Figure 5.2). Interestingly, Cys139 is conserved in almost all the Fur homologues compared in Figure 5.1, except for *P. aeruginosa*. The aromatic amino acids (tryptophan, tyrosine and phenylalanine) tend to be more ordered in proteins (Kissinger *et al.*, 1995). Cysteine and histidine are less common in disordered regions whereas charged residues like glutamate, aspartate and lysine are more likely to appear in disordered regions (Yang *et al.*, 2005).

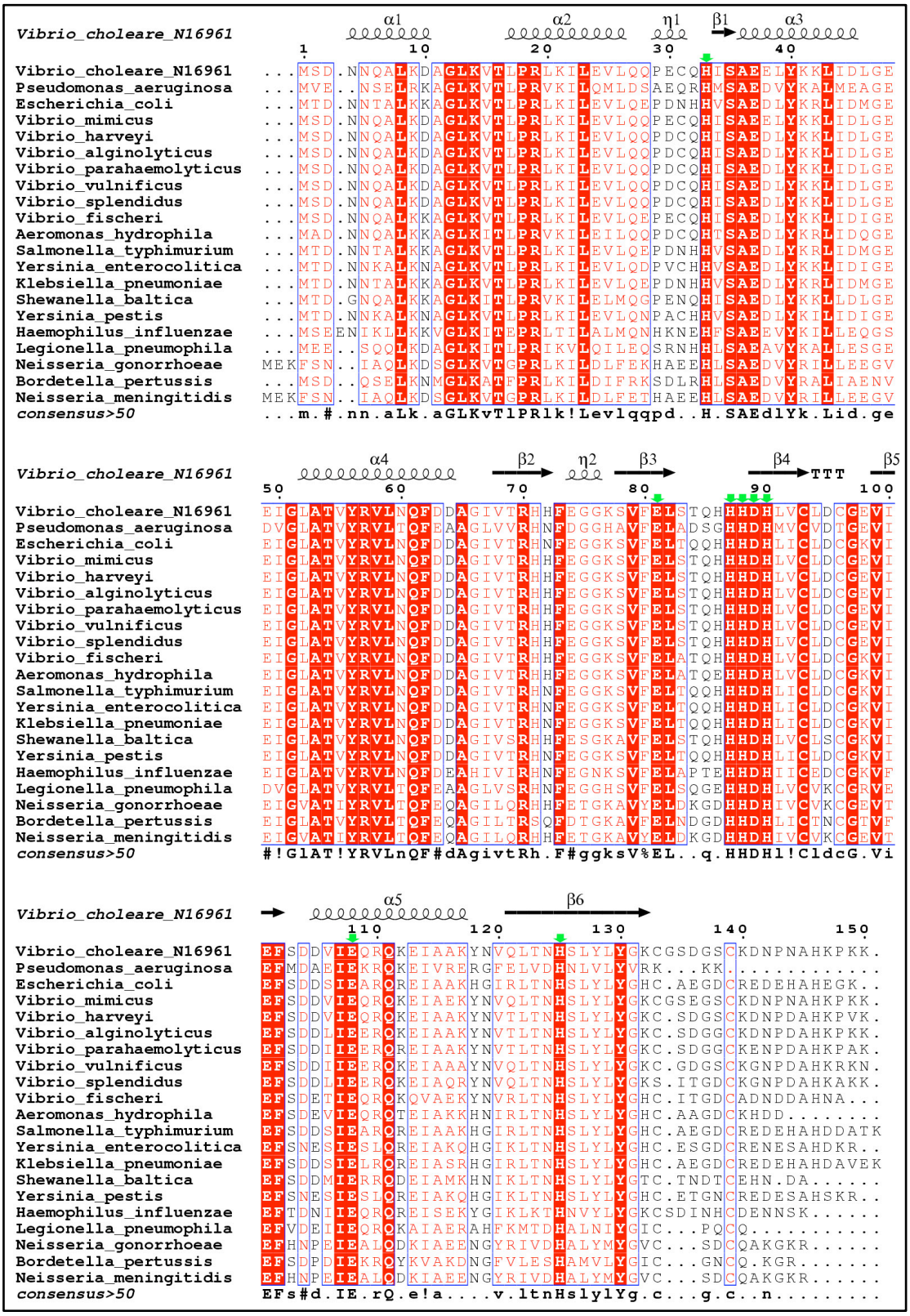


Figure 5.1 Multiple sequence alignment of *Vc*Fur homologues. The secondary structure of VC-Fur is shown on the top line. NCBI entry name and genome annotation name for the sequences are *Vibrio cholerae* N16961 Fur (NP_231738, VC2106) *Vibrio mimicus* (BAD24855, AB182655.1),

Vibrio harveyi (ABM89108, EF197913.1), *Vibrio alginolyticus* (ZP_01260898, V12G01_15725), *Vibrio parahaemolyticus* (NP_797212, VP0833), *Vibrio vulnificus* (NP_759186, VV1_0175), *Vibrio splendidus* (ZP_00992248, V12B01_21344), *Vibrio fischeri* (YP_204193, VF_0810), *Aeromonas hydrophila* (AAK28627, AF349468.1), *Escherichia coli* (NP_286398, Z0831) *Salmonella typhimurium* (NP_459678, STM0693), *Yersinia enterocolitica* (YP_001007162, YE2973), *Klebsiella pneumoniae* (YP_001334379, KPN_00706), *Shewanella baltica* (YP_001050871, Sbal_2509), *Yersinia pestis* (NP_406158, YPO2634), *Haemophilus influenzae* (ZP_00154690, Hflu103001406), *Legionella pneumophila* (YP_125780, lpl0414), *Pseudomonas aeruginosa* (NP_253452, PA4764), *Neisseria gonorrhoeae* (YP_002003120, NGK_2495), *Bordetella pertussis* (NP_881134, BP2507) and *Neisseria meningitidis* (YP_974323, NMC0197). Alignment was created by with CLUSTAL-W (Chenna *et al*., 2003) and the figure was produced by ESPRIPT (Gouet *et al*., 1999).

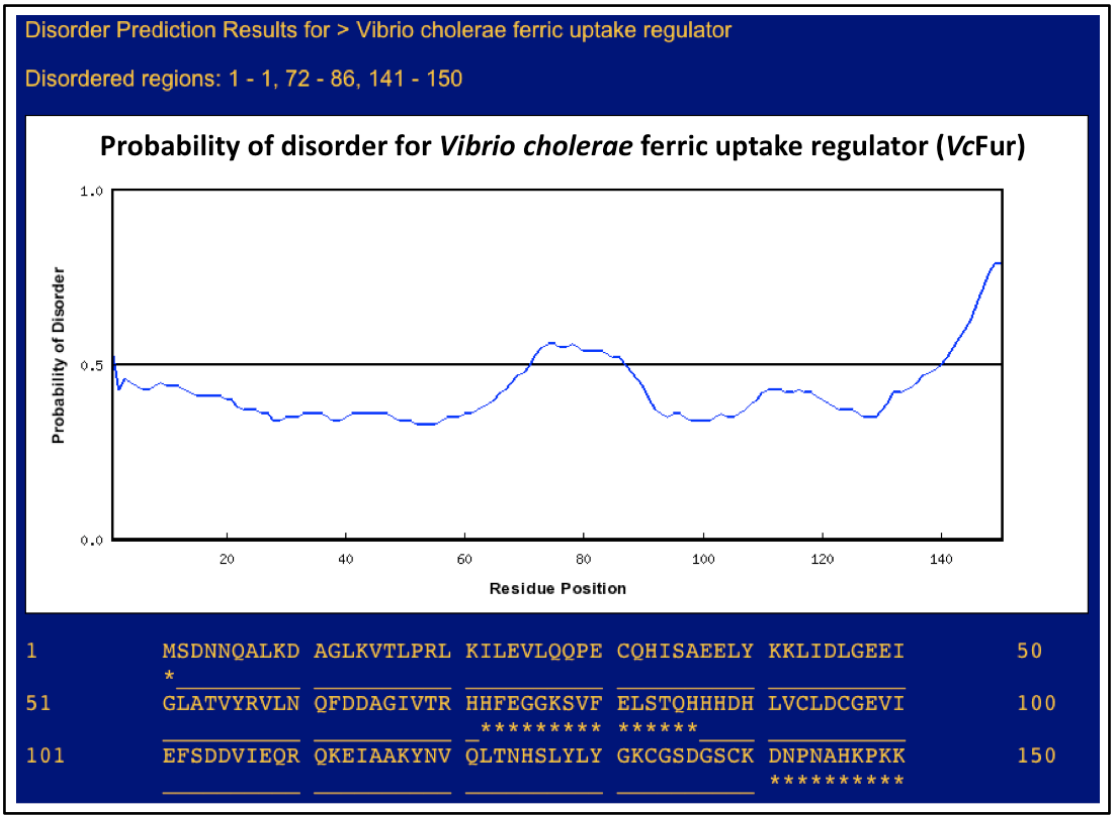


Figure 5.2 RONN analysis of *Vc*Fur

5.3 Cloning of *vc2106* gene

The *fur* gene at locus *vc2106* in the *V. cholerae* genome was successfully amplified using the polymerase chain reaction (PCR) from genomic DNA of *V. cholerae* O1 El Tor strain N16961 using standard protocols (Figure 5.3.A) (Section 2.4). Two primers commercially synthesized and purchased from VHBIO, forward (5′- CAGGAAAGTCCATGGCAGACAATAAC-3′) and reverse (5′- CGTAAAGAATTCGGTTATTTCTTCGGC-3′) were designed corresponding to the 5′ and complementary 3′ ends of *fur* gene with specific recognition sites for the restriction enzymes *Nco*I and *EcoR*I. The primers produce one mutation at position 2 of the protein from a serine to an alanine. After amplification, the PCR product was digested with *Nco*I and *EcoR*I enzymes and ligated with the similarly digested pEHISTEV vector (Section 2.6-7) (Liu and Naismith, 2009). Positive clones were verified by double digestion (Section 2.10.2) (Figure 5.3.B), and DNA sequencing at the University of Dundee (Section 2.10.3).

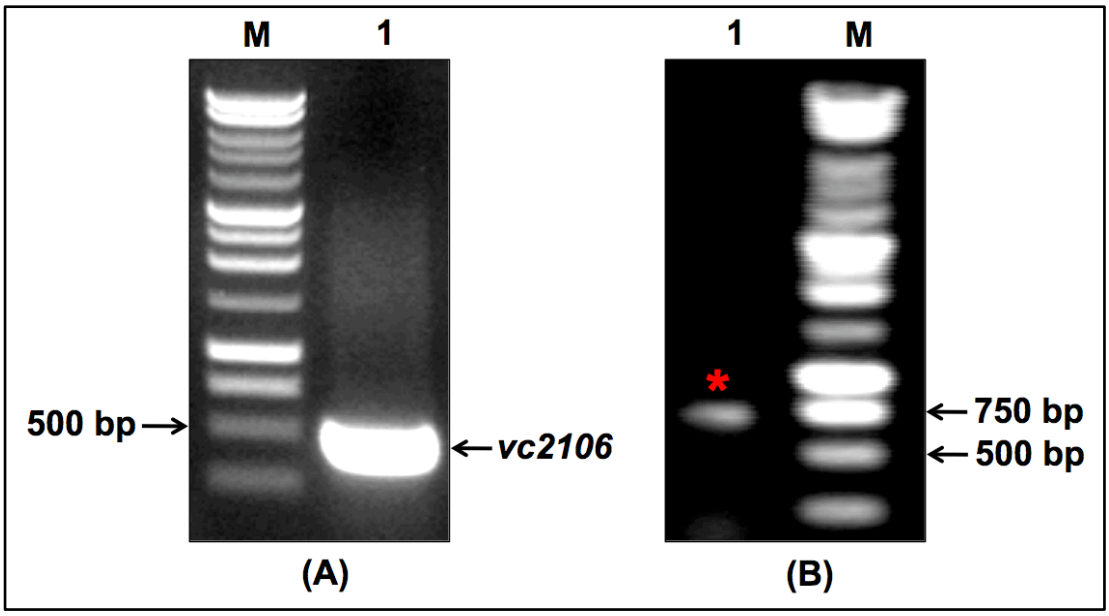


Figure 5.3 Agarose gel electrophoresis of *vc2106* gene products

(A) Amplification of *vc2106* gene from genomic DNA of *V. cholerae* El Tor O1. M, 1 kb DNA ladder (Promega); lane 1, amplification of *vc2106* gene product. **(B)** Colony PCR of cloning vector pLou3 contains *vc2106* gene. Lane 1, colony PCR product of pEHISTEV vector containing *vc2106* gene (indicated by red asterisk).

5.4 Protein expression, optimization of solubility and purification

The expression vector pEHISTEV containing the *vc2106* gene was transformed in *E. coli* BL21 (DE3) using standard methods. The expression, optimization of solubility and purification was done according to the procedure described earlier (Figure 5.4.A-E) (Section 2.13). The best soluble protein expression was achieved after 24 h growth at 25 °C with the addition of 0.5 mM IPTG. For crystallization trials, purified *Vc*Fur was pooled and dialysed overnight at 4 °C against 4 L of buffer containing 50 mM HEPES, 500 mM NaCl, 10 mM imidazole and 10% (v/v) glycerol, 5 mM DTT and 1 mM $ZnCl_2$. Next day, the sample was dialysed against the same buffer without

DTT and $ZnCl_2$ for 4 h at 4 °C with two buffer changes. Finally, the protein was dialysed in 10 mM HEPES, 50 mM NaCl, 5% (v/v) glycerol. Mass spectrometry was carried out to confirm the identity of *Vc*Fur (Figure 5.4.F).

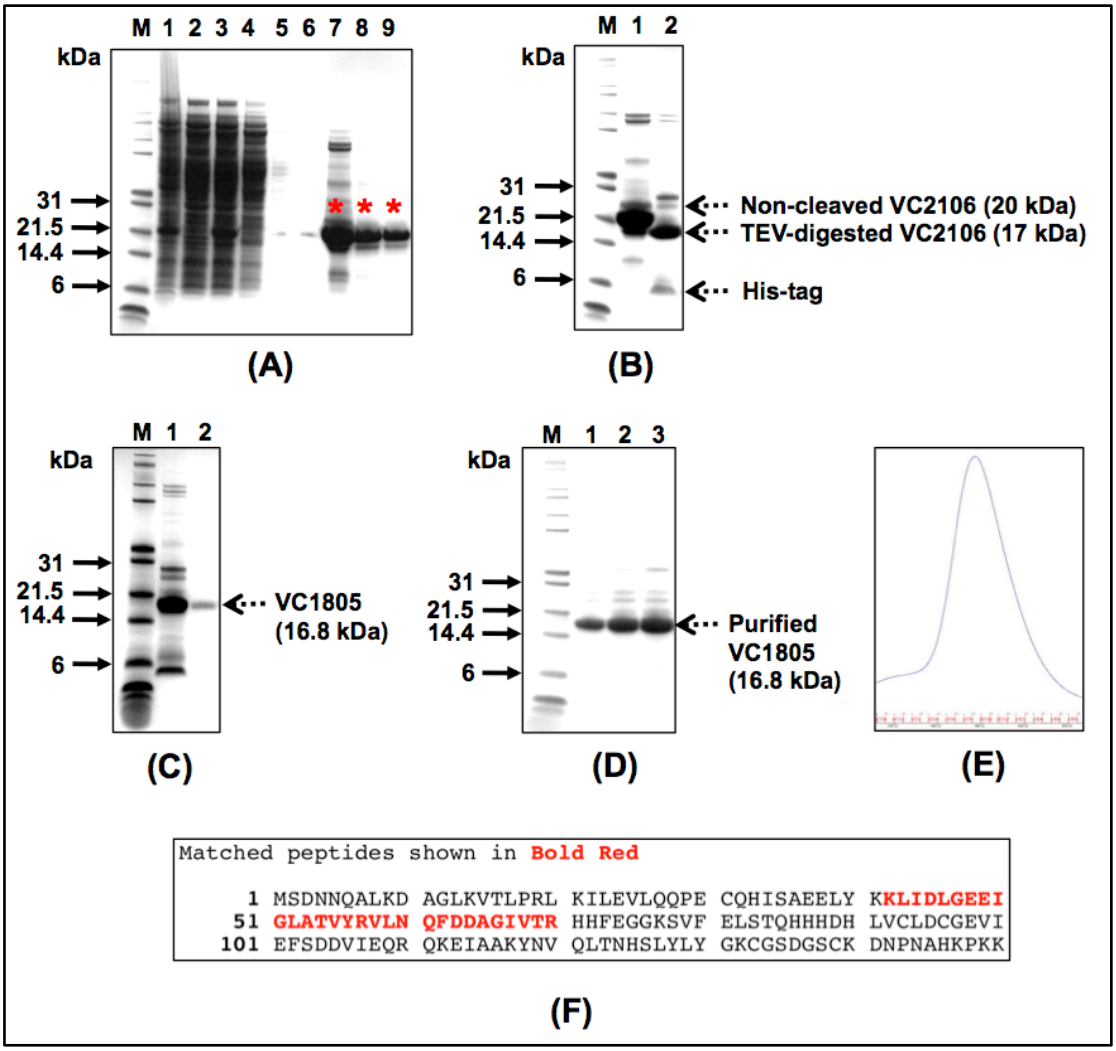


Figure 5.4 Expression and purification of *Vc*Fur (VC2106) protein

(A) First Ni^{2+}-affinity column. M, protein marker (Mark12™, Invitrogen); lane 1, cell lysate after IPTG induction; lane 2, supernatant after sonication; lane 3, flow-through (unbound protein) after loading sample on to the column; lane 4-6, column wash with the buffer containing 10 mM imidazole; lane 7-9, elution in buffer containing 350 mM imidazole. Bands indicated by the red asterisks are His-tagged VC2106. **(B)** TEV cleavage of his-tagged VC106. Lane 1, dialysed eluted protein from first Ni^{2+}-affinity column; lane 2, TEV digested sample of VC2106. **(C)** Second Ni^{2+}-affinity column. Lane 1, filtered TEV

digested sample; lane 2, VC106 without His-tag **(D)** Eluted VC2106 after gel filtration chromatography. Lane 1-3, fractions containing VC2106 protein. **(E)** Gel filtration peak corresponding to purified VC2106. **(F)** Matched peptides of VC2106 identified by mass spectrometry.

5.5 Crystallization, optimization and X-ray data collection of *Vc*Fur

The initial concentration of VcFur for crystallization was 7 mg/mL, as determined by the PCT. A total of six crystal screens were set up with three different concentrations which were 14 mg/mL, 10 mg/mL and 7 mg/mL. The screens were: The PEGs, Ammonium salt, pH clear, JCSG+, Classics (Qiagen) and Wizard I and II (Emerald Biosciences). Crystal plates were initially set up using the crystallization robot (Hamilton-Thermo Rhombix system), with 150 nl of buffer and 150 nl of protein per drop. After 4 weeks, initial hits with tiny rod shaped crystals were observed in two conditions of PEGs screen containing protein of 10 mg/mL (well number 55; 0.2 M calcium chloride dihydrate and 20% w/v PEG 3350; and well number 64; 0.2 M magnesium nitrate, and 20% w/v PEG 3350) (Figure 5.5.A-B). Crystal plates were set up at 10 mg/mL with a 192-condition optimization screen based on the initial hits. After 6 weeks, crystals suitable for X-ray analysis were produced in the condition 0.18 M magnesium nitrate and 16 % PEG 3350 (Figure 5.5.C). Crystals were cryoprotected in crystallization buffer with the addition of 20% (v/v) glycerol, and X-ray data to 2.6 Å were collected at 100K on an in-house X-ray facility, a Rigaku/MSC MicroMax-007HF rotating anode equipped with focusing optics and a Saturn 944+ CCD detector (Figure 5.5.D).

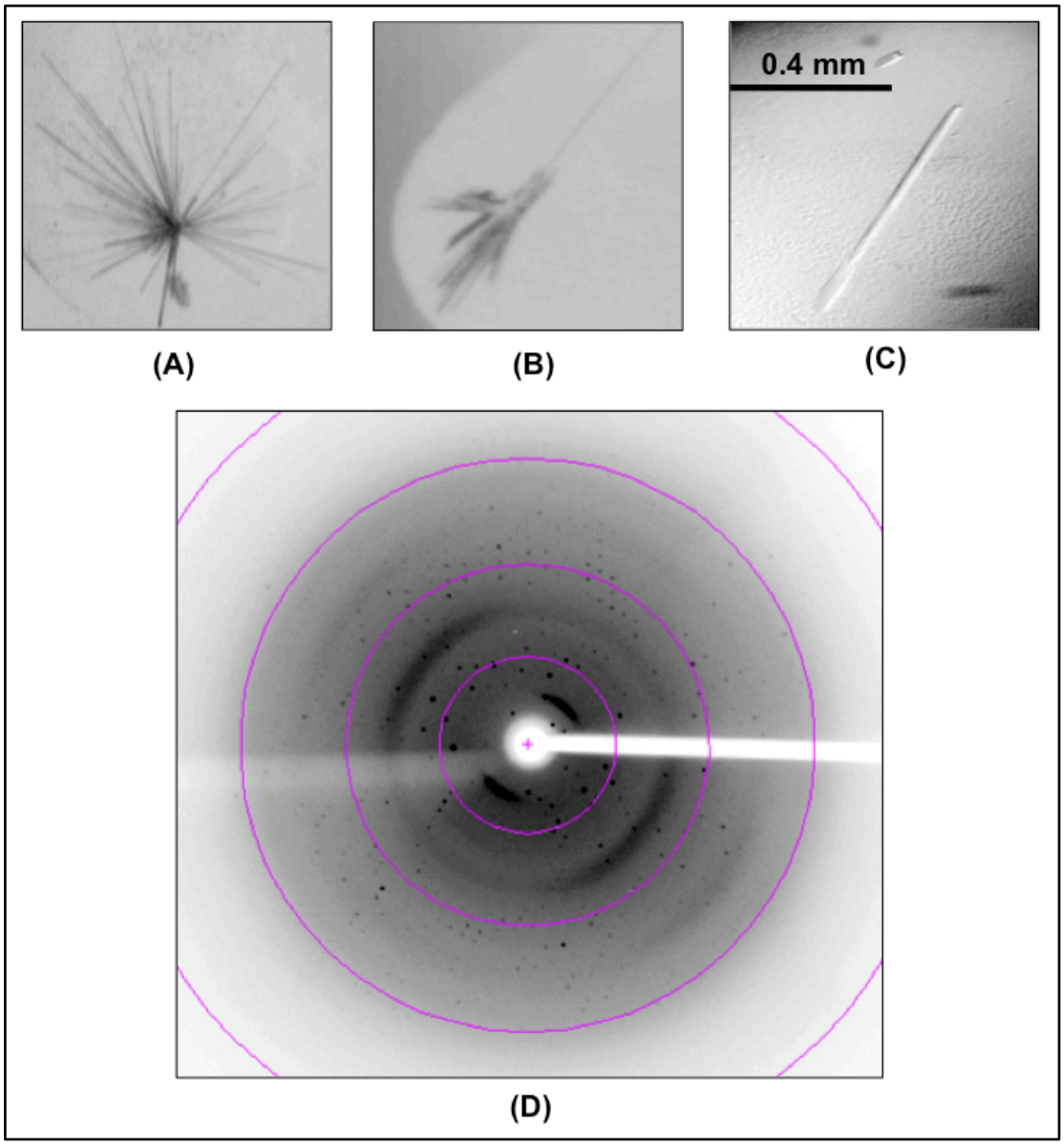


Figure 5.5 *Vc*Fur protein crystal and diffraction pattern

(A) Initial hits in PEGs screen (well number 55; 0.2 M calcium chloride dihydrate; 20% w/v PEG 3350). **(B)** Initial hits in PEGs screen (well number 64; 0.2 M magnesium nitrate, and 20% w/v PEG 3350). **(C)** Optimized crystal in 0.18 M magnesium nitrate and 16 % PEG 3350. **(D)** X-ray diffraction image of *Vc*Fur protein crystal. The resolution of the outer circle is 2.4 Å.

5.6 Structure solution and refinement of *Vc*Fur

X-ray data were integrated and scaled using HKL2000 (Otwinowski and Minor, 1997) and the CCP4 (CCP4., 1994) suite. The crystals belonged to the space group $P4_{1/3}2_12$. The unit cell suggested two monomers in the asymmetric unit giving a Matthews coefficient of 2.47 $Å^3$/Da with a 50% solvent content. The structure of *Vc*Fur was solved by molecular replacement using the program PHASER (McCoy *et al.*, 2007), by splitting the search model into two domains: the dimerization domain from *Pa*Fur (residues 84-134 from PDB code 1MZB), and the DNA-binding domain made up by superimposing the DNA-binding domains from *Ec*Fur (PDB code 2FU4) and *Pa*Fur (residues 1-83 from PDB code 1MZB). Zinc atoms were removed from the search models. A solution with Z-scores of 6.3 and 29.2 for the rotation and translation function steps, respectively, was obtained in the space group $P4_32_12$. The final model was built manually using the program COOT (Emsley and Cowtan, 2004) and the refinement was carried out using REFMAC(Murshudov *et al.*, 1997) and PHENIX (Adams *et al.*, 2002). The data collection and refinement statistics are summarized in Table 5.1. Atomic coordinates and structure factors have been deposited in the PDB (2w57). Part of the final $2F_o$-F_c electron density map is shown in Figure 5.6.

Table 5.1 Data collection and refinement statistics of *Vc*Fur crystal. Numbers in parentheses refer to highest resolution shell.

Data collection	**Crystal**
Space group	$P4_32_12$
Unit cell parameters (Å)	a = b = 88.88, c = 85.07
Resolution (Å)	62.9 - 2.6 (2.74 – 2.6)
R_{merge}[a]	0.128 (0.588)
Observed reflections	94,782
Unique reflections	10,961
Completeness (%)	100 (100)
$< I / \sigma(I) >$	16.6 (3.3)
Multiplicity	8.6 (8.6)
RFZ (rotation function Z-score)	6.3
TFZ (translation function Z-score)	29.2
Refinement	
No. of protein atoms	2114
No. of water atoms	80
Resolution limits (Å)	62.9 - 2.60
Number of used reflections	10,383
R factor[b]	0.207
Free R factor[c]	0.259
RMSD bond lengths (Å)	0.007
RMSD bond angles (°)	1.152
Average B-factor (Å^2) for chain A/B	
All protein atoms	31 / 29
DNA-binding domain (3-82)	26 / 27
Dimerization domain (83-133)	39 / 32
Zinc atoms Zn1, Zn2	41,43 / 28,32
Waters	34

[a]*R*merge = $\sum_h \sum_j | I_{hj} - \langle I_h \rangle | / \sum_h \sum_j$, where I_{hj} is the intensity of the *j*th observation of unique reflection *h.*

[b]Overall R factor $\sum h || F_{oh}| - |F_{ch} || / \sum h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes for reflection *h.*

[c]Free R factor is equivalent to overall R factor , but is calculated using 5% of reflections excluded from the maximum-likelihood refinement stages.

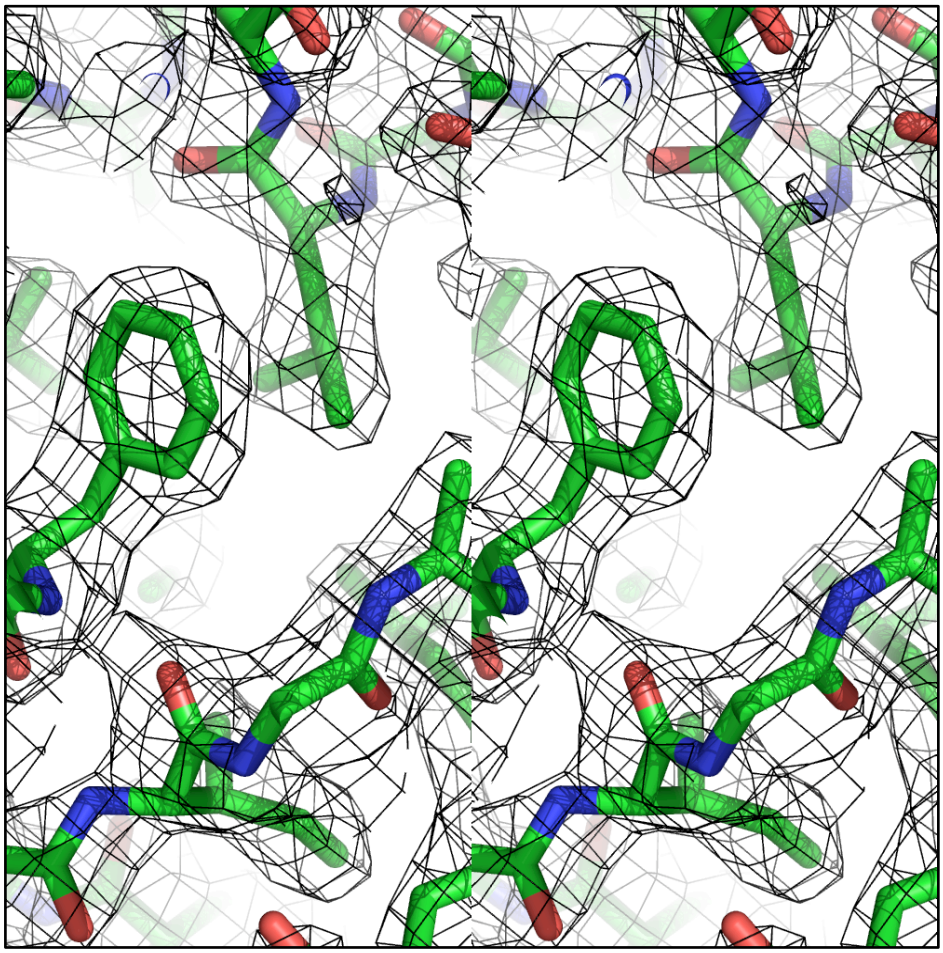


Figure 5.6 Stereo view of $2F_o$-F_c electron density map of *Vc*Fur. A region of the map is shown, contoured at 1σ. This and other figures were created with PyMOL (DeLano, 2007).

5.7 Structure of *Vc*Fur

The structure of *Vc*Fur reveals a dimer in the asymmetric unit. In each chain of the dimer, the first two amino acids and the last 17 amino acids (134-150) are not visible in the electron density map which is consistent with the RONN analysis (Figure 5.2). Each chain has two domains: the DNA-binding domain (residues 3-82) and the dimerization domain (residues 83-133), the latter of which has a disulfide bond connecting Cys93-Cys133 (Figure 5.7). The dimerization interface of *Vc*Fur has a buried surface area of ~1350Å^2, similar to the 1300 Å^2 reported for the same interface in *Pa*Fur (Pohl *et al*., 2003).

Superposition of the two *Vc*Fur chains gives an RMSD of 1.33 Å for 130 Cα atoms, however separate superposition of the DNA-binding and dimerization domains gives RMSDs based on Cα atoms of 0.48 Å and 0.99 Å, respectively, reflecting a slightly different orientation of the DNA-binding domains in each subunit relative to the dimerization domain. The crystallographic B-factor distribution throughout the dimer shows that the DNA-binding domain is relatively well-ordered compared to the dimerization domain, and this is reflected in the higher RMSD between the two dimerization domains. The electrostatic charge distribution over the dimer is shown in Figure 5.8.A-B, and reveals a largely negatively-charged surface apart from the helical regions involved in DNA recognition, particularly helix 4 (Figure 5.8.C). Two metal ions, refined as zincs, are present in each chain and have the same coordination in each chain. The electron density around each zinc in chain B is shown in Figure 5.9-10. The zinc in the dimerization domain, Zn1, is tetracoordinated by the following side chains where the distances refer to those observed in (chainA/ chainB): His87 (2.3/ 2.1 Å), Asp89 (2.0/ 1.9 Å), Glu108 (2.1/ 2.1 Å) and His125 (2.2/ 2.2 Å) (Figure 5.9 and Table 5.2), these are the equivalent ligands to those in *Pa*Fur. The zinc that connects the DNA-binding and dimerisation domains, Zn2, is tetracoordinated by His33 (2.1/ 2.0 Å), Glu81 (2.1/ 2.1 Å), His88 (2.1/ 2.1 Å) and His90 (2.4/ 2.2 Å) (Figure 5.10 and Table 5.3). The significant difference in comparison to *Pa*Fur, is that His88 replaces Glu100 (in *Pa*Fur) as a ligand to Zn2 which is discussed below.

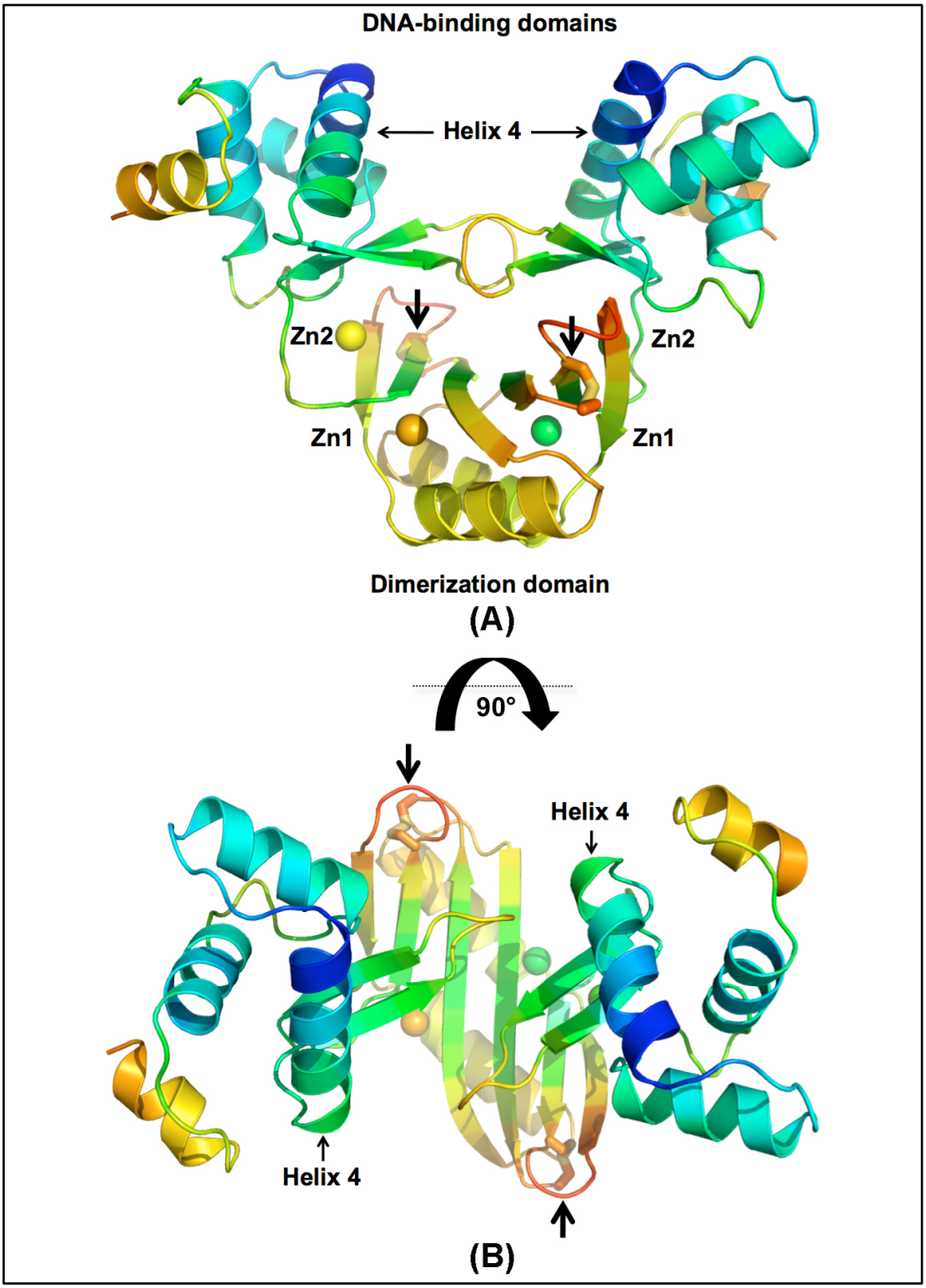


Figure 5.7 The structure of *Vc*Fur dimer

(A) Schematic view of the secondary structure and location of the four zinc atoms, coloured by crystallographic B-factor (dark blue, B~20Å^2 to red, B~50Å^2). Black arrows indicate disulphide bonds. **(B)** 90° rotation of Figure 5.5.A.

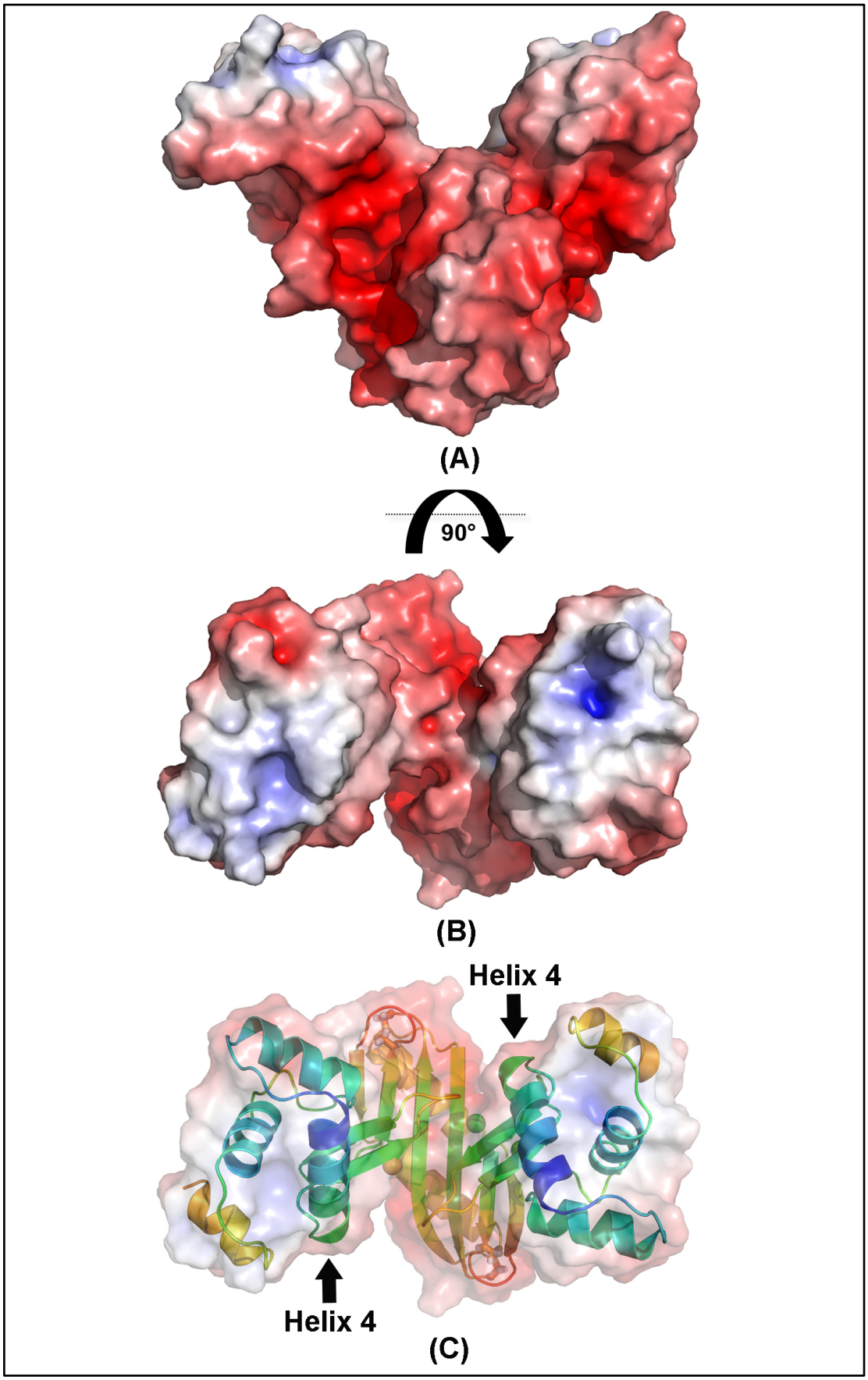


Figure 5.8 Electrostatic surface of *Vc*Fur

(A) Electrostatic surface of *Vc*Fur from -10 kT/e to +10 kT/e, calculated using APBS (Baker *et al.*, 2001). (**B)** 90° rotation of Figure 5.6.A. **(C)** Orientation of helix-4 on the surface.

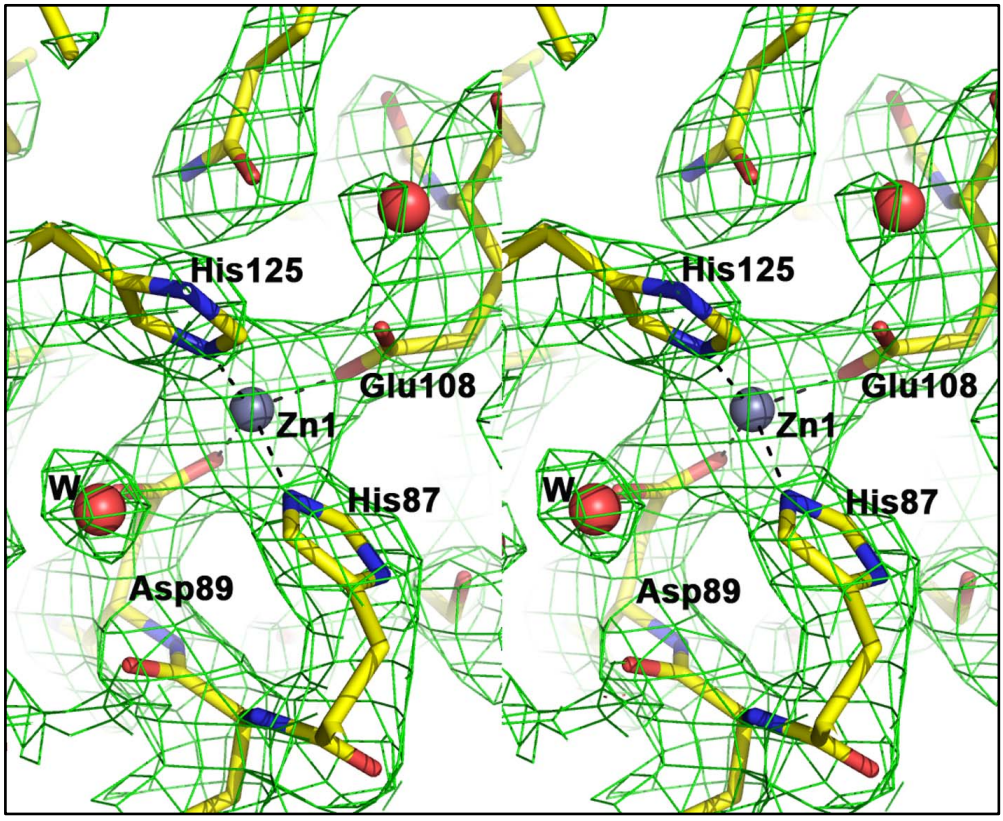


Figure 5.9 Metal binding ligands at Zn1 site of *Vc*Fur. Stereo view of the electron density around the zinc atoms in the Zn1 site of *Vc*Fur chain B. The maps represent a sigma-a weighted $2F_o$-F_c map contoured at 1σ. The atomic model represents the final refined structure. W denotes water molecule.

Table 5.2 Distances of Zn1 ligands in the crystal structure of *Vc*Fur

Zn1, chain A	**Distance (Å)**
Zn- Nε2, His87	2.3
Zn–Oδ2 Asp89	2.0
Zn–Oε2 Glu108	2.1
Zn–Nε2 His125	2.2
Zn1, chain B	**Distance (Å)**
Zn- Nε2, His87	2.1
Zn–Oδ2 Asp89	1.9
Zn–Oε2 Glu108	2.1
Zn–Nε2 His125	2.2

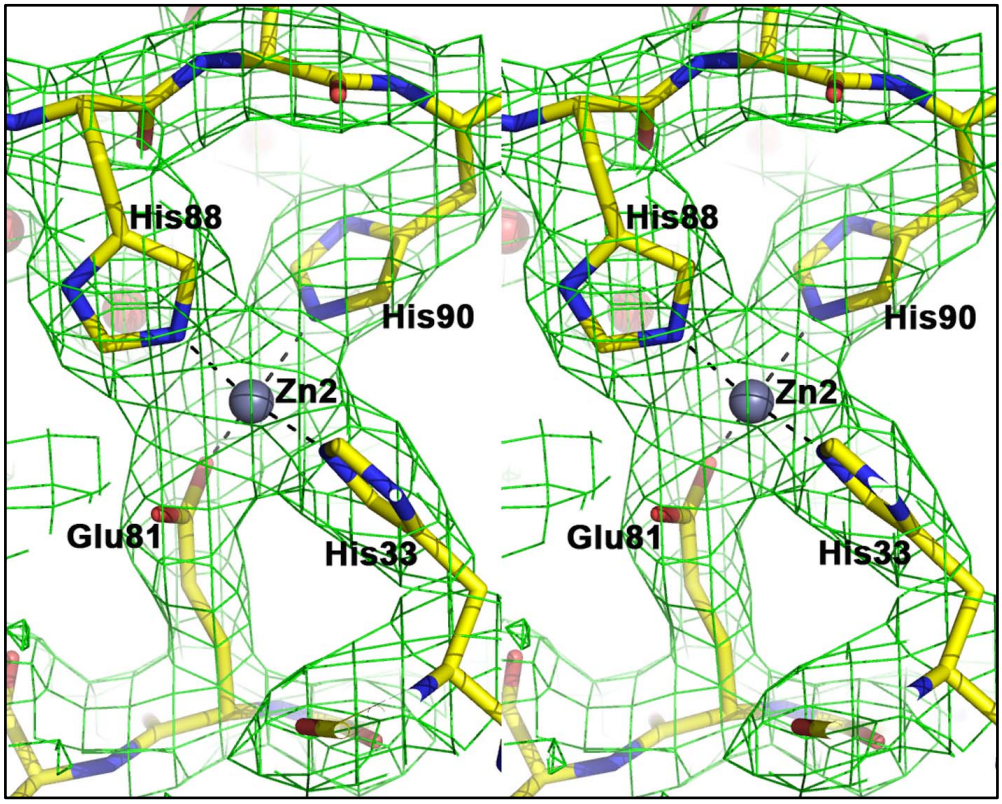


Figure 5.10 Metal binding ligands at Zn2 site of *Vc*Fur. Stereo view of the electron density around the zinc atoms in the Zn2 site of *Vc*Fur chain B. The maps represent a sigma-a weighted $2F_o$-F_c map contoured at 1σ. The atomic model represents the final refined structure.

Table 5.3 Distances of Zn2 ligands in the crystal structure of *Vc*Fur

Zn2, chain A	**Distance (Å)**
Zn–Nε2 His33	2.1
Zn–Oε2 Glu81	2.1
Zn–Nε2 His88	2.1
Zn–Nε2 His90	2.4
Zn2, chain B	**Distance (Å)**
Zn–Nε2 His33	2.0
Zn–Oε2 Glu81	2.1
Zn–Nε2 His88	2.1
Zn–Nε2 His90	2.2

Chapter 6

Analysis of *Vc*Fur and discussion

6. Analysis of *Vc*Fur and discussion

The *Vc*Fur structure is compared with the only other example of a Fur structure, *Pseudomonas aeruginosa* Fur. This reveals differences in metal binding coordination and the orientation of the DNA binding domain. Detailed structural comparison, metal content analysis and DNA binding studies of *Vc*Fur are presented in this chapter. EPR (Electron paramagnetic resonance spectroscopy) experiments of *Vc*Fur were designed to understand the movements of DNA binding domains in the presence of metal and DNA. Preliminary data from this experiment are also described in this chapter.

6.1 Overall comparison with *Pa*Fur

There is 50% sequence identity between *Vc*Fur and *Pa*Fur. The crystal structure of *Pa*Fur has a monomer in the asymmetric unit, the physiological dimer being generated by a crystallographic 2-fold axis. Superposition of the *Pa*Fur monomer with the A and B chains of *Vc*Fur gives an RMSD of 2.80 Å and 2.28 Å, respectively, reflecting the slightly different relative orientation of the DNA-binding domains in the *Vc*Fur dimer. A dramatically different orientation of the DNA-binding domains in *Vc*Fur compared to *Pa*Fur is revealed by superimposing the dimerization domains of a *Pa*Fur dimer onto the dimerization domains of the *Vc*Fur dimer (Figure 6.1). The structures of the individual domains of *Vc*Fur and *Pa*Fur are quite similar. The DNA-binding domains of *Vc*Fur (residues 3-82) and *Pa*Fur (residues 1-82) share 59% sequence identity and superimpose with an RMSD of 0.93 Å and 0.89 Å for chains A and B of *Vc*Fur, respectively (Figure 6.2). The dimerization domains of *Vc*Fur (residues 83-133) and *Pa*Fur (residues 83-135), which share 51%

sequence identity, superimpose with an RMSD of 1.12 Å and 1.07 Å for chains A and B of *Vc*Fur, respectively.

The program LSQKAB in the CCP4 (CCP4., 1994) suite was used to compare the relative positions of the DNA-binding domains in the *Vc*Fur and *Pa*Fur dimers. In chain A of *Vc*Fur, an almost pure 32° rotation (the angle between the centroid vector and the rotation axis being 87°) relates this DNA-binding domain to that in *Pa*Fur. In chain B of *Vc*Fur, a more complex relationship exists involving a 24° rotation around an axis that is 74° to the centroid axis. It may be significant that in the *Vc*Fur crystal chain B has many more crystal contacts than chain A. As the individual domains are structurally very similar, the program DynDom (Hayward and Berendsen, 1998) was used to analyze the domain movement between *Vc*Fur and *Pa*Fur. This suggested that the bending region encompasses residues 82-88 of *Vc*Fur, with the DNA-binding domain rotating of 30° relative to the dimerization domain, in agreement with the LSQKAB analysis. The most significant consequence of the change in the orientation of the DNA-binding domains is the decrease in the distance between helix 4 of each DNA-binding domain. The geometry between the two helices 4 of the dimer was analyzed with the program INTERHELIX (http://nmr.uhnres.utoronto.ca/ikura/resources/resources.html). In the *Pa*Fur dimer, the two helices are 33 Å apart and oriented 86º to one another. In the *Vc*Fur dimer the equivalent helices are 25 Å apart and oriented 103º to one another. If dimers are constructed from the *Vc*Fur A or B chains superimposed by their dimerization domains, then a dimer formed from two A chains has the helices 27 Å apart, oriented by 102º. A dimer formed from two B chains has the helices 23 Å apart, oriented by 108º. The change in

orientation is illustrated in Figure 6.1 by the change in distance between the Cα atoms of a conserved arginine on helix 4.

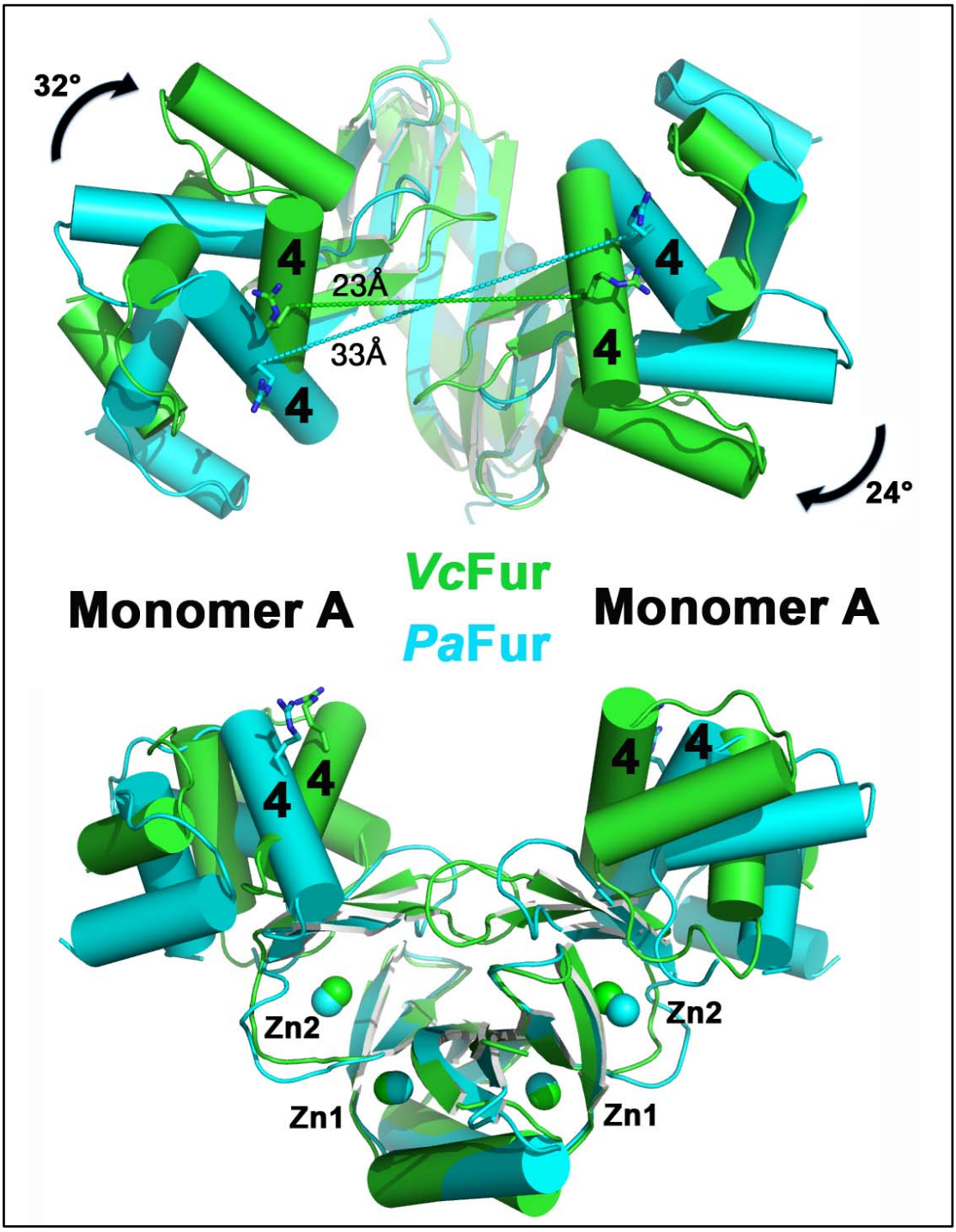


Figure 6.1 Superposition of *Vc*Fur with *Pa*Fur structure. Orthogonal views of the superposition of the *Vc*Fur (green) and *Pa*Fur (cyan) dimers. The dotted lines indicate the distances between the Cα atoms of the conserved arginine (Arg57 in *Vc*Fur) on the DNA-recognition helix 4 (indicated as 4 in the picture).

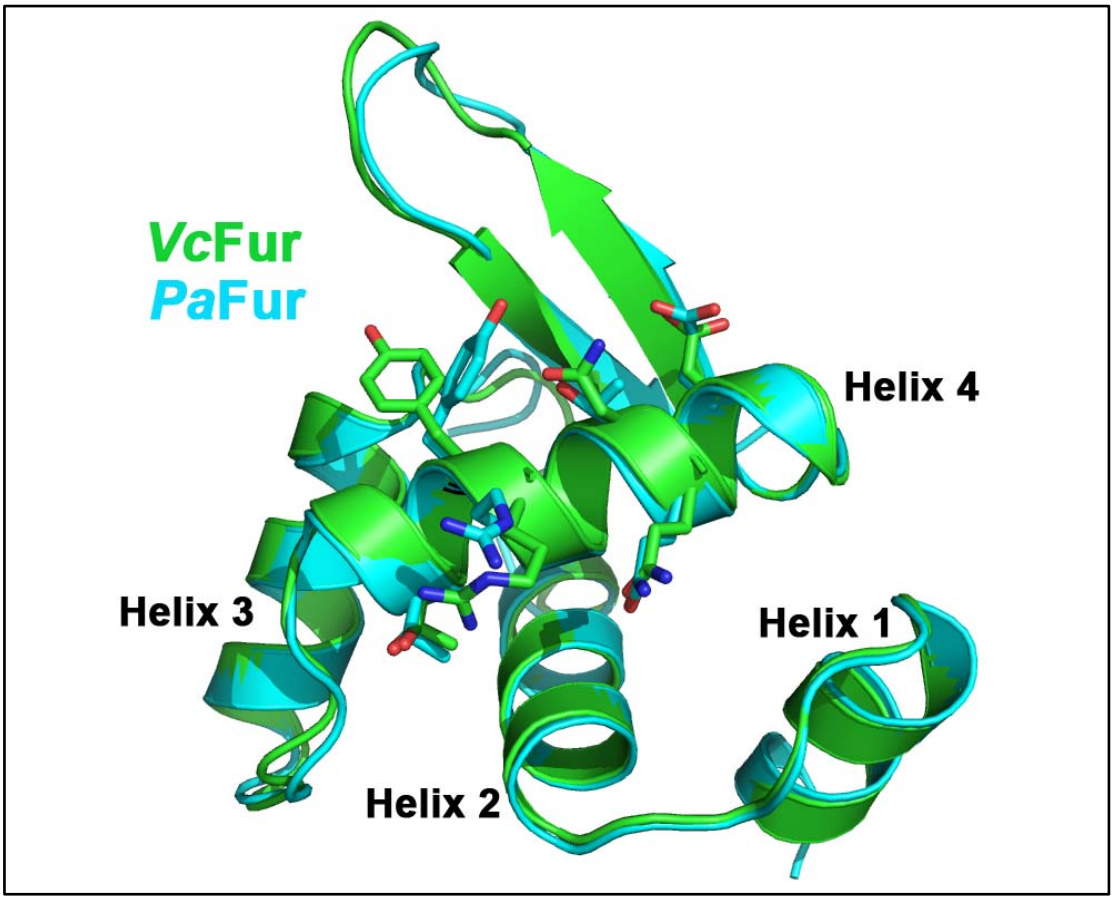


Figure 6.2 Superposition of the DNA-binding domains of *Vc*Fur with *Pa*Fur. Monomer A of *Vc*Fur and *Pa*Fur are green and cyan in colour, respectively.

6.2 Comparison of metal-binding sites of *Vc*Fur and *Pa*Fur

In Figure 6.3 illustrates the metal binding sites of *Vc*Fur (chain A) compared with *Pa*Fur (chain A) following superposition of the two monomers via their dimerization domains. The coordination of Zn1 involves the same conserved residues (two histidines, a glutamic acid and an aspartic acid), but in *Vc*Fur the zinc is tetracoordinated whereas in *Pa*Fur the zinc is hexacoordinated through a bidentate interaction with Asp88 and by water molecule. In *Vc*Fur, the equivalent aspartic acid, Asp89, makes only a single interaction with the zinc, and although there is a water molecule in a similar position, it is 3.5 Å away from Zn1. The distance between the Zn1 zincs in the two structures is

0.5 Å, and the main difference in the coordinated side chains is the position of His87 (His86 in *Pa*Fur). The second metal binding site, Zn2, shows significant differences. The distance between the zincs in this site is 0.9 Å. In *Pa*Fur, no electron density was observed beyond Cβ for the side chain of His87. In contrast, the equivalent residue in *Vc*Fur, His88, is ordered in both monomers and coordinates to Zn2 in place of the glutamic acid Glu100 in *Pa*Fur. The closest carboxylate oxygen of Glu101, the equivalent residue in *Vc*Fur, is 3.9 Å from Zn2. Comparison of *Pa*Fur with chain B of *Vc*Fur reveals a very similar picture, with minor changes around the Zn1 position and significant changes around Zn2, where the distance between the two zincs is 1.9 Å.

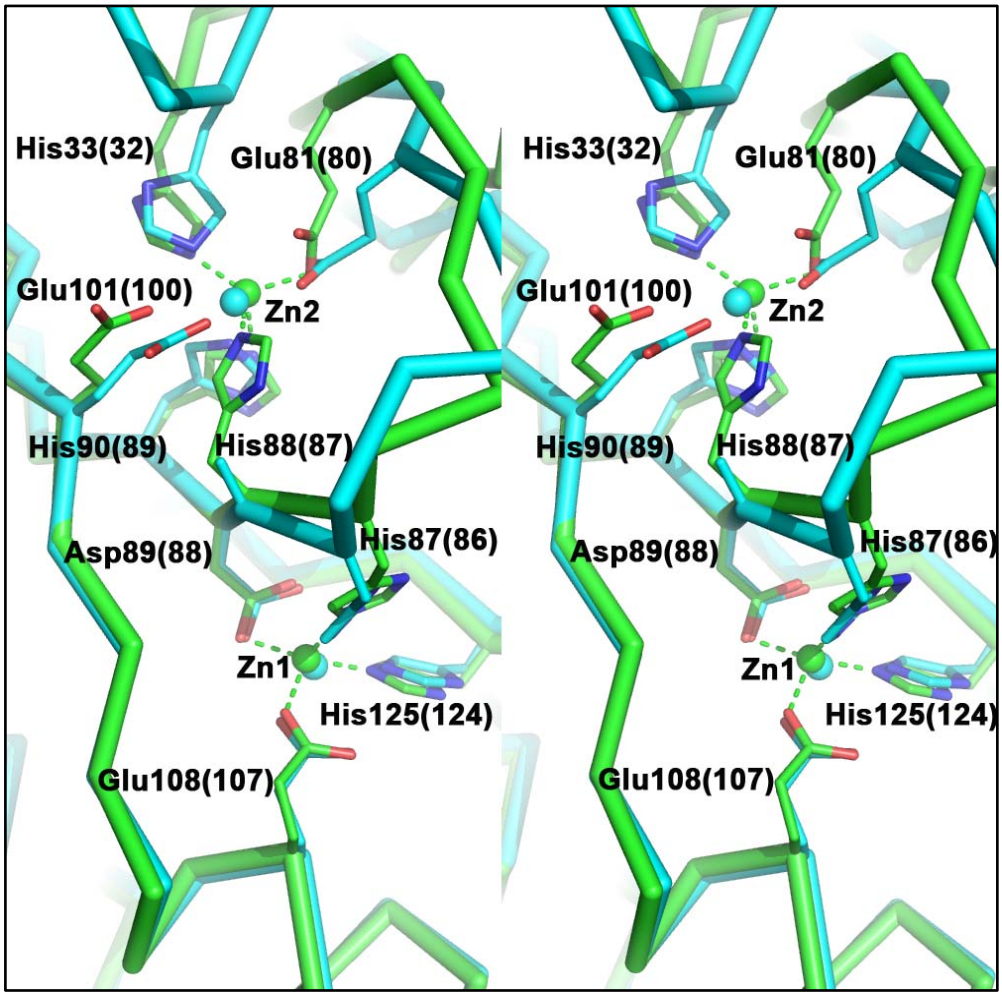


Figure 6.3 A comparison of the metal binding sites of *Vc*Fur and *Pa*Fur

The metal binding sites of *Vc*Fur coloured by green whereas *Pa*Fur is coloured cyan. The residue numbers are shown for *Vc*Fur, with the residue number for *Pa*Fur in parentheses. For clarity, dotted lines indicate the coordination interactions of the zincs in *Vc*Fur only. Note that no electron density was observed in *Pa*Fur for sidechain atoms of His87 beyond the Cβ.

6.3 Metal binding characteristics of *Vc*Fur

An analysis of *E. coli* Fur (*Ec*Fur) showed that it contains 2.1 mol of Zn(II) per mol of *Ec*Fur monomer (designated as Zn_2*Ec*Fur) when zinc was used in purification steps, and that one mol of Zn per monomer can easily be removed upon incubation/dialysis against EDTA (Althaus *et al.*, 1999). Metals analysis results of native and EDTA treated *Vc*Fur samples are shown in Table 6.1. When zinc was used in the purification of *Vc*Fur, ICP-MS analysis revealed 2.2 mol of Zn(II) per mol of Fur monomer (Zn_2*Vc*Fur) and also 0.1 mol of nickel per mol of Fur monomer. When *Vc*Fur was purified without using zinc in the purification steps, there was 0.9 mol of Zn(II) per mol of Fur monomer (Zn_1*Vc*Fur) and also 0.04 mol of nickel per mol of Fur monomer. The presence of nickel in *Vc*Fur indicates the purification of this protein through nickel column. In contrast to *Ec*Fur, when treated with EDTA, *Vc*Fur retains only 0.1 mol of Zn(II) per mol of *Vc*Fur monomer.

The EDTA treated *Vc*Fur, containing almost no zinc, was dialysed with three different metals individually (Table 6.2). In reconstitution with Zn(II), 2.3 mol of zinc was incorporated per mol of *Vc*Fur, consistent with the previous amount of zinc found after purification of *Vc*Fur in the presence of zinc. In the case of Fe(III) reconstitution, 1.6 mol of iron was incorporated per mol of *Vc*Fur. In the case of Mn(II), 2.3 mol of manganese was incorporated per mol of *Vc*Fur

which is quite unusual as *Ec*Fur has little affinity for manganese (Mills and Marletta, 2005).

Table 6.1 Metals analysis of native and EDTA treated *Vc*Fur samples

Added Zn(II) used in purification	EDTA	Mol of Zn/ mol of *Vc*Fur*	Mol of Ni/ mol of *Vc*Fur*
Yes	No	2.2±0.1	0.1±0.02
Yes	Yes	0.1±0.01	Not determined[a]
No	No	0.9±0.1	0.04±0.005
No	Yes	0.1±0.01	Not determined[a]

* Averages from triplicate studies

[a] < 0.01 mol of metals per mol of *Vc*Fur monomer

Table 6.2 Analysis of metal uptake into EDTA treated *Vc*Fur samples

Use of metals for reconstitution	Mol of metals/ mol of Fur*		
	Zn(II)	Fe(III)	Mn(II)
Zn(II)	2.3±0.1	Not determined[a]	Not determined[a]
Fe(III)	0.1±0.01	1.6±0.2	Not determined[a]
Mn(II)	0.1±0.01	Not determined[a]	2.1±0.2

* Averages from triplicate studies

[a] < 0.01 mol of metals per mol of *Vc*Fur monomer

6.4 Mobility shift assay

A DNA mobility shift assay was used to examine the interaction of *Vc*Fur with duplex DNA corresponding to the *fur* promoter of the gene *vc2694* (Mey *et al.*, 2005). Zn_1*Vc*Fur, that is the form containing one zinc per monomer, binds to the promoter (Figure 6.4.A). When $MnSO_4$, $ZnCl_2$ or $FeCl_3$ was added to Zn_1*Vc*Fur in the binding and running buffers, there was also a positive DNA mobility shift (Figure 6.4.B-D). When EDTA was used in the gel preparation and binding buffer with Zn_1*Vc*Fur, no shift was found (Figure 6.4.E), suggesting that apo-*Vc*Fur does not bind to DNA. Upon adding metal salts to

the apo-*Vc*Fur (150 μM of $ZnCl_2$, or 200 μM $MnSO_4$, or 150 μM of $FeCl_3$) in the binding and running buffer reaction, a mobility shift occurred (Figure 6.4.F-H). No shift was observed when non-promoter poly-A was used as a control (Figure 6.4.I).

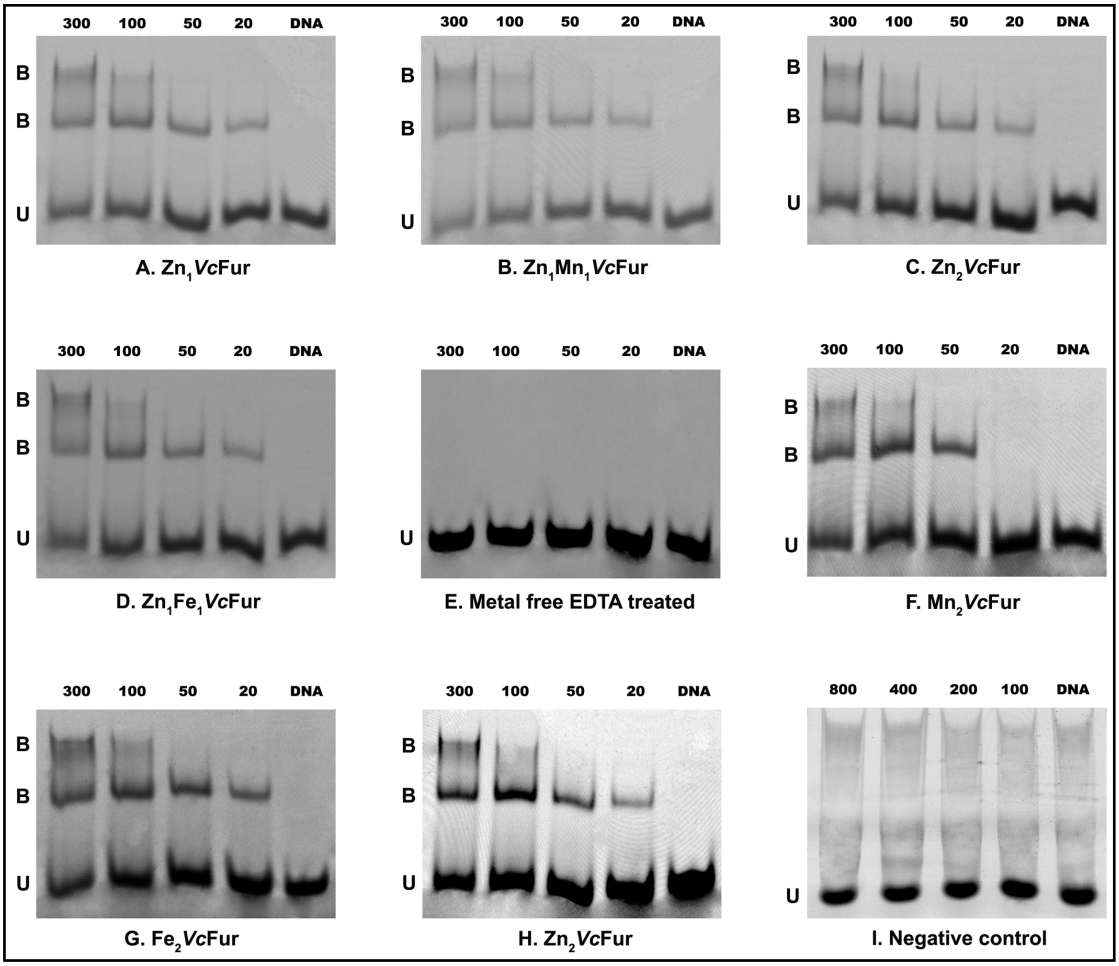


Figure 6.4 DNA mobility shift assays for *Vc*Fur. These assays show the effect of metals on the ability of *Vc*Fur to bind DNA. In each gel the rightmost lane is DNA alone. Upper bands (marked **B**) indicated protein-DNA binding, whereas lower bands (marked **U**) represent unbound fluorescent-labeled DNA. DNA concentrations were 50 nM. Protein concentrations decrease from left to right, and are indicated in nM

6.5 Electron paramagnetic resonance (EPR) spectroscopy of *Vc*Fur

EPR spectroscopy provides different tools to examine the structure and structural changes of large biomolecules. Continuous wave and pulsed EPR can measure up to 80 Å by measuring the magnetic dipole coupling between two unpaired electrons (Schiemann and Prisner, 2007). EPR PELDOR (pulsed electron–electron double resonance) experiments can be used to identify conformational changes of *Vc*Fur DNA binding domain in the presence of different metals and DNA. There are a total of five cysteines present in *Vc*Fur (Cys31, Cys93, Cys96, Cys133 and Cys139 in Figure 5.1). We observed from the crystal structure of *Vc*Fur that there is a disulphide bond between Cys93 and Cys133, which has been described in Section 5.7 and it is assumed that the Cys96 and Cys139 might also form a disulphide bond. So, Cys31 is the only free cysteine that can be spin labeled with MTSL. MTSL can attach to the free cysteine of the protein resulting in an increase in mass of 186.3 Da per spin label. The *Vc*Fur protein was spin labeled using the procedure described in section 2.24.6. ESI-TOF MS (electrospray-time of flight mass spectrometry) analysis shows that the native recombinant monomer of *Vc*Fur is 17,025 Da whereas the MTSL labeled monomer is 17,210 Da. (Figure 6.5.A). Tryptic digested fragments of the *Vc*Fur protein were also analyzed and confirmed that the spin label was present at position Cys31 (Figure 6.5.B). As *Vc*Fur is a dimer and only one Cys31 in the DNA binding domain of each monomer can be labeled, any movements of the DNA binding domains due to metal coordination or formation of *Vc*Fur-DNA complex domain can be precisely measured by analysis of intensities generated by MTSL. These EPR experiments are currently underway.

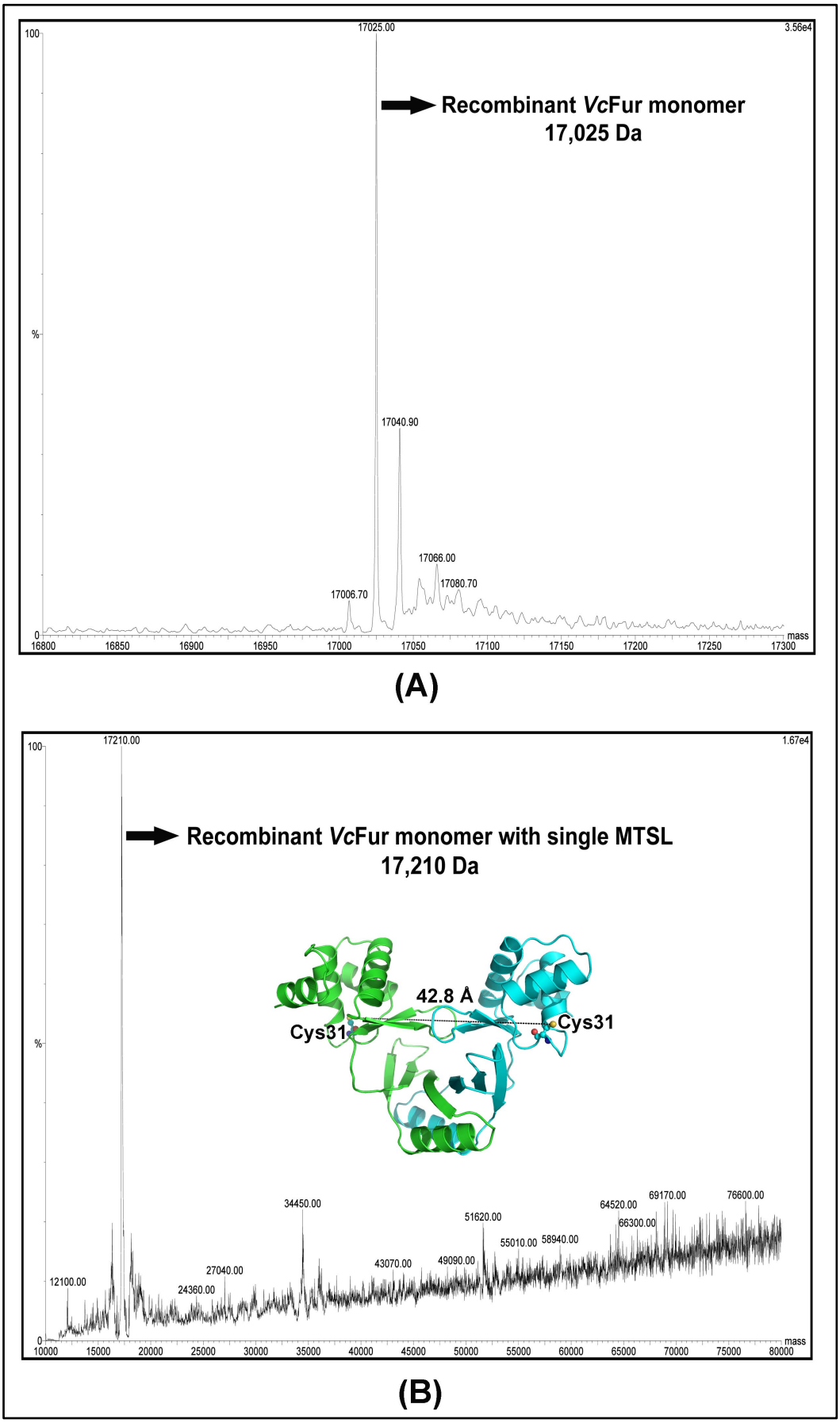


Figure 6.5 Spin labeling of *Vc*Fur

(A) Recombinant *Vc*Fur monomer peaks from mass spectrometry data. **(B)** MTSL labeled *Vc*Fur monomer sample. Inset is an image of *Vc*Fur showing the location of Cys31 and the distance between the sulphur atoms observed in the crystal structure.

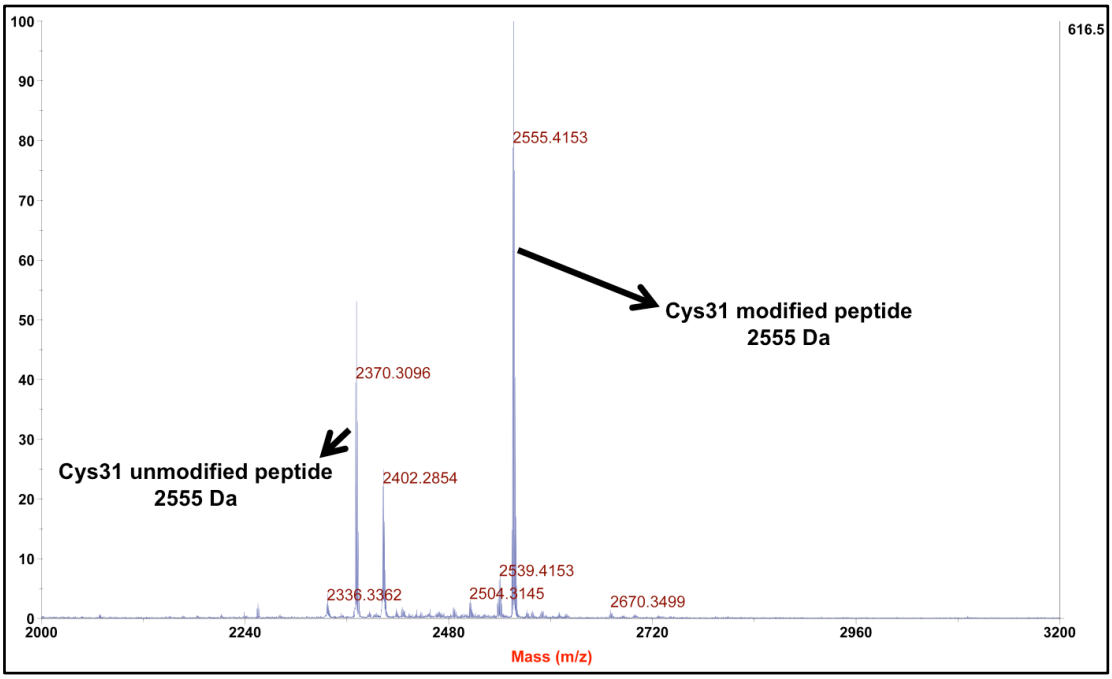


Figure 6.6 Confirmation of Cys31 spin label in *Vc*Fur

6.6 Discussion

6.6.1 Structure and metal ligands of *Vc*Fur

Given the key role of *Vc*Fur in the survival and pathogenesis of *V. cholerae*, we embarked on a structural investigation of *Vc*Fur with the ultimate aim of exploring the potential of developing an allosteric inhibitor of *Vc*Fur as a possible treatment for cholera. In this study, we present the crystal structure of *Vc*Fur to a resolution of 2.6 Å. An analysis of the metal binding characteristics of *Vc*Fur show that the protein can bind to promoter DNA in the presence of several divalent metals, and that the crystal structure contains two zincs per monomer. The residues coordinating the zincs are very similar

to those in *Pa*Fur, however the coordination details, particularly in Site 2, are different and this is possibly linked to a dramatic difference in the orientation of the DNA-binding domains relative to the dimerization domain.

Recombinant *Vc*Fur, when purified without the addition of metal salts, contains essentially one zinc per monomer (0.9 mol of Zn(II) per mol of *Vc*Fur, (Table 6.1), and this form, Zn_1*Vc*Fur, is able to bind to promoter DNA (Figure 6.4.A). The addition of Mn(II), Zn(II), or Fe(III) to the Zn_1*Vc*Fur form maintains the ability of *Vc*Fur to bind to promoter DNA (Figure 6.4.B-D). Zinc can be removed from *Vc*Fur by EDTA treatment (0.1 mol of Zn(II) per mol of *Vc*Fur), and this form does not bind to promoter DNA (Figure 6.4.E). Adding back metals to EDTA-treated *Vc*Fur produces 2.3 mol of Zn(II), 1.6 mol of Fe(III) or 2.1 mol of Mn(II) per mol of *Vc*Fur (Table 6.2), and these forms are able to bind to promoter DNA (Figure 6.4.F-H). As reported for other Fur proteins, *Vc*Fur shares the ability to bind DNA in the presence of different metals (Bagg and Neilands, 1987; de Lorenzo *et al.*, 1987). Analysis of metal binding to *Ec*Fur showed that in the presence of Zn(II) it contains 2.1 (Althaus *et al.*, 1999) or 2.3 (Mills and Marletta, 2005) mol of Zn (II) per mol of *Ec*Fur, and that EDTA could remove one mol of zinc per monomer. Removal of all zinc from *Ec*Fur required urea treatment in addition to EDTA (Althaus *et al.*, 1999). EDTA treatment of *Vc*Fur resulted in 0.1 mol of Zn(II) per mol of *Vc*Fur, similar to the value of 0.19 mol reported for *Pa*Fur (Lewin *et al.*, 2002). *Vc*Fur therefore contains two metal sites per monomer, and occupancy of at least one of these sites is necessary for DNA binding. The fact that Zn_1*Vc*Fur can bind to DNA suggests that the single zinc per monomer is binding to the regulatory metal binding site, and that this is the high affinity site that requires

EDTA to remove the metal. Other divalent metals can bind at both sites and promote binding to DNA.

The crystal structure of *Pa*Fur revealed the structural details of the metal binding sites for the first time for a member of the Fur family (Pohl *et al.*, 2003). It was suggested in the *Pa*Fur study that of the two zinc binding sites per monomer, the Zn(II) at site 1 (Zn1 in Figure 5.7.A and 5.11) is a low affinity metal binding regulatory site that can readily exchange iron causing a conformational change in the DNA binding domain of the protein, and that the Zn(II) at site 2 (Zn2 in Figure 5.7.A and 5.11) may be a structural metal binding site which helps to maintain the integrity of the protein (Pohl *et al.*, 2003). This interpretation has been challenged, based on an analysis of the reported mutagenesis of the ligands involved in binding the Zn(II) at site 2 in Fur from several species, and suggests that site 1 (Zn1) is the low affinity metal binding site with site 2 (Zn2) being the high affinity iron-sensing site (Lee and Helmann, 2007). This interpretation has been supported by a recent molecular dynamics and free energy calculation study of the binding of metals to *Pa*Fur that proposes site 2 as the regulatory iron-sensing site recognising a hexacoordinated Fe(II) in an octahedral environment (Ahmad *et al.*, 2008).This *in silico* study also suggested that His87 in *Pa*Fur might be an additional ligand of the metal at site 2. The structure of *Vc*Fur reported herein also reveals two metal binding sites per monomer, occupied by Zn(II). Zn1 at site 1 in *Vc*Fur is tetracoordinated by His87, Asp89, Glu108 and His125, residues that are conserved across all Fur proteins, including *Ec*Fur (Figure 5.1). Although the corresponding residues in *Pa*Fur coordinate the Zn(II), the coordination involves six ligands achieved through a bidendate interaction

with Asp88 and by a coordinated water molecule. Zn2 at site 2 in *Vc*Fur is tetracoordinated to His33, Glu81, His88 and His90 in contrast to *Pa*Fur where His32, Glu80, His89 and Glu100 tetracoordinate the Zn(II). His87 in *Pa*Fur (the equivalent of His88 in *Vc*Fur) is disordered in the crystal structure of *Pa*Fur, but the fact that it is observed as a ligand to the Zn(II) in *Vc*Fur lends support to its involvement in coordinating Fe(II) at the regulatory site (Ahmad *et al.*, 2008). Glu101 in *Vc*Fur (the equivalent of Glu100 in *Pa*Fur) is ordered, but too far away to coordinate to the Zn(II), however all five potential ligands His33, Glu81, His88, His90 and Glu101 are conserved across all Fur proteins and may therefore be involved in a hexacoordinated octahedral interaction with Fe(II). A His90Ala mutant of *Vc*Fur showed that His90 is critical for sensing iron (Lam *et al.*, 1994), as have mutational studies of the same residue in Fur from other species (Bsat and Helmann, 1999; Hall and Foster, 1996; Lewin *et al.*, 2002). Mutational studies on *Vibrio alginolyticus* Fur, which shares 94% sequence identity to *Vc*Fur, shows that mutation of His33 or His90, ligands of Zn2, completely inactivates Fur, whereas mutation of His87 and His125, ligands of Zn1, resulted in just a 2-fold reduction of Fur activity (Liu *et al.*, 2007). It can be concluded that site 1 (Zn1) plays a structural role and is not critical for activity, and site 2 (Zn2) is the regulatory site. A study of *Ec*Fur concluded that *in vivo* zinc concentrations are too low for Zn(II) to be a regulatory ligand, and that Fe(II) is the physiologically-relevant ligand that activtes Fur (Mills and Marletta, 2005). This is also likely to be the case for *Vc*Fur, but Mn(II) may also be a ligand given its ability to occupy both metal binding sites and activate DNA binding, however an analysis of metal concentrations in *V. cholerae* would be needed to confirm this.

There have been many reports, based on a study of *Ec*Fur, that four cysteines (Cys92, Cys95, Cys132, Cys137, in *Ec*Fur numbering) conserved throughout most of the Fur family are involved in coordination of zinc (Jacquamet *et al.*, 1998), specifically the Cys92XXCys95 motif (Coy *et al.*, 1994; Gonzalez de Peredo *et al.*, 1999), and that the redox state of these cysteines and coordination of zinc are essential for maintaining *Ec*Fur in a dimeric state (D'Autreaux *et al.*, 2007). *Pa*Fur appeared to be an exception as the second cysteine of this motif is a threonine (Figure 5.1). Despite having the conserved cysteines the structure of *Vc*Fur presented here does not reveal any metal binding site involving the cysteines, but shows that there is a disulphide linkage between Cys93 and Cys133 (equivalent to Cys92 and Cys132 in *Ec*Fur), in conflict with the disulphide analysis of the monomer to dimer transition of *Ec*Fur (D'Autreaux *et al*., 2007). It may be that this disulphide observed in *Vc*Fur would not form in the reducing environment of the cell. In the *Vc*Fur structure there is no electron density beyond Cys133, although mass spectrometry analysis confirms that the C-terminal residues are present in the crystal. The similarity between *Ec*Fur and *Vc*Fur in the number of metals bound per monomer, and the observation that the two zinc binding sites are conserved in the crystal structures of *Pa*Fur and *Vc*Fur, involving residues conserved across all Fur proteins, suggests that *Ec*Fur may also share the same two metal binding sites. How this squares with the reports on zinc binding to the conserved cysteines in *Ec*Fur is difficult to explain, although there is report that the cysteines in *Ec*Fur do not bind zinc (Saito *et al.*, 1991). Perhaps family members have subtle variations in metal binding sites that maintain structural integrity rather than regulatory roles.

6.6.2 DNA-binding model

Comparison of the crystal structures of *Vc*Fur and *Pa*Fur reveals a significant shift in the position of the DNA-binding domain relative to the dimerization domain, such that the mid-point distance between helix 4 in the dimer reduces from 33 Å in *Pa*Fur to between 23 and 27 Å in *Vc*Fur. For example the distance between the two Cα atoms of the conserved tyrosine in helix 4 of the dimer (Tyr55 in *Pa*Fur, Tyr56 in *Vc*Fur), that has been implicated in DNA recognition of the Fur-box (Tiss *et al.*, 2005), is 31 Å in *Pa*Fur and 21 Å in the *Vc*Fur. The consensus Fur-box sequences of *E. coli, P. aeruginosa* and *V. cholerae* are very similar: 5'- GATAATGATAATCATTATC -3', 5'- GATAATGATAATCATTATC -3' and 5'- GATAATGATAATNATTATC -3', respectively (de Lorenzo *et al.*, 1987; Mey *et al.*, 2005; Ochsner and Vasil, 1996). Although the Fur-box sequence is an inverted repeat sequence, there is extensive evidence that more than one Fur dimer binds to the Fur-box. One study described the Fur-box as a head-to-head-to-tail repeat of the simple hexamer GATAAT and proposed that this hexamer is the minimal recognition unit for Fur (Escolar *et al.*, 1998), and that longer operator regions can be formed by simple repeats of the hexamer (Escolar *et al.*, 2000). An alternative view of the Fur-box is that it contains a conserved 15-bp (7-1-7) inverted repeat presented twice within the 19-bp consensus sequence (Figure 6.7) (Baichoo and Helmann, 2002). This study showed that the stoichiometry of binding was one Fur dimer per 15-bp inverted repeat, and two Fur dimers per 19-bp Fur box, with each 7-1-7 inverted repeat predicted to bind a Fur dimer on opposite sides of the DNA. Such a model of binding was proposed for *Pa*Fur (Pohl *et al.*, 2003). The diptheria toxin repressor (DtxR) operator

sequence has been described as two similar 7-0-7 inverted repeats, and structural studies on DtxR binding to DNA shows two DtxR dimers binding to these repeats on either side of the DNA duplex (Figure 6.7) (Pohl *et al.*, 1999; White *et al.*, 1998).

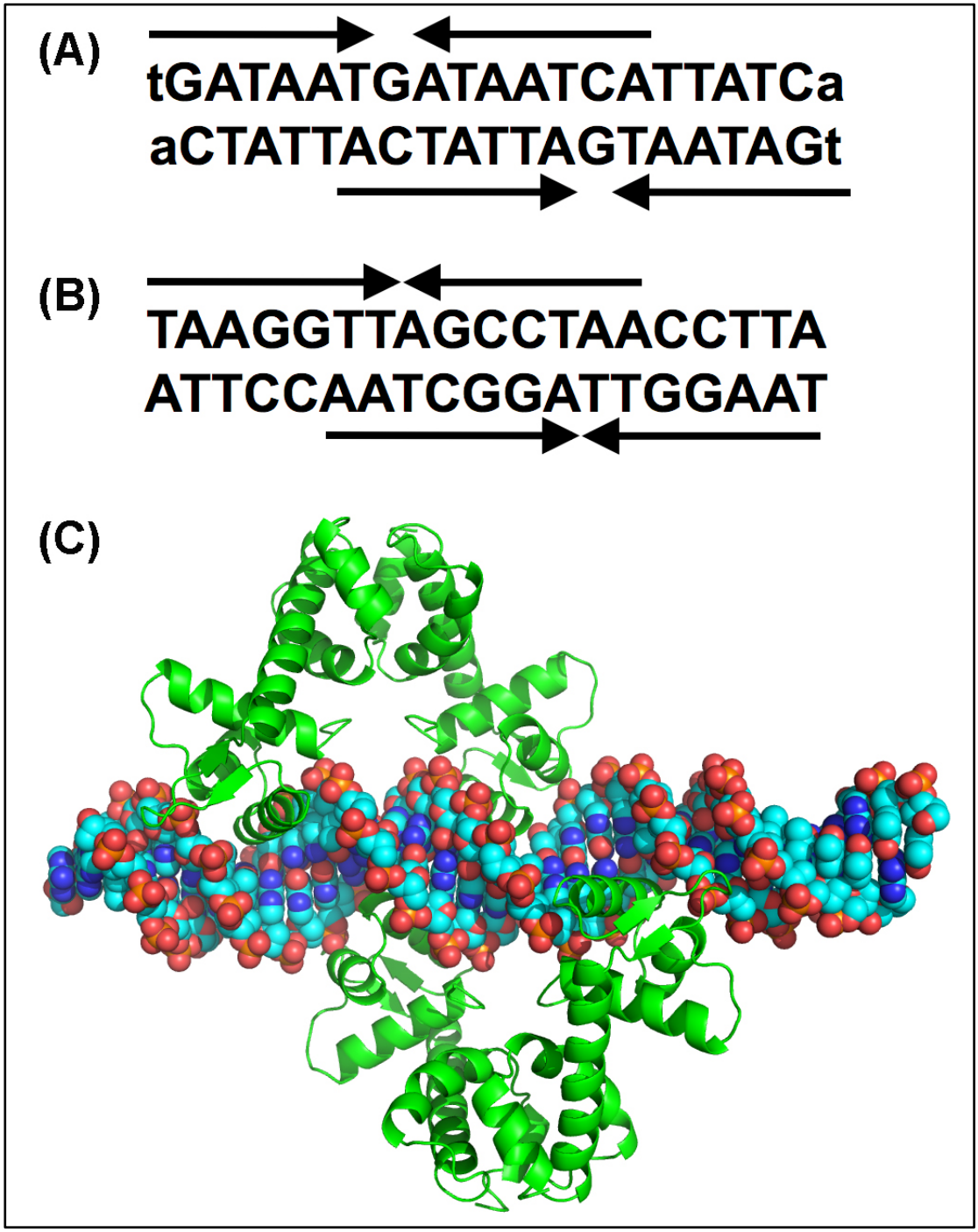


Figure 6.7 Comparison of possible Fur-DNA complex with the DtxR-DNA. Adapted from Baichoo *et. al*. (Baichoo and Helmann, 2002). **(A)** 19-bp Fur

box binding site with two overlapping 7-1-7 heptamer motifs. **(B)** Two overlapping imperfect inverted repeats that generate the 19-bp binding site for DtxR (Chen *et al.*, 2000). **(C)** DtxR-DNA complex (PDB code 1ddn) involves two dimers of DtxR bound to the operator DNA.

This study, when combined with the *Pa*Fur study, has shown that there is flexibility of the DNA-binding domains relative to the dimerization domain, with the distance between the two key DNA-binding helices (helix 4) varying from 23 Å to 33 Å. It could be envisaged that one particular form may be better stabilized through the hexacoordination of Fe(II) at site 2 (Zn2) involving the five conserved residues His33, Glu81, His88, His90 and Glu101 (*Vc*Fur numbering) with one of the glutamic acid residues forming a bidentate interaction, and this is supported by the recent molecular modeling studies (Ahmad *et al.*, 2008). In both the *Pa*Fur and *Vc*Fur structures zinc is only seen to form tetracoordination at site 2 with the consequence that in *Vc*Fur the DNA-binding domains are not in an optimal position for DNA-binding. Clearly the Zn(II) form can bind to promoter DNA, but the conformational changes required for association with DNA would provide an energetic penalty that would reduce the binding affinity in line with the reduced affinity seen for Zn(II) *Ec*Fur compared to Fe(II) *Ec*Fur (Mills and Marletta, 2005).

In conclusion, these studies on *Vc*Fur support the hypothesis that site 2 is the regulatory site where Fe(II), or possibly Mn(II), binds to control repression *in vivo*. The variation in the position of the DNA-binding domains compared to that observed in *Pa*Fur suggests that Zn(II) binding at site 2 (Zn2) may not lock the DNA-binding domains in the optimal position for DNA binding. The combined structural studies suggest that Fe(II) binding to site 2 would employ

five conserved residues to form a hexacoordinated interaction rather than the tetracoordinated interaction seen with Zn(II), and that these additional ligands would help generate the optimal orientation of the DNA-binding domains. A more complete understanding still awaits the elucidation of a Fur-DNA complex, ideally in the presence of iron, but these studies on *Vc*Fur provide further insights into this important family of gene activators and repressors.

6.7 Ongoing and future work

The structure of *Vc*Fur will definitely improve our understanding of this highly conserved metalloregulatory transcriptional factor. However, the structure resolution needs to be improved to better understand metal co-ordination of *Vc*Fur. As the bacterial cell cytoplasm is always in a reduced condition, the presence of disulfide bridges between cysteine residues Cys93 and Cys133, and between Cys96 and Cys139 in *Vc*Fur is questionable *in vivo*. Purified *Vc*Fur protein in the absence of DTT contains disulphide linkages, consistent with the results of spin labeling (Section 6.5). Further biochemical characterization of the protein with regard to the oxidation state of the cysteines, the oligomeric state of the protein as a function of this redox state and of the metal binding are underway. Improvements in mobility shift assay, by using gel purified double stranded DNA, are also in progress. Microcrystals of *Vc*Fur-DNA complex were observed in certain crystallization conditions and further optimizations are in progress. To investigate DNA binding domain movements of *Vc*Fur in the presence of different metals and DNA, electron paramagnetic resonance (EPR) experiments are underway in the University of St Andrews.

In *V. cholerae*, more than 100 genes can be regulated by *Vc*Fur and *fur* gene knockout strain showed significant reduction in virulence (Mey *et al*., 2005). Structural investigation and functional characterization of *Vc*Fur was carried out to design a potential inhibitor. Future approaches should aim to design a small molecule inhibitor to keep *Vc*Fur as an inactive form either by blocking the DNA or metal ligand binding sites. Equally, upregulation of *Vc*Fur could potentially block the expression of many genes involved in maintaining iron homeostasis in *V. cholerae*. Hence, another approach could be to design an allosteric activator to keep *Vc*Fur constitutively active. Future experiments will involve identification of lead compounds using fragment based screening, thermal shift assay, computer analysis, crystallography, synthetic chemistry and biochemical assay.

References

Adams, P.D., Grosse-Kunstleve, R.W., L.W., H., Ioerger, T.R., McCoy, A., Moriarty, N., Read, R., Sacchetttini, J., Sauter, N., and Terwilliger, T. (2002) PHENIX: building new software for automatic crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**: 1948-1954.

Ahmad, R., Brandsdal, B.O., Michaud-Soret, I., and Willassen, N.P. (2008) Ferric uptake regulator protein: Binding free energy calculations and per-residue free energy decomposition. *Proteins* **75**: 373-386.

Aldova, E., Laznickova, K., Stepankova, E., and Lietava, J. (1968) Isolation of nonagglutinable vibrios from an enteritis outbreak in Czechoslovakia. *J Infect Dis* **118**: 25-31.

Althaus, E.W., Outten, C.E., Olson, K.E., Cao, H., and O'Halloran, T.V. (1999) The ferric uptake regulation (Fur) repressor is a zinc metalloprotein. *Biochemistry* **38**: 6559-6569.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Altschul, S.F., and Koonin, E.V. (1998) Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. *Trends Biochem Sci* **23**: 444-447.

Andrews, S.C., Robinson, A.K., and Rodriguez-Quinones, F. (2003) Bacterial iron homeostasis. *FEMS Microbiol Rev* **27**: 215-237.

Bagg, A., and Neilands, J.B. (1987) Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry* **26**: 5471-5477.

Baichoo, N., and Helmann, J.D. (2002) Recognition of DNA by Fur: a reinterpretation of the Fur box consensus sequence. *J Bacteriol* **184**: 5826-5832.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**: 10037-10041.

Barua, D., and Paguio, A.S. (1977) ABO blood groups and cholera. *Ann Hum Biol* **4**: 489-492.

Bennett-Lovsey, R.M., Herbert, A.D., Sternberg, M.J., and Kelley, L.A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* **70**: 611-625.

Bereswill, S., Lichte, F., Vey, T., Fassbinder, F., and Kist, M. (1998) Cloning and characterization of the fur gene from *Helicobacter pylori*. *FEMS Microbiol Lett* **159**: 193-200.

Berish, S.A., Subbarao, S., Chen, C.Y., Trees, D.L., and Morse, S.A. (1993) Identification and cloning of a fur homolog from *Neisseria gonorrhoeae*. *Infect Immun* **61**: 4599-4606.

Bik, E.M., Bunschoten, A.E., Gouw, R.D., and Mooi, F.R. (1995) Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. *Embo J* **14**: 209-216.

Bork, P. (2000) Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res* **10**: 398-400.

Boyd, E.F., and Waldor, M.K. (2002) Evolutionary and functional analyses of variants of the toxin-coregulated pilus protein TcpA from toxigenic *Vibrio cholerae* non-O1/non-O139 serogroup isolates. *Microbiology* **148**: 1655-1666.

Braun, V., Hantke, K., and Koster, W. (1998) Bacterial iron transport: mechanisms, genetics, and regulation. *Met Ions Biol Syst* **35**: 67-145.

Brickman, T.J., and Armstrong, S.K. (1995) *Bordetella pertussis* fur gene restores iron repressibility of siderophore and protein expression to deregulated *Bordetella bronchiseptica* mutants. *J Bacteriol* **177**: 268-270.

Bruni, R., and Roizman, B. (1996) Open reading frame P--a herpes simplex virus gene repressed during productive infection encodes a protein that binds a splicing factor and reduces synthesis of viral proteins made from spliced mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 10423-10427.

Bsat, N., Herbig, A., Casillas-Martinez, L., Setlow, P., and Helmann, J.D. (1998) *Bacillus subtilis* contains multiple Fur homologues: identification of the iron uptake (Fur) and peroxide regulon (PerR) repressors. *Mol Microbiol* **29**: 189-198.

Bsat, N., and Helmann, J.D. (1999) Interaction of *Bacillus subtilis* Fur (ferric uptake repressor) with the dhb operator in vitro and in vivo. *J Bacteriol* **181**: 4299-4307.

Butler, T., Arnold, M., and Islam, M. (1989) Depletion of hepatic glycogen in the hypoglycaemia of fatal childhood diarrhoeal illnesses. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **83**: 839-843.

Carson, S.D., Thomas, C.E., and Elkins, C. (1996) Cloning and sequencing of a *Haemophilus ducreyi* fur homolog. *Gene* **176**: 125-129.

CCP4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760-763.

Chen, C.S., White, A., Love, J., Murphy, J.R., and Ringe, D. (2000) Methyl groups of thymine bases are important for nucleic acid recognition by DtxR. *Biochemistry* **39**: 10397-10407.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**: 3497-3500.

Clemens, J.D., Sack, D.A., Harris, J.R., Chakraborty, J., Khan, M.R., Huda, S., Ahmed, F., Gomes, J., Rao, M.R., Svennerholm, A.M., and *et al*. (1989) ABO blood groups and cholera: new observations on specificity of risk and modification of vaccine efficacy. *J Infect Dis* **159**: 770-773.

Colwell, R.R., Kaper, J., and Joseph, S.W. (1977) *Vibrio cholerae*, *Vibrio parahaemolyticus*, and other vibrios: occurrence and distribution in Chesapeake Bay. *Science* **198**: 394-396.

Colwell, R.R., Seidler, R.J., Kaper, J., Joseph, S.W., Garges, S., Lockman, H., Maneval, D., Bradford, H., Roberts, N., Remmers, E., Huq, I., and Huq, A. (1981) Occurrence of *Vibrio cholerae* serotype O1 in Maryland and Louisiana estuaries. *Appl Environ Microbiol* **41**: 555-558.

Colwell, R.R., and Spira, W.M. (1992) The ecology of *Vibrio cholerae*. In Cholera (D. Barua and W.B. Greenough, 3rd, eds). *Plenum Medical Book*: 107-127.

Coy, M., and Neilands, J.B. (1991) Structural dynamics and functional domains of the fur protein. *Biochemistry* **30**: 8201-8210.

Coy, M., Doyle, C., Besser, J., and Neilands, J.B. (1994) Site-directed mutagenesis of the ferric uptake regulation gene of *Escherichia coli*. *Biometals* **7**: 292-298.

Cvjetanovic, B., and Barua, D. (1972) The seventh pandemic of cholera. *Nature* **239**: 137-138.

D'Autreaux, B., Pecqueur, L., Gonzalez de Peredo, A., Diederix, R.E., Caux-Thang, C., Tabet, L., Bersch, B., Forest, E., and Michaud-Soret, I. (2007) Reversible redox- and zinc-dependent dimerization of the *Escherichia coli* fur protein. *Biochemistry* **46**: 1329-1342.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**: W375-383.

de Lorenzo, V., Wee, S., Herrero, M., and Neilands, J.B. (1987) Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor. *J Bacteriol* **169**: 2624-2630.

Deb, T.B., and Datta, K. (1996) Molecular cloning of human fibroblast hyaluronic acid-binding protein confirms its identity with P-32, a protein co-purified with splicing factor SF2. Hyaluronic acid-binding protein as P-32 protein, co-purified with splicing factor SF2. *J Biol Chem* **271**: 2206-2212.

DeLano, W. (2007) The Pymol Molecular Graphics System. De Lano Scientific, Palo Alto, CA, USA.

Delany, I., Rappuoli, R., and Scarlato, V. (2004) Fur functions as an activator and as a repressor of putative virulence genes in *Neisseria meningitidis*. *Molecular Microbiology* **52**: 1081-1090.

Desai, K., Loewenstein, P.M., and Green, M. (1991) Isolation of a cellular protein that binds to the human immunodeficiency virus Tat protein and can potentiate transactivation of the viral promoter. *Proc. Natl. Acad. Sci. U. S. A.* **88**: 8875-8879.

DiRita, V.J. (2001) Molecular basis of Vibrio cholerae pathogenesis. Principles of bacterial pathogenesis (Groisman, E. A., ed.): 457-508.

Donnenberg, M.S., and Kaper, J.B. (1991) Construction of an eae deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. *Infect Immun* **59**: 4310-4317.

Dziejman, M., Balon, E., Boyd, D., Fraser, C.M., Heidelberg, J.F., and Mekalanos, J.J. (2002) Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 1556-1561.

Egan, E.S., Fogel, M.A., and Waldor, M.K. (2005) Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Mol Microbiol* **56**: 1129-1138.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132.

Ernst, J.F., Bennett, R.L., and Rothfield, L.I. (1978) Constitutive expression of the iron-enterochelin and ferrichrome uptake systems in a mutant strain of *Salmonella typhimurium*. *J Bacteriol* **135**: 928-934.

Escolar, L., Perez-Martin, J., and de Lorenzo, V. (1998) Binding of the fur (ferric uptake regulator) repressor of *Escherichia coli* to arrays of the GATAAT sequence. *J Mol Biol* **283**: 537-547.

Escolar, L., Perez-Martin, J., and de Lorenzo, V. (1999) Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol* **181**: 6223-6229.

Escolar, L., Perez-Martin, J., and de Lorenzo, V. (2000) Evidence of an unusually long operator for the fur repressor in the aerobactin promoter of *Escherichia coli*. *J Biol Chem* **275**: 24709-24714.

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

Faruque, S.M., Asadulghani, Saha, M.N., Alim, A.R., Albert, M.J., Islam, K.M., and Mekalanos, J.J. (1998b) Analysis of clinical and environmental strains of nontoxigenic *Vibrio cholerae* for susceptibility to CTXPhi: molecular basis for origination of new strains with epidemic potential. *Infect Immun* **66**: 5819-5825.

Faruque, S.M., Chowdhury, N., Kamruzzaman, M., Ahmad, Q.S., Faruque, A.S., Salam, M.A., Ramamurthy, T., Nair, G.B., Weintraub, A., and Sack, D.A. (2003a) Reemergence of epidemic *Vibrio cholerae* O139, Bangladesh. *Emerg Infect Dis* **9**: 1116-1122.

Faruque, S.M., Kamruzzaman, M., Meraj, I.M., Chowdhury, N., Nair, G.B., Sack, R.B., Colwell, R.R., and Sack, D.A. (2003b) Pathogenic potential of environmental *Vibrio cholerae* strains carrying genetic variants of the toxin-coregulated pilus pathogenicity island. *Infect Immun* **71**: 1020-1025.

Faruque, S.M., and Mekalanos, J.J. (2003) Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends Microbiol* **11**: 505-510.

Feachem, R., Miller, C., and Drasar, B. (1981) Environmental aspects of cholera epidemiology. II. Occurrence and survival of *Vibrio cholerae* in the environment. *Trop Dis Bull* **78**: 865-880.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., and *et al*. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.

Fluit, A.C., and Schmitz, F.J. (2004) Resistance integrons and super-integrons. *Clin Microbiol Infect* **10**: 272-288.

Fullner, K.J., and Mekalanos, J.J. (1999) Genetic characterization of a new type IV-A pilus gene cluster found in both classical and El Tor biotypes of *Vibrio cholerae*. *Infect Immun* **67**: 1393-1404.

Gaboriaud, C., Juanhuix, J., Gruez, A., Lacroix, M., Darnault, C., Pignol, D., Verger, D., Fontecilla-Camps, J.C., and Arlaud, G.J. (2003) The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties. *J Biol Chem* **278**: 46974-46982.

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

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C.I. (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* **164**: 918-921.

Ghai, R., Waters, P., Roumenina, L.T., Gadjeva, M., Kojouharova, M.S., Reid, K.B., Sim, R.B., and Kishore, U. (2007) C1q and its growing family. *Immunobiology* **212**: 253-266.

Ghebrehiwet, B., Lim, B.L., Peerschke, E.I., Willis, A.C., and Reid, K.B. (1994) Isolation, cDNA cloning, and overexpression of a 33-kD cell surface glycoprotein that binds to the globular "heads" of C1q. *J Exp Med* **179**: 1809-1821.

Ghiran, I., Klickstein, L.B., and Nicholson-Weller, A. (2003) Calreticulin is at the surface of circulating neutrophils and uses CD59 as an adaptor molecule. *J Biol Chem* **278**: 21024-21031.

Glass, R.I., Huq, I., Alim, A.R., and Yunus, M. (1980) Emergence of multiply antibiotic-resistant *Vibrio cholerae* in Bangladesh. *J Infect Dis* **142**: 939-942.

Glass, R.I., Holmgren, J., Haley, C.E., Khan, M.R., Svennerholm, A.M., Stoll, B.J., Belayet Hossain, K.M., Black, R.E., Yunus, M., and Barua, D. (1985) Predisposition for cholera of individuals with O blood group. Possible evolutionary significance. *Am J Epidemiol* **121**: 791-796.

Gonzalez de Peredo, A., Saint-Pierre, C., Adrait, A., Jacquamet, L., Latour, J.M., Michaud-Soret, I., and Forest, E. (1999) Identification of the two zinc-bound cysteines in the ferric uptake regulation protein from *Escherichia coli*: chemical modification and mass spectrometry analysis. *Biochemistry* **38**: 8582-8589.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**: 305-308.

Haber, F., and Weiss, J. (1932) On the catalysis of hydroperoxide. *Naturwissenschaften* **20**: 948-950.

Hacker, J., Blum-Oehler, G., Muhldorfer, I., and Tschape, H. (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23**: 1089-1097.

Hall, H.K., and Foster, J.W. (1996) The role of fur in the acid tolerance response of *Salmonella typhimurium* is physiologically and genetically separable from its role in iron acquisition. *J Bacteriol* **178**: 5683-5691.

Hall, R.M., and Collis, C.M. (1995) Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol Microbiol* **15**: 593-600.

Halstensen, T.S., Mollnes, T.E., Garred, P., Fausa, O., and Brandtzaeg, P. (1992) Surface epithelium related activation of complement differs in Crohn's disease and ulcerative colitis. *Gut* **33**: 902-908.

Harkey, C.W., Everiss, K.D., and Peterson, K.M. (1994) The *Vibrio cholerae* toxin-coregulated-pilus gene tcpI encodes a homolog of methyl-accepting chemotaxis proteins. *Infect Immun* **62**: 2669-2678.

Hayward, S., and Berendsen, H.J. (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30**: 144-154.

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., and *et al*. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**: 477-483.

Heidrich, C., Hantke, K., Bierbaum, G., and Sahl, H.G. (1996) Identification and analysis of a gene encoding a Fur-like protein of *Staphylococcus epidermidis*. *FEMS Microbiol Lett* **140**: 253-259.

Herwald, H., Dedio, J., Kellner, R., Loos, M., and Muller-Esterl, W. (1996) Isolation and characterization of the kininogen-binding protein p33 from endothelial cells. Identity with the gC1q receptor. *J Biol Chem* **271**: 13040-13047.

ICDDRB (1993) Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. Cholera Working Group, International Centre for Diarrhoeal Diseases Research, Bangladesh. *Lancet* **342**: 387-390.

Islam, M.S., Drasar, B.S., and Bradley, D.J. (1990) Long-term persistence of toxigenic *Vibrio cholerae* O1 in the mucilaginous sheath of a blue-green alga, Anabaena variabilis. *J Trop Med Hyg* **93**: 133-139.

Islam, M.S., Drasar, B.S., and Sack, R.B. (1993) The aquatic environment as a reservoir of *Vibrio cholerae*: a review. *J Diarrhoeal Dis Res* **11**: 197-206.

Islam, M.S., Drasar, B.S., and Sack, R.B. (1994) Probable role of blue-green algae in maintaining endemicity and seasonality of cholera in Bangladesh: a hypothesis. *J Diarrhoeal Dis Res* **12**: 245-256.

Jacquamet, L., Aberdam, D., Adrait, A., Hazemann, J.L., Latour, J.M., and Michaud-Soret, I. (1998) X-ray absorption spectroscopy of a new zinc site in the fur protein from *Escherichia coli*. *Biochemistry* **37**: 2564-2571.

Jermyn, W.S., and Boyd, E.F. (2002) Characterization of a novel Vibrio pathogenicity island (VPI-2) encoding neuraminidase (nanH) among toxigenic *Vibrio cholerae* isolates. *Microbiology* **148**: 3681-3693.

Jiang, J., Zhang, Y., Krainer, A.R., and Xu, R.M. (1999) Crystal structure of human p32, a doughnut-shaped acidic mitochondrial matrix protein. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 3572-3577.

Jones TA, Z.J., Cowan SW, Kjeldgaard. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr Sect A* **47(Part2)**: 110-119.

Kantardjieff, K.A., and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**: 1865-1871.

Kaper, J.B., Morris, J.G., Jr., and Levine, M.M. (1995) Cholera. *Clin Microbiol Rev* **8**: 48-86.

Kapp, C. (2009) Zimbabwe's humanitarian crisis worsens. *The Lancet* **373**: 447.

Karaolis, D.K., Johnson, J.A., Bailey, C.C., Boedeker, E.C., Kaper, J.B., and Reeves, P.R. (1998) A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 3134-3139.

Kissinger, C.R., Parge, H.E., Knighton, D.R., Lewis, C.T., Pelletier, L.A., Tempczyk, A., Kalish, V.J., Tucker, K.D., Showalter, R.E., Moomaw, E.W., and *et al*. (1995) Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* **378**: 641-644.

Ko, W.C., Chuang, Y.C., Huang, G.C., and Hsu, S.Y. (1998) Infections due to non-O1 *Vibrio cholerae* in southern Taiwan: predominance in cirrhotic patients. *Clinical Infectious disease* **27**: 774-780.

Kovach, M.E., Shaffer, M.D., and Peterson, K.M. (1996) A putative integrase gene defines the distal end of a large cluster of ToxR-regulated colonization genes in *Vibrio cholerae*. *Microbiology* **142 (Pt 8)**: 2165-2174.

Krainer, A.R., Mayeda, A., Kozak, D., and Binns, G. (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. *Cell* **66**: 383-394.

Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**: 2256-2268.

Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**: 774-797.

Lahiri, M.N., Das, B.C., and Malik, K.S. (1939) The viability of *Vibrio cholerae* in natural water. *Indian. Med. Gaz* **74**: 742-744.

Lam, M.S., Litwin, C.M., Carroll, P.A., and Calderwood, S.B. (1994) *Vibrio cholerae* fur mutations associated with loss of repressor activity: implications for the structural-functional relationships of fur. *J Bacteriol* **176**: 5108-5115.

Laskowski, R.A., Watson, J.D., and Thornton, J.M. (2005) ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* **33**: W89-93.

Lattman, E. (2004) The state of the Protein Structure Initiative. *Proteins* **54**: 611-615.

Lee, J.W., and Helmann, J.D. (2007) Functional specialization within the Fur family of metalloregulators. *Biometals* **20**: 485-499.

Leoni, L., Ciervo, A., Orsi, N., and Visca, P. (1996) Iron-regulated transcription of the pvdA gene in *Pseudomonas aeruginosa*: effect of Fur and PvdS on promoter activity. *J Bacteriol* **178**: 2299-2313.

Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* **26**.

Lewin, A.C., Doughty, P.A., Flegg, L., Moore, G.R., and Spiro, S. (2002) The ferric uptake regulator of *Pseudomonas aeruginosa* has no essential cysteine residues and does not contain a structural zinc ion. *Microbiology* **148**: 2449-2456.

Lim, B.L., Reid, K.B., Ghebrehiwet, B., Peerschke, E.I., Leigh, L.A., and Preissner, K.T. (1996) The binding protein for globular heads of complement C1q, gC1qR. Functional expression and characterization as a novel vitronectin binding factor. *J Biol Chem* **271**: 26739-26744.

Lin, W., Fullner, K.J., Clayton, R., Sexton, J.A., Rogers, M.B., Calia, K.E., Calderwood, S.B., Fraser, C., and Mekalanos, J.J. (1999) Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 1071-1076.

Litwin, C.M., Boyko, S.A., and Calderwood, S.B. (1992) Cloning, sequencing, and transcriptional regulation of the *Vibrio cholerae fur* gene. *J Bacteriol* **174**: 1897-1903.

Liu, H., and Naismith, J.H. (2009) A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **63**: 102-111.

Liu, Q., Wang, P., Ma, Y., and Zhang, Y. (2007) Characterization of the *Vibrio alginolyticusfur* gene and localization of essential amino acid sites in fur by site-directed mutagenesis. *J Mol Microbiol Biotechnol* **13**: 15-21.

Long, F., Vagin, A.A., Young, P., and Murshudov, G.N. (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* **64**: 125-132.

Loos, M., Martin, H., and Petry, F. (1989) The biosynthesis of C1q, the collagen-like and Fc-recognizing molecule of the complement system. *Behring Inst Mitt*: 32-41.

Lucarelli, D., Vasil, M.L., Meyer-Klaucke, W., and Pohl, E. (2008) The Metal-Dependent Regulators FurA and FurB from *Mycobacterium Tuberculosis*. *Int J Mol Sci* **9**: 1548-1560.

Malhotra, R., Haurum, J., Thiel, S., Jensenius, J.C., and Sim, R.B. (1993) Pollen grains bind to lung alveolar type II cells (A549) via lung surfactant protein A (SP-A). *Biosci Rep* **13**: 79-90.

Marcus, H., Ketley, J.M., Kaper, J.B., and Holmes, R.K. (1990) Effects of DNase production, plasmid size, and restriction barriers on transformation of *Vibrio cholerae* by electroporation and osmotic shock. *FEMS Microbiol Lett* **56**: 149-154.

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

Matthews, D.A., and Russell, W.C. (1998) Adenovirus core protein V interacts with p32--a protein which is associated with both the mitochondria and the nucleus. *J Gen Virol* **79 (Pt 7)**: 1677-1685.

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

Meibom, K.L., Blokesch, M., Dolganov, N.A., Wu, C.Y., and Schoolnik, G.K. (2005) Chitin induces natural competence in *Vibrio cholerae*. *Science* **310**: 1824-1827.

Mekalanos, J.J. (1983) Duplication and amplification of toxin genes in *Vibrio cholerae*. *Cell* **35**: 253-263.

Mekalanos, J.J., Swartz, D.J., Pearson, G.D., Harford, N., Groyne, F., and de Wilde, M. (1983) Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development. *Nature* **306**: 551-557.

Mey, A.R., Wyckoff, E.E., Kanukurthy, V., Fisher, C.R., and Payne, S.M. (2005) Iron and fur regulation in *Vibrio cholerae* and the role of fur in virulence. *Infect Immun* **73**: 8167-8178.

Mhalu, F.S., Mmari, P.W., and Ijumba, J. (1979) Rapid emergence of El Tor *Vibrio cholerae* resistant to antimicrobial agents during first six months of fourth cholera epidemic in Tanzania. *Lancet* **1**: 345-347.

Mills, S.A., and Marletta, M.A. (2005) Metal binding characteristics and role of iron oxidation in the ferric uptake regulator from *Escherichia coli*. *Biochemistry* **44**: 13553-13559.

Morris, J.G., Jr., Takeda, T., Tall, B.D., Losonsky, G.A., Bhattacharya, S.K., Forrest, B.D., Kay, B.A., and Nishibuchi, M. (1990) Experimental non-O group 1 *Vibrio cholerae* gastroenteritis in humans. *J Clin Invest* **85**: 697-705.

Morris, K.M., Colten, H.R., and Bing, D.H. (1978) The first component of complement. A quantitative comparison of its biosynthesis in culture by human epithelial and mesenchymal cells. *J Exp Med* **148**: 1007-1019.

Murphy, E., Huwyler, L., and de Freire Bastos Mdo, C. (1985) Transposon Tn554: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants. *Embo J* **4**: 3357-3365.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255.

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

O'Shea, Y.A., Finnan, S., Reen, F.J., Morrissey, J.P., O'Gara, F., and Boyd, E.F. (2004a) The Vibrio seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in *V. vulnificus*. *Microbiology* **150**: 4053-4063.

O'Shea, Y.A., Reen, F.J., Quirke, A.M., and Boyd, E.F. (2004b) Evolutionary genetic analysis of the emergence of epidemic *Vibrio cholerae* isolates on the basis of comparative nucleotide sequence analysis and multilocus virulence gene profiles. *J Clin Microbiol* **42**: 4657-4671.

Ochsner, U.A., Vasil, A.I., and Vasil, M.L. (1995) Role of the ferric uptake regulator of *Pseudomonas aeruginosa* in the regulation of siderophores and exotoxin A expression: purification and activity on iron-regulated promoters. *J Bacteriol* **177**: 7194-7201.

Ochsner, U.A., Johnson, Z., Lamont, I.L., Cunliffe, H.E., and Vasil, M.L. (1996) Exotoxin A production in *Pseudomonas aeruginosa* requires the iron-regulated pvdS gene encoding an alternative sigma factor. *Mol Microbiol* **21**: 1019-1028.

Ochsner, U.A., and Vasil, M.L. (1996) Gene repression by the ferric uptake regulator in *Pseudomonas aeruginosa*: cycle selection of iron-regulated genes. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 4409-4414.

Ogden, C.A., deCathelineau, A., Hoffmann, P.R., Bratton, D., Ghebrehiwet, B., Fadok, V.A., and Henson, P.M. (2001) C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells. *J Exp Med* **194**: 781-795.

Ogierman, M.A., Zabihi, S., Mourtzios, L., and Manning, P.A. (1993) Genetic organization and sequence of the promoter-distal region of the tcp gene cluster of *Vibrio cholerae*. *Gene* **126**: 51-60.

Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Meth. Enzymol.* **276**: 307-326.

Pearson, G.D., Woods, A., Chiang, S.L., and Mekalanos, J.J. (1993) CTX genetic element encodes a site-specific recombination system and an intestinal colonization factor. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 3750-3754.

Philippe, N., Alcaraz, J.P., Coursange, E., Geiselmann, J., and Schneider, D. (2004) Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria. *Plasmid* **51**: 246-255.

Pohl, E., Holmes, R.K., and Hol, W.G. (1999) Crystal structure of a cobalt-activated diphtheria toxin repressor-DNA complex reveals a metal-binding SH3-like domain. *J Mol Biol* **292**: 653-667.

Pohl, E., Haller, J.C., Mijovilovich, A., Meyer-Klaucke, W., Garman, E., and Vasil, M.L. (2003) Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator. *Mol Microbiol* **47**: 903-915.

Pollitzer, R. (1954) Cholera studies. 1. History of the disease. *Bull World Health Organ* **10**: 421-461.

Pollitzer, R. (1959) Cholera. With a chapter on World incidence. *WHO, Geneva*. *Monograph series no. 43*.

Prince, R.W., Cox, C.D., and Vasil, M.L. (1993) Coordinate regulation of siderophore and exotoxin A production: molecular cloning and sequencing of the *Pseudomonas aeruginosa fur* gene. *J Bacteriol* **175**: 2589-2598.

Qadri, F., Bhuiyan, T.R., Dutta, K.K., Raqib, R., Alam, M.S., Alam, N.H., Svennerholm, A.M., and Mathan, M.M. (2004) 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.

Ramamurthy, T., Garg, S., Sharma, R., Bhattacharya, S.K., Nair, G.B., Shimada, T., Takeda, T., Karasawa, T., Kurazano, H., Pal, A., and *et al*. (1993) Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. *Lancet* **341**: 703-704.

Rapoport, G., and Dedonder, R. (1966) [Specificity of the levansucrase activity of *Bacillus subtilis*: clarification]. *Bull Soc Chim Biol (Paris)* **48**: 1311-1322.

Reidl, J., and Klose, K.E. (2002) *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiol Rev* **26**: 125-139.

Ribeiro, G., Viveiros, M., David, H.L., and Costa, J.V. (1997) Mycobacteriophage D29 contains an integration system similar to that of the temperate mycobacteriophage L5. *Microbiology* **143 (Pt 8)**: 2701-2708.

Richter, G., Fischer, M., Krieger, C., Eberhardt, S., Luttgen, H., Gerstenschlager, I., and Bacher, A. (1997) Biosynthesis of riboflavin: characterization of the bifunctional deaminase-reductase of *Escherichia coli* and *Bacillus subtilis*. *J Bacteriol* **179**: 2022-2028.

Roberts, R.J. (2004) Identifying protein function--a call for community action. *PLoS Biol* **2**: E42.

Sack, D.A., Lyke, C., McLaughlin, C., and Suwanvanichkij, V. (2001) Antimicrobial resistance in shigellosis, cholera and campylobacteriosis. In: WHO report (*Background document for the WHO Strategy for Containment of Antimicrobial Resistance*). Geneva: WHO: 1-51.

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

Safrin, S., Morris, J.G., Jr., Adams, M., Pons, V., Jacobs, R., and Conte, J.E., Jr. (1988) Non-O1 *Vibrio cholerae* bacteremia: case report and review. *Rev Infect Dis* **10**: 1012-1017.

Saito, T., Wormald, M.R., and Williams, R.J. (1991) Some structural features of the iron-uptake regulation protein. *Eur J Biochem* **197**: 29-38.

Salvador-Morales, C., Flahaut, E., Sim, E., Sloan, J., Green, M.L., and Sim, R.B. (2006) Complement activation and protein adsorption by carbon nanotubes. *Mol Immunol* **43**: 193-201.

Samadi, A.R., Huq, M.I., Shahid, N., Khan, M.U., Eusof, A., Rahman, A.S., Yunus, M., and Faruque, A.S. (1983) Classical *Vibrio cholerae* biotype displaces EL tor in Bangladesh. *Lancet* **1**: 805-807.

Satchell, K.J. (2003) Activation and suppression of the proinflammatory immune response by *Vibrio cholerae* toxins. *Microbes Infect* **5**: 1241-1247.

Schiemann, O., and Prisner, T.F. (2007) Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Q Rev Biophys* **40**: 1-53.

Schoolnik, G.K., and Yildiz, F.H. (2000) The complete genome sequence of *Vibrio cholerae*: a tale of two chromosomes and of two lifestyles. *Genome Biol* **1(3)**: 1016.1-1016.3.

Service, R.F. (2008) Structural biology. Protein structure initiative: phase 3 or phase out. *Science* **319**: 1610-1613.

Silva, T.M., Schleupner, M.A., Tacket, C.O., Steiner, T.S., Kaper, J.B., Edelman, R., and Guerrant, R. (1996) New evidence for an inflammatory component in diarrhea caused by selected new, live attenuated cholera vaccines and by El Tor and Q139 *Vibrio cholerae*. *Infect Immun* **64**: 2362-2364.

Sim, R.B., Moestrup, S.K., Stuart, G.R., Lynch, N.J., Lu, J., Schwaeble, W.J., and Malhotra, R. (1998) Interaction of C1q and the collectins with the potential receptors calreticulin (cC1qR/collectin receptor) and megalin. *Immunobiology* **199**: 208-224.

Simos, G., and Georgatos, S.D. (1994) The lamin B receptor-associated protein p34 shares sequence homology and antigenic determinants with the splicing factor 2-associated protein p32. *FEBS Lett* **346**: 225-228.

Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S., and *et al*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature* **406**: 151-159.

Sixma, T.K., Pronk, S.E., Kalk, K.H., Wartna, E.S., van Zanten, B.A., Witholt, B., and Hol, W.G. (1991) Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli*. *Nature* **351**: 371-377.

Snow, J., Frost, W.H., and Richardson, B.W. (1936) Snow on cholera, New York: Commonwealth Fund. *Oxford University Press*.

Sritharan, M. (2006) Iron and bacterial virulence. *Indian Journal of Medical Microbiology* **24**: 163-164.

Staggs, T.M., and Perry, R.D. (1991) Identification and cloning of a fur regulatory gene in *Yersinia pestis*. *J Bacteriol* **173**: 417-425.

Stojiljkovic, I., and Hantke, K. (1995) Functional domains of the *Escherichia coli* ferric uptake regulator protein (Fur). *Mol Gen Genet* **247**: 199-205.

Swerdlow, D.L., and Ries, A.A. (1993) *Vibrio cholerae* non-O1--the eighth pandemic? *Lancet* **342**: 382-383.

Tanamal, S. (1959) Notes on paracholera in Sulawesi (Celebes). *Am J Trop Med Hyg* **8**: 72-78.

Tange, T.O., Jensen, T.H., and Kjems, J. (1996) In vitro interaction between human immunodeficiency virus type 1 Rev protein and splicing factor ASF/SF2-associated protein, p32. *J Biol Chem* **271**: 10066-10072.

Terwilliger, T.C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55**: 849-861.

Terwilliger, T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* **56**: 965-972.

Terwilliger, T.C. (2003a) Automated side-chain model building and sequence assignment by template matching. *Acta Crystallogr D Biol Crystallogr* **59**: 45-49.

Terwilliger, T.C. (2003b) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* **59**: 38-44.

Tiss, A., Barre, O., Michaud-Soret, I., and Forest, E. (2005) Characterization of the DNA-binding site in the ferric uptake regulator protein from *Escherichia coli* by UV crosslinking and mass spectrometry. *FEBS Lett* **579**: 5454-5460.

Trucksis, M., Michalski, J., Deng, Y.K., and Kaper, J.B. (1998) The *Vibrio cholerae* genome contains two unique circular chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 14464-14469.

Vasil, M.L., and Ochsner, U.A. (1999) The response of *Pseudomonas aeruginosa* to iron: genetics, biochemistry and virulence. *Mol Microbiol* **34**: 399-413.

Vimr, E.R., Kalivoda, K.A., Deszo, E.L., and Steenbergen, S.M. (2004) Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* **68**: 132-153.

Waldor, M.K., Colwell, R., and Mekalanos, J.J. (1994) The *Vibrio cholerae* O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide virulence determinants. *Proc. Natl. Acad. Sci. U. S. A.* **91**: 11388-11392.

Waldor, M.K., and Mekalanos, J.J. (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**: 1910-1914.

Waldor, M.K., Rubin, E.J., Pearson, G.D., Kimsey, H., and Mekalanos, J.J. (1997) Regulation, replication, and integration functions of the *Vibrio cholerae* CTXphi are encoded by region RS2. *Mol Microbiol* **24**: 917-926.

Waldor, M.K., and RayChaudhuri, D. (2000) Treasure trove for cholera research. *Nature* **406**: 469-470.

Wandersman, C., and Delepelaire, P. (2004) Bacterial iron sources: from siderophores to hemophores. *Annu Rev Microbiol* **58**: 611-647.

Wang, Y., Finan, J.E., Middeldorp, J.M., and Hayward, S.D. (1997) P32/TAP, a cellular protein that interacts with EBNA-1 of Epstein-Barr virus. *Virology* **236**: 18-29.

Weir, E., and Haider, S. (2004) Cholera outbreaks continue. *CMAJ* **170**: 1092-1093.

White, A., Ding, X., vanderSpek, J.C., Murphy, J.R., and Ringe, D. (1998) Structure of the metal-ion-activated diphtheria toxin repressor/tox operator complex. *Nature* **394**: 502-506.

WHO (1995) Meeting on The Potential Role of New Cholera Vaccines in the Prevention and Control of Cholera Outbreaks during Acute Emergencies. Document CDR/GPV/95.1. World Health Organization, Geneva, Switzerland.

Wilderman, P.J., Vasil, A.I., Johnson, Z., Wilson, M.J., Cunliffe, H.E., Lamont, I.L., and Vasil, M.L. (2001) Characterization of an endoprotease (PrpL) encoded by a PvdS-regulated gene in *Pseudomonas aeruginosa*. *Infect Immun* **69**: 5385-5394.

Wyckoff, E.E., Mey, A.R., and Payne, S.M. (2007) Iron acquisition in *Vibrio cholerae*. *Biometals* **20**: 405-416.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**: 3369-3376.

Yu, L., Loewenstein, P.M., Zhang, Z., and Green, M. (1995) In vitro interaction of the human immunodeficiency virus type 1 Tat transactivator and the general transcription factor TFIIB with the cellular protein TAP. *J Virol* **69**: 3017-3023.

Zarembinski, T.I., Hung, L.W., Mueller-Dieckmann, H.J., Kim, K.K., Yokota, H., Kim, R., and Kim, S.H. (1998) Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 15189-15193.

Zhang, R.G., Scott, D.L., Westbrook, M.L., Nance, S., Spangler, B.D., Shipley, G.G., and Westbrook, E.M. (1995) The three-dimensional crystal structure of cholera toxin. *J Mol Biol* **251**: 563-573.

List of publications

1. **Sheikh MA**, Potter JA, Johnson KA, Sim RB, Boyd EF, Taylor GL. Crystal structure of VC1805, a conserved hypothetical protein from a *Vibrio cholerae* pathogenicity island, reveals homology to human p32. Proteins 2008; **71(3)**: 1563-1571.

2. **Sheikh MA**, Taylor GL. Crystal structure of the *Vibrio cholerae* ferric uptake regulator (Fur) reveals insights into metal coordination and flexibility of the DNA-binding domains. Molecular Microbiology doi:10.1111/j.1365-2958.2009.06718.x

3. **Sheikh MA**, Taylor GL. Crystal structures of VC0508 and VC0509 from a *Vibrio cholerae* seventh pandemic island II; towards understanding the hypothetical clusters in pathogenicity islands. Manuscript under preparation.

4. Jahan N, Potter JA, **Sheikh MA,** Botting, CH, Shirran SL, Westwood NJ, Taylor GL. Insights into the biosynthesis of the *Vibrio cholerae* major autoinducer CAI-1 from the crystal structure of the PLP-dependent enzyme CqsA. Manuscript under preparation.