Molecular Mechanism of CTXΦ phage uptake into *Vibrio cholerae*

by

Christopher G. Ford

B.Sc., Simon Fraser University, 2008

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Spring 2012

Approval

Name: **Christopher G. Ford**

Degree: **Master of Science**

Title of Thesis: ***Molecular Mechanism of CTXΦ phage uptake into Vibrio cholerae***

Examining Committee:

Chair: **Dr. Jonathan Choy**
Assistant Professor

Dr. Lisa Craig
Senior Supervisor
Associate Professor

Dr. Jamie K. Scott
Supervisor
Professor

Dr. Frederic F. Pio
Supervisor
Associate Professor

Dr. Ralph Pantophlet
Internal Examiner
Assistant Professor
Faculty of Health Sciences

Date Defended/Approved: April 11, 2012

Partial Copyright Licence

Abstract

Vibrio cholerae colonize the intestinal lumen and secrete cholera toxin, inducing a massive efflux of water and electrolytes, causing the diarrheal disease cholera. Non-pathogenic serogroups of *V. cholerae* are made pathogenic by infection with the filamentous cholera toxin bacteriophage Φ (CTXΦ). CTXΦ binding and uptake may occur via a similar mechanism to M13 phage infection of *E. coli*. By this model the minor coat protein of CTX, pIII, first binds to the toxin coregulated pilus (TCP) on the surface of *V. cholerae* followed by TolA in the periplasm of *V. cholerae*. To understand CTXΦ uptake we expressed and purified the N-terminal domain of CTXΦ pIII (N-pIII) and the C-terminal domain of TolA (TolA-C) and demonstrated an interaction between these protein domains using pull down assays. We solved the de novo crystal structure of N-pIII to 2.9 Å resolution and the structure of N-pIII in complex with TolA-C to 1.44 Å. CTXΦ N-pIII has a structure similar to its corresponding domain in M13 pIII proteins despite only 15% sequence identity, but surprisingly binds to TolA at a site distinct from that of M13 N-pIII. These crystal structures provide valuable insights for understanding the mechanism by which CTXΦ infects *V. cholerae*.

Keywords: pIII; TolA; filamentous phage; CTXΦ; *Vibrio cholerae*; toxin coregulated pili

Dedication

Dedicated to my daughter Lyvia Dawn Ford

Acknowledgements

Thank you to my supervisor, Dr. Lisa Craig, for her support and for the opportunities she has given me. Thank you to my supervisory committee, Dr. Jamie Scott and Dr. Frederic Pio for their time and guidance throughout my graduate career. Thank you to Kolappan Subramaniapillai for his guidance and time. Thank you to all the members, past and present, of the Craig Laboratory.

Table of Contents

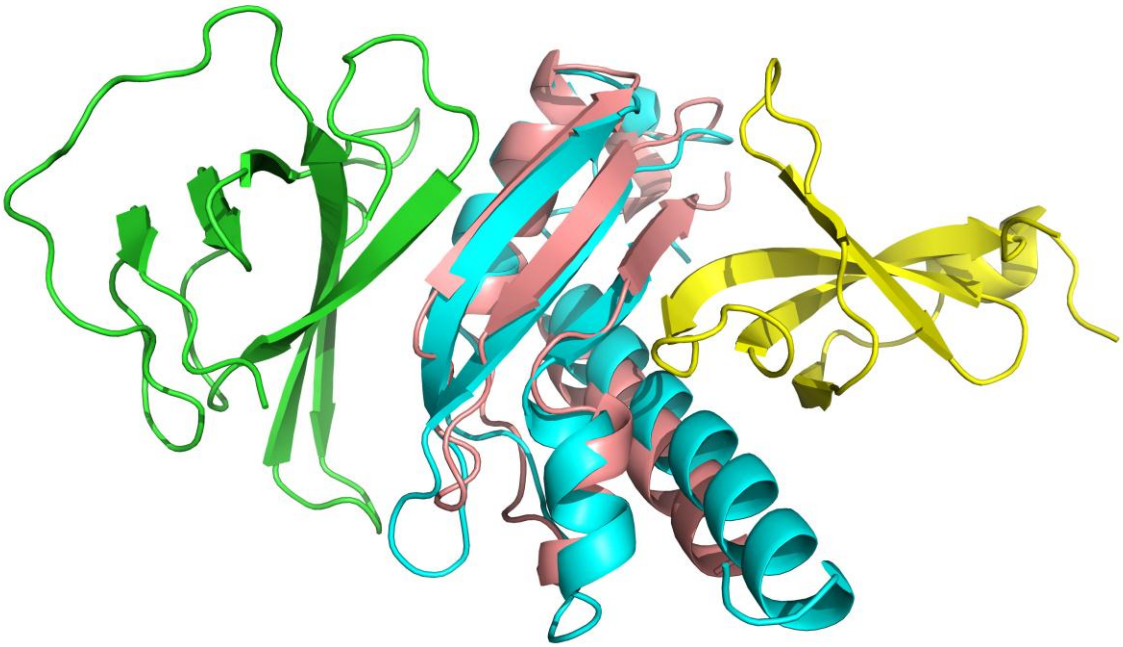
List of Tables

List of Figures

List of Abbreviation

Amp	Ampicillin
BMCD	Biological Macromolecule Crystallization Database
cAMP	Cyclic AMP
CT	Cholera toxin
DMSO	Dimethyl sulfoxide
GbpA	GlcNAc-binding protein A
IPTG	Isopropyl-beta-D-thiogalactopyranoside
LB	Luria-Bertani
MAD	Multiple wavelength anomalous diffraction
MPD	Methyl-2,4-pentanediol
MMCO	Molecular Mass cut off
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
RMSD	Root mean square deviation
RPM	Revolutions per minute
SAD	Single wavelength anomalous diffraction
ssDNA	Single stranded DNA
SSRL	Stanford Synchrotron Radiation Lightsource
TCP	Toxin coregulated pilus
TIISS	Type II secretion system

CTXΦ phage pIII binds to a novel site on TolA



1. Introduction

1.1. *Vibrio cholerae are agents of cholera disease*

Vibrio cholerae are gram negative, rod shaped, non-invasive bacteria that cause the often fatal diarrheal disease cholera (Figure 1.1). Hundreds of thousands of people are affected and thousands of deaths are reported to the World Health Organization (WHO) every year due to this illness. These numbers are believed to be highly under-reported for reasons ranging from countries only reporting laboratory confirmed cases to fears of travel and trade sanctions (WHO, 2006). Although cholera has nearly been eradicated in developed countries, developing countries fight this disease on an ongoing basis. Outbreaks in areas as the Ganges delta region of Bangladesh and India occur on a semi-annual basis. Epidemics can occur anytime but are especially devastating during desperate times such as natural disaster or war (Chaignat et al., 2008; Ganin, 2009, Nelson *et al.*, 2008).

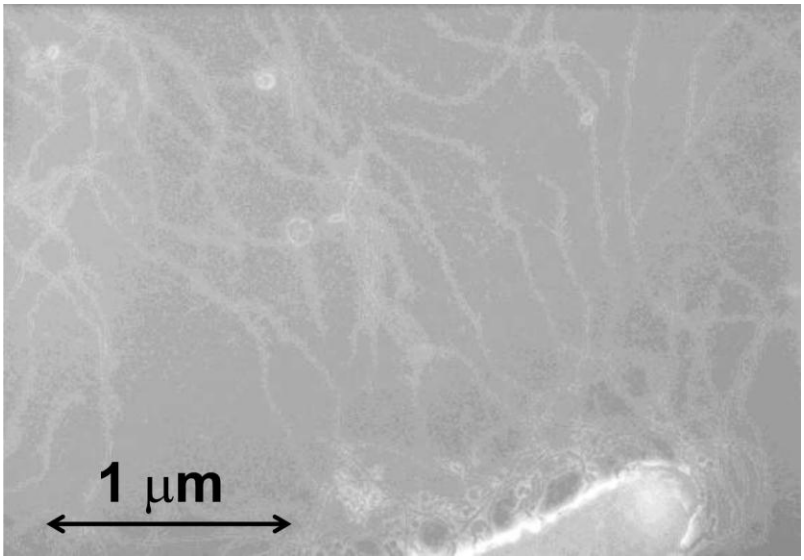


Figure 1.1. Vibrio cholerae with surface-displaye TCP

Gram negative rod shaped *Vibrio cholerae* bacteria (bottom) producing toxin corregulated pili (white strands). Photo courtesy of Dr. Lisa Craig, Simon Fraser University.

V. cholerae reside in brackish, riverine, and estuarine waters and can often be found associated with plankton (Colwell et al., 1977, 1994). Pathogenic bacteria are consumed in contaminated water and, to a lesser extent, food. The concentration of *V. cholerae* in consumed water and food sources seems to be the determining factor between asymptomatic infection and severe diarrheal disease (Righetto et al., 2012).

Although there are over 200 known *V. cholerae* serogroups based on Lipopolysaccharides O antigen, only two are pathogenic. Until a recent outbreak in 1993, all cholera disease were associated with serogroup O1 which can be further divided into two biotypes, classical and El Tor. The classical biotype was responsible for the first six global pandemics of cholera whereas the El Tor biotype is responsible for the current, seventh pandemic. The El Tor biotype was

originally distinguished from the classical biotype based on the production of soluble hemolysin. It has since been determined that the classical and El Tor biotypes are differentiable by a number of phenotypic tests. For example Classical strains are sensitive to the antibiotic polymyxin B, produce a negative reaction in a Voges-Proskauer test and fail to agglutinate chicken erythrocytes whereas El Tor strains are resistant to polymyxin B, produce a positive reaction in a Voges-Proskauer test and agglutinate chicken erythrocytes (Raychoudhuri et al., 2008). Genetically these biotypes have several notable differences such as sequence differences in the toxin co-regulated pilus (TCP) subunit gene *tcpA,* and in the B subunit gene of cholera toxin (CT). The classical biotype genome contains a repetitive sequence (RS) group called RS2 which codes for proteins responsible for phage integration and expression whereas the *El Tor* biotype is most often associated with the RS1 group which contains the full RS2 group plus *RstC*, a gene of unknown function (Keasler & Hall, 1993; Waldor, *et al.*, 1997). Serogroup O139 emerged in India in 1992 and is believed to have evolved by uptake of exogenous DNA by an O1 El Tor strain (Mooi & Bik, 1997).

Levels of *V. cholerae* in the environment appear to be self-regulated by serogroup specific lytic phage. As levels of serogroups O1 and O139 in the environment begin to increase rapidly during an outbreak, levels of their respective lytic vibriophages rise in response. Once lytic vibriophage concentrations are high enough, bacterial levels begin to drop rapidly and levels of lytic vibriophage eventually follow (Nelson et al., 2008, Faruque *et al.*, 2004).

After ingestion, pathogenic *V. cholerae* colonize the small intestine employing a combination of virulence factors. GlcNAc-binding protein A (GbpA) is a protein associated with the surface of *V. cholerae* cells. During infection, *V. cholerae* use GbpA to bind to N-acetylglucosamine on epithelial cell's (Kirn *et al.,* 2005). Bound to epithelial cells, bacteria can form many microscopic colonies of cells (microcolonies) mediated by TCP. TCP is a type IV pilus composed of thousands of copies of the pilin subunit TcpA organized in a helical arrangement and held together by extensive hydrophobic interactions between the N-terminal α-helices (Craig *et al*., 2003; Megli *et al.*, 2011). Type IV pili are found on many Gram negative bacteria such as enteropathogenic *E. coli*, enterotoxigenic *E. coli*, *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa* and at least one Gram positive bacterium, *Clostridium perfringens*. Different pili mediate virulence through a combination of many different functions such as adherence to host cells, colonization thru pilus-pilus interactions, secretion of colonization factors, antigenic variation by uptake of exogenous DNA and twitching motility from elongation and retraction of pili (Craig & Li. 2008). *V. cholerae* bacteria within microcolonies and bound to epithelial cells secrete cholera toxin and cause the severe diarrhea associated with cholera (Figure 1.2).

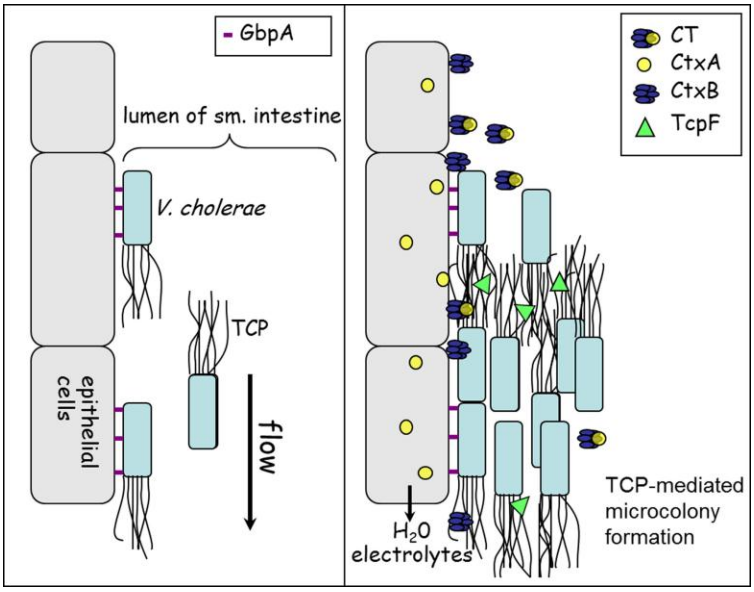


Figure 1.2. V. cholerae colonization of the small intestine

Left panel. After ingestion, V. cholerae use GbpA to adhere to N-acetylglucosamine on epithelial cells in the small intestine. Right panel. Bound to epithelial cells TCP mediates microcolony formation and *V. cholerae* secretes CT and TcpF. Image courtesy of Dr. Lisa Craig, Simon Fraser University.

1.2. Cholera toxin

Cholera toxin is an AB_5 toxin composed of a pentamer of B subunits (CtxB) bound to the catalytic A subunit (CtxA). Individual subunits are secreted across the bacterial inner membrane via the sec pathway. Once in the periplasm subunits are assembled into the AB_5 toxin, which are then exported across the outer membrane by the type II secretion system (TIISS). The B subunits interact with GM1 gangliosides on epithelial cells and the AB_5 toxins are endocytosed. Having previously undergone proteolysis at the bacterial surface CtxA now gets reduced, generating the ADP ribosylating enzyme A1, which can now gain access to the epithelial cell cytoplasm. A1 transfers an ADP-ribose to the α-

subunit of the heterotrimeric G-stimulatory (Gs) protein, preventing GTP hydrolysis and causing constitutive activation of adenylate cyclase. Elevated levels of cAMP act on cAMP dependent protein kinase PKA to cause the unregulated release of water and NaCl through the cystic fibrosis transmembrane conductance regulator (CFTR) and other membrane channels, resulting in the severe diarrhea and dehydration associated with cholera (Figure 1.3) (Mekalanos, 1983; Sanchez & Holmgren, 2008; Waldor & Mekalanos, 1996).

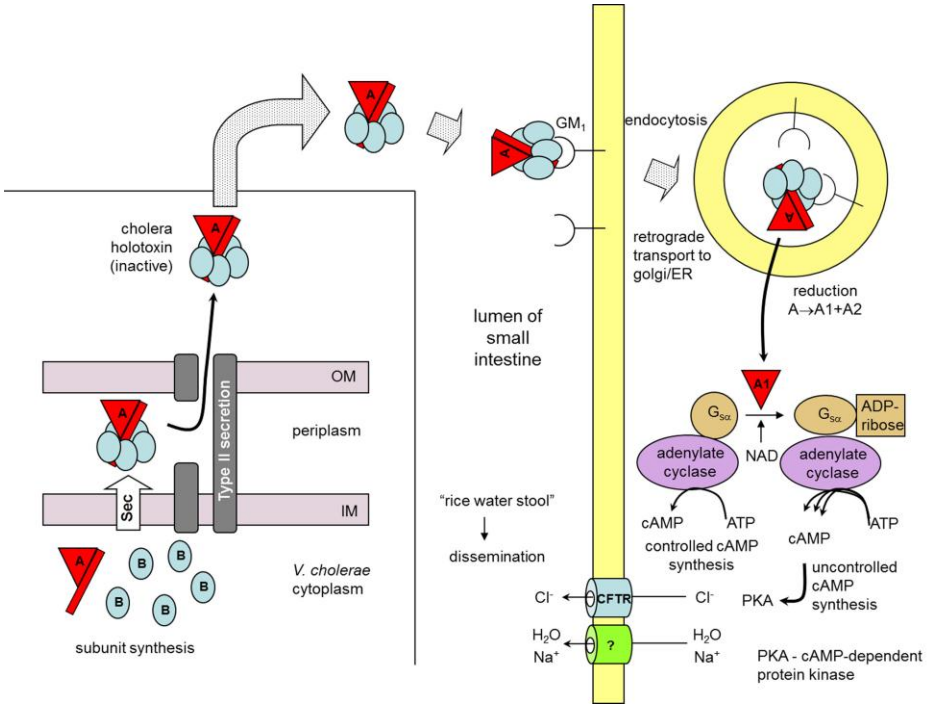


Figure 1.3. Cholera toxin effects

Cholera toxin A and B subunits are transported into the periplasm through the Sec pathway. The AB_5 complex is formed in the periplasm and secreted by type II secretion system. Extracellular CT binds to GM_1 on epithelial cells and is endocytosed. Proteolic cleavage of the A subunit releases the catalytic A_1 domain. A_1 ADP-ribosylates G proteins, constitutively activating adenylate cyclase. Over production of cAMP causes overactivation of PKA causing the massive efflux of water and electrolytes associated with cholera. Image courtesy of Dr. Lisa Craig, Simon Fraser University.

1.3. Cholera toxin phage infection

The genes encoding CT are present on the filamentous cholera toxin phage (CTXΦ) genome, which is integrated as a prophage into the *V. cholerae* genomes of O1 and O139 serogroups. Non-pathogenic serogroups of *V. cholerae* can be made toxigenic by infection with CTXΦ. The CTXΦ genome is very similar in structure and gene organization to the *E. coli* fd filamentous phage (Waldor & Mekalanos, 1996). CTXΦ has a single stranded, circular DNA 7 kbp long with a core region containing genes *cep, orfU, ace, zot, ctxA and* ctxB, which encode proteins pVIII, pIII, pVI, pI, CtxA and CtxB, respectively. Flanking the core sequence are one or more RS groups, which code for proteins responsible for expression of the phage genome and its integration into the *V. cholerae* chromosome. Genes *rstR, rstA and* rstB are found in both RS1 and RS2 groups and are responsible for phage repression, independent replication and phage DNA integration, respectively. Alternatively *rstC* is found only in RS1 groups and its function is currently unknown (Waldor & Mekalanos, 1996). The genome of classical *V. cholerae* contains a single copy of the full CTXΦ prophage genome and the RS2 group and the bacterium are unable to secrete CTXΦ phage. The presence of multiple tandem repeats of the CTXΦ prophage and / or the presence of the RS1 group in the *V. cholerae* genome, as is the case for the El tor biotype, confers upon *V. cholerae* the ability to secrete high titers of CTXΦ phage (Mekalanos *et al*. 1983, Kimsey *et al*.,1998, Davis & Waldor, 2000). Finally CTXΦ contains an 18 bp attRS1 sequence responsible for site specific

recombination through an identical 18 bp or a similar 17 bp site in the El tor genome (Figure 1.4) (Pearson *et al.*, 1993). The current model of CTXΦ phage infection of *V. cholerae* begins with CTXΦ phage binding to the high-affinity receptor, TCP *via* the minor coat protein, pIII. Phage are transported across the outer membrane *via* an unknown mechanism, where pIII now contacts TolA, a periplasmic *V. cholerae* protein with an inner membrane-spanning segment. The CTXΦ phage shed its protein coat in the inner membrane and releases the circular plus sense DNA strand into the cytoplasm. DNA replication using *V. cholerae* DNA polymerase produces a double-stranded plasmid or replicative fragment of the CTXΦ (pCTX) genome. pCTX integrates site-specifically into *V. cholerae* at the homologous region attRS1 often integrating as tandem prophage. CTX prophage and CTX plasmid DNA can both serve as templates for synthesis of genomic CTXΦ ssDNA. Phage proteins are translated from mRNA transcribed from the replicative fragment. Plus-sense single stranded circular phage DNA is packaged into CTXΦ phage particles at the inner membrane and secreted from the cell *via* the T2SS (Figure 1.5) (Davis *et al.*, 2000; Davis & Waldor, 2003).

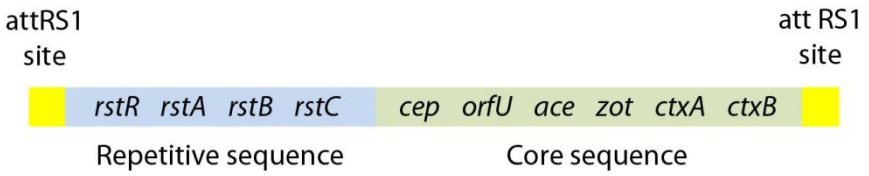


Figure 1.4. CTXΦ phage genome

Repetitive sequence (RS1) highlighted in blue, core sequence highlighted in green and attRS1 sites highlighted in yellow.

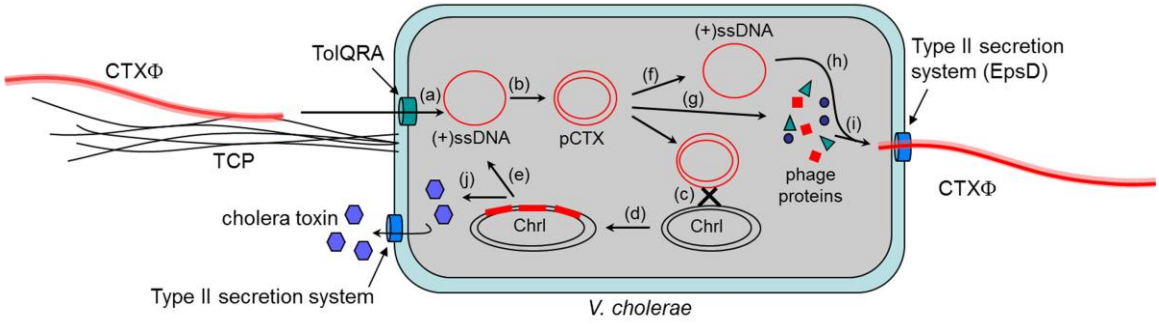


Figure 1.5. CTXΦ phage infection

(a) CTXΦ infects V. cholerae by binding to TCP then to the inner membrane/periplasmic TolQRA complex. The phage ssDNA loses its protein coat and is transported into the cytoplasm. (b) The replicative or plasmid form of CTXΦ (pCTX) DNA is generated by replication of the ssDNA. (c) pCTX integrates site-specifically into V. cholerae Chromosome I at the homologous region attRS1. (d) pCTX usually integrates as tandem prophage. (e) CTX prophage serve as templates to generate extra-chromosomal +ssDNA. (f) +ssDNA is also generated off pCTX for packaging into new phage. (g) The phage coat and secretion proteins are expressed from pCTX. (h,i) The +ssDNA is packaged into phage particles at the inner membrane and secreted though the T2SS. (j) Cholera toxin subunits CtxA and CtxB are synthesized from the ctxA-ctxB genes in the CTX prophage, assembled into holotoxin in the periplasm and secreted via the Type II secretion system.

Although little is known about the CTXΦ phage structure and infection mechanism, the fd phage is fairly well characterized with respect to its structure, genetics, and mechanism of infection of *E. coli*. Fd phage are composed of ~ 2700 copies of a small major coat protein (pVIII) composed of a transmembrane α-helix and a short amphipathic α-helix on the outside of phage (Im & Brooks III, 2004). pVIII forms the long cylindrical phage coat which packages the single stranded circular DNA genome with 3 – 5 copies each of the minor coat proteins pIII and pVI localized to one end and pVII and pIX localized to the other (Figure 1.6) (Wen *et al.*, 1997, Grant *et al.*, 1981, Lopez & Webster, 1982). Based on its amino acid sequence the pIII protein is divided into 3 domains, an N-terminal domain (D1), a central domain (D2) and a C-terminal domain (D3). These

domains are separated by glycine-rich flexible linkers (Figure 1.7) (Heilpern & Waldor, 2003). The C-terminal domain of the pIII protein is important for phage assembly and release from the bacterial outer membrane; the two N-terminal domains interact with the primary receptor of *E. coli*, the tip of the F pilus, and the secondary receptor, the periplasmic TolA protein (Bennett et al., 2011; Bennett & Rakonjac, 2006; Rakonjac *et al.*, 1999, Lorenz & Schmid, 2011, Lubkowski *et al.*, 1999).

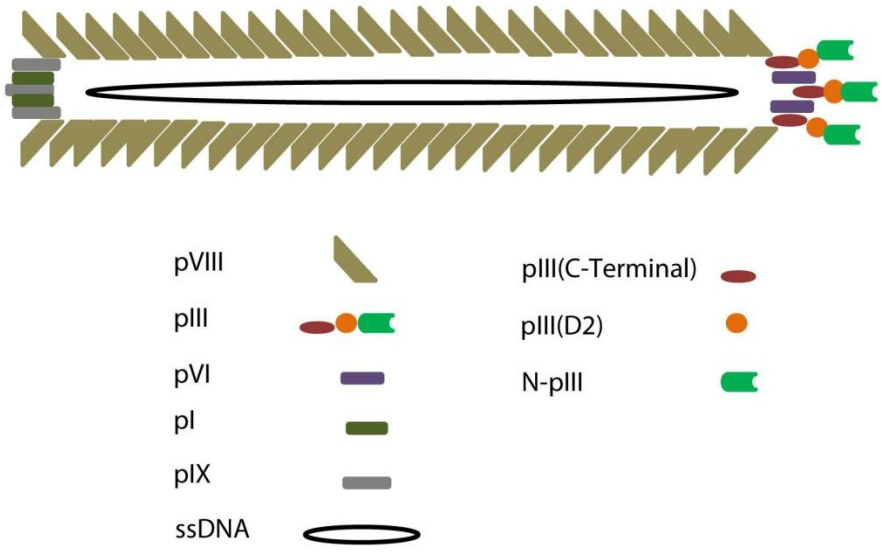


Figure 1.6. Filamentous phage composition

The major protein pVIII comprises the phage coat that encompasses the ssDNA. Minor coat proteins pIII and pVI localize to one end whereas pI and pXI localize to the other end.

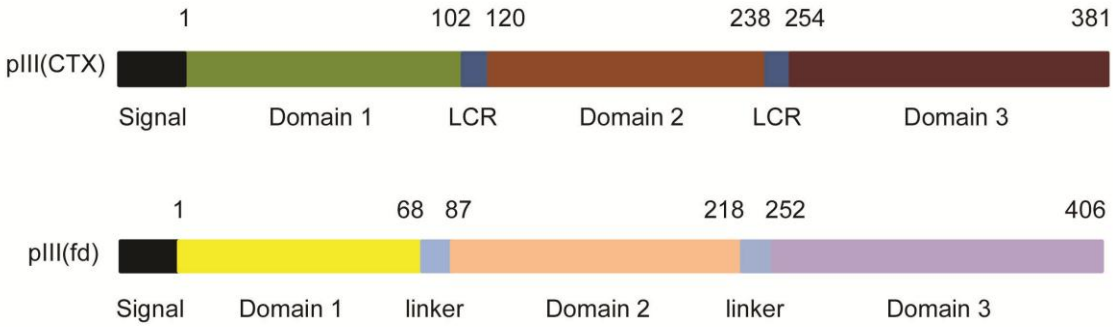


Figure 1.7. Sequence schematic of $pIII_{CTX}$ and $pIII_{fd}$

Colour coded sequence schematic of pIII from CTXΦ phage (top) and pIII from fd phage (bottom). Labelled above with the residue numbers separating the domains and below with the signal sequence, domain number and LCR/linker sequences.

The C-terminal end of D3 acts to anchor pIII in the pVIII phage coat, whereas all of D3 is required for release of newly assembled phage from the bacteria; D1 and D2 are required for infection (Davis *et al.*, 1985, Armstrong *et al.*, 1981, Rakonjac *et al.*, 1999, Bennett *et al.*, 2011). The structures of domains 1 and 2 of pIII from M13 and fd phages, and domain 1 of pIII from M13 and IF1 phages in complex with the C-terminal domain of TolA have been determined by X-ray crystallography (Lubkowski *et al.*, 1998, 1999, Holliger *et al.*, 1999, Lorenz *et al.*, 2011). D1 and D2 of $pIII_{fd}$ fold together and form extensive interactions, which are thought to be disrupted during the infection process. During infection $pIII_{fd}$ D2 binds to the tip of the F pilus of *E. coli* (Deng, Malik, & Perham, 1999; Deng & Perham, 2002 Armstrong *et al*., 1981), which induces a conformational change in $pIII_{fd}$ that exposes a previously buried binding domain in D1 (Lubkowski et al., 1998, 1999). This open, infectious conformation of $pIII_{fd}$ is thought to be maintained transiently due to *trans-to-cis* isomerization of proline 213 upon binding to the tip of the F pilus (Eckert & Schmid, 2007; Martin &

Schmid, 2003). F pili have been shown to extend and retract in R17 bacteriophage binding assays (Clarke *et al.*, 2008, Daehnel *et al.*, 2005). The current model suggests that upon fd binding, the pilus retracts into the cell, somehow transporting the phage across the outer membrane to present D1 to the inner membrane anchored, periplasm-spanning TolA protein (Bennett et al., 2011, Click & Webster, 1997, 1998). D1 binds to the C-terminal domain of TolA (TolA-C) and the TolQRA complex facilitates injection of the phage ssDNA into the bacterial cytoplasm (Figure 1.8) (Holliger *et al.*, 1999; Lubkowski *et al.*, 1999).

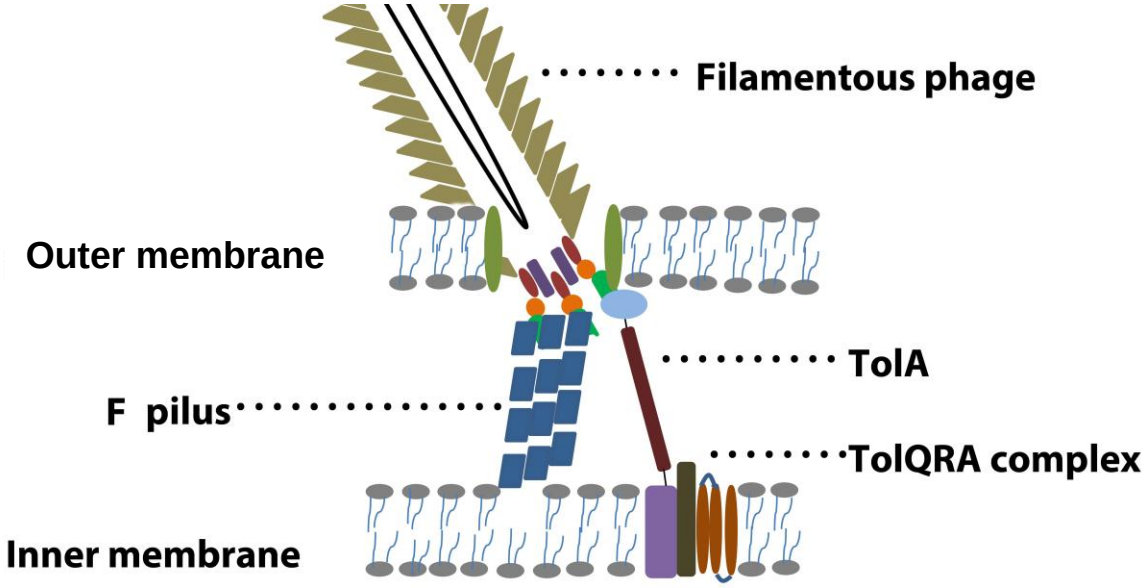


Figure 1.8. Filamentous phage infection

D2 (orange) of the minor coat protein pIII binds to F pilus, which brings pIII D1 (green) in close proximity to TolA. D1 of pIII bound to TolA facilitates the injection of phage genome into the cytoplasm assisted by the TolQRA complex.

The bacterial TolA is a cytoplasmic membrane protein known to give rigidity and structure to the outer membrane of Gram-negative bacteria (Karlsson *et al.,* 2006). The N-terminal end of TolA is anchored in the bacterial plasma

membrane whereas a central domain traverses the periplasm to a globular C-terminal domain that lies near the outer membrane (Figure 1.8). In conjunction with accessory membrane proteins TolR and TolQ, the TolA protein uses a proton motive force, which may mediate uptake of bacteriophage in a manner similar to TonB-facilitated uptake of siderophores and vitamin B12 into bacterial cells (Braun & Herrmann, 1993; Germon *et al.*, 2001; Witty *et al*., 2002).

Heilpern & Waldor (2003) demonstrated that fusing domain 1 of $pIII_{CTX}$ to pIII of fd phage confers *V. cholerae* infection to fd phage. In phage infection experiments they fused domain 1, 2 or 1 and 2 of $pIII_{CTX}$ to full length $pIII_{fd}$, or to domain 3 only, and measured the transduction frequencies of fd phage containing the different pIII fusions. They found that fd phage bearing any pIII-fusion, so long as that fusion contained domain 1 of $pIII_{CTX}$, infect *V. cholerae*. Phage that contained domain 3 of $pIII_{fd}$ fused to domains 1 and 2 of $pIII_{CTX}$ produced the highest transduction frequencies (Figure 1.9). Transduction frequencies of phage containing only domain 1 from $pIII_{CTX}$ in the fusion pIII were reduced by ~ 3 orders of magnitude whereas phage containing only domain 2 from $pIII_{CTX}$ in the fusion pIII did not infect *V. cholerae*. They also tested these pIII fusion phage on Tcp and TolA knockout strains of *V. cholerae* and found that transduction frequencies were reduced in Tcp knockout strains and completely abrogated in TolA knockout strains.

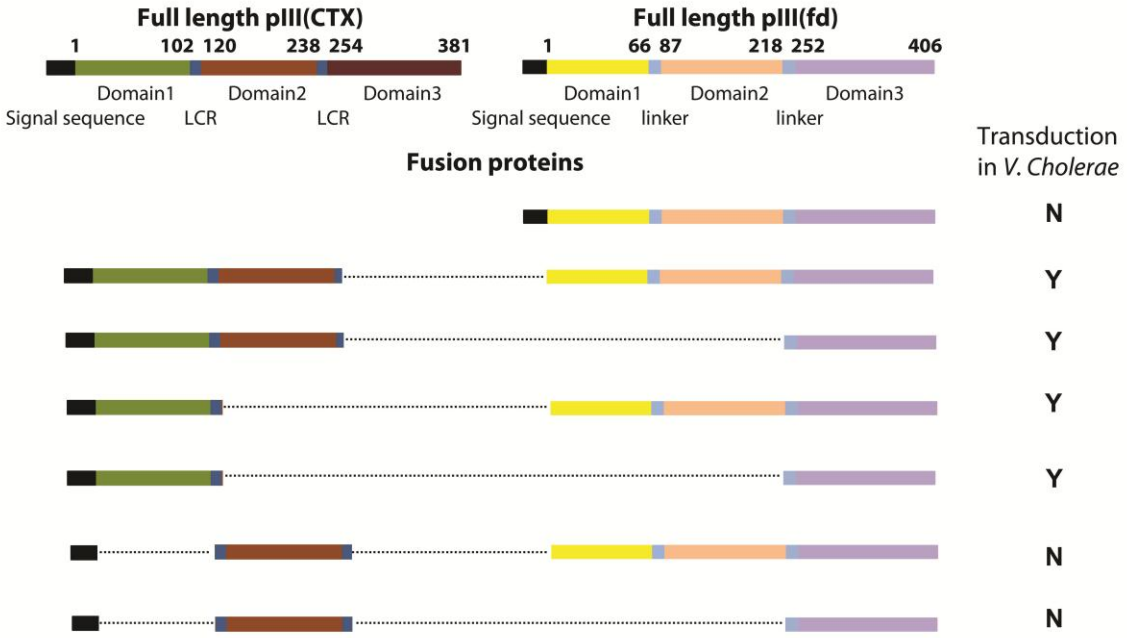


Figure 1.9. pIII fusion proteins cause fd phage infection in V. cholerae

Colour coded sequence schematic displays D1 (green), D2 (brown) and D3 (mauve) of pIII from CTXΦ phage and D1 (yellow), D2 (peach) and D3 (purple) of pIII from fd phage. A schematic of the fusion proteins is drawn below with the ability of a phage containing a specific fusion protein to infect *V. cholerae* listed as yes (Y) or no (N) to the right of the schematic.

1.4. pIII structures from M13 and IF1 phage

The first crystal structure of a pIII protein comprised D1 and D2 (residues 1 – 217) of the filamentous phage M13 and was solved to 1.46 Å resolution (Lubkowski, 1998) (PDB ID 1G3P). M13 pIII D1 and D2 together form a horseshoe-shaped protein, with the two domains linked by hydrophobic interactions among β-sheet loops (Figure 1.10). D1 (residues 1 – 68) is comprised of a 1.5-turn α helix (α1) followed by 4 β strands that form a tight beta–barrel-like motif. D1 is connected by a long, glycine-rich flexible linker to D2, which is composed of an irregular antiparallel β-sheet with long interstrand

loops that interacts with a short (2-turn) α-helix, α2. α2 is followed by an irregular loop leading to the C-terminal segment, which passes back along D1 to form the second strand of the 2-stranded β-sheet in this domain.

In 1999 the crystal structure of domains 1 and 2 (residues 2 – 217) of filamentous phage fd was solved using multiple isomorphous replacement techniques and refined to 1.9 Å resolution (Holliger et al., 1999) (PDB 2GP3). They report a structure very similar to $pIII_{M13}$ D1-D2, which is expected, as their sequence identity is over 90%. They further characterized the 2 beta stranded β-sheet, coloured green in Figure 1.10, as hinge region based on a structure that came out the same year of a fusion between $pIII_{M13}$ D1 and the C-terminal domain of *E. coli* TolA. The hinge regions, consisting of residues Lys88 to Pro123 and Tyr203 to Ala217, are believed to be responsible for a conformational change that releases the contact between D1 and D2 upon binding to Type IV pilus.

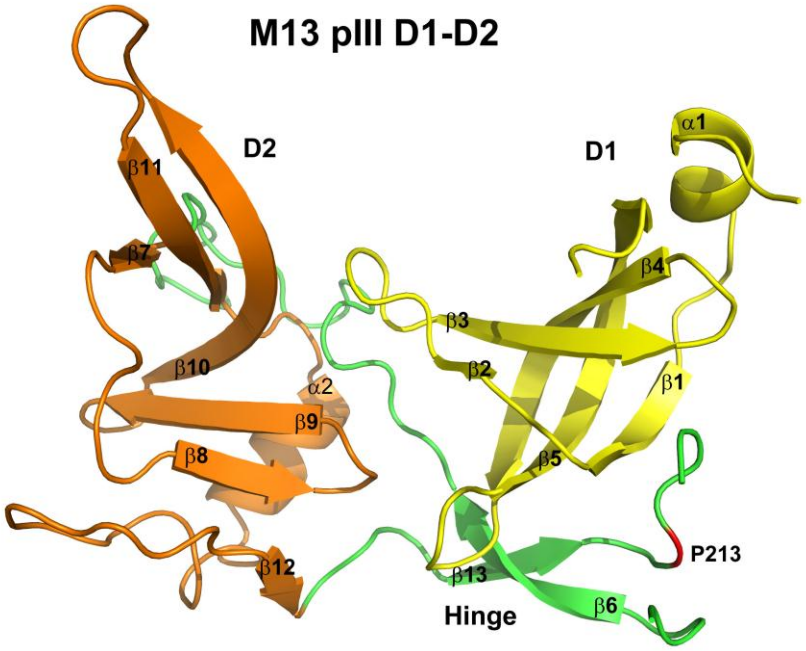


Figure 1.10. $pIII_{M13}$D1-D2

PDB # 1G3P; Crystal structure of M13 pIII D1-D2. D1 shown in yellow, D2 shown in orange, Hinge regions shown in green. α-helices and β-sheets are numbered N-terminally according to authors designations. Pro213 is coloured red.

The crystal structure of a fusion protein connecting D1 of M13 with the C-terminal domain of TolA was solved using multiple isomorphous replacement techniques and refined to 2.5 Å resolution (Figure 1.11) (Lubkowski *et al*., 1999) (PDB ID 1TOL). The fusion protein was created by fusing the C-terminal residues 295 – 421 of TolA to domain 1 and the first glycine rich linker sequence, residues 1 – 86, of $pIII_{M13}$. Topologies of pIII D1, whether fused to D2 or TolA, are very similar. Topologies between TolA and D2 are very different yet both interact with many of the same residues of D1. TolA interacts with the β-barrel at the site occupied by the hinge in the pIII D1-D2 structure (Figure 1.10). TolA also binds to pIII D1 *via* its extended β-sheet loops at a site overlapping the interface D1 forms

with D2. Neither of these sites are exposed in the pIII D1-D2 "pre-infection" complex. Both pIII D1 to D2 or to TolA interactions also show a very large buried surface area. Lubkowski *et al.* (1999) report a buried surface area of 2154 $Å^2$ when D1 is bound to D2 and of 1768 $Å^2$ when D1 is bound to TolA. These studies support a model in which pIII D2 keeps D1 concealed until D2 binds to the F pilus. This interaction induces a conformational change in the hinge region between pIII D1 and D2 that exposes D1, allowing it to bind with its receptor, TolA.

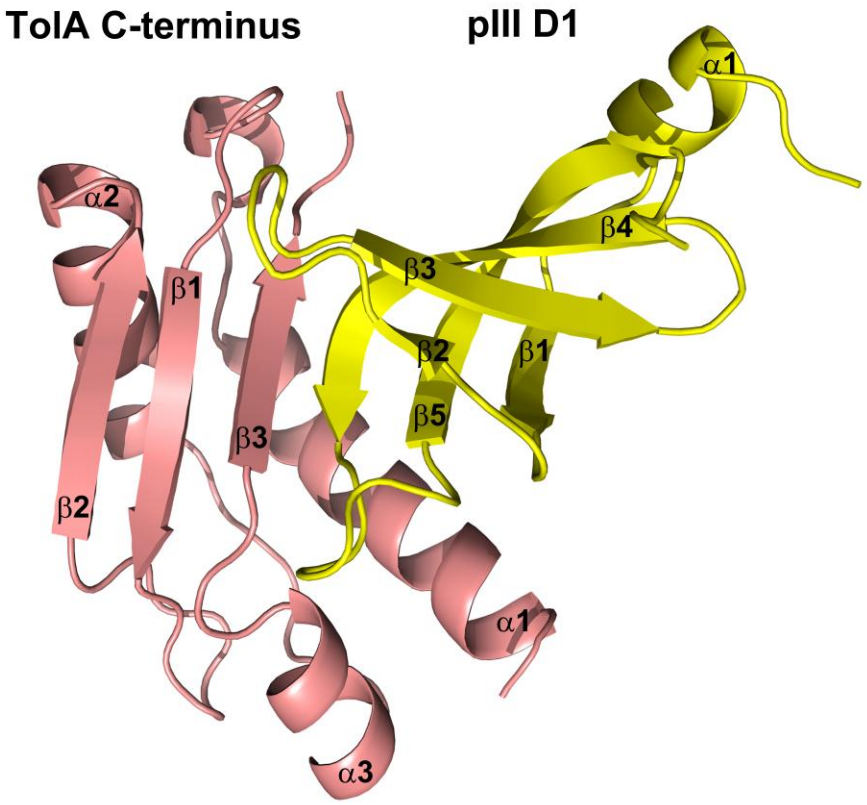


Figure 1.11. $pIII_{M13}$ D1 / TolA fusion protein

PDB ID 1TOL. pIII D1 shown in yellow, TolA shown in salmon. α helices and β strands are numbered N-terminally based on individual proteins.

Although no structures of uncomplexed TolA have been solved by X-ray crystallography, Deprez *et al.* (2005) solved the solution structure of the C-terminal end of *E. coli* TolA by NMR spectroscopy (PDB ID 1S62). Alignment of the averaged structure of C-terminal TolA with the C-terminal domain of TolA in complex with $pIII_{M13}$ D1 (PDB # 1TOL) shows very little conformational change, with a root mean square deviation (RMSD) between main chain atoms of 4.1 Å. These structures suggest that TolA undergoes little to no conformational change upon pIII binding.

More recently the crystal structure of pIII D1 from the filamentous phage IF1 bound to the C-terminal domain of TolA was solved (Lorenz *et al*., 2011). IF1 phage infect I-pilus-bearing bacteria. This group showed previously that although the IF1 infection mechanism appears to be similar to fd phage infection, D1 and D2 from IF1 do not interact to stabilize the protein as is the case with D1 and D2 from M13 and fd. They considered therefore that IF1 uses a different mechanism for infection (Lorenz & Schmid, 2011). Structural analysis revealed that despite the differences in domain interactions and a low, 31%, sequence identity, M13 and IF1 D1 structures and interactions with TolA are very similar (Figure 1.12). There are conserved backbone hydrogen bonds formed between β4 of D1 and β3 of TolA-C. Comparable hydrophobic interactions and salt bridges from M13 are conserved in IF1 (Lorenz et al., 2011). These structures demonstrate that although sequence similarity is low, both M13 and IF1 D1 bind to the same pocket of TolA via similar interactions.

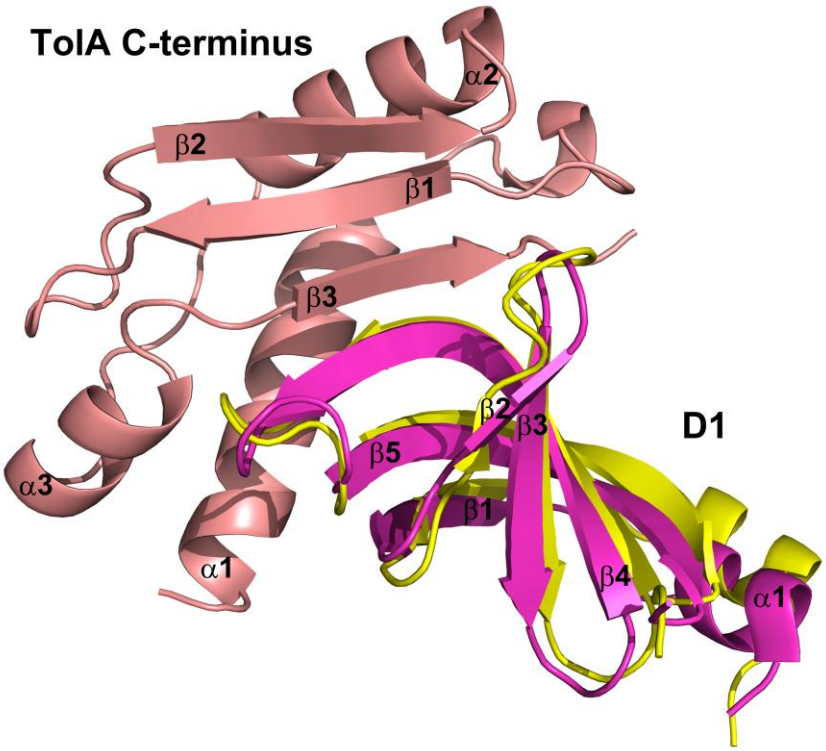


Figure 1.12. IF1 and M13 D1 alignment

IF1 D1, PDB# 2X9A, (magenta) and M13 D1, PDB# 1G3P, (yellow), bound to TolA-C (salmon).

Sequence identity between $pIII_{CTX}$ and $pIII_{M13}$ is low at 15% and D1 of $pIII_{CTX}$ is predicted by the secondary structure prediction program PORTER (Pollastri & McLysaght, 2005) to be predominantly α-helical. Despite these differences, the CTXΦ genome is very similar in structure and gene organization to that of the *E. coli* fd filamentous phage, and $pIII_{CTX}$ appears to be composed of three domains, similar to Ff phage pIII proteins, which are separated by serine and proline rich, low complexity regions. CTXΦ phage are believed to infect *V. cholerae* by binding of its minor coat protein $pIII_{CTX}$ to toxin co-regulated pilus (TCP) and TolA receptors, in a manner similar to *E. coli* infection by Ff phage (Heilpern & Waldor, 2000, 2003; Waldor & Mekalanos, 1996).

1.5. Research objectives

Cholera continues to be a great threat, devastating countries in desperate times. As phage uptake is extremely specific, understanding the molecular mechanism of uptake could uncover strategies to target *V. cholerae* specifically, promote uptake of other molecules, by *V. cholerae* and design new strategies for controlling outbreaks.

Because TolA is the only essential receptor for CTXΦ uptake, my objective was to investigate the interactions between $pIII_{CTX}$ and TolA.

To accomplish this I conceived of three experimental goals. The first was to express, purify and solve the crystal structure of the N-terminal domain of CTXΦ pIII, which, by analogy with the Ff pIII proteins, is expected to be the TolA-interacting domain. The second goal was to express and purify the C-terminal domain of TolA, and the third was to demonstrate and characterize the interactions between $pIII_{CTX}$ and TolA.

2. Materials and Methods

2.1. Bacterial strains and DNA constructs

TolA-C and pIII constructs were created by the polymerase chain reaction (PCR) using the primers listed in Appendix A and conditions listed in Appendix B. A gene fragment encoding residues 241 - 256 of TolA was amplified from the genomic DNA of *V. cholerae* O395 by PCR. A panel of pIII gene fragments was amplified from an El tor plasmid whose *ctxAB* genes were replaced with a kanamycin marker (pCTX-Km). The plasmid pET15b was chosen for protein expression in DE3 cells as it possesses a lac promoter and encodes a 6 x Histidine (His_6) tag followed by a thrombin digestion site 5' to the inserted gene. PCR products and empty pET15b plasmids (Appendix C) were digested by incubating them with NdeI fast-digest enzyme (Fermentas) for 3 hours at 37 °C, followed by 1 extra hour incubation after addition of BamHI. The digested PCR products were ligated with digested pET15b for 3 hours at 20 °C, using reagents listed in Appendix B3. Plasmid encoding the TolA-C constructs were used to transform *E. coli* BL21(DE3) cells, whereas plasmids encoding truncated pIII constructs were used to transform Rosetta-gami B cells for protein expression, *via* the $CaCl_2$ heat shock method at 42 °C (Huff *et al.* 1990). The longest

construct coding for residues 1 – 360 of pIII (D1-D3) was used as a template to design smaller, C-terminally truncated pIII constructs by engineering a stop codon into the sequence by site-directed mutagenesis using primers shown in Appendix D. All constructs were verified by DNA sequencing (GeneWiz).

2.2. Purification of *V. cholerae* TolA and CTXΦ-pIII domains

Bacterial cells transformed with the appropriate plasmid were grown over night (ON) in 200 ml Luria-Bertani (LB) broth containing 100 µg/ml Ampicillin (Amp) at 37 °C with shaking at 250 revolutions per minute (RPM). Ten-ml aliquots of ON culture were added to 1 L LB + Amp and grown at 37 °C to an OD_{600} of ~0.6. Protein expression was then induced by the addition of IPTG to 1 mM and cells were shaken 250 RPM, ON at 19 °C. Cells were pelleted by centrifugation at 4 000 X*g* for 30 min, the supernatant was discarded and the pellet was re-suspended in lysis buffer (Table 1). Cells were incubated in lysis buffer at room temperature for 1 hour and then lysed by sonication. Cellular debris was removed by centrifugation at 40 000 X*g* for 40 min. and the supernatant was filtered through a 0.45 µm membrane. The filtered supernatant was loaded onto a gravity column containing 2-3 ml of Ni-NTA beads (Qiagen) for purification of His_6-tagged proteins. The elution fraction was concentrated using a stirred cell concentrator (Amicon) with a 3 000 molecular mass cut off (MMCO) membrane and loaded into a Sepharose size exclusion column designed for proteins with masses between 1 and 100 kDa (GE Healthcare

HiPrep 26/60 Sephacryl S-100 HR). When His_6 tagged proteins were the desired product, fractions containing purified protein were concentrated again to a desired concentration, flash frozen in liquid nitrogen in various aliquots depending on their purpose, and stored at -80 °C. When His_6-cleaved purified proteins were the desired product, fractions containing the purified protein were collected and placed on a rocker at 4 °C ON with ~ 100 units of thrombin. Digested proteins were then run over Ni-NTA beads to remove undigested protein and free His_6 tags. The flow through and wash fractions from the Ni-NTA column were concentrated in an Amicon stirred cell concentrator and loaded on to a size-exclusion column to remove the thrombin. Fractions containing purified, His_6-cleaved protein were concentrated, flash frozen and stored as described above.

Table 1. Buffers

Buffer Name	**Buffer**	**pH**	**Salt**	**Other**
Lysis Buffer	Bis-Tris	6.5	100 mM NaCl	10% glycerol, 0.1% Tween, 5 mM Imidazole, Lysozyme, 1 Complete Protease Inhibitor Cocktail Tablet, EDTA Free
Purification Buffer	Bis-Tris	6.5	50 mM NaCl	
Protein Buffer	Bis-Tris	6.5	20 mM NaCl	
Wash Buffer	Bis-Tris	6.5	50 mM NaCl	10 mM Imidazole
Elution Buffer	Bis-Tris	6.5	50 mM NaCl	200 mM Imidazole

Seleno-methionine substituted N-pIII was created for crystal structure determination, in M9 minimal media with the required amino acids (Fan *et al*., 1997). Cells were grown ON in LB Amp at 37 °C at 250 RPM. Ten ml of ON culture for every 1 L of minimal media was centrifuged at 3 500 g for 20 min. The supernatant was discarded and the cells were re-suspended in 1 L of M9 minimal medium and returned to the shaker at 37 °C and 250 RPM. Once cells reached an OD_{600} of ~0.4, 60 mg/L of L-seleno-methionine, 100 mg/L of lysine, threonine, and phenylalanine and 50 mg/L of leucine, isoleucine and valine were added. The cells were incubated for 15 minutes at 37 °C shaking at 250 RPM, before IPTG was added to a concentration of 1 mM to induce protein expression. Cells were grown ON at 19 °C, shaking at 250 RPM. Cells were lysed and protein was purified and stored as described above.

2.3. Protein interaction assays

TolA-C and pIII D1 proteins were purified with and without N-terminal His_6 tags as described in Section 2.2. One mg of His_6-pIII D1 or His_6-TolA-C was loaded onto 250 µl of Ni-NTA beads. The beads were washed with 1 ml of wash buffer twice. Next 1 mg of the His_6 tag cleaved binding partner, TolA-C or pIII D1, respectively, was run over the Ni-NTA beads and washed twice. Finally the His_6 tagged proteins were eluted from the column with 1 ml elution buffer (Table 1).

2.4. Purification of a protein complex for crystal structure determination

His_6-TolA and N-pIII were purified as described in Section 2.2. For purification of the His_6-TolA-C / N-pIII complex, ~30 mg of purified His_6-TolA was loaded onto the Ni-NTA column (2 ml Ni-NTA beads) and the column was washed with 5 ml of purification buffer (Table 1). Next ~27 mg of N-pIII was loaded onto the column and the column was washed with 5 ml purification buffer. Proteins were eluted from the column with elution buffer and the elution fraction was concentrated in the Amicon stirred cell concentrator with a 10 000 MMCO membrane, and then loaded onto the Sepharose size exclusion column. The fractions containing the purified complex were concentrated and stored as described in section 2.2.

2.5. Crystallization and X-ray diffraction

All crystals were grown using hanging-drop vapor diffusion at 20 °C. Initial crystals were grown by mixing equal 2 µl volumes of purified protein in protein (His_6-TolA-C / N-pIII complex) or purification buffer (N-pIII) and 2 µl of well solution from the crystal screening kits. Trays were monitored weekly for crystal growth. Positive and promising conditions were further optimized by slight variations in well solution concentrations and pH. Optimized conditions were then treated with additive and detergent screens by mixing 2 µl volumes of purified

protein in protein or purification buffer, 1.6 µl of concentration adjusted well solution and 0.4 µl of additive or detergent. While many conditions produced crystals (Table 3), it was later found that the best diffracting crystals were from drops containing; 23.5 mg/ml, N-pIII-SeMet in purification buffer, 20% PEG 3350, 8% Tacsimate pH 5.0 and Fos-Choline-9 frozen in 25% glycerol and 27.1 mg/ml His_6-TolA-C / N-pIII complex in protein buffer, 25% PEG 6000 and 100 mM MES pH 6.0. Crystals developed after 2 weeks and were flash frozen in liquid nitrogen with (N-pIII) and without (His_6-TolA-C / N-pIII complex) up to 25% glycerol. Crystals were shipped to the Stanford Synchrotron Radiation Lightsource (SSRL) and data were collected on beamline 9-2.

2.6. Structure determination and refinement

X-ray diffraction data for N-pIII and the His_6-TolA-C / N-pIII complex crystals were collected remotely on SSRL beamline 9-2 at at 100 K using Blu-Ice software (Gonzales *et al.*, 2008). The initial test diffraction images were indexed and x-ray fluorescence scans were carried out to select the appropriate wavelengths for multiple and single wavelength anomalous diffraction data collection (MAD and SAD, respectively) by Blu-Ice. For N-pIII, two data sets at wavelengths corresponding to inflection point and high energy remote were collected from a crystal that diffracted to ~ 4 Å. We solved the structure with this MAD data set. Later as another SeMet N-pIII crystal was found to diffract greater than 3 Å, we collected a dataset at the peak wavelength for that crystal and used

the SAD 2.9 Å resolution data for refinement purpose. The MAD data set of N-pIII was processed and scaled by iMosflm and SCALA (Leslie, 1992, Collaborative Computational Project, Number 4, 1994). The highest resolution was cut off at 2.9 Å with an I/σI of 2.7 for this resolution shell. N-pIII single wavelength and His_6-TolA-C / N-pIII complex data were processed and scaled using the XDS suite (Kabsch, 1993) to resolutions 2.9 Å and 1.44 Å respectively. Matthews coefficients (Matthews, 1968) calculated using CCP4 (Collaborative Computational Project, Number 4, 1994), indicated three molecules per asymmetric unit for N-pIII and a single heterodimer for the His_6-TolA-C / N-pIII complex. Initial phases of N-pIII were determined by MAD phasing method using SOLVE (Terwilliger & Berendzen, 1999), and density modification by RESOLVE (Terwilliger, 2000) yielded an interpretable electron density map. The map was traced and model building was carried out in COOT (Emsley & Cowtan, 2004). A few cycles of rigid body refinement followed by tight NCS restrained refinement by REFMAC (Murshudov *et al.*, 1997) brought the R_{cryst} and R_{free} down to 0.417 and 0.447 respectively. Iterative cycles of tight main chain and loose side chain restrained refinement and model building improved the model and yielded R_{cryst} of 0.258 and R_{free} of 0.297. Then after a few cycles of TLS refinement (Murshudov et al., 1997), water oxygens, located using COOT, were included in the refinement. The final Rcryst and Rfree were 0.238 and 0.259 respectively.

The His_6-TolA-C / N-pIII complex structure was solved by the molecular replacement method using the N-pIII structure and TolA from *E. coli* (PDB id:

1TOL). A few cycles of the “map improvements by atoms update and refinement” option of ARP/wARP (Perrakis *et al.*, 1999) made a dramatic improvement to the map with clear density for both main chain and side chain atoms. The model was built manually in COOT and a few cycles of restrained refinement were carried out. Water oxygens were located in COOT and after checking them manually, WATERTIDY in CCP4 (Collaborative Computational Project, Number 4, 1994) was used to move the water oxygen coordinates to the symmetry related position nearest to the host protein molecule. After few cycles of refinement with water oxygens, the new map was checked for possible new density for more water oxygens, ligands and missing parts of the protein molecules. The refinement was stopped when there was no further improvement to the model and no new density in the difference map. Both N-pIII and His_6-TolA-C / N-pIII complex structures were validated using PROCHECK and MOLPROVITY (Laskowski *et al.*, 1993, Davis *et al.*, 2007). The crystal data and refinement statistics are summarised in Table (Table 2, Table 3).

3. N-pIII purification and structure determination

3.1. Introduction

The N terminal 101 residues of $pIII_{CTX}$ were demonstrated by Heilpern & Waldor (2003) to be essential for *V. cholerae* phage infection. Their experiments fused domain 1, domain 2 or domain 1 and 2 of $pIII_{CTX}$ to the full length $pIII_{fd}$ or to domain 3 of $pIII_{fd}$ and showed that, whereas fd phage are normally unable to infect *V. cholerae*, fd phage in which $pIII_{fd}$ fused to domain 1 of $pIII_{CTX}$ are able to *V. cholerae*. Experiments with domain 3 of $pIII_{fd}$ fused to domains 1 and 2 of $pIII_{CTX}$ showed the greatest transduction frequencies (Figure 1.9). These studies indicate that the N-terminal 101 residues of $pIII_{CTX}$ play a critical role in CTXΦ phage uptake in *V. cholerae* infection.

Although domain 1 (D1) from $pIII_{CTX}$ and $pIII_{fd}$ appear to have the same function in phage infection, our sequence analysis indicated minimal similarity between the two domains. Sequence alignment of the first domain from $pIII_{CTX}$ and $pIII_{fd}$ returns a low 15% identity. Secondary structure prediction using PORTER predict $pIII_{fd}$ D1 to be predominantly composed of β sheets, confirmed by protein structure, but predicts $pIII_{CTX}$ D1 to be composed predominantly of α-helices. Together these studies suggest that $pIII_{CTX}$ D1 has a similar function to

$pIII_{fd}$ D1 in phage infection but employs a novel structure and molecular mechanism.

To determine the molecular mechanism of pIII involvement in phage infection we created a variety of truncated pIII constructs (Appendix D), optimized the protein expression conditions and purified them. We screened several truncated pIII constructs against commercially available and lab developed crystal screening kits to search for crystal forming conditions. Conditions that formed N-pIII crystals were optimized and the structure was determined by X-ray crystallography.

3.2. N-pIII purification

To generate a soluble truncated $pIII_{CTX}$ protein, we created C-terminally truncated constructs of $pIII_{CTX}$ by PCR amplification of the $pIII_{CTX}$ gene using pCTX-Km for a template. The constructs were digested with NdeI and BamHI and inserted into a similarly digested pET15b plasmid. The plasmid provides IPTG-inducible protein expression and allows facile purification of the protein by metal-affinity chromatography *via* a cleavable His_6-tag (Appendix C). Most of the truncated pIII proteins expressed well in Rosetta-gami B (DE3) cells but any construct that contained the full D2 domain was insoluble. N-pIII, which comprises residues 1 – 139 and has an estimated molecular mass (MM) of 17.4 kDa with the His_6-tag, was the largest soluble protein generated (Figure 3.1).

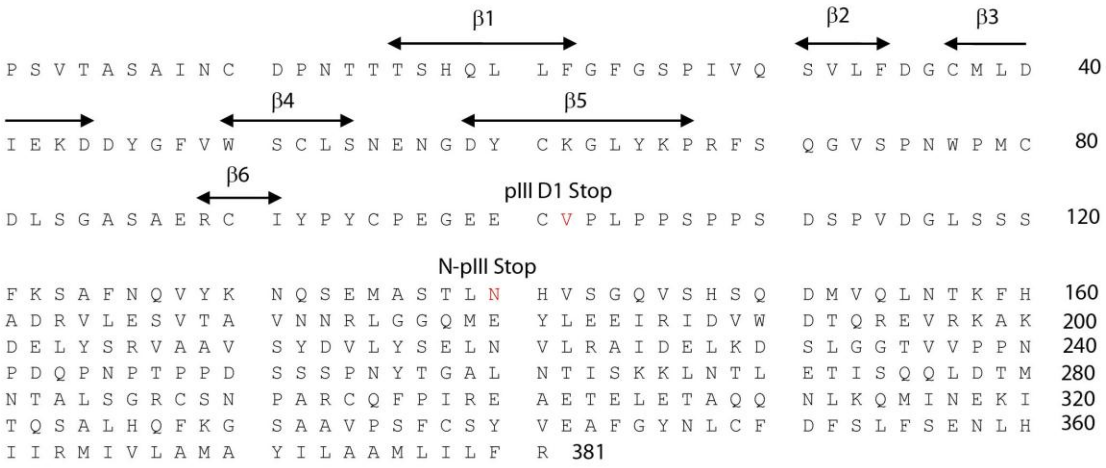


Figure 3.1. $pIII_{CTX}$ sequence showing truncated constructs and secondary structure of N-pIII

Mature pIII sequence. Mutations were made to generate stop codons for C-terminally truncated pIII proteins. The amino acids affected, V102 and N140, to generate pIII D1 and N-pIII respectively, are highlighted in red.

N-pIII was tested for large scale expression and purification under a range of conditions. Cells were grown at 37 °C to an OD of 0.3 OD_{600} – 0.8 OD_{600}. They were then induced with either 0.4 mM IPTG or 1 mM IPTG for 4 hours at 37 °C, for 6 hours at 30 °C or for 16 hours at 19 °C. Optimal conditions were determined to be growth at 37 °C to an OD_{600} of 0.6, at which point the cells were induced with 1 mM IPTG for 16 hours at 19 °C. For large scale purification cells were grown in 3 x 2.8 L flasks containing 1 L LB + Amp each. Between 13 g and 14 g of cell pellet were collected by centrifugation at 4 000 g for 30 min. The cell pellet was re-suspended in 30 ml of lysis buffer (Table 1) and incubated at room temperature for 1 hour. Cells were lysed by sonication on ice at 50% maximal intensity for 1 second bursts, 2 second recovery, for a total of 6 minutes. Cellular debris was removed by centrifugation at 40 000 g for 40 min. To purify N-pIII

from other soluble proteins, the supernatant was filtered through 0.45 µm membrane and loaded onto a gravity column containing 3 ml of Ni NTA beads. The column was washed twice with wash buffer (Table 1) then the bound proteins were eluted with elution buffer. The elution fraction was concentrated to under 5 ml with a 3 000 MMCO membrane in a stirred cell concentrator. In addition to N-pIII$_{CTX}$ (MM 17.4 kDa), all protein preparations from *E. coli* Rosetta-gami B (DE3) cells contained two distinct contaminating proteins after affinity purification, one at ~ 50 kDa and one at ~ 70 kDA. To remove these, proteins in the eluate, proteins were separated by size exclusion chromatography using 320 ml of a rigid allyl dextran/ bisacrylamide matrix on an Acta FPLC (GE Healthcare), producing six to seven 5 ml fractions containing highly purified His$_6$-N-pIII (Figure 3.2). The fractions were then incubated for 16 hours at 4 °C with 100 units of thrombin to remove the His$_6$ tag. Digested, proteins were loaded on a second Ni-NTA column to separate the His$_6$ tags and the His$_6$ tagged proteins from the His$_6$ cleaved N-pIII. The flow-through and the wash fractions, containing N-pIII and thrombin, were concentrated in the stirred cell concentrator and run through a second size exclusion column to separate N-pIII from thrombin (Figure 3.3). The 6, 5 ml fractions of purified N-pIII produced were concentrated to ~ 30 mg/ml, with a yield of 90 – 100 mg of pure His$_6$-cleaved protein per 3 L prep. Purified concentrated protein was separated into 104 µl aliquots. Aliquots not used within 48 hours were flash frozen in liquid nitrogen and stored at -80 °C.

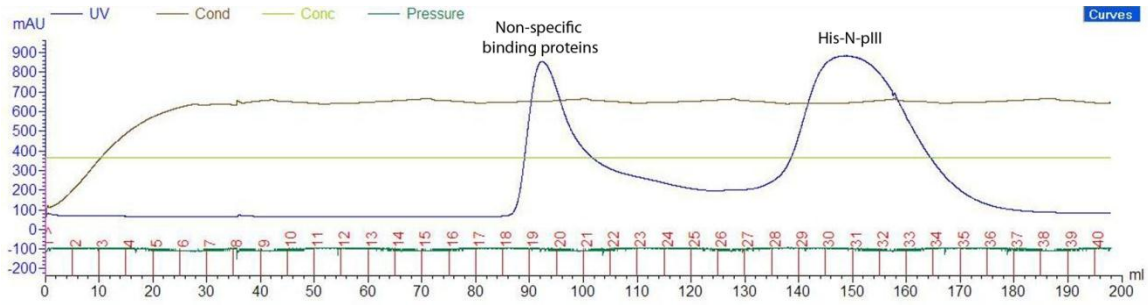


Figure 3.2. Purification of His_6-N-pIII by size exclusion chromatography

His_6-N-pIII size exclusion purification following affinity chromatography. A peak beginning just before 90 ml represents contaminants. The peak beginning just before 140 ml represents purified His_6-N-pIII.

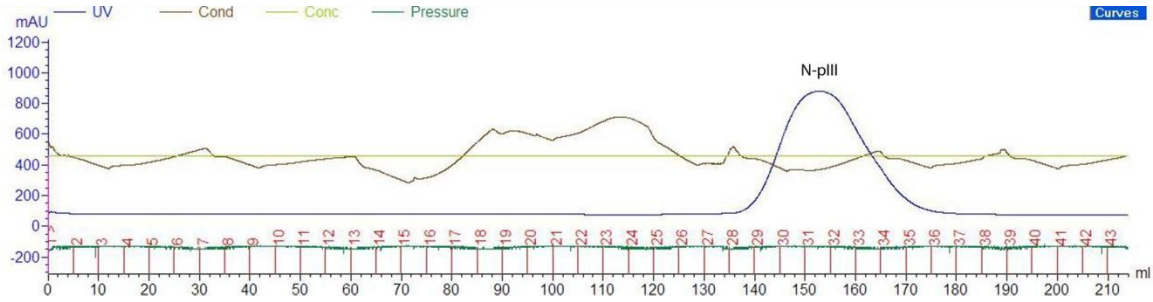


Figure 3.3. Purification of N-pIII by size exclusion chromatography

N-pIII size exclusion purification following thrombin digestion and affinity chromatography. The peak beginning just after 140 ml contains purified N-pIII.

3.3. Crystal screening

Our lab maintains a stock of 10 commercially available crystal screening kits from Hampton Research, Emerald BioSystems and Molecular Dimensions (Appendix G). Statistical analysis of crystal conditions from protein structures posted to the Protein Data Bank (PDB) and the Biological Macromolecule Crystallization Database (BMCD) indicate that PEG and ammonium sulfate based screens account for 38% and 37% of the conditions for X-ray crystallography solved structures respectively, followed by other salts at 19% and

finally organic solvents at 6%. The proportion of crystal structures solved by PEG based screens jumped to 71% when analysis was done for protein-protein complexes, with MM PEGs between 3 000 and 8 000 accounting for 80% of that total (Radaev S. & Sun P. D., 2002). Based on the high proportion of protein crystal structures solved by PEG and ammonium sulfate based screens, we manufactured supplemental screens that would screen a variety of concentrations of PEG 6000 or ammonium sulfate against a pH range and salt additives. Aliquots of 32.6 mg/ml of N-pIII in purification buffer were used to set up all our screening conditions using the hanging drop vapor diffusion method. Two conditions, Emerald Biosystems WIZ 1 screen condition # 39, containing 20% (w/v) PEG 1000, 0.1 M sodium phosphate/citric acid and 0.2 M lithium sulfate and Hampton Research PEG/ION 2 screen condition # 12, containing 20% w/v polyethylene glycol 3,350 and 8% v/v Tacsimate pH 5.0, showed promising crystals after two weeks with no further growth after 3 weeks.

Because N-pIII does not share sequence identity with any known protein, it was necessary to solve the structure *de novo* using MAD or SAD phasing. To accomplish this we prepared selenomethionine substituted N-pIII (N-pIII SeMet). This protein was purified, concentrated to 23.5 mg/ml, aliquoted and stored as described above. N-pIII SeMet was screened against a range of concentrations based on the two promising conditions listed above and crystals grew in ~ 3 weeks (Figure 3.4). Smaller crystals, grown under similar conditions, were tested to confirm the presence of protein crystals, and discount the possibility of salt

crystals, by treatment with Izit. As expected with protein crystals, Izit was able to enter the solvent channels between proteins and stain the crystals dark purple (Figure 3.5).

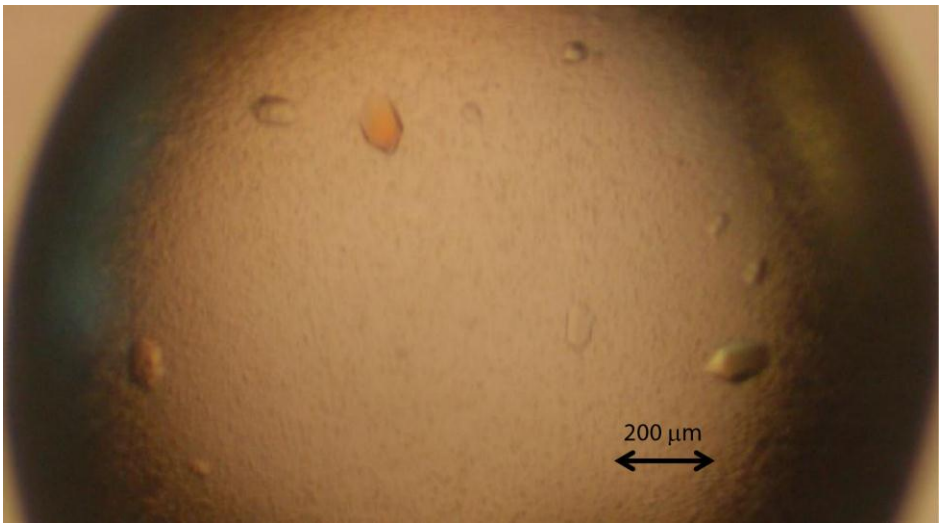


Figure 3.4. N-pIII SeMet crystals

N-pIII SeMet crystals grown in 20% (w/v) PEG 1000, 0.1 M sodium phosphate/citric acid and 0.2 M lithium sulfate

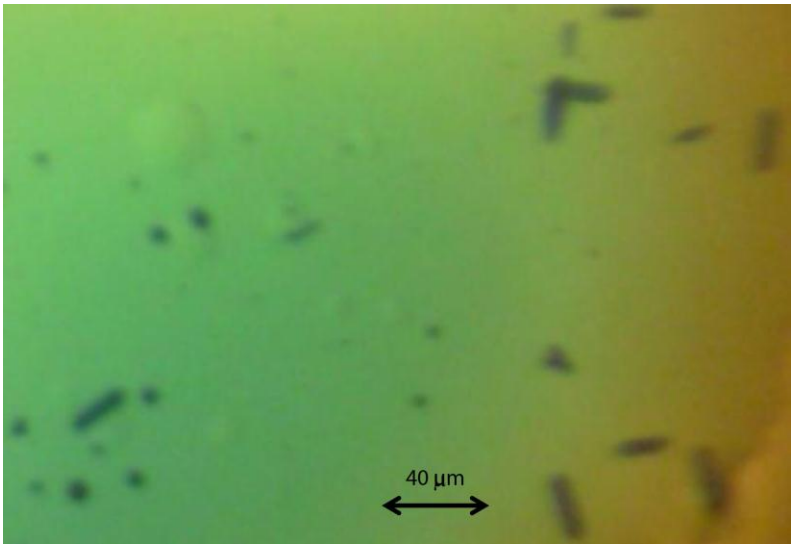


Figure 3.5. Izit test

Izit test on N-pIII crystals. Dark, Izit stained crystals indicate large solvent channels associated with protein crystals.

To further optimize the crystals, we used optimal conditions as the reservoir buffer, and screened these against Hampton Research additive and detergent screens. Many detergent screen conditions produced good diffraction quality crystals (Appendix G).

3.4. X-Ray data collection, structure determination and refinement

To solve the structure of SeMet N-pIII, crystals were flash frozen in 25% glycerol and shipped to the Stanford Synchrotron Radiation Lightsource (SSRL). X-ray diffraction data were collected remotely on beamline 9-2. The X-ray crystal structure of N-pIII SeMet was solved to 3.8 Å resolution by MAD methods using SOLVE/RESOLVE (Terwilliger & Berendzen, 1999, Terwilliger, 2000). The crystals belong to the hexagonal lattice system $P6_322$, with unit cell dimensions 126.25, 126.25, 128.27 Å. There are 3 molecules per asymmetric unit and a solvent content of 62.5%. The phases of SeMet N-pIII were transferred to a 2.9 Å SAD data set collected later during the same collection beam time, and an interpretable density electron map was generated. Amino acids 6 – 103 were fit using COOT (Emsley & Cowtan, 2004) (Table 2).

Table 2. Crystallographic data collection and refinement statistics for N-pIII

	N-pIII SeMet MAD		N-pIII SeMet SAD
Data Collection			
Beamline	SSRL 9-2		SSRL 9-2
Space group	$P6_322$		$P6_322$
Cell dimensions			
a, b, c (Å)	126.25, 126.25, 128.27		126.25, 126.25, 128.27
α, β, γ (°)	90, 90, 120		90, 90, 120
Resolution (Å)	3.8		2.9
Wavelength (Å)	0.918	0.9792	0.98
Completeness (%)	99.9/99.7	99.9/99.9	99.6/99.9
Observed reflections	210923	166181	584826
Unique reflections	20639	20639	25290
R_{sym} (%)	13.6/28.1	13.1/27.6	8.2/84.9
I/σ(I)	5.4/2.7	5.6/2.7	24.1/2.1
Mosaicity (°)	0.9	0.9	0.3
Refinement statistics			
Resolution limits (Å)			20 – 2.9
Z (Number of Molecules per Unit Cell)			36
R_{cryst} (%)			23.8
R_{free}(%)			25.9
Number of reflections used for refinement			13171
Number of atoms			
Protein			2251
Water			11
B-factor (Å^2)			
Protein			
Chain A			41.3
Chain B			46.3
Chain C			35.8
Water			
RMSD			
Bond lengths (Å)			0.012
Bond angles (°)			1.88
Ramachandran plot statistics (%)			
Favored			82.4
Allowed			10.3
Generously allowed			6.2

[a]Overall/last shell

[b]R_{sym} is the unweighted R value on I between symmetry mates.

[c]R_{cryst}= Σhkl||Fobs(hkl)|-|Fcalc(hkl)||/ Σhkl|Fobs(hkl)|.

[d]R_{free} is the cross validation R factor for 5% of reflections against which the model was not refined.

3.5. N-pIII structure

We solved the structure of N-pIII. The unit cell contains 3 molecules per asymmetric unit. Residues 6 – 103 for molecule A and residues 6 – 104 for molecules B and C of the asymmetric unit were resolved. N-pIII is a globular protein that resembles both a distorted β-barrel and a β-sandwich, with two β-sheets lying orthogonally to each other with several extended loops. Of the 6 β-strands that comprise N-pIII, β1, β3, β4 and β5 are longer strands, 6 – 9 residues long, whereas β2 and β6 are short strands of 3 – 4 residues. β1, β2, β3, β5 and β6 form an irregular β-barrel whereas β4 forms an antiparallel beta sheet with the ends of β1 and β5 not occupied in the barrel. There are many extended loops connecting distant strands. β5 and β6 are connected by a 21 residue loop whereas β1 and β2, and β3 and β4 are connected by 8 and 5 residue loops respectively (Figure 3.6).

N-pIII has eight cysteines forming four disulfide bonds within its first 101 amino acids. Cys10 links the N-terminus to the beginning of β3 through Cys37. Cys52 links β4 to β5 through Cys61. Cys80 links the extended loop between β5 and β6 to β6 through Cys90 and Cys95 forms a disulfide bond with Cys101 in the loop following β6 (Figure 3.6).

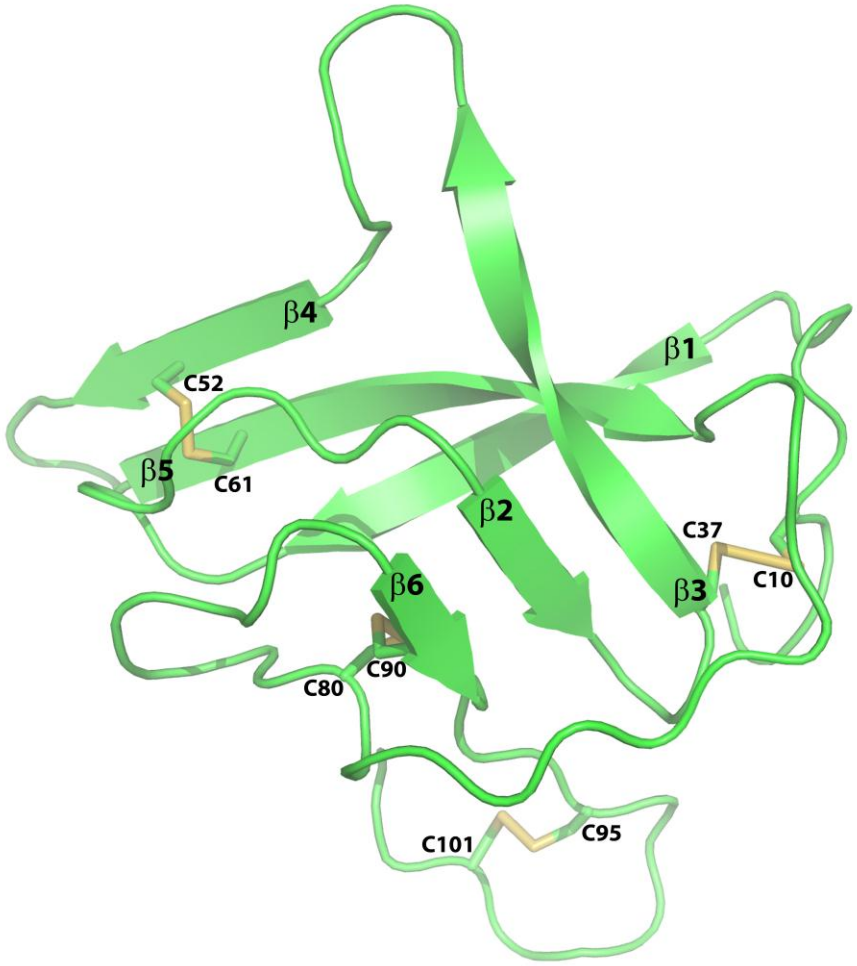


Figure 3.6. N-pIII structure for molecule B of the N-pIII asymmetric unit

Globular N-pIII with beta strands labeled 1 – 6 at the beginning of each strand. Disulfide bonds between C10 and C37, C52 and C61, C80 and C90 and between C95 and C101 are highlighted in yellow.

The three molecules per asymmetric unit align very well with a root mean square deviation (RMSD) between all main chain atoms of 1.00 Å, 1.401 Å and 1.621 Å for molecules A & B, A & C and B & C respectively. The RMSD for 75% of the main chain atoms is reduced to 0.634 Å, 1.178 Å and 1.212 when eliminating the more divergent N and C-terminal ends (calculated using COOT) (Figure 3.7). Most of the deviation between molecules A & B is attributed to the N

and C-terminal ends whereas molecule C has additional structural variability in the β4-β5 loop. The deviation at the terminal ends are to be expected whereas the β4-β5 loop of molecule C is constrained by crystal packing. Molecules A and B form symmetrical interactions, with β1 and β3 from molecule A forming comparable bonds with β3 and β1 of molecule B. Conversely molecule C packs into the extended loop between β5 and β6 of molecule B along the β4 strand and the loop following β1, apparently displacing the β4-β5 loop (Appendix F).

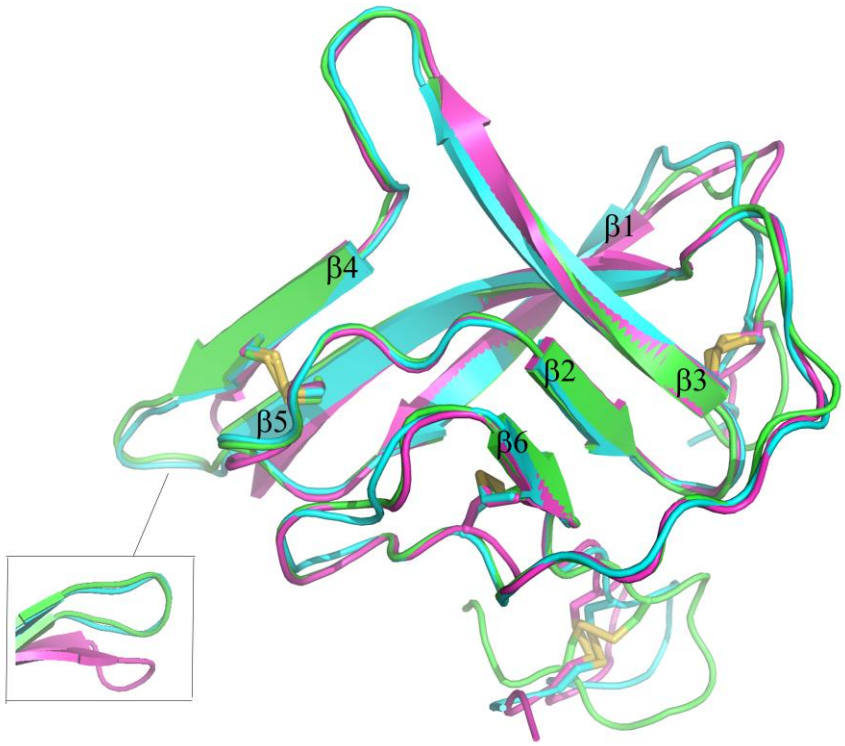


Figure 3.7. Alignment of the 3 molecules per asymmetric unit of N-pIII

Molecule A (blue), B (green) and C (pink) from the asymmetrical unit of N-pIII. Inset; the loop between β4 and β5 of molecule C folds away from the placement of the loop in molecule A and B.

The N-pIII structure was compared to the corresponding domain structure in fd and M13 phage pIII proteins. These structures align very well in spite of

having minimal amino acid sequence similarity. Whereas $pIII_{M13}$ D1 has an N-terminal α-helix followed by a loop, $pIII_{CTX}$ has an N-terminal loop. From there both proteins contain 5 β-strands organized into irregular β-barrels. Whereas the 5th β-strand completes the domain for $pIII_{M13}$, $pIII_{CTX}$ has an additional extended loop and β-strand. Domain 1 of $pIII_{M13}$ also contains corresponding disulfide bonds to the two N-pIII possesses prior to the end of β5, one joining the N-terminus to β3 and one joining β4 to β5. All previously known pIII D1 structures (M13, fd and IF1) have a very similar structure and share a common TolA binding domain. Structural analysis suggests N-pIII may bind TolA in a similar manner (Figure 3.8).

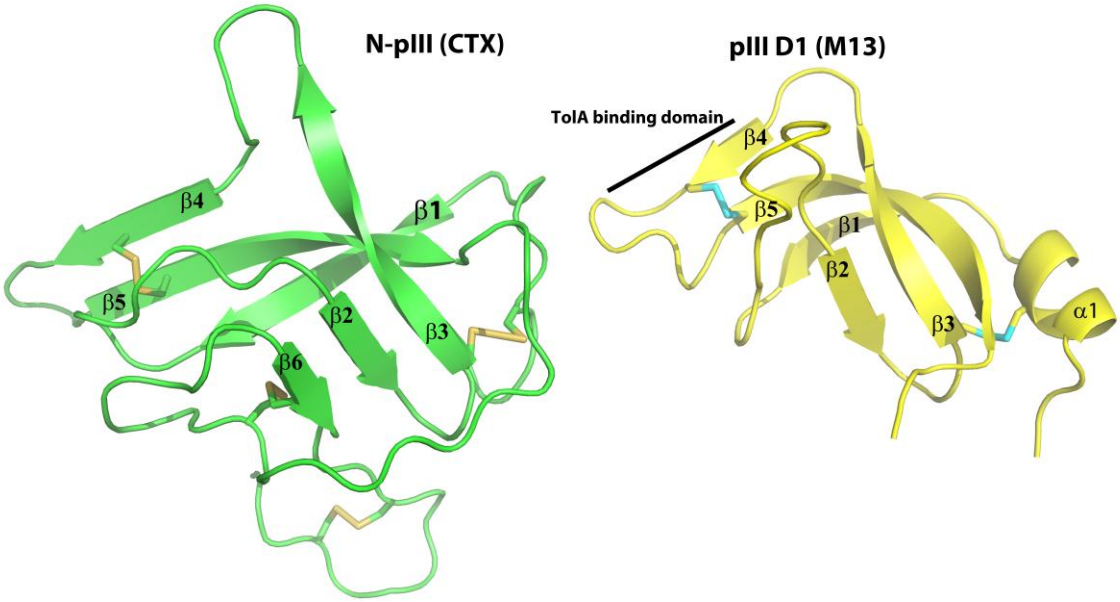


Figure 3.8. Comparison of N-$pIII_{CTX}$ and $pIII_{M13}$ D1 structures

N-$pIII_{CTX}$ (green) and $pIII_{M13}$ D1 (yellow) with β strands numbered accordingly. The TolA binding domain is labeled along β4 of $pIII_{M13}$.Disulfide bonds are indicated on N-$pIII_{CTX}$ in orange and on $pIII_{fd}$ D1 in blue.

M13 β4 has a stretch of 10 residues (NATGVVVCTG) shown to interact with TolA with a stretch of small, hydrophobic residues (VVVC) accounting for back bone hydrogen bonding between β4 of $pIII_{M13}$ and β3 of TolA.

Sequence analysis shows Cys52 (CTX) as the only possible identity in this region. The recently published structure of IF1 in complex with TolA demonstrated that despite low 30% sequence identities in domain 1 and none along the backbone interactions between β strands (IGIG sequence in IF1), the different proteins form an almost identical interaction. A small hydrophobic SCLS stretch of residues in β4 of $pIII_{CTX}$ could possibly play the same role in TolA binding.

3.6. Conclusion

I solved the de novo crystal structure of N-pIII SeMet using MAD phasing techniques of X-ray crystallography. The structure demonstrates that although the N-terminal domains of CTX and M13 pIII proteins have only 15% sequence identity, these domains have very similar topologies. Both proteins are predominantly composed of β-strands which form irregular β-barrels. The proteins also share two very similar disulfide bonds.

4. TolA, pIII binding interactions

4.1. Introduction

Domain 1 of pIII, from M13, fd and IF1 phage, bind to *E. coli* TolA, the essential receptor dictating phage uptake. $pIII_{CTX}$ appears to use a similar mechanism for uptake by *V. cholerae*. Molecular interactions between TolA of *e. coli* and $pIII_{M13}$ were demonstrated using a fusion protein of the C-terminal domain of TolA and the N-terminal domain of $pIII_{M13}$, linked by a glycine-rich linker sequence that separates domain 1 and 2 of $pIII_{M13}$ (Figure 1.11) (Lubkowski *et al.* 1999). Although the crystal structure shows extensive interactions between the C-terminal domain of TolA and D1 of $pIII_{M13}$, the possibility that the linker sequence induces the fit shown in this structure could not be completely discounted, until the complex structure between $pIII_{IF1}$ D1 and TolA was solved recently (Lorenz *et al*., 2011).

Heilpern and Waldor (2003) demonstrated that *V. cholerae* TCP knockouts reduced the transduction frequency of phage, whereas *V. cholerae* with a TolA knockout were not infected by phage within their detection limits. These results showed that $pIII_{CTX}$ alone can confer infection of *V. cholerae* to fd phage.

Although it has been shown that the N-terminal domain of pIII and the TolA protein are essential for CTXΦ phage uptake (Heilpern & Waldor, 2003), there is no direct evidence that N-pIII$_{CTX}$ binds to TolA. To identify an interaction between pIII$_{CTX}$ and TolA, we created constructs similar to those used in M13 and fd phage experiments (Lubkowski *et al*. 1999, Heilpern & Waldor, 2003). Sequence identity between the C-terminal domain of TolA used in the PDB 1TOL structure and the C-terminal domain of *V. cholerae* is 29% thus I designed a similar soluble C-terminal domain construct of *V. cholerae* TolA containing residues 241 – 356 (TolA-C) based on sequence analysis for interaction assays. Domain 1 of pIII$_{CTX}$ (pIII$_{CTX}$ D1) was designed to include all mature residues up to the first low complexity region. We used the pIII$_{CTX}$ D1 and TolA-C proteins in pull down experiments, showed stable interactions, and based on these results we purified the N-terminal domain of pIII$_{CTX}$ (N-pIII) in complex with TolA-C and solved the structure by X-ray crystallography.

4.2. His$_6$ tag pull down assay

To show interactions between pIII and TolA we created soluble truncated proteins containing the domains predicted to interact. pIII$_{CTX}$ D1 is a C-terminally truncated protein comprised of residues 1 – 101 of pIII$_{CTX}$ (Figure 3.1), generated and purified as described in Section 3.1. This construct is slightly shorter than the N-pIII form we crystallized, which comprised residues 1 – 139. TolA-C represents the C-terminal periplasmic domain of TolA, residues 241 – 356 of TolA (Figure

4.1). DNA encoding the TolA-C construct was PCR-amplified from the genomic DNA of *V. cholerae* O395, and inserted into the pET15b plasmid as described in Section 2.1. TolA-C protein was expressed in BL21(DE3) cells and purified as described in Section 2.2.

Ni-NTA beads have high affinity for His_6-tagged proteins. In a pull down assay using Ni-NTA beads, a His_6-tagged protein can be used as the “bait” with which to capture an interacting protein, the “prey”. The bait is immobilized on a Ni-NTA matrix and purified proteins or protein mixtures can be run over the column in order to capture a putative binding partner. We ran two separate experiments, one using His_6-TolA as the bait and His_6-cleaved N-pIII as the prey. The second experiment used His_6-N-pIII as the bait and His_6-cleaved TolA-C as the prey. SDS-PAGE gels confirmed that His_6-pIII D1 or His_6-TolA-C remained bound to the Ni-NTA beads until elution of the column with 200 mM imidazole. SDS-PAGE gels also confirmed that His_6 tag cleaved proteins pIII D1 and TolA-C ran through a Ni-NTA column unimpeded. To confirm binding interactions, we loaded either His_6-pIII D1 or His_6-TolA-C onto a Ni-NTA column. The column was washed twice with a wash buffer (Table 1) and then the untagged binding partner, TolA-C or pIII D1 respectively, was loaded onto the column. The column was washed twice with wash buffer and eluted with increasing concentrations of 20 mM, 200 mM and 500 mM imidazole for elution of non-specific binding, His_6 tagged proteins and excess His_6 tagged proteins respectively. SDS-PAGE gels revealed that untagged pIII D1 binds specifically to immobilized His_6-tagged

TolA-C, and that untagged TolA-C binds specifically to immobilized His_6-tagged pIII D1 (Figure 4.2). Thus, the pull down assay demonstrated a specific interaction between the N-terminal domain of CTXΦ pIII and the C-terminal domain of *V. cholerae* TolA.

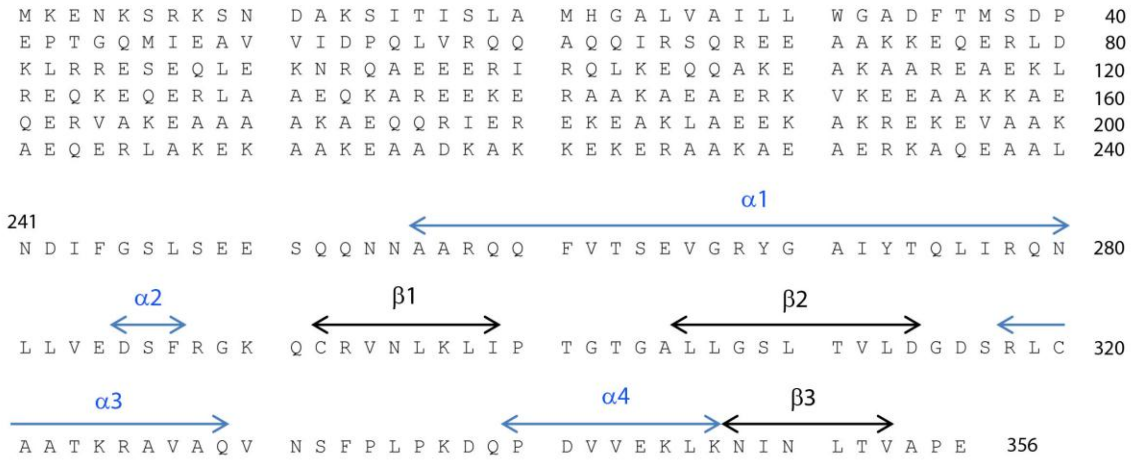
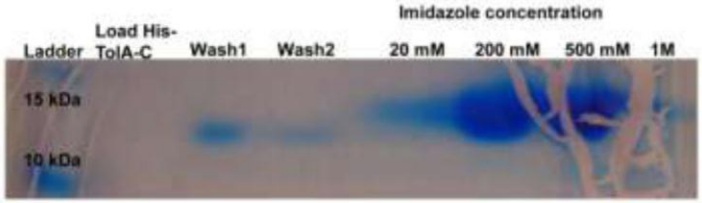


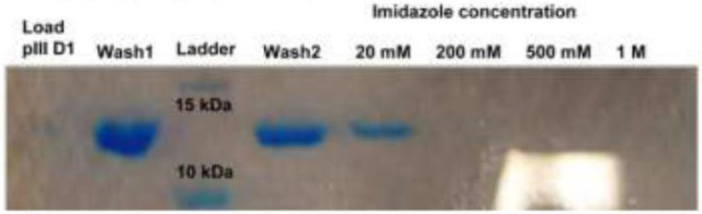
Figure 4.1. TolA sequence showing the truncated construct and secondary structure of TolA-C

V. cholerae O395 TolA sequence. TolA-C sequence (residues 241 – 356) begins on line 7 of the TolA sequence. β-strands and α-helices of TolA-C are highlighted in black and blue respectively.

A His-TolA-C Ni-NTA column



B pIII D1 Ni-NTA column



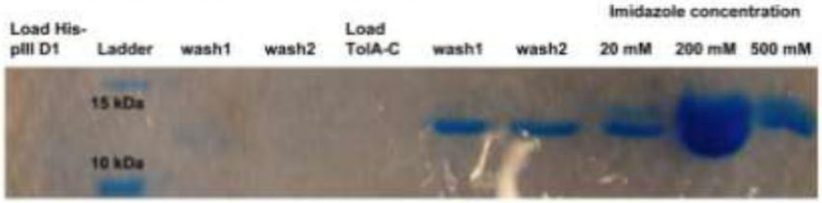
C His-TolA-C / pIII D1 Ni-NTA column

His-TolA-C MW = 14.95 kDa
pIII D1 MW = 11.49 kDa

D His-pIII D1 NTA-Ni column

E TolA-C NTA-Ni column

F His-pIII D1 / TolA-C NTA-Ni column



His-pIII D1 MW = 13.37 kDa
TolA-C MW = 13.07 kDa

Figure 4.2. TolA / pIII D1 pull down assay

All gels are labelled above the appropriate wells. The "Ladder" is with a mixture of unstained MM marker protein (Fermentas). Load is the flow through from adding the respective protein in purification buffer to the Ni column. Wash is the flow through from addition of 1 ml wash buffer. The elution flow throughs contained 20 mM, 200 mM and 500 mM of imidazole in wash buffer. Left panel; (A) pIII D1 pull down with His_6-TolA-C. His_6-TolA-C is immobilized on the Ni-NTA beads and eluted with 200 mM imidazole. (B) His_6-cleaved pIII D1 is washed from the column and not immobilied. (C) pIII D1 is captured by His_6-TolA-C. (D) Right panel; TolA-C pull down with His_6-pIII D1. His_6-pIII D1 is immobilized on Ni-NTA beads and eluted with 200 mM imidazole. (E) His_6-cleaved TolA-C is washed from the column and not immobilized. (F) TolA-C is captured by His_6-pIII D1.

4.3. His_6-TolA-C / N-pIII co-purification

Having shown that TolA-C and pIII D1 co-elute from a Ni-NTA column, we next purified a complex of the two proteins for crystallization studies. Thirty mg of purified His_6-TolA-C were loaded onto 2 ml of Ni-NTA beads in a gravity column followed by 30 mg of N-pIII. The complex was eluted with imidazone and loaded onto a size exclusion column. A single large peak was detected by UV light absorption at 280 nm for the fractions associated with protein around 30 MM, double the MM of either protein alone (Figure 4.3). This co-purification further demonstrates stable binding between the proteins, and indicates that the complex is a good candidate for complex crystal screening for X-ray crystallography.

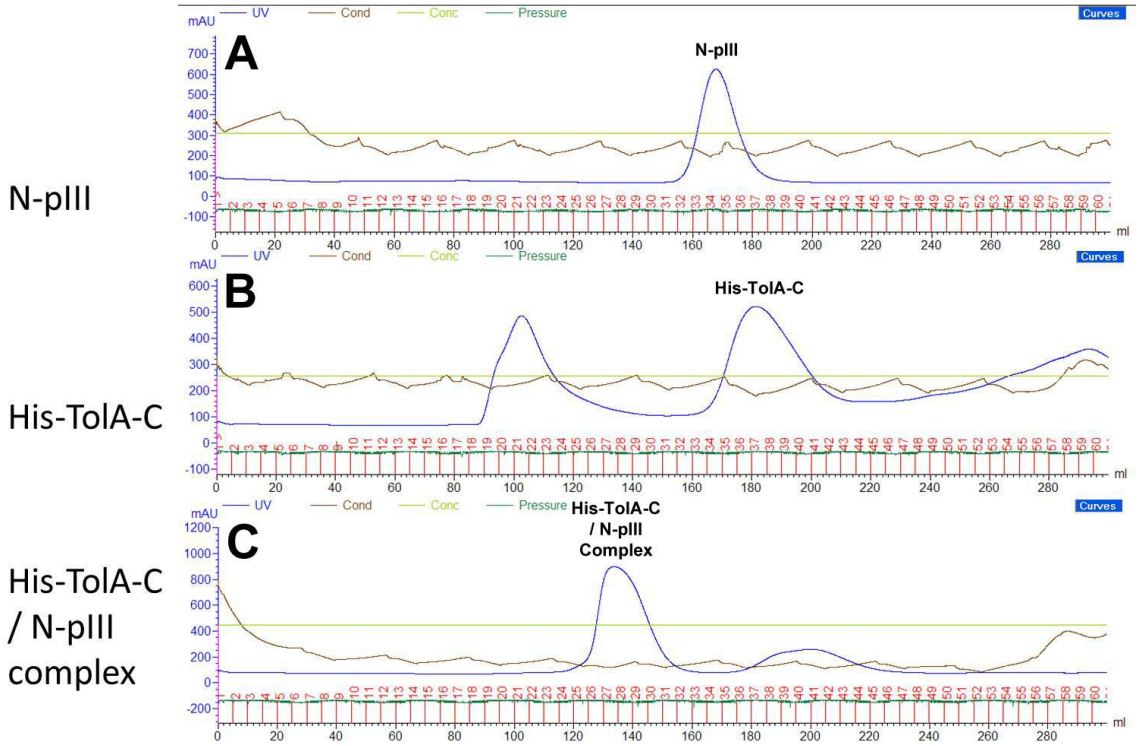


Figure 4.3. His_6-TolA-C and N-pIII co-elute from a size exclusion column

Elution profile from size exclusion chromatography column. The X axis indicates the volume of injected buffer at which point the protein elutes whereas the Y axis indicates the UV absorption of eluting proteins. (A) Purified N-pIII elution at ~ 170 ml. (B) Purified His_6-TolA-C elution at ~ 180 ml and impurities consistently found at ~ 100 ml due to non-specific binding to the Ni-NTA column. (C) His_6-TolA-C / N-pIII complex purification following pulldown of the above His_6-TolA-C and N-pIII proteins.

The purified His_6-TolA-C / N-pIII protein complex was concentrated to 27.1 mg/ml in protein buffer, flash frozen in liquid nitrogen and stored at -80 °C for crystal screening.

4.4. His_6-TolA-C / N-pIII complex crystal screening

To determine the molecular interactions between pIII and TolA we determined the X-ray crystal structure of the His_6-TolA-C / N-pIII complex. The purified His_6-TolA-C / N-pIII complex was screened against all the crystallization

screening conditions listed in Section 3.3. One condition, 25% PEG 6000, 100 mM MES pH 6.0, from the PEG 6000 vs pH screen generated in the lab, produced large reproducible crystals (Figure 4.4). We were unable to optimize further as the slightest variations to this conditions precluded crystal growth.

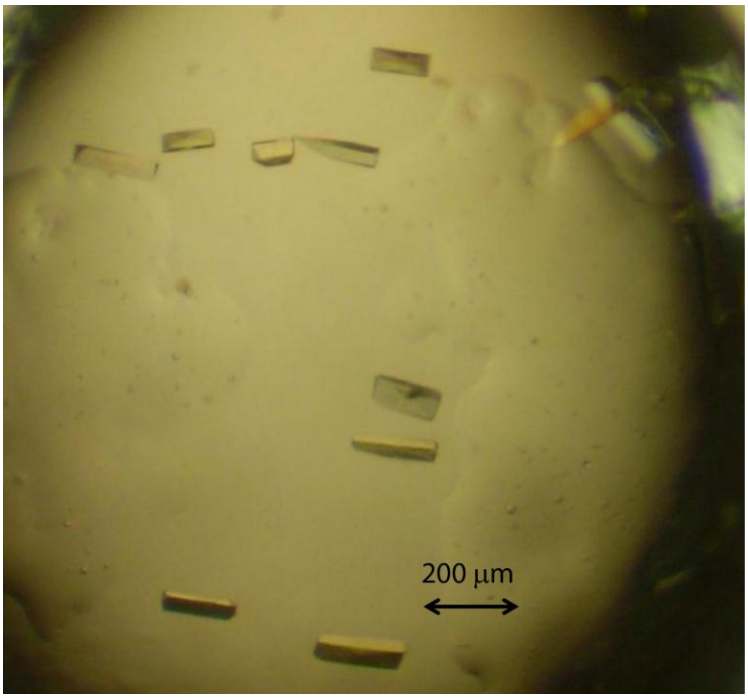


Figure 4.4. His_6-TolA-C / N-pIII crystals

His_6-TolA-C / N-pIII crystals in protein buffer, 25% PEG 6000 and 100 mM MES, pH 6.0

To freeze these crystals for shipping and X-ray diffraction we attempted to use cryo-protectants glycerol, ethylene glycol, PEG 4 000, dimethyl sulfoxide (DMSO) and 2-methyl-2,4-pentanediol (MPD), but slight changes in the drop condition resulted in immediate disintegration of the crystals. Thus, crystals were flash frozen in mother liquor, shipped to SSRL and data were collected as described in Sections 2.5 and 2.6.

4.5. Structure determination and refinement of the His_6-TolA-C / N-pIII complex

To solve the structure of the His_6-TolA-C / N-pIII complex, crystals were flash frozen in mother liquor and shipped to the Stanford Synchrotron Radiation Lightsource (SSRL). X-ray diffraction data were collected remotely on beamline 9-2. Using the N-pIII model, phases were determined for the N-pIII portion of the His_6-TolA-C / N-pIII complex by molecular replacement using PHASER (McCoy *et al.*, 2007). After proper orientation and positioning of N-pIII, phases were determined for TolA-C, using PDB 1TOL. The crystal belongs to orthorombic space group $p2_12_12_1$ with unit cell dimensions, 43.38, 46.16, 101.63 Å. There is 1 complex per asymmetric unit with a solvent content of 27.5% (Table 3). The data was resolved to 1.44 Å.

Table 3. Crystallographic data collection and refinement statistics for the His_6-TolA-C / N-pIII complex

	His-TolA-C / N-pIII
Data Collection	
Beamline	SSRL 9-2
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	43.38, 46.16, 101.63
α, β, γ (°)	90, 90, 90
Resolution (Å)	1.44
Wavelength (Å)	1.00
Completeness (%)	99.6 (95.5)
Observed reflections	257179
Unique reflections	37355
R_{sym} (%)	0.025 (0.163)
I/σ(I)	27.3 (4.4)
Mosaicity (°)	0.2
Refinement statistics	
Resolution limits (Å)	40.00 – 1.44
Z (Number of Molecules per Unit Cell)	4
R_{cryst} (%)	20.8
R_{free}(%)	21.5
Number of reflections used for refinement	35426
Number of atoms	
Protein	1728
Water	189
B-factor (Å^2)	
Protein	
Chain A	15.3
Chain B	14.9
Water	22.9
RMSD	
Bond lengths (Å)	0.005
Bond angles (°)	0.974
Ramachandran plot statistics (%)	
Favored	95.9
Allowed	4.1

aOverall/last shell

bR_{sym} is the unweighted R value on I between symmetry mates.

cR_{cryst}= Σhkl||Fobs(hkl)|-|Fcalc(hkl)||/ Σhkl|Fobs(hkl)|.

dR_{free} is the cross validation R factor for 5% of reflections against which the model was not refined.

4.6. His_6-TolA-C / N-pIII complex structure

In spite of only 29% sequence identity between the C-terminal domains of TolA from *E. coli* and *V. cholerae* their structures are very similar. *V. cholerae* TolA-C is an elongated globular domain with a very straight 24 residue N-terminal α-helix that interacts with an antiparallel 3 strand β-sheet flanked by 2 short α-helices. The N-terminal α-helix, α1, is followed by an irregular loop with a central type I turn. Next there is a β-hairpin comprising two of three β-sheet strands (β1 and β2). The polypeptide backbone exits the β-sheet at this point to form a 4-turn α-helix (α2), then a 3-turn α-helix (α3) separated by an extended loop. The C-terminal segment of TolA-C forms the third strand of the β-sheet, β3. Similarly *E. coli* TolA C-terminal domain begins with a 15 residue N-terminal α-helix followed by a very short 3_{10} helical motif. Next a β-hairpin leads to two smaller α-helices followed by the final strand of the antiparallel β-sheet (Lubkowski *et al.*, 1999).

N-pIII is essentially identical in the bound and unbound structures with 75% of main chain atoms (residues 16 - 89) producing an RMSD of 1.167 Å between bound N-pIII and molecule A of the unbound structure, 1.243 Å for molecule B and 0.608 Å for molecule C. The close RMSDs indicate that N-pIII undergoes no substantial conformational change upon binding to TolA-C. N-pIII interacts with TolA-C *via* three antiparallel strands of its β-barrel (β1, β5 and β6, together with their 3 extended loops). Together these β-strands form a concave

face that interacts with α2 and β2 of TolA-C, forming a continuous β-sheet between the two proteins. There is a large interface area between TolA-C and N-pIII with a buried surface area of 1489.6 $Å^2$ (Figure 4.5), calculated using PDBePisa on the European Bioinformatics Institute website (http://www.ebi.ac.uk/msd-srv/prot_int/). It is interesting to note that many of the residues responsible for the N-pIII β1 interaction with TolA β2 were the same residues generating the stable β1 to β3 interactions between molecule A and B of the unbound N-pIII asymmetric unit (Figure 4.5. inset).

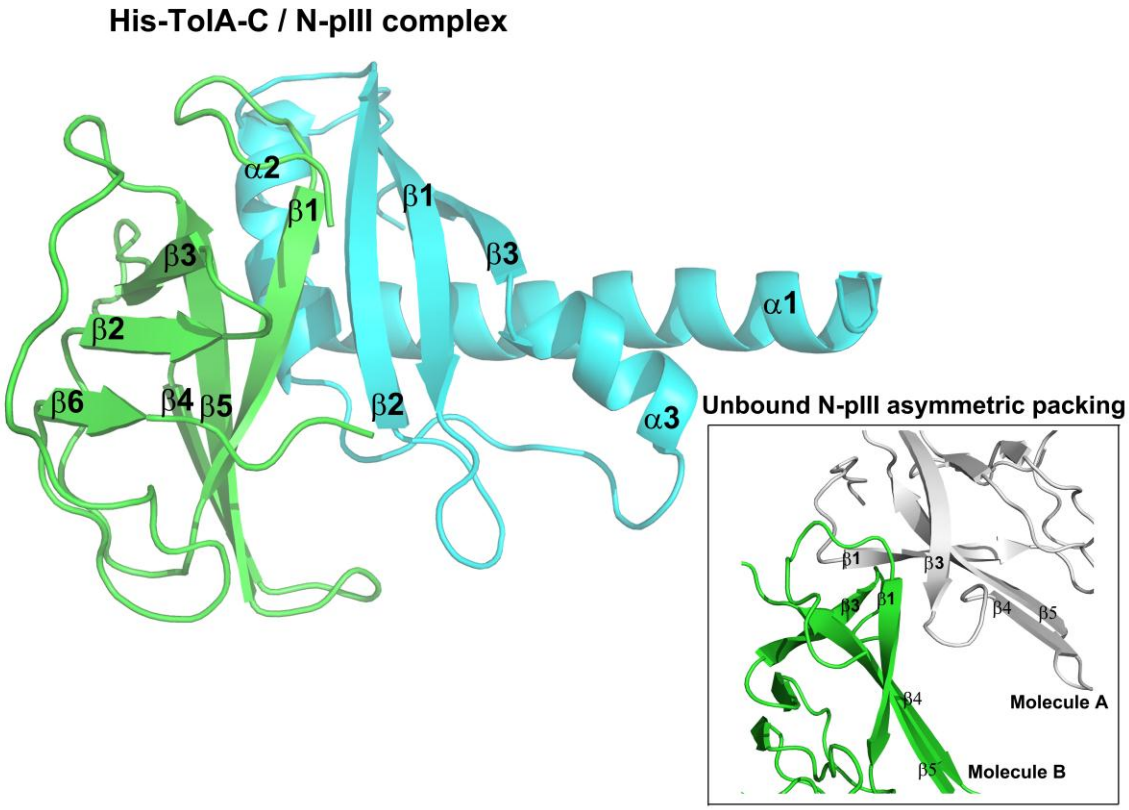


Figure 4.5. His_6-TolA-C / N-pIII complex structure

His_6-TolA-C / N-pIII continuous β sheet interaction. N-pIII (green) β strands and TolA-C (cyan) α-helices and β-strands are numbered at their N-terminal ends. Inset; Unbound N-pIII asymmetric crystal packing. Molecule A (grey) and molecule B (green) interact along β1 and β3.

Many interactions contribute to the N-pIII and TolA-C interface. B1 of N-pIII and β2 of TolA-C form several backbone hydrogen bonds (Figure 4.6 panel A) and a number of side chain hydrogen bonds (Figure 4.6 panel B). As the N-pIII β sheet folds around TolA, hydrogen bonds, salt bridges and van der Waals forces between N-pIII β1, β4 and β5 and their associated loops, and TolA-C β2 and α3 and their associated loops contribute to the interface (Table 4) (Figure 4.6 panel C). Val49 and Leu64 fit into a hydrophobic pocket in TolA created by Leu307, Leu310, Val327, Ala328 and the aliphatic side chain of Arg324 (Figure 4.6 panel D). Together, these interactions form the bulk of the interface between N-pIII and TolA-C.

Table 4. Salt bridge and hydrogen bond formation between N-pIII and TolA-C

Interaction	N-pIII Residue	TolA-C Residue	D1 (M13)	TolA (*e. coli*)
Hydrogen Bond	Thr15 BB	Val312 BB	Glu42 BB	Lys415 BB
	Thr15 SC	Val312 BB	Val 43 BB	Asp 413 BB
	Ser17 BB	Leu310 BB		
	Gln19 BB	Gly308 BB		
	Thr16 SC	Thr311 BB		
	Gly19 SC	Leu307 BB		
	Lys62 SC	Ala328 SC		
	Phe48 BB	Arg325 SC		
Salt Bridge	Glu42 SC	Arg324 SC	Glu50 SC	Arg365 SC
	Asp44 SC	Arg325 SC	Lys22 SC	Thr364 SC
			Asp28 SC	Lys415 SC
			Arg29 SC	Asp413 SC
BB = Back bone interaction SC = Side chain interaction				

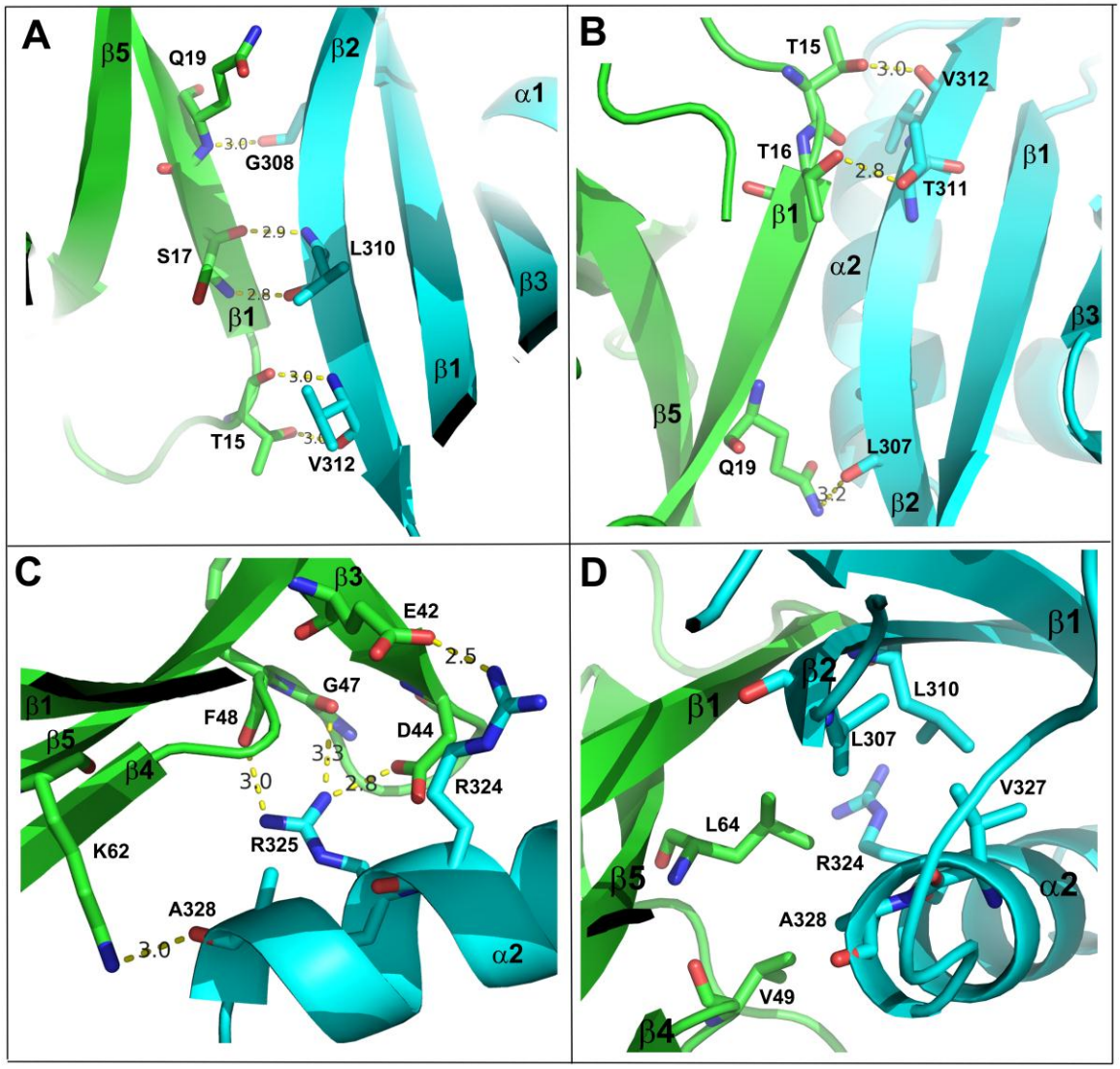


Figure 4.6. His_6-TolA-C / N-pIII complex interactions

His_6-TolA-C (cyan) and N-pIII (green) interactions. (A) Back bone hydrogen bonds and (B) side chain interactions between N-pIII β1 and TolA-C β2. (C) Side chain interactions between N-pIII β3, β4 and β5 and TolA-C α3. (D) Hydrophobic and van der Waals interactions between β1, β4 and β5 of N-pIII and β2 and α3 of TolA-C.

4.7. Comparison of $pIII_{CTX}$, $pIII_{fd}$, $pIII_{M13}$ and $pIII_{IF1}$ interactions with TolA

The *E. coli* TolA-C structure has been solved in complex with the N-terminal domains of the pIII minor coat proteins of M13 ($pIII_{M13}$) and IFI ($pIII_{IF1}$) filamentous phage (Lubkowski *et al.*, 1999, Lorenz *et al.*, 2011). These structures were found to be very similar and they shared an almost identical pIII-TolA interface. Alignment of the main chain atoms of TolA-C from the $pIII_{CTX}$ complex with those of TolA-C from the $pIII_{M13}$ complex produces an RMSD of 1.88 Å, indicating highly similar structures. As discussed in Chapter 3, despite low 15% sequence identity with $pIII_{M13}$ D1, N-$pIII_{CTX}$ shares a common topology with the pIII proteins solved to date. Yet remarkably, $pIII_{CTX}$ binds to the opposite side of TolA, compared with M13 and IFI pIII (Figure 4.7). It's interesting that all the pIII proteins form an extended 6 strand antiparallel β sheet with the TolA β-sheet. Whereas β4 of $pIII_{M13}$ or $pIII_{IF1}$ interacts with β3 along one edge of the TolA β-sheet and with the C-terminal end of α1, β1 of the $pIII_{CTX}$ β-sheet interacts with β2 on the opposite edge of the TolA β-sheet and with α2. Thus, the pIII proteins infecting *V. cholerae* and *E. coli* use completely different surfaces on the structurally similar pIII proteins to bind to distinct interaction sites on the structurally similar TolA proteins.

.

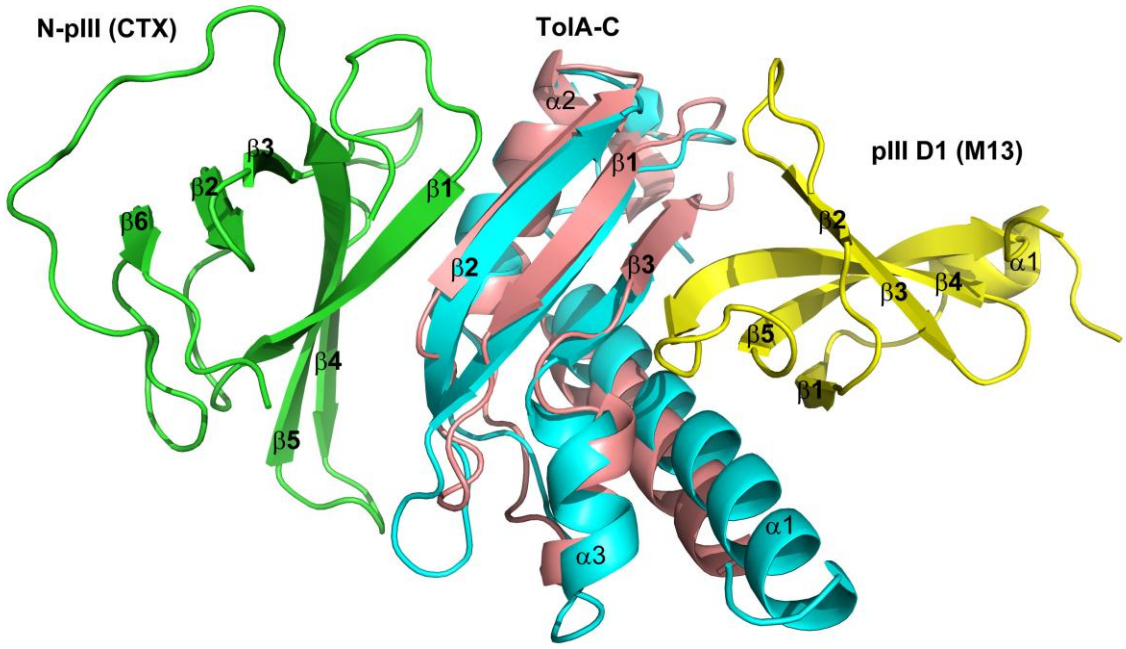


Figure 4.7. N-pIII$_{CTX}$ and pIII$_{M13}$ D1 binding to TolA

CTXΦ N-pIII (green) and M13 pIII D1 (yellow) bound to *V. cholerae* TolA-C (blue) or *e. coli* TolA-C (salmon) respectively.

4.8. Conclusion

In summary, we demonstrated a stable interaction between TolA-C and N-pIII based on a pull down assay and co-purification. To improve on our screening range we designed new screening kits for the lab, which allowed me to crystallize the TolA-C / N-pIII complex and solve its structure by molecular replacement techniques of X-ray crystallography. N-pIII$_{CTX}$ has now been shown to directly bind to TolA and the crystal structures show that N-pIII$_{CTX}$ shares a very similar topology to the almost identical pIII$_{M13}$ D1, pIII$_{fd}$ D1 and pIII$_{IF1}$ D1 topologies yet, interestingly, N-pIII$_{CTX}$ uses a unique pocket to bind to a unique site on TolA.

5. General Discussion and Conclusion

Ff phage is well characterized for phage display and molecular biology technologies and studies have suggested that a similar infection mechanism may apply to CTXΦ phage infection of *V. cholerae*. We have provided structural evidence that the CTXΦ phage, as with fd phage infection, binds to TolA but uses a different binding mechanism. An understanding of this mechanism can help to provide targets for treating and preventing cholera disease as well as a way to target *V. cholerae* cells specifically.

We used binding assays and X-ray crystallography techniques to determine the molecular interactions between pIII and TolA at an atomic level. We found that although N-pIII$_{CTX}$ and pIII$_{M13}$ have a low sequence identity and N-pIII$_{CTX}$ was predicted to have a predominantly α-helical secondary structure, the N-pIII$_{CTX}$ structure can be aligned very well with the 3 known structures of pIII proteins found in filamentous phage that infect *E. coli*. Most fascinating was our finding that although all these pIII proteins have the same function in phage uptake and have very similar topologies, pIII$_{CTX}$ binds to the opposite side of the β-sheet in the C-terminal domain of TolA and that pIII$_{CTX}$ uses the opposite side of its β-sheet for these binding interactions.

Binding of pIII to opposing sides of TolA with no conformational change in the TolA C-terminal domain or pIII N-terminal domain supports a model for phage uptake whereby pIII binding to TolA leaves the structure of TolA relatively unaffected so that TolA can act as a sort of fulcrum to induce unfolding of domains 2 and 3 of the pIII protein. Unfolded or stretched pIII exposes a portion of the pIII C-terminal domain required for insertion into the inner membrane. Insertion of the C-terminal domain of pIII could subsequently lead to insertion of pVIII into the inner membrane and injection of the ssDNA, assisted by the TolQRA complex, into the cytoplasm (Bennett & Rakonjac, 2006). As pIII are found in ~ 5 copies grouped together at one end of the phage (Gray *et al*. 1981, Lopez & Webster, 1982), it is also possible that more than one interaction is essential for translocation of the phage DNA across the inner membrane. Two to five TolQRA complexes may cluster, or be recruited upon initial binding, and interact with the same number of pIII proteins. The TolQRA function would be to peel apart the pIII proteins, the tension of TolA pulling pIII in opposing directions may be essential to uncoat the phage at the inner membrane (Figure 5.1).

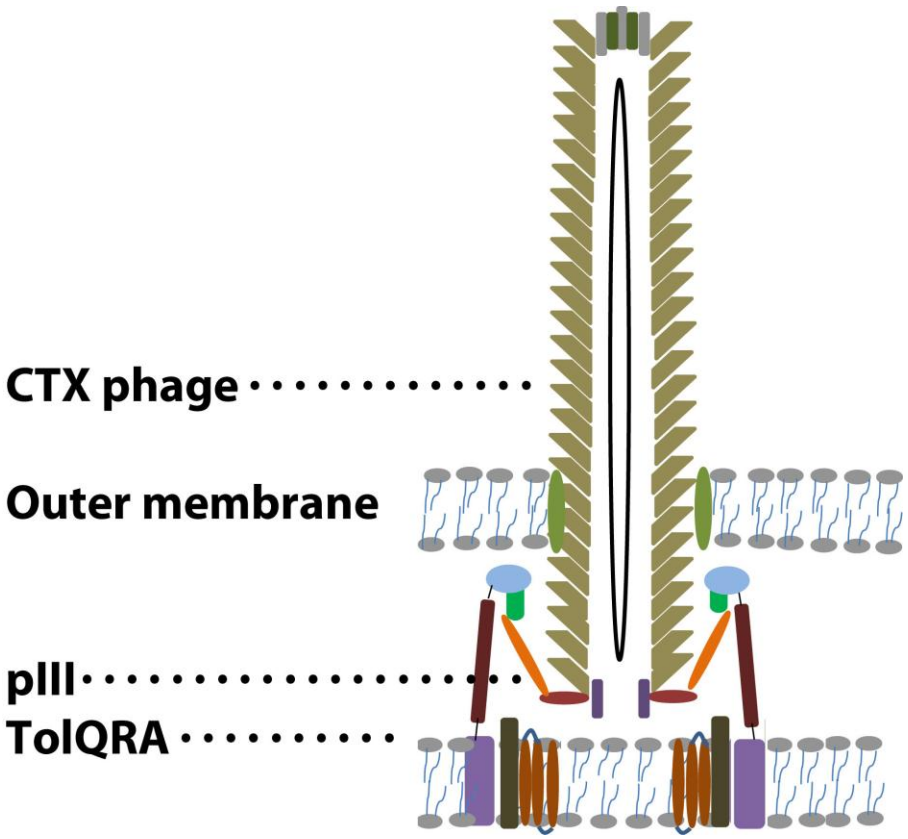


Figure 5.1. Proposed TolA function in CTXΦ phage uptake

Once domain 1 of up to 5 pIII proteins are bound up to TolQRA complexes, TolA could act as a leaver to peel the pIII proteins and subsequently the pVIII proteins apart, facilitating their insertion into the membrane.

Ours and other studies have shown that CTXΦ and Ff phage pIII proteins bind TolA (Hecht *et al.* 2010, Housden *et al.* 2005, Lazdunski et al., 1998). pIII proteins on their own may be used in drug therapy to target gram negative bacteria with high specificity. An N-$pIII_{CTX}$-$pIIID1_{M13}$ fusion protein may be capable of binding both *V. cholerae* and *E. coli* TolA. TolA is essential for outer membrane integrity and its C-terminal domain has been shown to interact with periplasmic proteins TolB and Pal (Bonsor *et al.* 2009). It's possible that exposure to large amounts of pIII may harm Gram negative bacteria by preventing TolA:TolB interactions or rendering the bacteria susceptible to

antibiotics such as vancomycin, which disrupt peptidoglycan synthesis. More interesting would be to include N-pIII with vancomycin or β-lactam antibiotics to compound their destabilizing effects on the bacterial cell wall. It would be interesting to see if N-pIII attached to the pore forming C-terminal domain or the TolA unfolding domain of colicin A proteins would target and kill bacterial cells while bypassing the complex outer membrane entry mechanism required for colicin toxicity (Housden *et al*., 2005, Hecht *et al*., 2010)

There are many ways in which this research should be continued. Due to the drastic decrease in phage infectivity in the absence of the central domain of pIII (Heilpern & Waldor, 2003), it seems the central domain plays a crucial role in bacterial specificity and infection. Soluble pIII constructs containing the central domain could be used for structural analysis by X-ray crystallography and functional assays, such as pull down assays with TcpA or isothermal calorimetry with TCP. Overcoming insolubility of pIII proteins containing the central domain could further elucidate CTXΦ phage uptake. My first approach would be to utilize many of the constructs created recently but not thoroughly tested. There were several newly designed constructs with N and C-terminal truncations that were never tested for solubility, which may prove to be good candidates (Appendix D). These experiments would examine the intramolecular interactions in domains 1, 2 and 3 of pIII.

As we have shown here that despite low sequence identities D1 of CTXΦ and M13 phage have very similar topologies, domain 2 may be similar as well.

Using the solved structure of D1 and D2 of $pIII_{M13}$ as a template we could identify and delete larger, more flexible loops that may be causing instability in D2 of $pIII_{CTX}$. We could substitute residues in areas of interest with cysteines for crosslinking experiments to test for accessibility and to support tertiary structure predictions. Introducing a disulfide bond between domains 1 and 2 could stabilize a $pIII_{CTX}$ protein containing D1 and D2. Introducing cysteines in the TolA binding domain of N-pIII and testing for crosslinking with TolA can indicate whether N-pIII is always accessible, as is the case in $pIII_{IF1}$, or if the TolA binding surface of N-pIII is buried until TCP binding, as is the case in $pIII_{M13}$ and $pIII_{fd}$. These studies could be supplemented by binding experiments between phage to pili. Using surface plasmon resonance we could fix purified TCP to a chip and CTXΦ phage, containing full length or N-terminally truncated pIII, would be run over the chip. In this experiment we could compare TCP binding to wild type pIII and N-terminally truncated pIII mutants designed to contain domains 2 and 3 only and even domain 3 only. Other deletions could subsequently be made to narrow in on the pIII TCP binding domain. It may even be found that without the N-terminal domain pIII is unable to bind TCP.

Current models for filamentous phage infection of *E. coli* suggest phage pIII proteins bind tip to tip with bacterial F-pilus. Pilus retractand bring pIII D1 in contact with TolA (Bennett & Rakonjac, 2006, Deng *et al.*, 1999, Jacobson, 1972). There is no evidence to date to describe how pIII crosses the outer membrane to initiate contact with TolA. Because of this and as TCP have never

been shown to retract, determining the role of the outer membrane secretin protein TcpC in phage infection could identify how phage are able to cross the outer membrane. Phage infection studies by myself and others in the lab show that *tcpA* knockout *V. cholerae*, which are unable to produce TCP, are infected by CTXΦ phage at a rate roughly 3 orders of magnitude lower than wild type *V. cholerae*, consistent with previously reported data (Heilpern & Waldor, 2003), illustrating the central role of TCP in CTXΦ phage infection. Phage infection in the absence of TCP allows us to study the role of TcpC in phage infection without attributing a possible loss of infection to lack of TCP. Generation of a TcpC/TcpA double knockout followed by phage infection assays could help to determine whether phage gain access to the periplasm through the TcpC outer membrane secretin, quite possibly by TCP retraction.

It would also be interesting to visualize binding interactions between phage and TCP. Previously our lab has raised antibodies against a peptide fragment of pVIII and visualized interactions with biotinylated TCP by electron microscopy of gold labelled secondary antibodies. The studies suggested phage and TCP interact lengthwise and not tip to tip or as well as tip to tip. These micrographs showed TCP and CTXΦ phage have very similar appearances of relatively smooth featureless surfaces. It would be interesting to show that these surfaces interact non-specifically but that the critical interaction is between TCP and the pIII end of CTXΦ phage. Antibodies raised against purified N-pIII should be able to clearly label the pIII end of CTXΦ phage to demonstrate the presence

or absence of tip to tip binding and to visualize lengthwise interactions. If the end of CTXΦ phage containing pIII was found to be proximal to the bacteria along TCP this may suggest that pIII use non-specific binding along the smooth surfaces of TCP and phage to localize to the outer membrane. If, on the other hand, the pIII containing end of CTXΦ phage was only found associated with the tip of TCP, this would demonstrate tip to tip binding and a strong argument could be made for TCP retraction.

Using fluorescently labelled R17 bacteriophages, F-pilus were shown to retract by confocal microscopy (Daehnel *et al.*, 2005, Clarke *et al*., 2008). Alexa Fluor® Succinimidyl Esters (Invitrogen) conjugates to primary amines of proteins and can be used to label purified bacteriophage. Labelling CTXΦ phage in a similar manner could serve to visualize uptake of phage and could possibly also show TCP retraction. An extension of these fluorescence studies could be to label phage containing N-terminally truncated pIII proteins, engineered to contain only domains 2 and 3. These studies could demonstrate binding of phage to TCP without phage infection.

As it is well established that *V. cholerae* continues to affect millions of people every year, that CTXΦ phage uptake confers toxicity to non-pathogenic *V. cholerae* and that pIII is responsible for phage uptake, these and other studies are crucial in contributing to our understanding of *vibrio* evolution and may reveal strategies for anti-*vibrio* therapeutics.

References

Armstrong, J., Perham, R. N., and Walker, J. E. (1981). Domain structure of bacteriophage fd adsorption protein. FEBS Lett. 135, 167±172.

Bennett, N. J., Gagic, D., Sutherland-Smith, A. J., & Rakonjac, J. (2011). Characterization of a dual-function domain that mediates membrane insertion and excision of Ff filamentous bacteriophage. [Research Support, Non-U.S. Gov't]. J Mol Biol, 411(5), 972-985. doi: 10.1016/j.jmb.2011.07.002

Bennett, N. J., & Rakonjac, J. (2006). Unlocking of the filamentous bacteriophage virion during infection is mediated by the C domain of pIII. [Research Support, Non-U.S. Gov't]. J Mol Biol, 356(2), 266-273. doi: 10.1016/j.jmb.2005.11.069

Bonsor DA, Hecht O, Vankemmelbeke M, Sharma A, Krachler AM, Housden NG, Lilly KJ, James R, Moore GR, Kleanthous C. (2009). Allosteric beta-propeller signalling in TolB and its manipulation by translocating colicins. EMBO J. 2009 Sep 16;28(18):2846-57. Epub 2009 Aug 20.

Braun, V., & Herrmann, C. (1993). Evolutionary relationship of uptake systems for biopolymers in Escherichia coli: cross-complementation between the TonB-ExbB-ExbD and the TolA-TolQ-TolR proteins. [Comparative Study Research Support, Non-U.S. Gov't]. Mol Microbiol, 8(2), 261-268.

Chaignat, C. L., Monti, V., Soepardi, J., Petersen, G., Sorensen, E., Narain, J., & Kieny, M. P. (2008). Cholera in disasters: do vaccines prompt new hopes? [Research Support, Non-U.S. Gov't]. Expert Rev Vaccines, 7(4), 431-435. doi: 10.1586/14760584.7.4.431

Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. J. Bacteriol. 179, 6464–6471.

Click, E. M. & Webster, R. E. (1998). The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage f1 during infection. J. Bacteriol. 180, 1723–1728.

Clarke, M., Maddera, L., Harris, R. L., & Silverman, P. M. (2008) F-pili dynamics by live-cell imaging. 17978–17981 PNAS November 18, 2008 vol. 105 no. 46

Colwell, R. R., & Huq, A. (1994). Environmental reservoir of Vibrio cholerae. The causative agent of cholera. [Research Support, U.S. Gov't, Non-P.H.S. Review]. Ann N Y Acad Sci, 740, 44-54.

Colwell, R., Kaper, J., Joseph. S. W. Vibrio cholerae, Vibrio parahaemolyticus, and other Vibrios: occurrence and distribution in Chesapeake Bay. Science 198: 394-396 (1977).Craig, L., Pique, M. E., & Tainer, J. A. (2004). Type IV pilus structure and bacterial pathogenicity. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review]. Nat Rev Microbiol, 2(5), 363-378. doi: 10.1038/nrmicro885

Craig, L., Li, J., Type IV pili: paradoxes in form and function. (2008) Current opinion in structural Biology 18:267-277

Craig L, Pique ME, Tainer JA. Type IV pilus structure and bacterial pathogenicity. (2004). Nat Rev Microbiol. May;2(5):363-78.

Craig L, Taylor RK, Pique ME, Adair BD, Arvai AS, Singh M, Lloyd SJ, Shin DS, Getzoff ED, Yeager M, Forest KT, Tainer JA. (2003). Type IV pilin structure and assembly: X-ray and EM analyses of Vibrio cholerae toxin-coregulated pilus and Pseudomonas aeruginosa PAK pilin. Mol Cell. 2003 May;11(5):1139-50.

Daehnel K, et al. (2005) Fluorescence assays for F-pili and their application. Microbiology 151:3541–3548.

Davis, N. G., Boeke, J. D., and Model, P. (1985). Fine structure of amembrane anchor domain. J. Mol. Biol. 181, 111±121.

Davis, B. M., Moyer, K. E., Boyd, E. F., & Waldor, M. K. (2000). CTX prophages in classical biotype Vibrio cholerae: functional phage genes but dysfunctional phage genomes. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. J Bacteriol, 182(24), 6992-6998.

Davis, B. M., & Waldor, M. K. (2000). CTXΦ contains a hybrid genome derived from tandemly integrated elements. PNAS, 97(15), 8572-8577.

Davis, B. M., & Waldor, M. K. (2003). Filamentous phages linked to virulence of Vibrio cholerae. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review]. Curr Opin Microbiol, 6(1), 35-42.

Deng, L. W., Malik, P., & Perham, R. N. (1999). Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli. [Research Support, Non-U.S. Gov't]. Virology, 253(2), 271-277. doi: 10.1006/viro.1998.9509

Deng, L. W., & Perham, R. N. (2002). Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli. [Research Support, Non-U.S. Gov't]. J Mol Biol, 319(3), 603-614. doi: 10.1016/S0022-2836(02)00260-7

Deprez, C., Lloubes, R., Gavioli, M., Marion, D., Guerlesquin, F., & Blanchard, L. (2005). Solution structure of the E.coli TolA C-terminal domain reveals conformational changes upon binding to the phage g3p N-terminal domain. [Research Support, Non-U.S. Gov't]. J Mol Biol, 346(4), 1047-1057. doi: 10.1016/j.jmb.2004.12.028

Eckert, B., & Schmid, F. X. (2007). A conformational unfolding reaction activates phage fd for the infection of Escherichia coli. [Research Support, Non-U.S. Gov't]. J Mol Biol, 373(2), 452-461. doi: 10.1016/j.jmb.2007.07.060

Faruque, S. M., Naser, I. B., Islam, M. J., Faruque, A. S. G., Ghosh, A. N., Balakrish, G., Nair, Sack, D. A. & Mekalanos, J. J., Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages, PNAS, 102(5), 1702-1707.

Fan, Q. R., Mosyak, L., Winter, C. C., Wagtmann, N., Long, E. O., & Wiley, D. C. (1997). Structure of the inhibitory receptor for human natural killer cells resembles haematopoietic receptors. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. Nature, 389(6646), 96-100. doi: 10.1038/38028

Ganin, V. S. (2009). [Cholera and war]. [Historical Article]. Voen Med Zh, 330(9), 83-88.

Germon, P., Ray, M. C., Vianney, A., & Lazzaroni, J. C. (2001). Energy-dependent conformational change in the TolA protein of Escherichia coli involves its N-terminal domain, TolQ, and TolR. [Research Support, Non-U.S. Gov't]. J Bacteriol, 183(14), 4110-4114. doi: 10.1128/JB.183.14.4110-4114.2001

Grant RA, Lin TC, Konigsberg W, Webster RE. (1981). Structure of the filamentous bacteriophage fl. Location of the A, C, and D minor coat proteins. J Biol Chem. 1981 Jan 10;256(1):539-46

Gray CW, Brown RS, Marvin DA. (1981) Adsorption complex of filamentous fd virus. J Mol Biol. 1981 Mar 15;146(4):621-7.

Hecht O, Zhang Y, Li C, Penfold CN, James R, Moore GR. Characterisation of the interaction of colicin A with its co-receptor TolA. FEBS Lett. 2010 Jun 3;584(11):2249-52. Epub 2010 Apr 29.

Heilpern, A. J., & Waldor, M. K. (2000). CTXphi infection of Vibrio cholerae requires the tolQRA gene products. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. J Bacteriol, 182(6), 1739-1747.

Heilpern, A. J., & Waldor, M. K. (2003). pIIICTX, a predicted CTXphi minor coat protein, can expand the host range of coliphage fd to include Vibrio cholerae. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. J Bacteriol, 185(3), 1037-1044.

Holliger, P., Riechmann, L., & Williams, R. L. (1999). Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 A: evidence for conformational lability. J Mol Biol, 288(4), 649-657. doi: 10.1006/jmbi.1999.2720

Housden, N. G., Loftus, S. R., Moore, G. R., James, R., & Kleanthous, C. (2005). Cell entry mechanism of enzymatic bacterial colicins: porin recruitment and the thermodynamics of receptor binding. [Research Support, Non-U.S. Gov't]. Proc Natl Acad Sci U S A, 102(39), 13849-13854. doi: 10.1073/pnas.0503567102

Huff JP, Grant BJ, Penning CA, Sullivan KF. (1990). Optimization of routine transformation of Escherichia coli with plasmid DNA. Biotechniques. 1990 Nov;9(5):570-2, 574, 576-7.

Im W, Brooks CL 3rd (2004). De novo folding of membrane proteins: an exploration of the structure and NMR properties of the fd coat protein. J Mol Biol. 2004 Mar 26;337(3):513-9.

Jacobson A., (1972). Role of F Pili in the Penetration of Bacteriophage fl. J. Virol. 1972, 10(4):835.

Karlsson F, Malmborg-Hager AC, Borrebaeck CA. Escherichia coli TolA tolerates multiple amino-acid substitutions as revealed by screening randomized variants for membrane integrity and phage receptor function. FEMS Microbiol Lett. 2006 Jun;259(1):81-8.

Keasler, S. P., & Hall, R. H. (1993). Detecting and biotyping Vibrio cholerae O1 with multiplex polymerase chain reaction. [Letter]. Lancet, 341(8861), 1661.

Kirn TJ, Jude BA, Taylor RK. (2005) A colonization factor links Vibrio cholerae environmental survival and human infection. Nature. 2005 Dec 8;438(7069):863-6.

Lazdunski, C. J., Bouveret, E., Rigal, A., Journet, L., Lloubes, R., & Benedetti, H. (1998). Colicin import into Escherichia coli cells. [Review]. J Bacteriol, 180(19), 4993-5002.

Lopez J, Webster RE. (1982). Minor coat protein composition and location of the A protein in bacteriophage f1 spheroids and I-forms. J Virol. 1982 Jun;42(3):1099-107.

Lorenz, S. H., Jakob, R. P., Weininger, U., Balbach, J., Dobbek, H., & Schmid, F. X. (2011). The filamentous phages fd and IF1 use different mechanisms to infect Escherichia coli. [Research Support, Non-U.S. Gov't]. J Mol Biol, 405(4), 989-1003. doi: 10.1016/j.jmb.2010.11.030

Lorenz, S. H., & Schmid, F. X. (2011). Reprogramming the infection mechanism of a filamentous phage. [Research Support, Non-U.S. Gov't]. Mol Microbiol, 80(3), 827-834. doi: 10.1111/j.1365-2958.2011.07617.x

Lubkowski, J., Hennecke, F., Pluckthun, A., & Wlodawer, A. (1998). The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. Nat Struct Biol, 5(2), 140-147.

Lubkowski, J., Hennecke, F., Pluckthun, A., & Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. Structure, 7(6), 711-722.

Martin, A., & Schmid, F. X. (2003). A proline switch controls folding and domain interactions in the gene-3-protein of the filamentous phage fd. [Research Support, Non-U.S. Gov't]. J Mol Biol, 331(5), 1131-1140.

Megli CJ, Yuen AS, Kolappan S, Richardson MR, Dharmasena MN, Krebs SJ, Taylor RK, Craig L. (2011). Crystal structure of the Vibrio cholerae colonization factor TcpF and identification of a functional immunogenic site. J Mol Biol. 2011 Jun 3;409(2):146-58. Epub 2011 Apr 1

Mekalanos, J. J. (1983). Duplication and amplification of toxin genes in Vibrio cholerae. [Comparative Study Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. Cell, 35(1), 253-263.

Mooi, F. R., & Bik, E. M. (1997). The evolution of epidemic Vibrio cholerae strains. [Research Support, Non-U.S. Gov't Review]. Trends Microbiol, 5(4), 161-165. doi: 10.1016/S0966-842X(96)10086-X

Nelson, E. J., Chowdhury, A., Flynn, J., Schild, S., Bourassa, L., Shao, Y., LaRocque, R. C., Calderwood, S. B., Qadri, F., Camilli, A., Transmission of Vibrio cholerae is antagonized by lytic phage and entry into the aquatic environment. PLoS Pathog. 2008 Oct;4(10):e1000187. Epub 2008 Oct 24.

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

Pollastri G, McLysaght A. (2005) Porter: a new, accurate server for protein secondary structure prediction. Bioinformatics. 2005 Apr 15;21(8):1719-20. Epub 2004 Dec 7.

Rakonjac, J., Feng, J., & Model, P. (1999). Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. J Mol Biol, 289(5), 1253-1265. doi: 10.1006/jmbi.1999.2851

Raychoudhuri, A., Mukhopadhyay, A. K., Ramamurthy, T., Nandy, R. K., Takeda, Y., & Nair, G. B. (2008). Biotyping of Vibrio cholerae O1: time to redefine the scheme. [Research Support, Non-U.S. Gov't Review]. Indian J Med Res, 128(6), 695-698.

Righetto, L., Casagrandi, R., Bertuzzo, E., Mari, L., Gatto, M., Rodriguez-Iturbe, I., & Rinaldo, A. (2012). The role of aquatic reservoir fluctuations in long-term cholera patterns. Epidemics, 4(1), 33-42. doi: 10.1016/j.epidem.2011.11.002

Sasmal D, Guhathakurta B, Bhattacharya SK, Pal CR, Datta A. (1996). Role of cell-associated N-acetyl-D-glucosamine specific haemagglutinin in the adhesion of Vibrio cholerae O1 to rabbit intestinal epithelial cells in vitro. FEMS Immunol Med Microbiol. 1996 Feb;13(2):101-5.

Sanchez, J., & Holmgren, J. (2008). Cholera toxin structure, gene regulation and pathophysiological and immunological aspects. [Review]. Cell Mol Life Sci, 65(9), 1347-1360. doi: 10.1007/s00018-008-7496-5

Sergei Radaev & Peter D. Sun. (2002). Crystallization of protein±protein complexes. J. Appl. Cryst, 35, 674-676.

Waldor, M. K., & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. [Comment Research Support, U.S. Gov't, P.H.S.]. Science, 272(5270), 1910-1914.

Waldor, M. K., Rubin, E. J., Pearson, G. D., Kimsey, H., & Mekalanos, J. J. (1997). Regulation, replication, and integration functions of the Vibrio cholerae CTXphi are encoded by region RS2. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. Mol Microbiol, 24(5), 917-926.

Wen, Z. Q., Overman, S. A. & Thomas, G. J., Jr (1997). Structure and interactions of the single-strandedDNA genome of filamentous virus fd: investigation by ultraviolet resonance Raman spectroscopy. Biochemistry,36, 7810–7820.

Witty, M., Sanz, C., Shah, A., Grossmann, J. G., Mizuguchi, K., Perham, R. N., & Luisi, B. (2002). Structure of the periplasmic domain of Pseudomonas aeruginosa TolA: evidence for an evolutionary relationship with the TonB transporter protein. [Research Support, Non-U.S. Gov't]. EMBO J, 21(16), 4207-4218. American Psychological Association (APA). (2010). *Publication manual of the American Psychological Association.* Washington, DC: Author.

Appendices

Appendix A. PCR primers

Table A1. Primers

Construct	Forward Primer	Reverse Primer	Template
pIII(D3)	GGGAATTC CATATG CCATCGGTAACGGCTTCCG	CGTA GGATCC TTAGTG CAG GTTTTCAGAAA AGAGGGAG	*pCTX-Km*
N-pIII	GGC TTC GAC TCT C TAA CA TGT CAG TGG TCA GG	CCT GAC CAC TGA CAT G TTA GA GAG TCG AAG CC	pIII(D3) plasmid
TolA-C	CGTA GAAGAC TA CTCG AAT GAT ATT TTT GGC AGC TTG AGT GAA G	CGTA GGATCC TTA TTC AGG TGC TAC GGT TAA ATT AAT ATT CTT TAG	*V. cholerae* O395
D1	GGGAATTC CATATG CCATCGGTAACGGCTTCCG	CGTA GGATCC TTA GCA CTC TTC CCC CTC AGG	*pCTX-Km*
D1-D2	GGGAATTC CATATG CCATCGGTAACGGCTTCCG	CGTA GGATCC TTA AAC GAC AGT CCC ACC GAG TC	*pCTX-Km*
D1-D3	GGGAATTC CATATG CCATCGGTAACGGCTTCCG	CGTA GGATCC TTA CGA AGG CAC CGC CGC CG	*pCTX-Km*
D2	GGGAATTC CATATG TCG TTT AAG TCT GCG TTC AAT CAG GTC	CGTA GGATCC TTA AAC GAC AGT CCC ACC GAG TC	pCTX-Km
D1Δ20	GGGAATTC CATATG TCC GCC ATC AAT TGT GAT CCT AAT ACT ACT ACG	CGTA GGATCC TTA GCA CTC TTC CCC CTC AGG	*pCTX-Km*
D1/TolA	D1 - GGGAATTC CATATG CCATCGGTAACGGCTTCCG TolA - CGTA GAAGAC TA CTCG AAT GAT ATT TTT GGC AGC TTG AGT GAA G	D1 - CGTA GAAGAC TA CGA GCT GCT CAG CCC ATC AAC AGG G TolA - CGTA GGATCC TTA TTC AGG TGC TAC GGT TAA ATT AAT ATT CTT TAG	*pCTX-Km / V. cholerae* O395
M1	CAG TCA ACA ATC GTT TGG GTG GGA TTG ATG TTT GGG ATA CGC AAC G	CGT TGC GTA TCC CAA ACA TCA ATC CCA CCC AAA CGA TTG TTG ACT G	pIII(D2) plasmid
M2	GTC AAC AAT CGT TTG GGT GGG CAA AAC GGA AAG ATT GAT GTT TGG GAT ACG C	GCG TAT CCC AAA CAT CAA TCT TTC CGT TTT GCC CAC CCA AAC GAT TGT TGA C	pIII(D2) plasmid

Construct	Forward Primer	Reverse Primer	Template
M3	GTC ACC GCA GTC AAC AAT CGT CAA AAC GGA AAG GTT TGG GAT ACG CAA CGG	CCG TTG CGT ATC CCA AAC CTT TCC GTT TTG ACG ATT GTT GAC TGC GGT GAC	pIII(D2) plasmid
M4	CAG TCA ACA ATC GTT TGG GTG GGA TTG ATG TTT GGG ATA CGC AAC G	CGT TGC GTA TCC CAA ACA TCA ATC CCA CCC AAA CGA TTG TTG ACT G	pIII(D1-D2) plasmid
M5	GTC AAC AAT CGT TTG GGT GGG CAA AAC GGA AAG ATT GAT GTT TGG GAT ACG C	GCG TAT CCC AAA CAT CAA TCT TTC CGT TTT GCC CAC CCA AAC GAT TGT TGA C	pIII(D1-D2) plasmid
M6	GTC ACC GCA GTC AAC AAT CGT CAA AAC GGA AAG GTT TGG GAT ACG CAA CGG	CCG TTG CGT ATC CCA AAC CTT TCC GTT TTG ACG ATT GTT GAC TGC GGT GAC	pIII(D1-D2) plasmid
M8	CCA TGA ACA CGG CGC TAT AAG GGC GCT GTA G	CTA CAG CGC CCT TAT AGC GCC GTG TTC ATG G	pIII(D3) plasmid
M10	GTC ACC GCA GTC AAC AAT CGT CAA AAC GGA AAG GTT TGG GAT ACG CAA CGG	CCG TTG CGT ATC CCA AAC CTT TCC GTT TTG ACG ATT GTT GAC TGC GGT GAC	M8 plasmid
D2-D3.1	GGGAATTC CATATG GTT TGG GAT ACG CAA CGG GAG G	CGTA GGATCC TTAGTG CAG GTTTTCAGAAA AGAGGGAG	pIII(D3) plasmid
D2-D3.2	GGGAATTC CATATG TAC GAT GTG CTT TAT AGC GAG CTT AAT GTC C	CGTA GGATCC TTAGTG CAG GTTTTCAGAAA AGAGGGAG	pIII(D3) plasmid
D2-D3.3	GGGAATTC CATATG ACA GGG GCG CTT AAT ACC ATC TC	CGTA GGATCC TTAGTG CAG GTTTTCAGAAA AGAGGGAG	pIII(D3) plasmid

Appendix B. PCR and ligation conditions

Table B1. PCR components

Component	Volume added	Final concentration
2X KAPA HiFi HotStart ReadyMix	12.5 µl	1 X
Template DNA	0.5 µl	50 ng
Forward primer (10 µM)	0.75 µl	0.3 µM
Reverse primer (10 µM)	0.75 µl	0.3 µM
ddH_2O	10.5 µl	

Table B2. PCR conditions

Step	Temperature	Time	Number of cycles
Initial denaturation	95 °C	2 min	1
Denaturation	98 °C	20 sec	30 for PCR amplification 25 for site directed mutagenesis
Primer annealing	59 °C for PCR amplification 70 °C for site directed mutagenesis	30 sec	
Extension	72 °C	1 min for PCR amplification 7 min for site directed mutagenesis	
Final extension	72 °C	2 min	1
Storage	4 °C	Forever	

Table B3. Ligation components

Component	Amount added
T4 DNA Ligase (New England BioLabs)	1 µl
T4 DNA Ligase Reaction Buffer (10X)	1 µl

Component	**Amount added**
50 mM Tris-HCl, 10 mM MgCl2, 1 mM ATP, 10 mM Dithiothreitol, pH 7.5 @ 25°C	
Digested pET15b	100 ng
Digested PCR product	20 – 25 ng
ddH_2O	Top up to 10 µl

Appendix C. pET15b

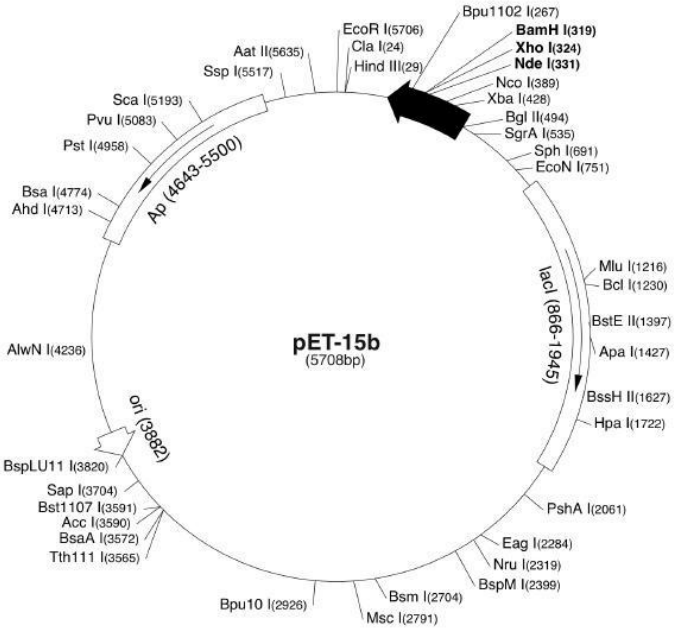
Novagen

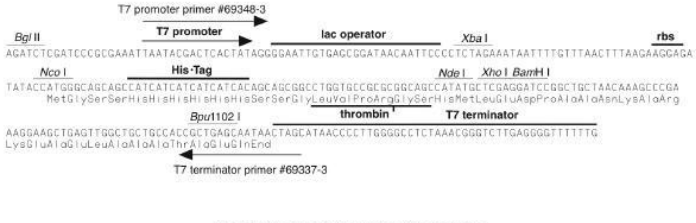
pET-15b Vector

TB045 5/99

The pET-15b vector (Cat. No. 69661-3) carries an N-terminal His•Tag® sequence followed by a thrombin site and three cloning sites. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below.

pET-15b sequence landmarks	
T7 promoter	463-479
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites	
(*Nde* I - *Bam* H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500





pET-15b cloning/expression region

Appendix D. Protein expression and solubility

Protein Name	Amino acid	Expression	Soluble	Notes
N-pIII	1-139	Rosetta-gami B (DE3)	Yes	AKA M7 and pIII(M7), solved structure
D1	1-102	Rosetta-gami B (DE3)	Yes	Crystals diffracted to ~ 3.5 Å
D1-D2	1-238	Rosetta-gami B (DE3)	No	Refolding tests unsuccessful
D1-D3	1-335	Rosetta-gami B (DE3)	No	Refolding tests unsuccessful
pIII(D3)	1-360	Rosetta-gami B (DE3)	No	
D1/TolA	pIII 1-120, TolA 241-356	Rosetta-gami B (DE3)	Yes	Crystal screening unsuccessful
TolA-C	241-356	BL21 (DE3)	Yes	Solved structure
D1Δ20	6-102	Rosetta-gami B (DE3)	Yes	Crystal screens
D2	120-238	BL21 (DE3)	No	All D2 only constructs require further testing
M1	120-176, 186-238	BL21 (DE3)	No	
M2	120-176, QNGK 186-238	BL21 (DE3)	No	
M3	120-173, QNGK 188-238	BL21 (DE3)	No	
M4	1-176, 186-238	Rosetta-gami B (DE3)	No	
M5	1-176, QNGK, 186-238	Rosetta-gami B (DE3)	No	
M6	1-173, QNGK, 188-238	Rosetta-gami B (DE3)	No	
M8	1-176, 186-284	Rosetta-gami B (DE3)	No	
M10	1-173, QNGK, 188-284	Rosetta-gami B (DE3)	No	
D2-D3.1		Rosetta-gami B (DE3)	Expression testing suggested soluble proteins	Further testing required
D2-D3.2		Rosetta-gami B (DE3)	Expression	Further testing required

			testing suggested soluble proteins
D2-D3.3		Rosetta-gami B (DE3)	Untested
M7Δ20	6-139	Rosetta-gami B (DE3)	Untested

Appendix E. crystal growth conditions

Protein	Concentration	Buffer	Well Conditions	Additive/Detergent
N-pIII	32.5 mg/ml	Purification	20% (w/v) PEG-1000, phosphate-citrate pH 4.2, 0.2 M Li_2SO_4	
N-pIII SeMet	23.5 mg/ml	Purification	19% (w/v) PEG-1000, phosphate-citrate pH 4.2, 150 mM Li_2SO_4	
	23.5 mg/ml	Purification	17% (w/v) PEG-1000, phosphate-citrate pH 4.2, 225 mM Li_2SO_4	
	23.5 mg/ml	Purification	15% (w/v) PEG-1000, phosphate-citrate pH 4.2, 250 mM Li_2SO_4	
	23.5 mg/ml	Purification	15% (w/v) PEG-1000, phosphate-citrate pH 4.2, 275 mM Li_2SO_4	
	23.5 mg/ml	Purification	20% (w/v) PEG-1000, phosphate-citrate pH 4.2, 0.2 M Li_2SO_4	390.0 mM Nonanoyl-N-hydroxyethylglucamide
	23.5 mg/ml	Purification	20% PEG 3350, 11% Tacsimate pH 5.0	
	23.5 mg/ml	Purification	20% PEG 3350, 8% Tacsimate pH 5.0	
	23.5 mg/ml	Purification	23.5 mg/ml	Purification
	23.5 mg/ml	Purification	8% Tacsimate pH 5.0	18.0 mM n-Decyl-b-D-maltopyranoside
	23.5 mg/ml	Purification	8% Tacsimate pH 5.0	10% w/v ANAPOE®-X-114
	23.5 mg/ml	Purification	8% Tacsimate pH 5.0	395.0 mM FOS-Choline®-9
His_6-TolA-N-pIII complex	27.1 mg/ml	Protein	25% PEG 6000, 100 mM MES pH 6.0	

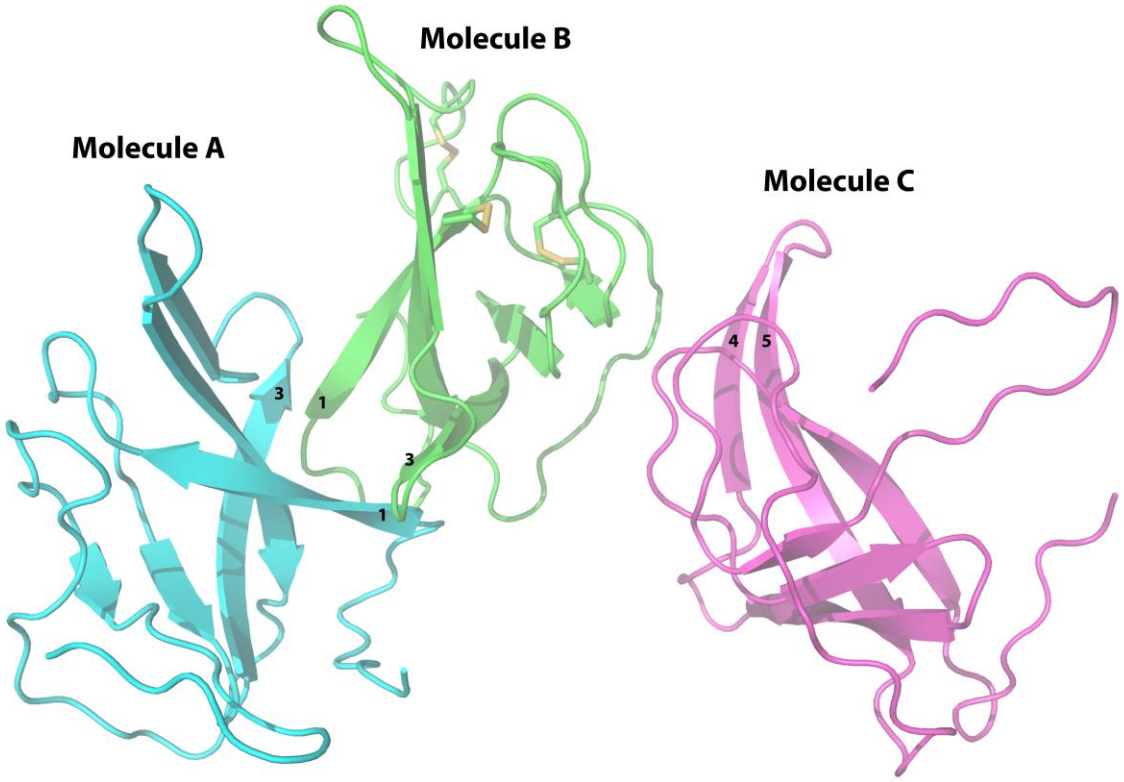
Appendix F. Crystal Packing

N-pIII crystal packing

A P6322 space group with 3 molecules per asymmetric unit contains 12 asymmetric units and thus, 36 molecules per unit cell. Four asymmetric units bind closely to form a chain of 12 tightly associated N-pIII molecules, illustrated best by chain A of Figure 3.6. The 3 chains of 12 molecules interact distally to form the 36 molecule unit cell.

A closer look at chain A shows two distinct inter chain interactions uniting the 4 asymmetric units. The left and right lobe of a chain is composed of two asymmetric units bound extensively by forming continuous beta sheets between the two 'A' molecules and between the two 'B' molecules. The two lobes are then joined by interactions between extended loops of the central 'C' molecules.

The 3 molecules per asymmetric unit align very well with a root mean square deviation (RMSD) between main chain atoms of 1.00 Å, 1.401 and 1.621 for molecules A & B, A & C and B & C respectively (calculated by COOT). Molecules A and B form interactions between β1 from one molecule and the loop connecting β3 and β4 of the connecting molecule whereas β4 and the loop following β1 of molecule C interact with the extended loop between β5 and β6 of molecule B. It's interesting that there isn't a fourth molecule to form a molecule D to molecule A interaction symmetric to the molecule B to molecule C interaction (Figure 3.8). For our purpose this lack of symmetry proves to be beneficial as a molecule D to molecule A interaction could sterically inhibit the joining of the two lobes essential in forming the P6322 crystals.

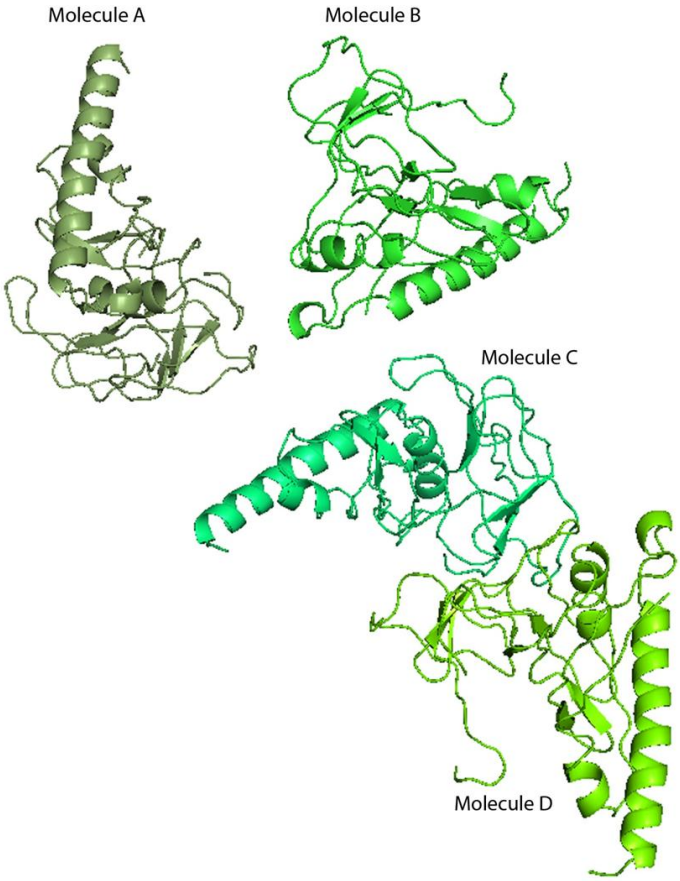


N-pIII asymmetric unit

Molecular interactions between the three N-pIII molecules in the asymmetric unit. Molecules A and B interact between β1 and the loop connecting β3 and β4 of the other molecule whereas β4 and the loop following β1 of molecule C interact with the extended loop between β5 and β6 of molecule B.

His6-TolA-C / N-pIII complex crystal packing

The P212121 space group with 1 complex per asymmetric unit contains 4 molecules per unit cell.



His6-TolA-C / N-pIII complex unit cell

The 4 His6-TolA-C / N-pIII molecules per unit cell.

Appendix G. Crystal screen conditions

G1. Hampton Screens

1 Crystal Screen

2 Crystal Screen 2

3 PEG / ION Screen 1

4 PEG / ION Screen 2

5 Detergent Screen HT

6 Additive Screen

G2. Emerald BioSystems Screens

1 Wizard I

2 Wizard II

3 Wizard III

4 Wizard IV

G3. Molecular Dimensions Screens

1 MD1-29-P1

2 MD1-29-P2

G4. Tacsimate

Tacsimate is a unique crystallization reagent developed exclusively by Hampton Research. Tacsimate is composed of a mixture of titrated organic acid salts. Tacsimate contains 1.8305 M Malonic acid, 0.25 M Ammonium citrate tribasic, 0.12 M Succinic acid, 0.3 M DL-Malic acid, 0.4 M Sodium acetate trihydrate, 0.5 M Sodium formate, and 0.16 M Ammonium tartrate dibasic1. This mixture is titrated to the appropriate pH using sodium hydroxide and is available in pH 4, 5, 6, 7, 8, or 9 reagent formulations.

G5. Crystal Producing Screens

PEG/Ion 2 Screen™ HR2-098 Reagent Formulation

Tube #	Salt	Buffer ◊	Polymer
1.	0.1 M Sodium malonate pH 4.0	None	12% w/v Polyethylene glycol 3,350
2.	0.2 M Sodium malonate pH 4.0	None	20% w/v Polyethylene glycol 3,350
3.	0.1 M Sodium malonate pH 5.0	None	12% w/v Polyethylene glycol 3,350
4.	0.2 M Sodium malonate pH 5.0	None	20% w/v Polyethylene glycol 3,350
5.	0.1 M Sodium malonate pH 6.0	None	12% w/v Polyethylene glycol 3,350
6.	0.2 M Sodium malonate pH 6.0	None	20% w/v Polyethylene glycol 3,350
7.	0.1 M Sodium malonate pH 7.0	None	12% w/v Polyethylene glycol 3,350
8.	0.2 M Sodium malonate pH 7.0	None	20% w/v Polyethylene glycol 3,350
9.	4% v/v Tacsimate pH 4.0	None	12% w/v Polyethylene glycol 3,350
10.	8% v/v Tacsimate pH 4.0	None	20% w/v Polyethylene glycol 3,350
11.	4% v/v Tacsimate pH 5.0	None	12% w/v Polyethylene glycol 3,350
12.	8% v/v Tacsimate pH 5.0	None	20% w/v Polyethylene glycol 3,350
13.	4% v/v Tacsimate pH 6.0	None	12% w/v Polyethylene glycol 3,350
14.	8% v/v Tacsimate pH 6.0	None	20% w/v Polyethylene glycol 3,350
15.	4% v/v Tacsimate pH 7.0	None	12% w/v Polyethylene glycol 3,350
16.	8% v/v Tacsimate pH 7.0	None	20% w/v Polyethylene glycol 3,350
17.	4% v/v Tacsimate pH 8.0	None	12% w/v Polyethylene glycol 3,350
18.	8% v/v Tacsimate pH 8.0	None	20% w/v Polyethylene glycol 3,350
19.	0.1 M Succinic acid pH 7.0	None	12% w/v Polyethylene glycol 3,350
20.	0.2 M Succinic acid pH 7.0	None	20% w/v Polyethylene glycol 3,350
21.	0.1 M Ammonium citrate tribasic pH 7.0	None	12% w/v Polyethylene glycol 3,350
22.	0.2 M Ammonium citrate tribasic pH 7.0	None	20% w/v Polyethylene glycol 3,350
23.	0.1 M DL-Malic acid pH 7.0	None	12% w/v Polyethylene glycol 3,350
24.	0.2 M DL-Malic acid pH 7.0	None	20% w/v Polyethylene glycol 3,350
25.	0.1 M Sodium acetate trihydrate pH 7.0	None	12% w/v Polyethylene glycol 3,350
26.	0.2 M Sodium acetate trihydrate pH 7.0	None	20% w/v Polyethylene glycol 3,350
27.	0.1 M Sodium formate pH 7.0	None	12% w/v Polyethylene glycol 3,350
28.	0.2 M Sodium formate pH 7.0	None	20% w/v Polyethylene glycol 3,350
29.	0.1 M Ammonium tartrate dibasic pH 7.0	None	12% w/v Polyethylene glycol 3,350
30.	0.2 M Ammonium tartrate dibasic pH 7.0	None	20% w/v Polyethylene glycol 3,350
31.	2% v/v Tacsimate pH 4.0	0.1 M Sodium acetate trihydrate pH 4.6	16% w/v Polyethylene glycol 3,350
32.	2% v/v Tacsimate pH 5.0	0.1 M Sodium citrate tribasic dihydrate pH 5.6	16% w/v Polyethylene glycol 3,350
33.	2% v/v Tacsimate pH 6.0	0.1 M BIS-TRIS pH 6.5	20% w/v Polyethylene glycol 3,350
34.	2% v/v Tacsimate pH 7.0	0.1 M HEPES pH 7.5	20% w/v Polyethylene glycol 3,350
35.	2% v/v Tacsimate pH 8.0	0.1 M Tris pH 8.5	16% w/v Polyethylene glycol 3,350
36.	None	0.07 M Citric acid, 0.03 M BIS-TRIS propane / pH 3.4	16% w/v Polyethylene glycol 3,350
37.	None	0.06 M Citric acid, 0.04 M BIS-TRIS propane / pH 4.1	16% w/v Polyethylene glycol 3,350
38.	None	0.05 M Citric acid, 0.05 M BIS-TRIS propane / pH 5.0	16% w/v Polyethylene glycol 3,350
39.	None	0.04 M Citric acid, 0.06 M BIS-TRIS propane / pH 6.4	20% w/v Polyethylene glycol 3,350
40.	None	0.03 M Citric acid, 0.07 M BIS-TRIS propane / pH 7.6	20% w/v Polyethylene glycol 3,350
41.	None	0.02 M Citric acid, 0.08 M BIS-TRIS propane / pH 8.8	16% w/v Polyethylene glycol 3,350
42.	0.02 M Calcium chloride dihydrate, 0.02 M Cadmium chloride hydrate, 0.02 M Cobalt(II) chloride hexahydrate	None	20% w/v Polyethylene glycol 3,350
43.	0.01 M Magnesium chloride hexahydrate 0.005 M Nickel(II) chloride hexahydrate	0.1 M HEPES sodium pH 7.0	15% w/v Polyethylene glycol 3,350
44.	0.02 M Zinc chloride	None	20% w/v Polyethylene glycol 3,350
45.	0.15 M Cesium chloride	None	15% w/v Polyethylene glycol 3,350
46.	0.2 M Sodium bromide	None	20% w/v Polyethylene glycol 3,350
47.	1% w/v Tryptone	0.05 M HEPES sodium pH 7.0	12% w/v Polyethylene glycol 3,350
48.	1% w/v Tryptone	0.05 M HEPES sodium pH 7.0	20% w/v Polyethylene glycol 3,350

◊ Buffer pH is that of a 1.0 M stock prior to dilution with other reagent components: pH with HCl or NaOH.

PEG/Ion 2 Screen contains forty-eight unique reagents. To determine the formulation of each reagent, simply read across the page.

34 Journey
Aliso Viejo, CA 92656-3317 U.S.A.
Tel: (949) 425-1321 • Fax: (949) 425-1611
E-mail: tech@hrmail.com
Website: www.hamptonresearch.com

HAMPTON RESEARCH
Solutions for Crystal Growth

Wizard: Classic 1-Tubes

EB-WIZ-1

No.	Precipitation Reagent	Buffer	pH	Salt
1	20% (w/v) PEG 8000	0.1M CHES/Sodium hydroxide	9.5	
2	10% (v/v) 2-propanol	0.1M HEPES/Sodium hydroxide	7.5	0.2M Sodium chloride
3	15% (v/v) Ethanol	0.1M CHES/Sodium hydroxide	9.5	
4	35% (v/v) 2-methyl-2,4-pentanediol	0.1M Imidazole/Hydrochloric acid	8.0	0.2M Magnesium chloride
5	30% (v/v) PEG 400	0.1M CAPS/Sodium hydroxide	10.5	
6	20% (w/v) PEG 3000	0.1M Sodium citrate/Citric acid	5.5	
7	10% (w/v) PEG 8000	0.1M MES/Sodium hydroxide	6.0	0.2M Zinc acetate
8	2.0 M Ammonium sulfate (dibasic)	0.1M Sodium citrate/Citric acid	5.5	
9	1.0 M Ammonium phosphate (dibasic)	0.1M Sodium acetate/Acetic acid	4.5	
10	20% (w/v) PEG 2000 MME	0.1M Tris base/Hydrochloric acid	7.0	
11	20% (v/v) 1,4-butanediol	0.1M MES/Sodium hydroxide	6.0	0.2M Lithium sulfate
12	20% (w/v) PEG 1000	0.1M Imidazole/Hydrochloric acid	8.0	0.2M Calcium acetate
13	1.26 M Ammonium sulfate (dibasic)	0.1M Sodium cacodylate/Hydrochloric acid	6.5	
14	1.0 M Sodium citrate	0.1M Sodium cacodylate/Hydrochloric acid	6.5	
15	10% (w/v) PEG 3000	0.1M Imidazole/Hydrochloric acid	8.0	0.2M Lithium sulfate
16	2.5 M Sodium chloride	0.1M Potassium phosphate (dibasic)/Sodium phosphate (monobasic)	6.2	
17	30% (w/v) PEG 8000	0.1M Sodium acetate/Acetic acid	4.5	0.2M Lithium sulfate
18	1.0 M Potassium/sodium tartrate	0.1M Imidazole/Hydrochloric acid	8.0	0.2M Sodium chloride
19	20% (w/v) PEG 1000	0.1M Tris base/Hydrochloric acid	7.0	
20	1.6 M Potassium phosphate (dibasic)/0.4 M Sodium phosphate (monobasic)	0.1M Imidazole/Hydrochloric acid	8.0	0.2M Sodium chloride
21	20% (w/v) PEG 8000	0.1M HEPES/Sodium hydroxide	7.5	
22	10% (v/v) 2-propanol	0.1M Tris base/Hydrochloric acid	8.5	
23	15% (v/v) Ethanol	0.1M Imidazole/Hydrochloric acid	8.0	0.2M Magnesium chloride
24	35% (v/v) 2-methyl-2,4-pentanediol	0.1M Tris base/Hydrochloric acid	7.0	0.2M Sodium chloride
25	30% (v/v) PEG 400	0.1M Tris base/Hydrochloric acid	8.5	0.2M Magnesium chloride
26	10% (w/v) PEG 3000	0.1M CHES/Sodium hydroxide	9.5	
27	0.8 M Potassium phosphate (dibasic)/1.2M Sodium phosphate (monobasic)	0.1M CAPS/Sodium hydroxide	10.5	0.2M Lithium sulfate
28	20% (w/v) PEG 3000	0.1M HEPES/Sodium hydroxide	7.5	0.2M Sodium chloride
29	10% (w/v) PEG 8000	0.1M CHES/Sodium hydroxide	9.5	0.2M Sodium chloride
30	1.26 M Ammonium sulfate (dibasic)	0.1M Sodium acetate/Acetic acid	4.5	0.2M Sodium chloride
31	20% (w/v) PEG 8000	0.1M Sodium phosphate/Citric acid	4.2	0.2M Sodium chloride
32	10% (w/v) PEG 3000	0.1M Potassium phosphate (dibasic)/Sodium phosphate (monobasic)	6.2	
33	2.0 M Ammonium sulfate (dibasic)	0.1M CAPS/Sodium hydroxide	10.5	0.2M Lithium sulfate
34	1.0 M Ammonium phosphate (dibasic)	0.1M Imidazole/Hydrochloric acid	8.0	
35	20% (v/v) 1,4-butanediol	0.1M Sodium acetate/Acetic acid	4.5	
36	1.0 M Sodium citrate	0.1M Imidazole/Hydrochloric acid	8.0	
37	2.5 M Sodium chloride	0.1M Imidazole/Hydrochloric acid	8.0	
38	1.0 M Potassium/sodium tartrate	0.1M CHES/Sodium hydroxide	9.5	0.2M Lithium sulfate
39	20% (w/v) PEG 1000	0.1M Sodium phosphate/Citric acid	4.2	0.2M Lithium sulfate
40	10% (v/v) 2-propanol	0.1M MES/Sodium hydroxide	6.0	0.2M Calcium acetate
41	30% (w/v) PEG 3000	0.1M CHES/Sodium hydroxide	9.5	
42	15% (v/v) Ethanol	0.1M Tris base/Hydrochloric acid	7.0	
43	35% (v/v) 2-methyl-2,4-pentanediol	0.1M Potassium phosphate (dibasic)/Sodium phosphate (monobasic)	6.2	
44	30% (v/v) PEG 400	0.1M Sodium acetate/Acetic acid	4.5	0.2M Calcium acetate
45	20% (w/v) PEG 3000	0.1M Sodium acetate/Acetic acid	4.5	
46	10% (w/v) PEG 8000	0.1M Imidazole/Hydrochloric acid	8.0	0.2M Calcium acetate
47	1.26 M Ammonium sulfate (dibasic)	0.1M Tris base/Hydrochloric acid	8.5	0.2M Lithium sulfate
48	20% (w/v) PEG 1000	0.1M Sodium acetate/Acetic acid	4.5	0.2M Zinc acetate

Well #	Detergent	Synonyms	MW	CMC (mM)	[Actual]	Type [1]
25. (C1)	ANAPOE® -$C_{12}E_{10}$	Polyoxyethylene(10)dodecyl ether / 3,6,9,12,15,18,24,27,30-decaoxadotetracontan-1-ol	~ 627	0.20	10% w/v	N
26. (C2)	Sucrose monolaurate	β-D-Fructopyranosyl-α-D-glucopyranoside monododecanoate / Lauroyl sucrose / Dodecanoyl sucrose / Sucrose monododecanoate	524.61	0.30	3.0 mM	N
27. (C3)	CYMAL® -6	6-Cyclohexyl-1-hexyl-β-D-maltoside	508.61	0.56	5.6 mM	N
28. (C4)	n-Undecyl-β-D-maltoside	n-Undecyl-β-D-maltopyranoside	496.60	0.59	5.9 mM	N
29. (C5)	ANAPOE® -X-405	TRITON® X-405 / α-[4-(1,1,3,3-Tetramethylbutyl)phenyl]-ω-hydroxy-poly(oxy-1,2-ethanediyl)	~ 1967	0.81	10% w/v	N
30. (C6)	TRITON® X-100	Octylphenoxypolyethoxyethanol / Polyethylene Glycol-p-isooctylphenyl Ether	650.0	0.90	9.0 mM	N
31. (C7)	ANAPOE® -$C_{10}E_8$	Polyoxyethylene(6)decyl ether / 3,6,9,12,15,18-hexaoxaoctacosan-1-ol	~ 423	0.90	10% w/v	N
32. (C8)	n-Decyl-β-D-thiomaltoside	1-s-Decyl-β-D-thiomaltoside / n-Decyl-β-D-thiomaltopyranoside	498.64	0.90	9.0 mM	N
33. (C9)	Octyl maltoside, fluorinated	(1H, 1H, 2H, 2H-Perfluorooctyl)-β-D-maltopyranoside	688.39	1.02	10.2 mM	N
34. (C10)	ANAPOE® -$C_{10}E_9$	Polyoxyethylene(9)decyl ether / α-Decyl-ω-hydroxy-poly(oxy-1,2-ethanediyl)	~ 555	1.3	10% w/v	N
35. (C11)	Big CHAP, deoxy	N,N-bis-(3-D-Gluconamidopropyl)deoxycholamide	862.10	1.4	14.0 mM	N
36. (C12)	n-Decyl-β-D-maltoside	n-Decyl-β-D-maltopyranoside	482.57	1.8	18.0 mM	N
37. (D1)	LDAO	Lauryldimethylamine-N-oxide / DDAO / N,N-Dimethyl-1-dodecanamine-N-oxide / n-Dodecyl-N,N-dimethylamine-N-oxide	229.41	2.0	20.0 mM	N
38. (D2)	n-Decanoylsucrose	α-D-Glucopyranoside / β-D-Fructofuranosyl Monodecanoate / Sucrose Monocaprate	496.55	2.5	25.0 mM	N
39. (D3)	n-Nonyl-β-D-thioglucoside	n-Nonyl-β-D-thioglucopyranoside	322.47	2.9	1.5 mM	N
40. (D4)	n-Nonyl-β-D-thiomaltoside	n-Nonyl-β-D-thiomaltopyranoside	484.61	3.2	32.0 mM	N
41. (D5)	CYMAL® -5	5-Cyclohexyl-1-pentyl-β-D-maltoside	494.58	5.0	50.0 mM	N
42. (D6)	n-Nonyl-β-D-maltoside	n-Nonyl-β-D-maltopyranoside	468.54	6.0	60.0 mM	N
43. (D7)	n-Nonyl-β-D-glucoside	n-Nonyl-β-D-glucopyranoside	306.40	6.5	65.0 mM	N
44. (D8)	HEGA® -10	Decanoyl-N-hydroxyethylglucamide	379.50	7.0	70.0 mM	N
45. (D9)	MEGA -10	Decanoyl-N-methylglucamide	349.47	7.0	70.0 mM	N
46. (D10)	C_8E_5	Pentaethylene glycol monooctyl ether / Octyl pentaethylene glycol ether / Octylpentaglycol / 3,6,9,12,15-Pentaoxatricosan-1-ol	350.50	7.1	71.0 mM	N
47. (D11)	CYMAL® -4	4-Cyclohexyl-1-butyl-β-D-maltoside	480.55	7.6	76.0 mM	N
48. (D12)	C_8E_4	Tetraethylene glycol monooctyl ether / Octyl tetraethylene glycol ether / Octyltetraglycol / 3,6,9,12-Tetraoxaeicosan-1-ol	306.44	8.0	80.0 mM	N

[1] **N** = Non ionic / **I** = Ionic / **Z** = Zwitterionic / **SL** = Synthetic Lipid

HAMPTON RESEARCH
Solutions for Crystal Growth