The crystal structure of the VCC heptameric pore and its implications in membrane targeting

By

Swastik De

Faculty Advisor: Dr. Richard Olson

A Dissertation submitted to the Faculty of Wesleyan University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Middletown, Connecticut May 2016
Submission of Graduate Student Work to Olin Library, WesScholar, and UMI/Proquest
TABLE OF CONTENTS

Table of contents ..................................................................................................................................................... I

List of figures ........................................................................................................................................................... IX

List of tables ........................................................................................................................................................... XI

Acknowledgements .................................................................................................................................................. XII

Dedication ............................................................................................................................................................... XV

Abstract ...................................................................................................................................................................... XVI

CHAPTER 1: INTRODUCTION

CHAPTER 1.1: VIBRIO CHOLERAE: DISEASE AND PATHOGENESIS

*Vibrio cholerae* .......................................................................................................................................................... 2

Outbreaks of cholera ................................................................................................................................................... 4

Cholera in America .................................................................................................................................................... 4

Cholera in the rest of the world ................................................................................................................................ 5

Life-cycle of *V. cholerae* ...................................................................................................................................... 6

Molecular factors behind cholera pathogenesis ...................................................................................................... 7

Colonization of the intestine by *V. cholerae* .......................................................................................................... 9

The ToxR regulon – the master switch of virulence ............................................................................................. 15

Cholera Toxin ........................................................................................................................................................... 17
Mechanism of action of cholera toxin ................................................................. 20
Lipopolysaccharides (LPS) of V. cholerae ............................................................... 22

CHAPTER 1.2: PORE-FORMING TOXINS (PFTs)

Bacterial toxins ........................................................................................................ 24
Pore-forming toxins ................................................................................................. 24
Classification of PFTs ............................................................................................. 25
  α-pore forming toxins (α-PFTs) ............................................................................ 26
  β-pore forming toxins (β-PFTs) .......................................................................... 31
Membrane insertion mechanism .............................................................................. 34
Cellular responses to pore formation ..................................................................... 36
Role of PFTs in infection .......................................................................................... 39

CHAPTER 1.3: VIBRIO CHOLERAE CYTOLYSIN (VCC) AND PATHOGENESIS

Vibrio cholerae : a brief history .............................................................................. 41
Identification of VCC as a pathogenic agent .......................................................... 42
Effects of VCC on eukaryotic cell lines ................................................................. 43
Effects of VCC in C. elegans .................................................................................. 46
Structure of the VCC monomer .............................................................................. 47

CHAPTER 1.4: MEMBRANE BINDING OF VCC
The lipid membrane and cholesterol .................................................................52

Cholesterol interactions in PFT family .............................................................55

Cholesterol dependent cytolyssins (CDCs) and their dependence on cholesterol for
membrane binding and insertion ........................................................................55

The TL motif .......................................................................................................65

Cholesterol-interacting domains in other membrane-associated proteins ........65

Steroidogenic acute regulatory protein (StAR)-related lipid transfer (START)
domains .................................................................................................................66

Sterol-sensing domains (SSDs) .............................................................................68

Cholesterol recognition/interaction amino acid consensus sequence (CRAC) motif
......................................................................................................................................69

Inverted CRAC (CARC) motif ...............................................................................72

Tilted peptide motif ...............................................................................................73

Cholesterol consensus motif (CCM) .....................................................................74

Membrane binding of VCC ..................................................................................74

Effect of cholesterol .............................................................................................74

Effect of lipids ........................................................................................................79

Oligomerization and pore formation of VCC ......................................................82

CHAPTER 1.5: GLYCANS AND LECTINS
Glycans: an overview..................................................................................................................84

Different types of glycans...........................................................................................................85

N-linked glycans ..........................................................................................................................86

O-linked glycans ..........................................................................................................................89

Glycosylphosphatidylinositol (GPI) anchors ..............................................................................91

Glycosphingolipids (GSL) ...........................................................................................................92

Lectins : a brief history ..................................................................................................................92

Molecular aspects of carbohydrate binding .................................................................................94

Flexibility .......................................................................................................................................94

Multivalency ..................................................................................................................................95

Energetics of binding .....................................................................................................................96

Old classification of lectins ...........................................................................................................97

Monosaccharide binding lectins .................................................................................................97

Oligosaccharide binding lectins .................................................................................................99

Modern classification of lectins ...................................................................................................100

R-type lectins ...............................................................................................................................100

L-type lectins ...............................................................................................................................103

P-type lectins ...............................................................................................................................106

C-type lectins ...............................................................................................................................107
I-type lectins .................................................................................................................................................. 109
Galectins.......................................................................................................................................................... 111
Microbial lectins .............................................................................................................................................. 114
Viral lectins and hemagglutinin ........................................................................................................................ 114
Bacterial lectins .................................................................................................................................................. 114
Lectins and bacteria .......................................................................................................................................... 115
Lectin domains in *V. cholerae* ...................................................................................................................... 118
*V. cholerae* mannose/glucose specific hemagglutinin (MSHA) ................................................................. 118
Neuraminidase .................................................................................................................................................. 119
Cholera Toxin (CT) ......................................................................................................................................... 119
*Vibrio cholerae* cytolysin (VCC) .................................................................................................................. 120
Biofilm associated protein 1 (Bap1) and rugosity and biofilm structure modulator C (RbmC) .................. 123

CHAPTER 2: STRUCTURE OF THE VCC HEPTAMER

Introduction ....................................................................................................................................................... 125
Materials and Methods .................................................................................................................................. 129
Assembly of the VCC oligomer ....................................................................................................................... 129
Crystallization and data collection .................................................................................................................. 129
Structure solving and refinement ................................................................................................................... 130
Results ........................................................................................................................................ 131
VCC heptamer structure ............................................................................................................. 131
Structural rearrangements during assembly ............................................................................. 136
VCC membrane interactions ....................................................................................................... 138
The VCC pore .............................................................................................................................. 140
Discussion .................................................................................................................................... 142

CHAPTER 3: MEMBRANE BINDING OF VCC

Introduction ................................................................................................................................... 149

Materials and Methods .............................................................................................................. 153
Construction of VCC mutants ...................................................................................................... 152
Protein expression and purification .............................................................................................. 154
Isolation of primary human polymorphonuclear leukocytes and culture of primary human immune cells ...................................................................................................................... 155
VCC cytotoxicity assays ............................................................................................................... 156
Hemolysis assays .......................................................................................................................... 156
Isothermal titration calorimetry (ITC) .......................................................................................... 157
Analytical ultracentrifugation ....................................................................................................... 158
Liposome preparation .................................................................................................................... 158
SDS-PAGE gel-based oligomerization assays .............................................................................. 159
Surface plasmon resonance ................................................................. 159
Results .................................................................................................. 160
Activity of VCC against primary and derived cell lines ......................... 160
Construction and quality control of VCC rim-domain alanine-scanning mutations .................................................................................................................................................................................. 164
Hemolysis assays of VCC rim-domain alanine-scanning mutants .............. 166
Additional mutagenesis of rim-domain residues ..................................... 167
The interplay between sugar-binding mutants and rim-domain mutants .... 169
The effect of rim-domain mutations on activity against liposomes ............ 170
Surface plasmon resonance (SPR) studies of membrane binding by VCCWT and alanine-scanning mutants .................................................................................................................. 177
Discussion ............................................................................................. 178

CHAPTER 4: GLYCAN BINDING BY VCC

Introduction .......................................................................................... 187
Materials and Methods ......................................................................... 195
Construction of RbmC lectin constructs .................................................. 195
Construction of VCC lectin mutants ....................................................... 195
Glycan-chip assay .................................................................................. 196
Fluorescence binding assay .................................................................... 197
LIST OF FIGURES

1.1 Life cycle of the human intestinal pathogen *Vibrio cholerae* ...............................6

1.2 Structure of cholera toxin......................................................................................18

1.3 Schematic showing the molecular structure of the gram-negative envelope .......22

1.4 Examples of the structures of α-channel forming proteins .................................29

1.5 Representative examples of the structures of β-pore forming proteins ............30

1.6 Cytotoxicity of VCC ..........................................................................................51

1.7 Representative structures of cholesterol dependent cytolysins .......................60

1.8 Two types of cholesterol-interacting sequences .................................................60

1.9 Effect of cholesterol on VCC’s pore forming activity ......................................79

1.10 Different types of conventional representations used for glycans ......................85

1.11 Early steps of the N-linked glycan synthesis pathway ................................87

1.12 Types of N-linked glycans and complex O-GalNAc glycans ..........................88

1.13 Structural motifs found in the lectin family ......................................................113

2.1 Structure of the VCC heptamer ........................................................................133

2.2 Comparison of the VCC protoxin structure and a protomer from the VCC heptamer ........................................................................................................134

2.3 Key residues at the interprotomer interface .....................................................135
3.1 VCC/membrane interactions .................................................................162
3.2 Cytotoxicity and hemolysis data ...............................................................163
3.3 Analytical ultracentrifugation of selected VCC mutants ..........................165
3.4 ITC experiments measuring VCC binding to asialofetuin .........................171
3.5 Alanine-scanning mutagenesis .................................................................172
3.6 L/Y mutants and glycan double mutants ..................................................175
3.7 Liposome gel-based assays .....................................................................176
3.8 SPR analysis of rim-domain mutations .....................................................179
4.1 Comparison between the β-prism domains of VCC, Bap1, and RbmC ............194
4.2 Glycan screening of the RbmC 1st and 2nd lectin domains identified several
glycans with strong binding to RbmC β-prism domains. ...............................201
4.3 Representative binding curves for RbmC 2nd lectin domain obtained using ITC
and fluorescence assays ..............................................................................205
4.4 Crystal structure of the RbmC 2nd lectin domain bound to trimannose.........210
4.5 Structure aided mutagenesis in the VCC β-prism lectin domain. .................214
4.6 Sequence alignment of the β-prism domains from Vibrio spp. ....................215
4.7 Evolutionary relationships between the different β-prism domains in Vibrio spp.
..................................................................................................................216
LIST OF TABLES

3.1 HD50 values for all mutants described in this study ........................................180

4.1 Summary of the binding affinities of the β-prism domains for various ligands .208
4.2 Thermodynamic parameters of the three β-prism domains with methyl-α-
mannose, trimannose and asialofetuin .................................................................209

4.3 X-ray and refinement statistics ..........................................................217
Acknowledgements

Undertaking the PhD has proved to be an extremely stimulating, rewarding, and life changing process for me. During my time as a graduate student, there were a number of individuals to whom I am indebted for their kind support and guidance. These people not only helped me remain focused on the goal, but also made this journey invigorating and enjoyable at the same time.

I would like to thank my advisor, Dr. Rich Olson, for allowing me to join his lab and providing me the opportunity to do research. I consider myself extremely fortunate to have Rich as my mentor. He always encouraged me to think independently and acted as a catalyst for my intellectual growth as a scientist. It’s been a pleasure working with him during the entirety of this dissertation, which is also a defining part of my life as a graduate student. In addition to academic or research collaborations, I deeply value the close link that has formed between me and Rich over the span of this work. From my first bowling lessons to first-time beer making, I have learned much more from him than what this thesis contains. I quite simply could not have asked for a better person as my advisor. I am confident that without him this PhD would not have been achievable.

I would like to thank my committee members for their advice throughout the tenure of my PhD life. I want to thank Dr. Ishita Mukerji for teaching me fluorescence and other biophysical techniques from scratch, which proved to be instrumental in finishing this thesis work. I also thank you for letting me work in your
lab and use any instrument as I required. I would also like to thank Dr. Don Oliver for his suggestions and support throughout my stay at Wesleyan. Whenever I knocked at his door with a question (from what is a membrane to how to write a cover letter and apply for jobs), he took the time to answer any query I had even though he was very busy. Thanks should also be bestowed upon Dr. Manju Hingorani for her kind advice and support. Manju made sure that I could nurture myself in Wesleyan by continuously encouraging me to be analytical about science and by providing unlimited chocolates.

I would also like to thank other Wesleyan professors, especially Dr. Irina Russu, Dr. Mike McAlear, Dr. Amy McQueen, Dr. Michelle Murolo, Dr. Scott Holmes, Dr. Bob Lane, Dr. Erika Taylor, Dr. David Beveridge, Dr. Francis Starr, and Dr. Rex Pratt for their teaching and help throughout the years. Additionally, I would like thank the non-academic assistants including Anika Dane, Maureen Snow, Jeff Gilardy, Horace Chambers, Doug Allen, and Bill Nelligan for all of their help.

Dr. Anwesha Bhattacharya, Karen V. Meiman, Dr. Miho Sakato, and Dr. Lorry Grady provided invaluable assistance at various stages of this work. Being the first and only grad student in my lab during that time, I don’t have words to fully appreciate their kindness.

I also owe a great deal of gratitude to the past and present members of the Olson lab including, Kelsey, Beth, Sergei, Robert, Lee, Sophie, Shu, Li Lin, Adele, Jinsol, Chloe, Shada, Ben, Charlie, Mitchell, Ken, Allison, and Katie. You guys made
me feel like I am at home in the lab. Between the lab cleaning, bowling, defrosting, tours, and parties, I have spent some of the best time of my life. Special thanks are due for Katie Kaus, for always being the helping lab-mate and for making the lab a fun place. I don’t think I could have asked for a better lab mate than you.

At Wesleyan, I have made some great friends, who have always been there whenever I needed them. Brandon Case, Sudipta Lahiri, Tithi Banerjee, Katie Kaus, Sergei Pourmal, Sarah Mustafa, Jenny Michtavy, Pritam Mukherjee, Jun Huang, Anna Rogers, Joyce Noble, and Legairre Radden made the Wesleyan experience very enjoyable. I am particularly thankful to Brandon and Sudipta for your friendship, for all those dinner parties, and for those ever-refreshing hallway chats. The distractions you guys provided proved to be crucial for keeping myself sane.

I am deeply thankful to my family for their unconditional love and support. My parents and grandparents have always provided the means for me to enrich myself and helped me to be whoever I am today. Finally, a very special thank you to Jyoti Gupta, without whom I wouldn’t have the courage to embark on the path I am walking now. Her love, care, and encouragements kept me focused till the end.

Swastik De

Middletown, CT, USA
DEDICATION

I dedicate this thesis to all my teachers
ABSTRACT

*Vibrio cholerae* cytolysin (VCC) is a pore-forming toxin (PFT) produced by the human gastrointestinal pathogen *Vibrio cholerae* that acts as an important accessory toxin during pathogenesis. VCC is secreted as a water-soluble monomer, targets cells using cell-surface receptors, and assembles into membrane-spanning channels in the lipid bilayer. These receptor molecules include membrane proteins, membrane-associated glycan molecules, cholesterol, and other lipid molecules. VCC was previously shown to have a strong preference for the presence of cholesterol, ceramide, and sphingomyelin in the membrane. Additionally, VCC interacts with glycoconjugates through its β-prism lectin domain. However, structural information regarding the transmembrane pore, the molecular mechanism of VCC-membrane binding, and the precise mechanism of VCC-glycan interactions is not known.

To understand the dynamic process of channel formation, we solved the structure of the detergent-solubilized VCC oligomeric pore to 2.9 Å resolution using X-ray crystallography. Comparison of the heptameric pore structure with the already published VCC monomer structure provides a clear picture of domain rearrangements that occur as a result of the assembly process. Following the clues from the heptamer structure and to better understand the nature of membrane interactions, we systematically scanned the rim region surface of VCC using alanine mutagenesis. In doing so, we identified a number of residues responsible for lipid and cholesterol binding. In a separate investigation, we looked at the molecular mechanism of carbohydrate binding and illustrated the importance of the β-prism domain in the
process. Our study revealed that the VCC β-prism domain shows specificity towards the pentasaccharide core of complex N-linked glycans with nM affinity. Using a combination of structural and functional approaches, we probed the glycan binding process of the VCC β-prism lectin domain. To further understand its specificity, we used glycan fragments to footprint its essential binding motif and measured affinities by ensemble binding assays. By utilizing the X-ray structure of a smaller piece of N-glycan bound to a β-prism lectin domain and structure aided mutagenesis, we developed a molecular model for VCC-β-prism lectin binding to cell surface glycans. The studies reported in this dissertation will provide a template for β-PFT assembly, a functional map of VCC-membrane interactions, and a precise mechanism for cell surface glycan binding by VCC. Knowledge from this investigation not only enables us to understand the mechanism of PFTs, but may also lead to development of drugs or vaccines against pathogenic organisms.
CHAPTER 1

INTRODUCTION
SECTION 1- VIBRIO CHOLERAE: DISEASE AND PATHOGENESIS

Vibrio cholerae

Cholera is an ancient disease with references that date back to 5th century BC in Sanskrit literature. It spread outside the Indian subcontinent around 1817 and, since then, has caused a number of pandemics and epidemics. Vibrio cholerae was first reported and named by Pacini (Kaper et al., 1995), in 1854, who found that the intestines of cholera patients contain a comma-shaped bacteria. It was further described in 1883 by Robert Koch who described that cholera is caused by these comma-shaped organisms, which he called Kommabazillen (Kaper et al., 1995). The name Vibrio cholerae was adopted in honor of Pacini’s remarkable work many years after Koch’s discovery.

Vibrio cholerae is a Gram-negative, non-spore forming, rod-shaped, facultative anaerobic, oxidase positive, nitrate reductase negative bacteria with a distinctive single polar long flagellum (Gorchev and Ozolins, 1984). Among different V. cholerae serotypes, V cholerae O1 is the most prevalent, which is named due to its ability to agglutinate the O1 blood-group 1 specific antiserum (Finkelstein, 1996). Vibrio cholera has an incubation period of 2-3 days (range 6 hours to 5 days) and could be lethal if not treated in time (Daniels & Shafaie, 2000). V. cholerae O1 has been identified as a major cause of multiple epidemics and pandemics of cholera. Wild-type O1 do not contain the cholera toxin (CT, the main cause behind the watery diarrhea characteristics of cholera) gene, and therefore, are mostly non-pathogenic. However, in recent years many CT- O1 strains have been found to cause diarrhea or
extra-intestinal infections (Daniels & Shafaie, 2000). O1 strains can be classified based on biotype, serotype and toxin production and include the following: El Tor and Classic (based on biotype), Inaba, Ogawa or Hikojima (based on serotype), toxigenic or nontoxigenic (based on toxin production) (Morris, 2003; Stroeher et al., 1992). The El Tor biotype is the major cause of the seventh cholera pandemic and named after the place it was first isolated, El Tor (Raithu) in Egypt, and differs from the classical cholera strain in hemolysin production, expression of mannose specific hemagglutinin, polymyxin sensitivity, resistance to different bacteriophages, and differential regulation of the cholera toxin (CT) and toxin coregulated pilus (TCP) etc. (Finkelstein, 1996; DiRita et al., 1996).

*V. cholerae* O139 or the Bengal strain (1992) was identified as a major cholera causing strain and is speculated to contain more epidemic-causing potential than any other strain. This serogroup is a perfect genetic hybrid between O1 and non-O1 strains. In terms of virulence, O139 is almost indistinguishable from O1 and contains both the CT and toxin coated pilus (TCP). The O139 can be distinguished from O1 by the lack of O1 LPS production and absence of O1 antigen. O139 strains also produce a capsule made up of polysaccharides, which are characteristic of non-O1 strains (Kaper et al., 1995). Bioinformatics studies show that the O139 serogroup originated from O1 El Tor by deletion of an O-antigen biosynthesis gene and by acquiring genes from a non-pathogenic serogroup. Also, *V. cholerae* has higher survivability in the marine environment compared to other fecal bacteria, such as *E. coli* and fecal streptococci (Davies et al., 1995). Because of their rapid
transformation, horizontal gene transfer capability, and high adaptability, *V. cholerae* is likely to lead to more outbreaks in the future. Additionally, due to recent changes in global climate, cholera may affect new areas of the world where it has never been seen before.

**Outbreaks of cholera**

**Cholera in America**

Most cases of cholera in the United States are acquired during foreign travel, but some domestic cases have also been reported. The predicted fatality rate is less than 1% when proper treatments are available to the patient and high fatality rates usually result from late or inappropriate medical care. The standard treatment for cholera patients consists of administering oral rehydration salts (ORS) or an oral electrolyte rehydration solution. In severe cases, ORS may not be enough and intravenous fluid and antibiotics (like ciprofloxacin) may be needed (Usubutun et al., 1997).

The only major cholera epidemic in the South American continent originated in 1991 in Peru where approximately 4,500 cases/day were reported. Because of quick medical attention and available treatment, the mortality rate never crossed 1% of the total cases. This epidemic rapidly spread in nearby countries like Colombia, Brazil, Chile, and Ecuador (Gorchev and Ozolins, 1984).
Recently, more than 700,000 unofficial cases of cholera (Salvador et al., 2015) have been estimated by ground workers in Haiti after the earthquake in 2010, resulting in more than 8,500 deaths, showing the vulnerability of the modern world to this ancient disease. WHO statistics show that the highest number of cholera cases were documented in the United States in 1997 and affected more than 40,000 individuals. Although the number of cholera cases has since decreased, still greater than 200 people in the United States and around 180 in Mexico suffer from cholera every year (WHO statistics).

**Cholera in the rest of the world**

The majority of worldwide cholera cases are reported in countries with a poor or inadequate clean drinking water supply such as in developing countries or areas affected by natural or man-made disasters like flood, earthquake or refugee crisis. In 2013, more than 129,000 cholera cases were documented by the WHO around the globe. In the same year, five countries that had the highest reported number of cholera cases were Haiti (58,809 cases), Democratic Republic of Congo (26,944 cases), Somalia (6,864 cases), Angola (6,655 cases), and Nigeria (6,600 cases). With conservative estimates, the WHO predicts as many as 1.4-4.3 million cholera cases causing 28,000 to 142,000 deaths every year, worldwide. For the last decade and a half, Africa accounted for at least 50% of the total number of cases worldwide, with the number reaching as high as 90% in some years (2001-2009). In the last few years, the disease profile has changed, and in 2013, the American continent became the largest source of cholera infection with 47% of all the cases worldwide (with 43%
reported from Africa in the same year). These numbers clearly suggest a changing
trend in the epidemiology of cholera and more new countries may suffer from cholera
infection in the future.

**Life-cycle of *V. cholerae***

The life cycle of *Vibrio cholerae*, the causative agent of cholera, consists of
stages in both the human body and in the marine environment. After being ingested
by human hosts, *V. cholerae* enters the gut (Nelson et al., 2009) and as the bacterium
is highly susceptible to gastric acids, very few bacteria survive passage through the
stomach (Salvador et al., 2015). Once in the small intestine, the bacteria use their
motility to penetrate the mucus gel layer to reach the epithelial cell surface.

Figure 1.1 *Life cycle of the human intestinal pathogen Vibrio cholerae*. The
human host is typically infected through ingestion of contaminated food or water.
The bacteria escape from the human body in watery diarrhea and continue their
life cycle in aquatic environments.

Here, using ‘non-commitment’ adhesin molecules, the bacterial cells attach to the
epithelial cell surface, and if the environment is favorable, other adhesin factors are
expressed to begin the colonization of the small intestine. After construction of a
biofilm (Teschler et al., 2015; Heithoff and Mahan, 2004; Waters et al., 2008), several genes related to the virulence (or required for pathogenicity) are substantially overexpressed and lead to the characteristic symptoms of cholera. At this stage, the bacteria again alter their gene expression profile and the ‘escape response’ begins (Nelson et al., 2009; Schild et al., 2008). This results in motile, but non-chemotactic hyperinfective cells, which leave the human host through watery diarrhea. Once outside the human host, V. cholerae can either find a chitinous surface and form a biofilm or transform into so-called active, but non-culturable state searching for a suitable surface (Colwell et al., 1985; Colwell et al., 1996).

When outside the human host, V. cholerae bacteria associate themselves with a variety of organisms including, cyanobacteria (Anabaena variabilis), phytoplankton (Skeletonema costatum), filamentous green algae (Rhizoclonium fontanum), water hyacinths (Eichornia crassipes), eastern oysters (Crassostrea virginica), aquatic arthropod (Gerris spinolae), and Atlantic blue crab (Callinectes sapidus) (Islam et al., 1989; Martin & Bianchi, 1980; Islam et al., 1989; Hood et al., 1981; Spira et al., 1981; Shukla et al., 1995; Huq et al., 1986). In the marine environment, V. cholerae awaits for a suitable human host to infect and continue the cycle.

**Molecular factors behind cholera pathogenesis**

A number of proteins are important for intestinal colonization by *Vibrio cholerae* in mouse and rabbit models. Some of these include toxin-coregulated pili (TCP), accessory colonization factors (ACFs), mannose-fucose resistant
hemagglutinin, TolC, outer membrane porins, biotin and purine synthesis enzymes, iron-regulated OMP protein (IrgA), LPS O-antigen, LPS core region and the regulatory proteins like ToxR/ToxS, TcpP/TcpH and ToxT (Reidl & Klose, 2002). Amongst these proteins, the TCP or the pilus facilitates the colony formation and is, therefore, absolutely necessary for the colonizat process. Coincidentally, TCP also acts as a receptor for the CTXϕ bacteriophage, which is responsible for the cholera pathogenicity. The virulence genes of Vibrio cholerae are either located in Vibrio pathogenicity island (VPI), are scattered in the chromosome, or reside in the CTXϕ phage, which contains the classical cholera toxin (CT) gene. Some of the important virulence genes that are located in the VPI cluster include acfABCD (unknown function), aldA (aldehyde dehydrogenase), tagA (ToxT-activated gene), tcpA (pili), tcpPH (transmembrane regulatory proteins) and toxT (transcriptional activator for virulence genes) while genes like ace (accessory enterotoxin), cep (core-encoded pili), ctxAB (CT toxin), orfU (homolog of an M13 gene), zot (zonula occludens toxin) are located in the CTXϕ phage. Chromosomal genes like flrC (flagellar transcription regulator), irgA (iron-regulated outer membrane protein), msh genes (pili), omuUT (outer membrane porins), rtxA (repeats in toxin), toxRS (transmembrane regulatory proteins), wav genes (LPS core synthesized) and vps genes (exopolysaccharide synthesis) have also been associated with the virulence of Vibrio cholerae (Reidl & Klose, 2002).

Bacteria-phage interactions are a dominant and significant factor in bacterial evolution. In many cases, the survival of bacteria depends on the horizontal transfer
of genetic material through phage infection. *Vibrio cholerae* is no exception and perhaps one of the best-known examples of phage dependent emergence of virulence and toxicogenic strains (Faruque & Mekalanos, 2012).

*Vibrio cholerae* is able to transform itself to adapt to the environment through various processes. One of the most important adaptive processes is infection by the lysogenic CTXϕ bacteriophage, which carries the classical cholera toxin or CT gene, the main causative factor for cholera. CTXϕ is a single stranded filamentous DNA bacteriophage, which infects the bacterium without killing it. After binding to the bacterium’s surface, the CTXϕ bacteriophage injects its DNA into the cell where it either stays as a chromosome integrated lysogen or as a plasmid replication form (RFs). Another way *Vibrio cholerae* can acquire new genes is through its ability to uptake naked DNA from the environment by its natural competent machinery. In this process, the bacterium uses a macromolecular complex primarily composed of a type IV pilus along with other components. Once the bacterium uptakes the ssDNA, it recombines with the bacterial chromosome and becomes a part of the bacterial genome (Matthey & Blokesch, 2016).

**Colonization of the intestine by Vibrio cholerae**

*Vibrio cholerae* infects human body through contaminated food or water and exists either as single free-living cells, as microcolonies or as a part of a biofilm. *V. cholerae* cells are extremely sensitive to stomach acid and studies have shown that ~$10^{11}$ cells are required for infection in healthy volunteers (Cash et al., 1974).
Interestingly, *V. cholerae* sometimes infects the human host in a dormant vegetative state termed as a viable, but nonculturable (VBNC) state. VBNC cells potentially have much higher acid resistance and were able to infect human volunteers more easily (Salvador et al., 2015). The microcolonies of *V. cholerae* are generally in a hyperinfectious state that confers resistance to low pH. A similar hyperinfectious state has also been identified in freshly shed bacteria from the human host. Bacterial cells associated with biofilms have also been shown to be hyperinfectious and are characterized by a very low infectious dose (Merrell et al., 2002).

Once *V. cholerae* cells reach the intestine, the motility of each cell becomes a crucial factor in the ability to penetrate the mucus layer over the epithelial lining. This motility dependence of infection is more important in the case of colonization of the proximal section of the small intestine (Liu et al., 2008). Studies with mutant *V. cholerae* strains also established the importance of chemotaxis in the colonization process by changing the distribution of the bacterium in the human gut (Butler & Camilli, 2004). In a rabbit model, chemotaxis related gene mutations involving *vspR*, *pomA*, and *cheA* resulted in hypercolonization phenotype (Fu et al., 2013). Further studies also suggest that the bacterial type VI secretion system (T6SS) dependent inter-bacterial communication is necessary for the infection process (Fu et al., 2013). Several studies have also investigated the relationship between bile composition and *V. cholerae* chemotaxis, reporting that virulence gene expression is correlated with the bile acids (Gunn, 2000; Gupta & Chowdhury, 1997).
Even after reaching the small intestine, the bacteria has to penetrate the physical barrier of mucin before it can reach the epithelium. This can be a challenging task for the bacterium as the mucus barrier is ~50–75 times the body length of the *V. cholerae* cells (Almagro-Moreno et al., 2015; Flemstrom & Isenberg et al., 2001). To overcome this, *V. cholerae* deploys soluble mucinase complexes composed mainly of the HapA hemagglutinin/protease to degrade the polysaccharide and proteins in the mucin barrier (Booth et al., 1983; Silva et al., 2003; Zhu & Mekalanos, 2003). There is evidence that another metalloprotease TagA may be involved in the direct modification of the host cell surface molecules and can thereby control the attachment process (Szabady et al., 2011). Non-mucinase enzymes, such as neuraminidase (NanH), also play an important role in this step by unmasking receptors for other virulence factors to work (Galen et al., 1992).

After the bacterium successfully penetrates the mucin layer and is in close proximity of the epithelial lining, it starts producing adhesion molecules that help it attach to the surface of epithelial cells. Using deletion strains in mouse models, at least five important adhesion factors have been identified, including the cholera flagellum (Attridge & Rowley, 1983), Mam7 (Multivalent adhesion molecule 7) (Krachler et al., 2011), GbpA (GlcNac binding protein) (Kirn et al., 2005), OmpU (Outer membrane porin U) (Sperandio et al., 1995) and FrhA (Syed et al., 2009).

The *V. cholerae* flagellum has been demonstrated to be a vital component for *in vitro* attachment and *in vivo* colonization. Although the presence of the flagellum is absolutely necessary for the processes, no dependence on motility of the flagellum
has been established. Attridge et al. hypothesized that a functional flagellum works as a scaffold for different adhesion factors to bind and promote the adhesion or attachment process (Attridge & Rowley, 1983).

The outer membrane protein Mam7 is another key factor that controls host cell binding in two different ways. Firstly, Mam7 interacts with host cell fibronectin, and secondly, it engages with phosphatidic acid residues on the membrane. Using nematode and mammalian host cell models, Krachler et al. found that MAM7 provides a competitive benefit for different gram negative pathogenic strains including *Vibrio cholera*, by increasing their chances of attaching to host cell surfaces (Krachler et al., 2011).

In 2005, a small 53 kDa protein named GbpA was recognized to be required for efficient colonization of human intestine as well as for attachment to zooplankton surfaces. GbpA, which is a GlcNac-sensitive chitin-binding protein, interacts with both GlcNac moieties on human cell surface proteins and bacterial membrane-associated receptor and essentially crosslinks the bacterial cell to the human one. It has also been speculated that the secreted form of GbpA may bind to human cell surfaces and reserve those surfaces for the bacterium, giving *V. cholerae* a competitive edge over other species (Kirn et al., 2005).

Outer membrane porins of *V. cholerae* are known to be another important protein for the life cycle and survivability of the bacterium. The 38 kDa porin OmpU was additionally revealed to be an *in vivo* immunogen, which strongly interacts with host fibronectin and an Arg-Gly-Asn tripeptide. Complete inhibition of host intestinal
colonization by several *V. cholerae* strains has been achieved using anti-OmpU antibodies, indicating Omp U’s role in the colonization process (Sperandio et al., 1995). However, further experimentation is needed to completely understand the mechanism of this observation.

Although the exact function of FrhA in cholera pathogenesis is not clear, mutational studies have demonstrated a loss of intestinal colonization in an infant mouse model. FrhA, which binds to chitin, has been predicted to enhance intestinal colonization and biofilm formation by helping the bacteria attach to the epithelial cells more efficiently (Syed et al., 2009).

These nonspecific adhesion molecules function as attachment factors for the bacterium to host cell surfaces. Interestingly, this arrangement provides non-permanent binding, and if the attached surface lacks nutrients, the bacterium can detach to find a better place for prolonged attachment. Once the bacterium finds a good surface to attach using non-specific adhesins, it can produce specific adhesins required for a committed attachment, paving the way for biofilm formation (Salvador et al., 2015).

Recent studies revealed that *V. cholerae* uses N-acetylneuraminic acid (sialic acid) (Ghosh et al., 2011) and N-acetylglucosamine (GlcNac) (Almargo-Moreno et al., 2009) as carbon sources conferring another advantage to the bacterium, as both are available in excess in the human gut. *V. cholerae* also uses glucose as carbon source and controls the uptake of glucose through the expression of the small RNA TarA, a member of the virulence regulatory cascade (Richard et al., 2010). In a
nutshell, *V. cholerae* adheres first to the host cell surface in a non-permanent way by the use of non-specific adherins. Once the bacterium activates the virulence cascade and finds a specific carbon source (glucose), attachment to the epithelial cell surface radically strengthens. In the absence of the specific carbon source and virulence cascade, *V. cholerae* delays the committed attachment step until it reaches a surface with favorable nutrients for the later stages of its life cycle.

Right after the bacterium commits to an attachment surface, it becomes less motile, and starts reproducing and expressing virulence factors. In this phase of the bacterial life cycle, the pilus plays multiple essential roles in microcolony formation. Based on the type of surface, *V. cholerae* uses two different types of pili – Toxin-coregulated pilus (TCP) for biotic surfaces like the human gut and mannose-sensitive hemagglutinin (MSHA) for abiotic surfaces like borosilicate (Watnick et al., 1999). Unfortunately, little is known about the MSHA pilus and its mechanism of function. TCP, on the other hand, acts as a receptor for the CTXϕ bacteriophage, helps in secretion of the colonization factor TcpF and is required for the formation of viable microcolonies (Kirn et al., 2003). The role of TCP in *V. cholerae* pathogenicity has been shown by studies in human volunteers where ingestion of strains with mutant TCP did not lead to any diarrhea and the bacteria were also absent in stool samples (Herrington et al., 1988). Interestingly, no anti-TCP antibodies can be identified in volunteers infected with the wild-type cholera strains. This is perplexing as anti-TCP
antibodies are successful in protecting infant mice from *V. cholerae* infections (Hall et al., 1991; Sun et al., 1990).

Other accessory factors are involved in the colonization process, but are not very well understood, including MFRHA, core encoded pilus, soluble protease and other OMPs (Kaper et al., 1995). LPS and polysaccharide capsule are also involved in the adherence process. Purified LPS and antibodies against LPS are able to significantly stop the adherence of the bacterium to rabbit mucosa and polysaccharide capsules, inhibiting 50% of the binding to the Caco cells (Freter et al., 1976; Johnson et al., 1993). Eventually biofilm proteins like RbmA, RbmC and Bap1 are expressed, which also have an instrumental role in forming the mature biofilm.

**The ToxR regulon – the master switch of virulence**

Research by several groups has extensively characterized multiple systems involved in regulating the virulence cascade in *V. cholerae*. The three major regulators are the ToxR regulon, iron-dependent regulation and in vivo regulation, which is not well understood. There are also other virulence factors such as neuraminidases or different hemagglutinins, which are not regulated by the above-mentioned factors. However, a tight regulation of the virulence cascade is essential for *V. cholerae* to adapt to new environments by optimizing the expression of the virulence factors. This regulation process probably confers great adaptability to *V. cholerae*, allowing it to thrive in diverse conditions including marine environments and the human intestine.
ToxR plays the role of the master control of virulence in *V. cholerae*. ToxR is a 32 kDa transmembrane protein, which binds to a 7-base pair DNA sequence signaling the overexpression of the *ctxAB* genes producing the classical cholera toxin (CT) (Miller & Mekalanos, 1984; Miller et al., 1987). The significance of ToxR in the virulence and pathogenicity of cholera was established in a human volunteer-based study where ToxR mutant strains were completely unable to cause any symptoms of cholera (Herrington et al., 1988). Another protein ToxS, which is also a small transmembrane protein, is speculated to increase the stability of the ToxR monomers and to aid the functional dimer formation (DiRita & Mekalanos, 1991). The ToxR regulon controls the expression of at least 17 different virulence genes in *V. cholerae*. Many models predict that ToxR does not regulate the expression of all of its regulon-members directly, rather it does so by controlling the expression of a protein called ToxT. Therefore, the ToxR regulon is extremely hierarchical in its function. First, ToxR activates ToxT and CT and second, the regulator ToxT controls the expression of other genes in the regulon like TcpA, TcpI, AldA, TagA etc (DiRita et al., 1991). Parsot et al. described that ToxR shares its promoter region with a heat shock protein and is thereby modulated through the heat shock response (Parsot & Mekalanos, 1990). This environmental regulation ensures that after entering the human body, the virulence genes do not turn on before the bacterium crosses the stomach and reaches a much favorable surface like small intestine.

The human intestine is a low iron environment and thus possesses a threat to any bacterium wishing to colonize it (Sciortino & Finkelstein, 1983). In *V. cholerae*,
a low iron environment induces the expression of a specific group of outer membrane porins (OMPs) while repressing another group. Along with OMPs many other proteins like hemolysins, IrgA and vibriobactin also get expressed to combat the low iron environment. Hemolysins are generally secreted and function by lysing host cells to access the iron trapped inside. IrgA is an OMP, which functions in the import of iron to deal with such environments. Vibriobactin is a siderophore that imports iron into the cell through a 74 kDa OMP after binding it extracellularly (Stoebner et al., 1992). Studies suggest that vibriobactin is a member of a vast network of iron transport systems, which together regulate iron metabolism in the cell. However, when iron is freely available, a protein named Fur binds to an operator sequence in an iron-dependent manner blocking the transcription of iron-regulated genes (Calderwood & Mekalanos, 1987).

Sciortino et al. first described how the gene expression profile of *V. cholerae* is drastically different *in vivo* than *in vitro* and identified a number of proteins (ranging from 29 kDa - ~200 kDa) only present in *in vivo* cultures (Sciortino & Finkelstein, 1983; Jonson et al., 1989). Further experiments have shown that the expression of ToxR regulon products and the iron-regulated genes are unchanged between *in vivo* and *in vitro* conditions, indicating that *in vivo* regulation is not dependent on ToxR regulon or the environmental iron status (Jonson et al., 1989).

**Cholera toxin**

S. N. De in 1959 first described that cell-free culture filtrates contain an enterotoxin able to cause fluid accumulation in rabbit intestines (De et al., 1959).
Almost a decade later, this enterotoxin was purified and named as choleragen (later identified as A subunit) and choleragenoid (later identified as B subunit) by Finkelstein et al. (Finkelstein & LoSpalluto, 1969). Further biochemical studies showed that this toxin is composed of two separate subunits – A (for toxic-active) and B (for binding) (Holmgren, 1981). Cholera Toxin (CT) is composed of soluble globular A subunit (CTA) and pentameric B (CTB) subunits. Bioinformatic analysis identified another toxin from Escherichia coli named heat labile enterotoxin or LT, which shares ~80% sequence identity with CT. In both toxins, subunit A is exclusively responsible for the enzymatic activity. The CTA subunit also contains an ADP-ribosyltransferase center, which catalyzes the transfer of an ADP-ribose group from NAD\(^+\) to the G\(_{\text{su}}\) protein involved in GPCR cell signaling (O’Neal et al., 2005). This post-translational modification results in a massive efflux of water and salt.

Figure 1.2 **Structure of classical Cholera Toxin.** A, Structure of cholera toxin and B, top-down view of the pentameric B-subunit. The images are rendered from PDB ID 1XTC and 1FGB, respectively, using PyMOL.
molecules into the intestinal lumen initiating watery diarrhea, the main symptom of cholera. The subunit B monomers of CT assemble into homopentamers and bind to a specific ganglioside called GM1. The CTB homopentamer can interact with GM1 with or without the A subunit, but the CTB monomer has no activity towards GM1 (De Wolf et al., 1981). The first structure of an AB5 type toxin, of which CT is a member, was published in 1991 from *E. coli* (LT holotoxin) (Sixma et al., 1991). Within a few years, the structure of the cholera toxin B-pentamer was solved bound to the GM1 pentasaccharide (Merritt et al., 1994). The complete structure of isolated and purified cholera toxin was determined in 1995 by Zhang et al. The structures of the CT and LT are almost identical within the A and B subunits with the exception of the A2 domain. The two domains of the A subunit (A1 and A2) are connected through a long linker containing a cleavage site and a disulfide bond. The A1 domain contains the catalytic part of the toxin and the catalytic site has been identified to be a well-defined groove in the N-terminal domain of the A subunit (CTA1 domain). On the other hand, the A2 domain interacts with all the B-subunits by projecting its C-terminal into the central pore created by the B-pentamer and helps anchor the enzymatic A1 subunit to the B-pentamer (Sanchez et al., 2011). The CTB monomer is a small protein with only 103 residues and a single disulfide bond. When the monomers assemble into a pentamer, only ~39% of each monomer’s surface becomes water accessible, making an extremely stable assembly. The pentamer forms an inner pore along its five-fold symmetry axis with an average diameter of 11 Å (C-terminal) – 16 Å (N-terminal) and is primarily lined by hydrophobic residues. Even in the
absence of the enzymatic subunit, the B₅ alone is capable of binding to the GM1 by specifically interacting with the terminal Gal-Nac-Gal residues (Fukuta et al., 1988). Each monomer of the B₅ assembly contains a functional GM1 binding site with affinities ranging from $10^{-9}$ to $10^{-10}$ M (Chaudhuri et al., 2009).

**Mode of action of cholera toxin**

*V. cholerae* produces CT in a ToxR regulated manner as a result of a Lux OPQ-dependent quorum sensing signal. Once the polypeptides are synthesized they are transported into the periplasmic space through a Sec-dependent translocation pathway. When completely folded in the periplasmic space, CT gets secreted into the intestinal lumen using a type II secretion system (T2SS) in its final AB₅ form (Hirst et al., 1984; Steve et al., 2011). There, the B-pentamer binds to the GM1 receptor and helps dock the entire protein complex to epithelial cells (Merritt et al., 1994; Merritt et al., 1995). Data indicate a slight preference of the CT for cells in G₀/G₁ phase perhaps due to the cell cycle dependent expression of GM1 (Majoul et al., 2002). As GM1 moieties are generally localized in lipid raft type structures, after binding to the GM1 the entire toxin gets internalized through either a clathrin-mediated or an ARF6 (ADP ribosylation factor 6) dependent pathway (Fujinaga et al., 2003; O’Neal et al., 2005). This internalized complex gets trafficked to the endoplasmic reticulum (ER) directly or through the Golgi-associated retrograde transport to the ER (Feng et al., 2004), where the CTA becomes free from the CTB and binds to the Erd2p (ER-associated degradation or degradosome pathway) receptor using its C-terminal KDEL sequence (Hazes & Read, 1997; Wernick et al., 2010). The rest of the protein gets
degraded during this step and only the N-terminus of the A subunit (CTA1) escapes the degradation process (Tsai & Rapoport, 2002). CTA1 is the enzymatic domain of CT that enhances the function of adenylate cyclase in the cell by ribosylating $G_{as}$ (O’Neal et al., 2005). The $G_{as}$ subunit dissociates from the $G_{b}g_{y}$ and stimulates adenylate cyclase by physically associating with it. This results in an elevated cAMP concentration inside the cell causing a decrease in sodium uptake and an increase in chloride efflux through the cystic fibrosis trans-membrane conductance regulator (CFTR) (Sonawane et al., 2007). These, together, result in an extreme water loss from the intestine, which in turn causes the watery diarrhea characteristic of the disease cholera. Some other studies have pointed out that CT may promote the release of serotonin from the neuroendocrine cells of the epithelial lining of the intestinal lumen (called enterochromaffin cells) and vasointestinal peptides from epithelial neuronal cells (Lundgen, 2002; Sanchez & Holmgren, 2011). Both of these processes may have a tremendous impact on the human body, but the precise effects of these secretions in cholera are still unknown.
Lipopolysaccharides (LPS) of Vibrio cholerae

Vibrio cholerae LPS, like LPS from other pathogenic bacteria, shows endotoxin properties, antigenic activities, and immune response activation qualities (Chatterjee et al., 2003). All gram-negative bacteria have three distinct layers around the cytoplasm – inner layer, outer layer, and the peptidoglycan layer. LPS can be found attached to the outer layer through its LipidA region whereas the core-PS region and the O-antigen polysaccharide (O-PS) region are projected outwards. The
Lipid A is part of the LPS responsible for its endotoxic properties and the core-PS and the O-PS cause the immunogenic response and vibriocidal antibody production in the host.

The functional roles of LPS are not only limited to the obvious antigenic and endotoxic properties of the molecule, but also perform diverse biological functions. Firstly, LPS helps the bacteria to survive in a harsh marine and host environment by aiding the formation of biofilms. Secondly, LPS performs a key role in adherence to epithelial cell surfaces thereby helping in the intestinal colonization process (Hankins et al., 2012). Also, as a receptor for phage, it performs an important role in population control in V. cholerae (Chatterjee et al., 1984; Maiti et al., 1971).

V. cholerae LPS, in model organisms, demonstrated various endotoxic properties by causing a pyrogenic reaction, local thrombosis (Shwartzmann reaction), mitogenic effects and even death (in mice). Not surprisingly, LPS also possesses characteristics of an adjuvant as described by Kabir et al. (Kabir et al., 1982). V. cholerae LPS acts directly on human cells and alters the morphology of cells resulting in stimulated chemotactic behavior as observed in neutrophil suspensions. The importance of the change in cell morphology and chemotaxis are not clear and more research is needed to understand the bigger picture.
SECTION 2– PORE-FORMING TOXINS

**Bacterial toxins**

Bacteria produce and release a variety of toxic materials that target and kill other bacteria and host cells. Bacterial toxins can be classified into two different groups- endotoxins and exotoxins. Endotoxins are heat stable liposaccharide (LPS) molecules (Figure 1.3), which get released from the outer leaflet of dead or damaged gram negative bacterial outer membranes (Wang and Quinn, 2010). The general structure of endotoxins or LPS consists of lipid A, core sugar and the O-antigens (Raetz et al., 1990). LPS molecules are generally not essential for bacterial growth or colonization, rather they act solely as virulence factors as first identified in *E. coli, S. typhi* and *V. cholerae* infections.

In contrast to endotoxins, exotoxins are proteinaceous in nature and are actively secreted into the environment. These toxins are highly antigenic and can be functionally divided into three categories - exotoxins with enzymatic activity, superantigens and pore forming toxins (PFTs). Examples of exotoxins involved in infection and human diseases include exotoxins from *Staphylococcus aureus, Bacillus cereus, Bacillus anthracis*, etc. (Dinges *et. al.*, 2000; Bottone, 2010; Lowe and Glomski, 2012).

**Pore-forming toxins**

Pore forming toxins (PFTs) or channel forming toxins are one of the largest groups of bacterial cytotoxic factors and are crucial for virulence in a number of
human pathogens (Ferdinand et al., 2013). *Staphylococcus aureus, Escherichia coli, Mycobacterium tuberculosis* and *Bacillus anthracis* are all pathogenic bacteria where PFTs are absolutely required for the infection and progression of the disease. It is worth mentioning that although the main focus of this chapter is bacterial pore forming toxins, many different forms of living organisms use PFTs including plants, insects, reptiles and marine invertebrates (Parker and Feil, 2005). PFTs work by forming pores in the plasma membrane of the target cells. Plasma membranes or cell membranes constitute the physical barrier between the cellular cytoplasm and the extracellular environment. Many bacterial toxins have developed mechanisms to use the membrane architecture to allow interactions with membranes followed by insertion of parts of the toxin into the lipid bilayer. PFTs are almost universally present in the bacterial kingdom and studying them could not only help prevent the spread of pathogenic bacteria, but potentially also provide novel targets for antibiotics.

**Classification of PFTs**

PFTs can be classified in multiple ways. However, the most commonly used classification is based on the type of structure that forms the membrane spanning pore. Based on this criteria, PFTs can be divided into two classes – α-PFTs (α-helical channels) and β-PFTs (β-barrel channels).
**α-pore forming toxins (α–PFTs)**

The best-characterized members of this family include *Escherichia coli* colicins, which kill other *E. coli* cells (Cascales et al., 2007). In addition to the presence of the colicin gene in the colicin-secreting *E. coli*, the bacteria also contain an immunity protein, which protects them from the effects of the toxin. Following secretion from the bacteria, colicins find target cells by binding to surface receptors before being translocated through the outer membrane. After entering the cell, the toxin kills the cell either by forming a pore in the membrane or by functioning as a nuclease. The pore forming region of colicins is composed of an α-helical hairpin with a hidden, central hydrophobic part. Interaction with the membrane unmasks the hydrophobic patch and the channel formation begins through local unfolding of the toxin (Slatin et al., 1994; Parker and Feil, 2005). This has been speculated to be a common mechanism not only for colicins, but also for other α-PFTs.

Another well-characterized α-pore forming protein is the diphtheria toxin (DT), which is one of the most efficient toxin molecules known to mankind (Yamaizumi et al., 1978). DT is a secreted toxin with three individual domains, each of which has a distinct function (Choe et al., 1992). The receptor binding domain (R-domain) helps the toxin bind to target cells by interacting with an epidermal growth factor-like precursor, which results in the endocytosis of the entire toxin molecule (Naglich et al., 1992). Once inside the endosome the second, transmembrane domain (T-domain), undergoes a pH-mediated conformational change to form a membrane spanning pore (Mindell et al., 1994) in the endosomal membrane. The third domain or
the catalytic domain (C-domain) is the only part of the toxin that enters the cytosol and stops the entire translation process by ribosylating EF-2 molecules (Weiss et al., 1995). The DT C-domain is extremely efficient at catalyzing this reaction and only one toxin molecule per cytosol is sufficient to kill a single cell (Yamaizumi et al., 1978).

The ClyA family of α-PFTs are speculated to be important virulence factors and are found in pathogens such as *E. coli, Salmonella enterica* and *Shigella flexneri* (Hunt et al., 2010). As all these bacteria produce multiple types of toxins and virulence factors, identifying the mechanism of toxicity for a particular type of ClyA toxin is extremely challenging and still not clearly understood. ClyA PFTs contain one soluble and one transmembrane region, both with mostly an elongated α-helical structure (Wallace et al., 2000; Mueller et al., 2009). Usually these toxins undergo massive structural rearrangements after interacting with cholesterol rich membranes resulting in oligomerization and insertion of the amino terminal amphipathic α-helix into the membrane. NheA (non-hemolytic tripartite enterotoxin), another member of this family from *Bacillus cereus*, could potentially form a β-barrel membrane-spanning pore (Ganash et al., 2013). However, further studies are needed to understand the mechanism and the significance of this observation.

Cry toxins are insecticidal agents produced by bacteria during sporulation events. These toxins interact with GPI-anchored aminopeptidases (Knight et al., 1995) or glycolipids (Barrows et al., 2006) and partition themselves into the specific lipid microdomains for oligomerization (Gomez et al., 2007; Bravo et al., 2007). The
general structure of Cry toxins shows that the N-terminal domain (domain I) forms the pore-forming region after binding the membrane through its lectin like domain II (Iacovache et al., 2008). Cry toxins also contain a third domain (Li et al., 1991) whose exact function is unknown, but has been speculated to play some role in protein stabilization and receptor binding (Gomez et al., 2006). Further studies are required to understand the extent of function of each domain in the toxicity caused by Cry proteins.

Another group that belongs to the α-PFT family are the actinoporins, which are produced by sea-anemones. The two best-studied actinoporins are equinatoxin II (from Actinia equinia) (Anderluh et al., 2005) and sticholysin II (from Stichodactyla helianthus) (Mancheno et al., 2003). Both of these toxins bind to the lipid membrane as monomers by targeting sphingomyelin rich or phase separated membranes (Kristan et al., 2009; Barlic et al., 2004). These toxins contain β-sandwich regions flanked by two α-helices. One of these helices changes conformation and inserts into the membrane along with 7 other protomers.
Figure 1.4. **Examples of structures of α-pore-forming toxins.** Representative structures of 6 toxins from the α-PFT family are shown in cartoon representation and colored by B-factors. The images were made using PyMOL.
Figure 1.5 **Representative examples of structures of β-pore-forming toxins.** Structures of eight members of the β-pore forming family are shown using cartoon representation and colored by B-factors. Structures shown in panel A-F depict the monomeric soluble-state of the respective PFTs, while G, and H depict oligomeric assembled forms.
**β-pore forming toxins (β-PFTs)**

Aerolysins are toxins produced by bacteria belonging to the family *Aeromonas*, which is known to cause severe gastrointestinal disease in humans (Leclerc et al., 2002). Most of the aerolysins are produced by the species *A. hydrophila*, *A. trota*, and *A. salmonicida* and share 80% sequence identity among them (Howard et al., 1987). Aerolysins are synthesized as a protoxin that is activated after the removal of a C-terminal sequence (Abrami et al., 1998). The activated aerolysins then bind to GPI-anchored proteins (Nelson et al., 1997) on cell membranes and form the membrane spanning voltage-gated channels that ultimately lead to cell lysis or death. X-ray crystal structures and other structural studies have shown four distinct domains that together form an L-shaped molecule. Domains 1 and 2 are essential for cell membrane binding as they interact with glycan complexes (Hong et al., 2002) and GPI-anchored proteins (Mackenzie et al., 1999), respectively. Domain 3 is responsible for oligomerization and pore formation in the membrane. Domain 4 acts as a switch by being proteolytically chopped off before the toxin oligomerizes (Iacovache et al., 2006). Recently, using X-ray crystallography, Cryo-EM, and molecular dynamics, Degiacomi et al. described how the pre-pore complex of aerolysin undergoes a swirling motion and a vertical collapse before forming the mature membrane spanning channel (Degiacomi et al., 2013). Toxins related to aerolysin can be found in a variety of lifeforms like *Clostridium septicum* (α-toxin), *C. perfringes* (ε-toxin), Cnidaria (hydrolysins) and even plants like *Enterolobium contorliquum* (enterolobin) (Ballard et al., 1993; Sher et al., 2005; Sousa et al., 1994).
Anthrax toxin is another deadly β-pore forming protein produced by *Bacillus anthracis* and is almost solely responsible for mortality due to anthrax infection (Young and Collier, 2007). This toxin is a binary toxin that consists of three exotoxins – edema factor (EF), lethal factor (LF) and protective antigen (PA). PA forms the transmembrane channel and is absolutely necessary for the progress of the infection. Without PA, both the EF and LF lack any toxic activity and are totally harmless when administered in animal models (Pezard et al., 1991). As evident from the name, PA is the most immunogenic of the three virulence factors and functions as a translocation machine for EF and LF to enter the cytosol. LF is a Zn\(^{+2}\) dependent MAP kinase protease and EF is a Ca\(^{+2}\) and calmodulin dependent adenylate cyclase.

After the PA binds to either tumor endothelial marker-8 (TEM8) or capillary morphogenesis protein 2 (GMG2) (Scobie et al., 2003; Bradley et al., 2001), it is proteolyzed losing a 20 kD fragment. This activates the toxin (PA\(_{63}\)) to form a prepore (Lacy et al., 2004). The LF and the EF bind to the prepore complex (Mogridge et al., 2002) and are internalized into the endosome as a ternary protein complex. Inside the endosome, the low pH activates the prepore through a multistep process and the prepore converts itself to a mature transmembrane pore (Jiang et al., 2015). The LF and EF then use this pore to translocate to the cytosol catalyzing their corresponding reactions to cause cell death.

Cholesterol dependent cytolysins (CDCs) are a massive family of β-PFTs with more than 20 identified members (Tweten et al., 2001). The unique feature that groups all CDCs together is the absolute requirement of cholesterol in forming large
transmembrane pores consisting of 20-100 monomers, another characteristic of CDCs. All members of this family share 40-80% sequence similarity (Parker and Feil, 2005) and generally contain a C-terminal undecapeptide sequence known to be involved in membrane binding (Sekino-Suzuki et al., 1996; Jacobs et al., 1998). Cholesterol is absolutely required for the pore formation process, however it is not necessary for the membrane-binding step (Giddings et al., 2003). In some studies, membrane binding occurs independent of cholesterol and through a protein receptor, as revealed in intermedilysin (Nagamune et al., 1996).

The most studied CDC is perhaps Perfringolysin O (PFO) from Clostridium perfringens. The water-soluble monomer structure of the toxin was solved by Rossjohn et al. and revealed an overall elongated rod shaped morphology (Rossjohn et al., 1997). Like aerolysins, PFO also contains four distinct domains rich in β-sheet type structures. Domain 1 is responsible for oligomerization, domain 2 likely helps in prepore formation, domain 3 contains the transmembrane segment of the toxin and domain 4 possesses the cholesterol-rich membrane-binding activity (Shepard et al., 2000; Heuck et al., 2002; Shimada et al., 2002). Because of the cholesterol binding activity of domain 4 of PFO, researchers are trying to use it as a potential probe for cholesterol-containing lipid-rafts (Yoshiko et al., 2004).

Staphylococcus aureus produces another variety of PFT, which includes α-hemolysin and the leukocidins (Gouaux et al., 1997). The leukocidins are bicomponent toxins that require both a class F protein (LukF-PV, HlgB, LukD, LukF-I) and a class S protein (LukS-PV, LukE, LukM, HigA, HlgC, LukS-I) to form the
mature pore (Kaneko et al., 2004; Olson et al., 1999). All F proteins and S proteins share significant sequence identity within the same group (71-79% for F and 59-79% for S), but the similarity fall down to 20-30% across groups (Prevost et al., 2001). α-hemolysin is secreted by the bacterium as a monomer, which after binding to the receptor on the membrane (Wilke et al., 2010), forms a homo-oligomeric heptameric channel in the target cell membrane. Overall, α-hemolysin, LukF, LukS and HγII have similar molecular weights, isoelectric points, common central glycine patches and contain a significantly conserved hydrophobic core.

The recent discovery that ADAM-10, a Zn^{2+} dependent metalloprotease, is the receptor of α-hemolysin paved the way to examine a new function of the toxin in infection (Wilke et al., 2010). This new understanding of the toxin-ADAM10 interaction does not only enable a more detailed understanding of the protein-lipid interaction, but also provides an insight into how receptor binding and pore formation alter the cellular signaling. It is known that S. aureus produces multiple virulence factors and that they act in concert, however how the interplay between different toxins leads to pathogenesis has yet to be answered.

**Membrane insertion mechanism**

The membrane insertion mechanisms for pore-forming toxins has been poorly understood. For colicins, the “umbrella model” was proposed where membrane-spanning helices insert in a sequential manner causing the pore to open up like an
umbrella upon completion (Lakey and Slatin, 2001). Other α-PFTs have distinct domains responsible for each step of the pore formation. Generally it is thought that membrane bound receptors bring monomeric toxin molecules in close proximity to each other, thereby changing a search from a 3-dimensional to a 2-dimensional space problem (Iacovache et al., 2008). Some studies have speculated that toxins like diphtheria or equinatoxin undergo a local unfolding event before getting inserted into the membrane, a phenomenon termed the “molten globule model” of membrane insertion. Many of the toxins from the α-PFT family regulate their pore formation activity through the pH of the environment they are in. Only at a proper pH (generally acidic) do key acidic amino acids become neutrally charged (less hydrophilic), ultimately helping in membrane insertion. Relatively recently, another interesting and controversial aspect of pore formation by some α-PFTs, like colicins and equinatoxin, came to light. Both toxins are shown to form pores comprising of both lipids and protein molecules with a significantly varying stoichiometry between lipids and proteins (Sobko et al., 2004; Anderluh et. al. 2004). Further studies are needed to understand the significance of this observation.

Some common features amongst the β-PFTs include proteolytic activation (aerolysin, anthrax) and major domain rearrangements while forming a transmembrane pore. All members of the CDC and anthrax toxin family exhibit a very similar transition of a loop or α-helix region into transmembrane β-sheets. In the cases of S. aureus α-hemolysin or aerolysins, the tip of the hairpin loop in the membrane-spanning region has been speculated to play a key role in the insertion
mechanism. In both alpha as well as beta pore forming toxins, membrane components play crucial roles in the pore formation process. Several membrane components help in binding of monomers in local lipid microdomains (Waheed et al., 2001) or directly promote the oligomerization process (Abrami et al., 1999). In addition, some components also aid in permeabilization of lipid bilayer through phase inversion (Alonso et al., 2000).

**Cellular responses to pore formation**

Pore forming toxins are essential virulence factors for many bacterial infections like *Staphylococcus aureus, Clostridium speticum, Bacillus anthracis* and *Corynebacterium diphtheria* (O’Callaghan et al., 1997; Kennedy et al., 2005; Orlik et al., 2005; Hoch et al., 1985). These toxins form membrane-spanning pores and destroy the transmembrane ion gradients. Eukaryotic cells have evolved to deal with these pores since *trans*-bilayer pores are commonly found due to the action of toxins and various natural physical processes.

These natural pores are induced by mechanical stress (Karatekin et al., 2003) and thermal fluctuations (Benz et al., 1975). The dynamics of pores formed due to mechanical stress is solely dependent on the surface tension (for growing) and the line tension (for shrinking) of the membrane (Brochard-Wyart et al., 2000). These types of transient pores are almost omnipresent in cell membranes and are resealed by the line tension of the bilayer as seen in giant vesicle systems (Sandre et al., 1999).
Thermally induced membrane pores are the indirect effect of thickness fluctuations due to thermal undulation of the membrane. The energy barrier for a rupture or pore to appear due to thickness variation is very high (~91 k_B T) (Popescu et al., 1991), but once formed these elliptical holes might advance to a meta-stable state lethal to the cell (Popescu et al., 1991). The influence of parameters like temperature, lipid composition, cholesterol content, lipid head-group size, and hydrophobic core thickness on naturally occurring membrane pores has been studied extensively, but further studies are needed to ascertain the significance of pore formation and resealing on cellular morphology and health (Movileanu and Popescu, 2004).

In order to survive, cells have to detect pores, which have already formed on the membrane or a faulty constitutively open channel (which essentially acts as a pore in the membrane) and react accordingly to stop the effect of the osmotic shock. Cells have evolved for millions of years to deal with such a scenario, as naturally occurring pores sometimes do not get resealed quickly. In those cases, mainly a general p38 mitogen activated protein kinase (MAPK) dependent pathway rescues the cell from the osmotic shock and promotes cell survival (Husmann et al., 2006). In the event of pore formation by PFTs, cells have learnt to use the same pathways to combat against these foreign molecules. Other mechanisms like the caspase-1 dependent SREBP pathway for survival and the Ca^{+} mediated NF-κB (Dragneva et al., 2001) pathway for inflammations are major cellular responses to PFTs. In a recent study, Kao et al. described in Caenorhabditis elegans that along with p38, another MAPK dependent pathway called JNK (c-Jun N-terminal kinase) acts as an important part of the so
called core PFT defense network (Kao et al., 2011). In the same study, the authors also found that almost 0.5% (106 of ~20,000) of the entire *C. elegans* genome is involved in protecting cells in case of a PFT attack. All of these *hpo* (hypersensitive to pore forming toxin) genes together (Kao et al., 2011) constitute the cellular non-immune defense (CNID) against the PFTs (Aroian and Goot, 2007).

The cellular responses against large pore forming CDCs has been examined in detail by various groups. Upon infection by CDCs, the cell membrane shows two different phenotypes. Firstly, the infected membrane areas are endocytosed and Ca\(^{2+}\) dependent resealing begins. The endocytosed vesicles containing PFT pores are trafficked for complete degradation (Corrette et al., 2012). Some evidence suggests that sphingomyelin performs multiple roles in the process by initiating endocytosis and the membrane repair process (Tam et al., 2010). Secondly, the infected cells display membrane blebbing and shedding events, generally associated with apoptosis (Kerr et al., 1972), but in these cases, these processes protect the cells from osmotic lysis (Babiychuk et al., 2011). Inside cells, CDCs instigate a variety of effects, which include, but are not limited to, MAPK/IL-8 dependent immune response (Dogan et al., 2011), TLR-4 mediated inflammation (Dogan et al., 2011), disruption of the endolysosomal network (Cassidy and O’Riordan, 2013), and a change in mitochondrial morphology facilitated by Ca\(^{2+}\) (Stavru et al., 2011).

Compared to CDCs, other PFTs like aerolysin, α-hemolysin, and *Vibrio cholerae* cytolysin (VCC) form much smaller pores and can evade the destruction process for longer thereby promoting extended ion dysregulation across the
membrane (Gonzalez et al., 2011). Otherwise, smaller PFTs are suspected to exhibit similar effects on the membrane and cytoplasm.

Role of PFTs in infections

Inside the infected host, PFTs not only lyse host cells for nutrients and protection, but also abolish the compactness of the epithelial and endothelial linings helping the bacteria to colonize the host. PFTs also disrupt the host’s immune response to infection by lysing immune cells, helping the bacteria form colonies or biofilms, inhibiting immune cells to target colonies, and by taking over host cellular defenses (Los et al., 2013). An increasing number of studies have described the role PFTs play in bacterial colony formation. PFTs aid bacterial growth by weakening epithelial barriers, causing inflammation, triggering vascular effects, and by modifying the extracellular matrix (Los et al., 2013). Toxins liberate essential nutrient molecules by either lysing host cells or by promoting programmed cell deaths as seen in vivo for PLY and VCC (Garcia-Suarez et al., 2007; Saka et al., 2008). Toxins like α-hemolysin, PLY and ALO take a step further and unsettle cellular junctions causing barrier dysfunction, a vital step in many infections including S. pneumoniae (Orihuela et al., 2004) and S. aureus (Diep et al., 2008).

In the case of PLY, the PFT is responsible for the critical step of penetrating the blood-brain barrier, but has no role in bacterial growth (Orihuela et al., 2004; Hirst et al., 2004). PVL is shown to be essential for bacterial growth in the lungs and blood of
mice probably through complement activation (Los et al., 2013). In *V. cholerae*, the VCC and MARTX (Multifunctional Autoprocessing Repeats-in-Toxin) toxins promote the colonization of the small intestine and can cause lethality in mouse models (Olivier et al., 2007; Olivier et al., 2009).
SECTION 3– VIBRIO CHOLERAE CYTOLYSIN (VCC) AND PATHOGENESIS

Vibrio cholera: a brief history

*Vibrio cholera* produces a variety of toxins and exoenzymes to aid the bacteria in infection and pathogenesis. *Vibrio cholerae* cytolsin, or VCC (also called HlyA, El Tor hemolysin, *V. cholerae* hemolysin, entero-cytolysin or EC), is one of these virulence factors synthesized from the hlyA gene on chromosome 2. VCC was isolated and identified as a thermolabile direct hemolysin with cytotoxic and cardiotoxic properties by Honda et al. from a *V. cholerae* El Tor strain (Honda and Finkelstein, 1979). Many groups at that time studied the microbiology of strains containing hemolytic capabilities and found that the hemolytic agent was secreted from cells, but very little biochemical information was published (Feeley and Pittman, 1963; Felsenfeld, 1964). Primary characterization of the protein showed many striking functional similarities with known lethal toxins from *Vibrio parahaemolyticus* and other bacterial species (Honda and Finkelstein, 1979). The identification of the hemolysin molecule was very exciting because for a long time no precise information regarding the nature and the role of the hemolysin was available after Robert Koch isolated the hemolytic cholera (El Tor) stains almost a century ago. The spread of the hemolysin genes was thought to be limited to a few El Tor strains until Brown and Manning discovered the presence of hlyA in all the O1 and non-O1 strains tested (Brown and Manning, 1985). The authors also described that many of
the strains with the hlyA gene behave as non-hemolytic strains under common test conditions because of differences in gene regulation.

**Identification of VCC as a pathogenic agent**

Due to the presence of a number of known toxin molecules in *V. cholerae*, it was previously thought that VCC does not play a key role in the pathogenicity and virulence of the bacteria. This notion changed when Ichinose et al. reported severe enterotoxic abilities of the toxin using animal models and proposed a role in gastroenteritis caused by CT− strains (Ichinose et al., 1987). In this study, purified VCC was able to induce massive mucous and bloody fluid accumulation when added to ligated rabbit ileal loops or when injected intraintestinally in infant rabbits. Similar fluid buildup was also observed following oral administration of the toxin in suckling mice along with a noticeable histochemical change in the mucosa (Ichinose et al., 1987). In another study, deletion of the HlyA gene led to a ~30-fold increase in the LD₅₀ in infant mouse models (Willaims et al., 1993). Similarly in humans, a number of gastroenteritis cases have been reportedly caused by CT− HlyA+ hemolytic strains of *V. cholerae* indicating the pathogenic role of the hemolysin in humans (Honda et al., 1988; Morris et al., 1984; Zitzer et al., 1997). However, both CT and HlyA negative strains resulted in rather mild watery diarrhea when tested in human volunteers signifying role of other virulence factors in the disease (Levine et al., 1987).
**Effects of VCC on eukaryotic cell lines**

In 1990, Hall et al. determined the correct molecular weight of the VCC protein to be ~80 kD. They were also the first to suggest that the precursor of VCC undergoes proteolytic cleavage and identified the region that is cleaved (Hall and Drasar, 1990). Further studies have shown that VCC forms hydrophilic pores in the membrane and that the pore-forming activity of the toxin is responsible for lysing erythrocytes and other cultured cells (like CHO, Vero, MDCK, BHK-21, L-929, L-41, HEp-2) (Zitzer et al., 1993). Work on α-hemolysin from *Staphylococcus aureus* has also established membrane permeabilization as a common mechanism of hemolysis and cell damage by hemolysin molecules (Walev et al., 1993; Walev et al., 1994). Later, VCC membrane spanning pores were identified as oligomeric SDS-stable aggregates with an approximate pore size of 0.7 nm (Zitzer et al., 1995; Sathyamoorthy et al., 1997).

In intestinal 407 cells, VCC exhibits acute toxicity and leakage of cytosolic potassium ions (K⁺) followed by a drastic exhaustion of the cellular ATP level. Remarkably, VCC pores were also found to be somewhat impervious to calcium ions (Ca⁺) and did not cause any measurable DNA degradation in intestinal cells unlike α-hemolysin (Zitzer et al., 1997; Jonas et al., 1994). When similar assays were done in Caco-2 cells, VCC stimulated major chloride (Cl⁻) secretion from the cytosol without changing the paracellular permeability (Debellis et al., 2009). This efflux of chloride ions is dependent on epithelial chloride channels and independent of sodium (Na⁺) movement through membrane. In theory, the VCC mediated efflux of chloride would...
stimulate sodium and water secretion, which are the main components of fluid accumulation in the intestinal track, making VCC a direct diarrhogenic agent (Debellis et al., 2009).

A new interaction between the VCC toxin and host was recognized when it was found that the so called V. cholerae vacuolating protein is HlyA (Coelho et al., 2000; Mitra et al., 2000). Vacuolization caused by VCC in vero cells follows a similar mechanism to aerolysin and is endoplasmic reticulum (ER)-mediated in the beginning (1-4 hour incubation). Over a duration of up to 16 hours, the smaller early vacuoles transform into massive endosome-like multivesicular acidic compartments (Figueroa-Arredondo et al., 2001), called autophagosomes. As a part of the prosurvival pathways, these autophagosomes are associated with destruction of cellular macromolecules and organelles under conditions of starvation or stress. In Cho cells, VCC not only leads to autophagosome-mediated vacuolization, but also colocalizes to the vacuole membrane (Gutierrez et al., 2007). This autophagic response helps the cell to survive upon infection by VCC, and obstructing this pathway causes cell death more rapidly (Gutierrez et al., 2007). It has been proposed that this phenomenon contributes to the pathogenicity of V. cholerae by indirectly blocking ER mediated trafficking and ion transport in intestinal cells, which may in turn negatively affect the inflammatory or defense responses in the human lower alimentary canal (Figueroa-Arredondo et al., 2001). The cellular reactions to VCC has opened up a new frontier of research to investigate the potential cytotoxic roles of the toxin in the gastro-intestinal cells and thus, the pathogenicity of V. cholerae.
Although Zitzer et al. did not observe any VCC induced DNA destruction in the intestinal 407 cell line, similar studies in Caco-2 cells displayed a completely different outcome. Saka et al. performed a detailed investigation of the effect of VCC on the Caco cell line and observed that VCC is capable of instigating a major DNA-fragmentation event, which is generally coupled with apoptosis (Saka et al., 2008). This DNA fragmentation event was found to be directly dependent on VCC ion channel activity and blocking the channel always prevented DNA damage. Identical results were obtained in experiments conducted using HeLa, CHO, and COS-7 cell lines. Upon detailed examination, several attributes of programmed cell death such as a surge in sub-diploid DNA content and caspase-3 activation were also found to be associated with the aftermath of VCC infection. Using electron microscopy, Saka et al. additionally confirmed that VCC is capable of causing apoptosis-related features like chromatic condensation, DNA-destruction, and membrane blebbing in cells both in vitro as well as in vivo. Additionally, VCC has been shown to induce inflammation in human intestinal epithelial cells (Ou et al., 2009) and cause rapid mortality when administered in mice (Olivier et al., 2007). The pro-inflammatory effect of VCC has also been demonstrated by studies using purified proteins, where lower concentrations of VCC alone are able to induce the production of IL-8 and TNF-α without any cell death. However, the levels of IL-8 and TNF-α decrease rapidly at higher VCC concentrations along with a sharp surge in the transepithelial electric resistance (TER) (Ou et al., 2009).
PrtV, a metalloprotease involved in the host response against bacterial infection in humans and *Caenorhabditis elegans* (Ou et al., 2009; Vaitkevicius et al., 2006), actively contributes in the response against VCC infection. Moreover, VCC is degraded by PrtV when LB broth is used, but usage of an epithelial lumen-like BHI medium inhibits the protease, allowing more time for the toxin to work.

**Effect of VCC in *C. elegans***

Identifying the precise response to the infections caused by VCC in humans has always been a challenge for obvious reasons. *C. elegans*, on the other hand, provides a simpler platform for studying the effects of not only VCC, but several other pore forming toxins. A genome wide analysis of gene expression in response to VCC was performed by Sahu et al. and at least fifty five (55) genes in *C. elegans* were identified with an over two-fold increase in expression when infected with *hlyA*+*V. cholerae*. Many of these genes were previously characterized as genes associated with the *C. elegans* innate immune response and when silenced, seven out of the nine genes caused significantly more fatality against *hlyA*+*V. cholerae* strains. In addition, the activation of blocked UPR (ABU) genes, which are a part of prion-like glutamine/asparagine (PQN) clusters, were also induced upon VCC infection. ABU/PQN genes have been traditionally linked to the unfolded protein response (UPR) pathway, which is instrumental in regulating *C. elegans* immune response against bacterial infection (Haskins et al., 2008; Sahu et al., 2012). At least 40% of
the PQN genes (twenty nine out of seventy one) and 80% of the ABU genes (nine out of eleven) in *C. elegans* showed disproportionate expression profiles between infection with *hlyA*<sup>+</sup> and *hlyA*<sup>−</sup> *V. cholerae*, and at least two of these genes resulted in higher mortality when knocked out (Sahu et al., 2012). When compared with the effects of the insecticidal toxin Cry5B, more than 10% of the genes (274 out of 2078) exhibiting major differential expression were found to be common between HlyA and Cry5B infections. Another similarity between Cry5B and VCC is that the hypoxia-inducible factor-1 (HIF-1) mediated hypoxia pathway is activated in response to both toxins immediately after infection, which is not surprising as the autonomous hypoxia pathway is expected to activate as a protective measure against pore forming toxins (Bellier et al., 2009).

**Structure of the VCC monomer**

VCC is synthesized as an inactive pro-toxin and requires proteolytic removal of a small ~15 kD pro-domain for activity. *V. cholerae* uses the soluble hemagglutinin/protease, which is involved in activation of cholera toxin (CT) (Booth et al., 1984), for this proteolytic cleavage reaction (Nagamune et al., 1996). This activation process is not dependent on a precise cleavage site nor protease, but can instead be activated by a variety of exogenous proteases like thermolysin, trypsin, α-chymotrypsin, and papain - to name a few. The target for all proteases was identified to be a loop (residues 118-133) connecting the N-terminal pro-domain with the core
cytolysin domain of the toxin. In recent years, a member of the human metalloprotease family called ADAM-17, or the tumor necrosis factor α converting enzyme, has been identified as other proteases that activate VCC (Valeva et al., 2004).

A number of structure-related questions regarding VCC were answered when Olson et al. crystallized and solved the structure of the VCC monomer to 2.3 Å (Olson and Gouaux., 2005). The structure correlated well with previous speculations and predictions based on sequence homology (Olson et al., 2003). Overall, the protein contains four domains – an N-terminal pro-domain, a cytolysin domain, a β-trefoil lectin domain and a β-prism lectin domain. The pro-domain, which may be related to the Hsp90 chaperone, acts as a switch to prevent pore formation. As mentioned above, proteolytic removal of this domain is necessary for the activation of the toxin, which is a common phenomenon among pore forming toxins like anthrax PA channel, diphtheria toxin, and aerolysin (Petosa et al., 1997; Choe et al., 1992; Parkaer et al., 1994). As discussed previously, the proteolytic activation event is carried out by proteases from V. cholera, proteases on the host cell surfaces or by in vitro proteases like trypsin (Nagamune et al., 1996; Olson and Gouaux 2005; Valeva et al., 2004). The pro-domain is essential for the proper secretion of the toxin in the extracellular space and is also found to be important for refolding of the protein when denatured (Nagamune et al., 1997). Upon proteolytic cleavage of the 15 amino acid long loop region, the pro-domain remains attached to the rest of the protein, as seen in chromatographic analysis (De and Olson, unpublished data). Removal of this domain
causes structural transitions, which completely alters the unfolding profile of the mature toxin (Paul and Chattopadhyay, 2011).

The cytolysin domain forms the core of the protein and exhibits similarity to the Staph toxins (Olson et al., 1999; Olson and Gouaux, 2005). The cytolysin domain additionally contains the so-called pre-stem region with two β-sheet structures that penetrate the membrane and form the mature channel. The pore forming pre-stem in VCC is longer than its counterpart in both LukF and α-hemolysin by three and five amino acids, respectively, but shorter than the anthrax loop. The membrane proximal region of the cytolysin domain is also known as the rim domain and in other PFTs is involved in binding to lipid or protein receptors (Galdiero and Gouaux, 2004). However, the VCC rim domain doesn’t contain the putative lipid binding site seen in α-hemolysin, but single amino acid mutations in the rim domain causing a drastic loss of activity have been reported (Paul and Chattopadhyay, 2012). The third domain of VCC is the β-trefoil domain, but which displays sequence similarity to the lectin domains of Ricin (Montfort et al., 1987). The β-trefoil domain is present in many toxins from the Vibrionaceae family and at least in one of them, V. vulnificus hemolysin (VVH), this domain is active in binding to carbohydrate moieties (Kaus et al., 2014). Lastly, the fourth domain is the active β-prism lectin domain, which targets cell surface carbohydrate molecules and increases the local concentration of the toxin on the membrane (Saha et al., 1997; Zhang et al., 1999). The β-prism lectin domain, like the β-trefoil domain, has a pseudo-three fold symmetry (Olson and Gouaux, 2005) and can be found in related toxins from other Vibrio species.
Although once speculated to bind terminal β-galactosyl residues, recent studies (described in this thesis) identified N-linked glycan heptasaccharide cores as specific binding targets (Levan et al., 2013). My hypothesis is that by binding to N-glycans, the β-prism lectin domain helps the toxin target membranes. This explains the massive loss of hemolytic activity in β-prism lectin domain truncated VCC (Olson and Gouaux, 2005; Levan et al., 2013). Some studies have proposed additional roles for the β-prism lectin domain in oligomerization and the pore formation process (Rai and Chattopadhyay, 2013; Dutta et al., 2010), but in my opinion no definitive evidence exists yet.

In brief, VCC is a PFT produced by the human gastro-intestinal pathogen *Vibrio cholerae* to lyse human cells and is suspected to be a major virulence factor in diarrhea causing CT TCP *V. cholerae* infections. Although it was long thought to have no involvement in the pathogenicity of the disease, VCC has now been shown to be a potent diarrhogenic agent in both human intestinal cell lines and in mouse models (Debellis et al., 2009). VCC induces fluid accumulation and Ca\(^+\) efflux via a plethora of molecular mechanisms. At low concentrations, VCC instigates apoptosis, vacuole and autophagosome formation. Infected cells respond to VCC infection by overexpressing inflammatory molecules like IL-8 and TNF-α and can undergo a hypoxia pathway as a protective measure (Khilwani et al., 2015). Continuous battles like these between PFTs and infected cells determine not only the fate of the cell, but also the progress of the disease in the infected animal.
The crystal structure of the VCC monomer is known and provided a great deal of insight into the structural aspect of the toxin. However, many questions regarding the oligomeric assembly process and the structural aspect of the final assembled pore was unknown. My studies with the heptameric assembly will add to this knowledge and help answer the questions regarding the structural aspects of the membrane bound pore.

Figure 1.6. Cytotoxicity of VCC. In this set of experiments, the effect of VCC was tested on Caco-2 cells, a human epithelial colorectal adenocarcinoma cell line. The image in the first panel shows cultured cells exposed to buffer alone while the second panel shows cells exposed to VCC (75 ng/mL). Trypan blue uptake was used as a measure of cell viability. Visual comparison revealed higher number of blue cells in the presence of VCC. Experiment conducted and data analyzed by Anand Sitaram (unpublished data from Sitaram and Aroian, 2016).
SECTION 4– MEMBRANE BINDING OF VCC

The lipid membrane and cholesterol

Cholesterol (greek chole – bile, greek stereos – solid, IUPAC ol -- alcohol) is an absolutely indispensible constituent of the animal plasma membrane. It was first discovered by M.E. Chevreul in 1815 as a major constituent of human cholelith or gallstone (Goluszko and Nowicki, 2005). Later, cholesterol was characterized as an amphipathic molecule containing four hydrocarbon rings and a single hydroxyl group (-OH) at the 3β position. As a consequence of this chemical structure, cholesterol experiences immense hydrophobic repulsion by solvent and prefers to partition with hydrophobic molecules like lipids. Along with hydrophobic repulsion, intermolecular interactions provide a greater energetic stability to the cholesterol micelle structure resulting in a remarkably low (25-40 nM) critical micelle concentration (CMC) in water (Haberland and Reynolds, 1973).

Cholesterol is responsible for a number of structural functions like regulating membrane fluidity, upholding diffusion barrier characteristics and modulating the thickness of the membrane. Consequently, it is dispersed very systematically throughout the membrane (Haberland and Reynolds, 1973; Gimpl, 2010; Yeagle, 1985). Moreover, in addition to structural functions cholesterol performs a variety of biochemical roles in the cell. For example, cholesterol influences the activity and stability of membrane proteins (Gimpl, 2010), acts as a substrate for various
biochemical enzymes (Hanukoglu, 1992), and takes part in bile and steroid hormone
(Burger, 2000; Gimpl, 2010) production inside the animal body.

The fluid mosaic model for the lipid bilayer was postulated based on the
thermodynamic aspects of the membrane components (Singer and Nicolson, 1972).
The idea of ordered membrane microdomains, or lipid rafts, with a high content of
cholesterol and sphingomyelins, was primarily developed to describe protein
targeting and sorting in the membrane. Over the years, many reports have been
published in support of the lipid raft model and >100 proteins associated with lipid
rafts have been identified (Simons and Van, 1988; Simons and Ikonen, 1997; Simons
and Toomre, 2000; Simons et al., 2000; Fielding and Fielding, 2000; Chatterjee and
Mayor, 2001; Goluszko and Nowicki, 2005; Lingwood and Simons, 2010). The direct
proof of variable microdomains came from electron microscopic image analysis of
frozen membranes and they were postulated to be in the liquid-ordered (L_0) state.
Although these small dynamic domains were primarily thought to be involved in
protein sorting, new evidence has allied them with additional functions like signal
transduction, cholesterol trafficking, and vesicular transport. Some proteins with
characterized or identified cholesterol interacting motifs include caveolin, cholesterol
dehydrogenase, cholesterol oxidases, the retinoic acid-related orphan receptor
(RORα), the Sigma-1 receptor, the TSPO translocator protein, the β2-adrenergic
receptor, oxytocin receptors, and nicotinic acetylcholine receptors (Gimpl, 2010).
Most of these examples play authoritative and essential roles in various crucial
cellular pathways. Furthermore, as these microdomain structures are predominantly
present in all cell membranes, many pathogens have evolved to use them as a platform for membrane recognition, assembly and cellular escape (Goluszko and Nowicki, 2005). Pathogens like HIV from *Helicobacter pylori* and virulence factors like cholesterol dependent cytolysins (CDC) from pathogenic bacteria have exhibited firm requirements for the presence of these microdomains for successful infection. Some studies have also reported evidence of self-associated cholesterol crystalline domains in mammalian smooth muscle cell membranes associated with the regulation of ion channels and DNA replication (Mason et al., 2003).

Although the function of cholesterol microdomains in regulating cellular pathways is universally accepted, the present lipid raft model has been challenged by many critics (Munro, 2003; Shaw, 2006; Kenworthy, 2008). Firstly, the modeled average size of the lipid rafts is ~20 nm, which can hold only 60-80 lipid molecules and ~20 membrane proteins, too small when compared to the many different functions they perform. Secondly, the half-lives of these microdomains is measured to be around 100 nanoseconds, which is too fast compared to the turnover rates of most enzymes associated with rafts. Thirdly, the lipid raft model is unable to explain raft heterogeneity. As an alternative model, a protein-mediated creation of larger microdomains has been proposed where similar structures are formed by protein-protein interactions and protein-driven cholesterol sequestration in the membrane (Shaw, 2006; Gimpl, 2010). The major difference between this new raft model and the old lipid raft model is that it permits protein molecules in the raft to organize a reformation of the microdomains after they are disintegrated by the natural propensity
of the lipids. This in turn can clarify how slow enzymes can function in nanosecond timescale rafts. Additionally, it can also explain the size paradox by allowing communication between proteins from neighboring lipid rafts and the existence of different kinds of rafts by protein dependent variability in lipid structures (Shaw, 2006).

**Cholesterol interactions in PFT family**

**Cholesterol dependent cytolysins (CDCs) and their dependence on cholesterol for membrane binding and insertion**

Among the β-PFT family, which forms a β-barrel transmembrane channel, there is a large group of toxins absolutely dependent on cholesterol for their activity. Although originally identified as virulence factors and thought to be present only in Gram-positive pathogens, recent discoveries have recognized multiple Gram-negative bacteria with cholesterol dependent cytolysins (CDC). At present, at least twenty species of Gram-positive bacteria are shown to produce CDCs including *Arcanobacterium*, *Bacillus*, *Clostridium*, *Listeria*, and *Streptococcus* and at least two Gram-negative bacteria - *Desulfobulbus* and *Enterobacter* (Heuck et al., 2009; Tweten, 2005; Hotze and Tweten, 2012; Hotze et al., 2013). Two characteristics of these group of toxins separate them from the rest of the β-PFTs. Firstly, their utter dependence on cholesterol to form membrane spanning pores in target membranes and secondly, the remarkably large β-barrel pores (~250 Å in diameter) (Heuck et al., 2009). These toxins, like other β-PFTs, are generally produced as water-soluble
monomers and secreted into the extracellular space through a type II secretion system (Tweten, 2005). Although most of the CDCs have a cleavable signal sequence required for their secretion, some exceptions like *Streptococcus pneumoniae, S. mitis* and *S. pseudoneumoniae* have been identified in recent years. Upon binding the target membrane, these toxin monomers form oligomeric pores in membranes and are crucial for pathogenicity in bacterial infections caused by *Streptococcus intermedius* and *Clostridium perfringens* (Tweten, 2005). Even after the first CDC was discovered more than a century ago (Heuck et al., 2009), many questions regarding the molecular mechanism of membrane binding and pore formation are yet to be answered.

A study of the core sequences (without the signal peptide) of twenty-eight CDC proteins has revealed a remarkable 28 to 99% similarity, and up to 74% identity, among them. Across different genera the similarities and identities decrease, but toxins like Pneumolysin (PLY from *S. pneumoniae*), Perfringolysin (PFO from *C. perfringens*) and Anthrolysin (ALO from *B. anthracis*) share a striking 67-88% similarity and 46-72% identity (Heuck et al., 2009). As a result of this sequence similarity, all CDCs exhibit similar membrane binding and pore formation activities and similar overall 3D structures. Interestingly, cholesterol rich membranes are not the only prerequisite for CDCs binding to target cells. For example, intermedilysin (ILY from *S. intermedius*) specifically binds to a membrane bound receptor, CD59, in addition to requiring cholesterol rich membranes for pore formation (Giddings et al., 2004). On the other hand, cholesterol in the membrane is an absolute constraint for PFO binding to membranes (Heuck et al., 2009). Either way, after targeting a
membrane, CDC monomers oligomerize and form a pre-pore complex on the membrane with ~35-50 monomers. Monomeric CDCs, like PLY, oligomerize even in absence of a membrane at very high concentrations (~µM), but overall the pre-pore formation step is highly dependent on the composition of the membrane.

Cholesterol and its relationships with different membrane components has been investigated in detail. The presence of cholesterol in the membrane is known to increase the thickness as well as the stiffness of the bilayer, both of which are instrumental in membrane-protein interactions. Furthermore, cholesterol, being a cone shaped molecule and thereby occupying a larger surface area, regulates membrane fluidity, changes the lateral segregation of membrane lipids, and enhances the line-tension (energy in unit length of the membrane) of the membrane. All of the above phenomena are responsible for the dynamic health of the biological membranes. Biochemically, cholesterol aids in different biosynthetic pathways, protein sorting in the ER and golgi, COPI-mediated protein transport, and in maintaining the rigidity of the apical membrane. Some types of cholesterol-rich domains are known to have immiscible characteristics with other membrane components (Sankaram and Thompson et al., 1991). Although controversial, cholesterol along with sphingomyelin, forms lipid-rafts on membranes providing an opportunity for nanocluster formation necessary for function and stability (Goswami et al., 2008; Simons and Ikonen, 1997). In principle, cholesterol in the membrane interacts and forms stoichiometric complexes with membrane phospholipids and upon concentration increase, forms a homogeneous phase with lipid molecules. Additional
cholesterol cannot achieve a stable interaction with the membrane, which is why it is called free cholesterol with a propensity to occasionally leave the bilayer (Radhakrishnan and McConnell, 1999; Mason et al., 2003; Heuck et al., 2009).

The paradigm of CDCs requiring cholesterol for membrane binding/pore formation mainly came mainly from two separate observations. Firstly, pre-incubation with cholesterol prevents hemolysis by CDCs (Tweten, 2005) and secondly, many CDCs bind to cholesterol molecules in direct cholesterol-binding assays (Johnson et al., 1980; Tweten, 2005). Later, direct evidence of this dependence came from using PFO as a model CDC (Iwamoto et al., 1987; Iwamoto et al., 1990). Questions about the stoichiometry of binding is still under investigation. In some cases (PFO) cholesterol dependence is so strict that binding occurs when incubated with ≥30% cholesterol rich membranes, but not with membranes containing <30% cholesterol (Heuck et al., 2000). This indicates a complex mechanism of binding. Flanagan et al. proposed that the state of cholesterol in the membrane and the association capacity of lipids are key factors behind such observations. Accordingly, when the amount of cholesterol surpasses the phospholipid’s capacity to interact, excess cholesterol exists as free cholesterol, with which CDCs interact (Flanagan et al., 2009). Additionally, PFO binding to the membrane is independent of lipid packing and the presence of detergent-resistant membrane fractions, but is dependent on the presence of sphingomyelin (SM). In a cell membrane scenario, this could mean that not only phospholipids and glycolipids, but also membrane bound proteins are playing significant roles in determining the amount of free cholesterol. The chemical
properties of lipids, like the length of the hydrophobic chains, the degree of unsaturation, and type of polar head group are also vital factors in this process (Ohvo-Rekila et al., 2002). In a cell membrane, the amount of free cholesterol will be governed by not only the amount of cholesterol-phospholipid complexes, but also the amount of membrane protein complexes and cholesterol-protein interactions (Epand, 2006).

CDC monomers generally consist of four domains as seen in crystal structures (Rossjohn et al., 1997; Tweten, 2005). The role of individual domains has been studied in membrane binding and pore-formation by various groups. The overall shape of PFO monomers are flat. The N-terminal domain contains four β-sheets with a characteristic 90° bend in the center of the domain. The transmembrane domains (TMH-1 and TMH-2) are situated just next to the β-sheets, and after pre-pore formation, insert into the membrane to form the channel. TMH-1 is less solvent exposed compared to the TMH-2 in the monomer structure. Domains 1, 2 and 3 perform functions like oligomerization and pore formation, whereas domain 4, which adopts a β-sandwich topology, is responsible for membrane binding and cholesterol dependence (Rosado et al., 2008). Recently, a small two amino acid sequence pair (Thr-Leu) was identified and shown to constitute the cholesterol recognition motif in PFO, SLO, ILY (Ferrand et al., 2010).

Once initial binding of the CDC to the membrane is complete, monomers form multiple stronger interactions with the membrane through amino acid side chains and start diffusing on the membrane surface to find other monomers (Cho and
The undecapeptide of domain 4, which is present in all CDCs sequenced, undergoes a change in conformation with some of the aromatic residues.

Figure 1.7 **Representative structures of cholesterol dependent cytolysins (CDCs).** A, PFO monomer, B, Intermedilysin monomer, and C, SLO monomer bound to its receptor CD59 (red).

Figure 1.8 **Two examples of cholesterol-interacting sequences.** A, Undecapeptide (shown in the stick representation) of perfringolysin O (PFO) and B, Start domain with phosphatidylcholine (purple) molecule bound.
penetrating into the membrane. This penetration was detected in PFO using three tryptophans in the undecapeptide as markers, revealing that at least two of these tryptophans (W438 and W439 in PFO) are involved in a major conformational change (Nakamura et al., 1998; Seiko-Suzuki et al., 1996). The comparable undecapeptide sequence from ILY (GATGLAWEPWR) is different than PFO (ECTGLAWEWR) and these changes contribute indirectly towards the binding of CD59 instead of cholesterol (Polekhina et al., 2005). The further characterization of the role of the undecapeptide in cholesterol binding came from substitution studies of the ILY undecapeptide with the PFO sequence, which showed ILY binding to the cholesterol rich membrane just like other CDCs (Nagamune et al., 2004). Moreover, domain 4 exhibited identical conformations and interactions between its loops and the undecapeptide when either free cholesterol or a cholesterol rich membrane was used in the assays (Heuck et al., 2007). Surprisingly, these observations do not hold true when epicholesterol, an enantiomer of cholesterol, was used instead of cholesterol (Heuck et al., 2007; Heuck et al., 2009). Instead, titrating epicholesterol into the membrane results in PFO binding to the membrane even with as little as 19% cholesterol as long as the total sterol concentration remains at 48% (Flanagan et al., 2009). Additional roles of the undecapeptide region have been proposed (Polekhina et al., 2005) based on structural and mutational studies (Heuck et al., 2000) including conformational coupling between domains 3 and 4. In the case of SLO and PFO, binding was observed even with cholesterol microcrystals (Heuck et al., 2007; Heuck et al., 2009).
After binding to the membrane, CDCs form large transmembrane pores with a minimum of 70 β-strands (TMH-1 and 2) inserted into the membrane. In contrast to the smaller β-PFTs, which form heptameric pores and insert 14 β-hairpin loops, the CDCs perform a massive task of coordinated insertion of many β-strands. Two mechanisms for large pore formation have put forward by various groups. One hypothesis is that the CDCs first form smaller assemblies (like dimers) and start inserting TMHs into the membrane. These smaller oligomers start growing and transition from small to large pores gradually (Palmer et al., 1998). Other groups have actively proposed that PFO only makes large channels through membranes (Shepard et al., 2000). Attempts at isolating smaller pores have not succeeded, but Hotze et al. did successfully isolate a large PFO prepore complex (Hotze et al., 2001). All evidence points toward formation of a large prepore complex on the membrane before pore formation followed by simultaneous co-insertion of all β-hairpin loops into the membrane. The prepore must achieve a considerable size before membrane insertion and pore formation can commence (Heuck et al., 2003; Tweten, 2005), but how this size limit is enforced during the process is a mystery.

One way by which PFTs avert premature pre-pore formation in solution or on improper cell surfaces is through proteolytic activation of the toxin by membrane bound proteases. This is seen in the anthrax PA channel, aerolysins, Vibrio cholerae cytolysin, etc. (Gordon et al., 1995; Gordon et al., 1997; Milne et al., 1994; Olson and Gouaux, 2003). CDCs, on the other hand, do not undergo proteolytic cleavage and rely heavily on pre-pore formation to regulate the pore formation process. A key
conformational change in the domain 3 β5α1 region (the fifth β-strand followed by an α-helix) regulates the oligomerization process by blocking the monomer-monomer interaction surface when not bound to a cholesterol rich membrane. The β5α1 normally interacts with the β4 strand of domain 3, which makes the monomer-monomer interaction in the prepore and the pore. Upon binding to the cell surface and before pre-pore formation the β5 strand breaks off from the β4 strand and swings away from that surface, thereby allowing the monomer-monomer interaction to happen (Ramachandran et al., 2004; Tweten, 2005).

Once the prepore structure is stabilized and has crossed a certain threshold of size, TMH1 and 2 undergo a massive change in conformation to form the membrane-spanning pore. Consequently, an α-helical TMH sequence transitions into a β-strand in order to form the channel. Similar transitions have been seen in amyloid-β-peptides, but are relatively uncommon in the pore forming toxin field. Firstly, TMHs cease all interactions with domain 2 (for TMH1) and domain 3 (for TMH2). The disentanglement of TMHs results in an altering of the twist of domain 3 and help the TMHs adopt an extended conformation necessary for the α-β switch. The mechanism by which the TMHs proceeds from here to membrane insertion is not well understood, but data suggest that TMHs must embrace a local disordered or completely disordered conformation before the structural transition occurs. The insertion process of TMHs is also dependent on monomer-monomer interactions and other conformational changes in the monomer (Hotze et al., 2002), like the coupling between domains 3 and 4. A cooperative mechanism for the insertion step is desirable
as the penalty for inserting single amino acids into the membrane falls by 0.5 kcal/mol/residue when potential backbone hydrogen bonds are satisfied (White and Wimley, 1999). The interactions between polar amino acid residues in the hairpin ends and phospholipid headgroups aid the insertion process in two different ways: first, by providing thermodynamic free energy for the insertion, and second, by tethering the ends of the hairpins to the outer faces of the bilayer (Shepard et al., 1998; Shatursky et al., 1999; Tweten, 2005). As two TMHs from each monomer get inserted into the membrane, the intermolecular alignment of the two adjacent TMHs presents additional challenges for cohort insertion. Two residues in PFO (Tyr-181, Phe-318), one from each neighboring TMH, preserve the π-stacking alignment. Mutating either of them causes an arrest in the insertion process (Ramachandran et al., 2004; Hotze et al., 2001). The concluding step of the pore formation process is the construction of a β-barrel transmembrane channel. Among the many models that have been proposed for this step (Ramachandran et al., 2002; Shatursky et al., 1999; Tweten, 2005), the most promising one is the vertical collapse model. In this model the monomers stand upright on the membrane when binding and when forming the pre-pore, but experience a 35-40 Å fall from an upright to collapsed position (Czajkowsky et al., 2004; Ramachandran et al., 2005; Tilley et al., 2005). Tweten proposed that the extra space of the upright position and may enable the toxin to coordinate the creation of a preinsertion β-barrel structure, lowering the energetic cost of membrane insertion and aiding in mature channel formation.
The TL motif

A TL motif was identified in the pore forming toxin PFO (Perfringolysin O) and shown to consist of only two amino acids. This threonine-leucine pair causes an almost absolute loss of activity when both are mutated to alanine (Farrand et al., 2010). The conservative substitutions of leucine to isoleucine or threonine to serine also exhibits >95% loss in activity. Remarkably, when the TL motif was mutated to LT, the toxin also did not display meaningful activity. In the CDCs like PFO, SLO, and ILY this TL motif may recognize and bind to cholesterol before the undecapeptide insertion can happen, making the TL motif a decisive step in pathogenicity (Farrand et al., 2010).

Cholesterol-interacting domains in other membrane-associated proteins

For proteins to bind to membranes, cholesterol is an excellent choice of a receptor. In eukaryotic cells >90% of cholesterol is found in the cell membrane where the cholesterol content locally can be >50%. In eukaryotic membranes, cholesterol is postulated to form lipid-raft like structures along with sphingomyelin. Sterol rich crystalline domains have been shown to be important in cellular activities (Epand, 2006; Mason et al., 2003). Although the existence of raft-like structures is still debated by the scientific community, the discovery of caveolae in primary rat adipocytes, with almost identical cholesterol and sphingolipid content as raft structures, made the argument for lipid rafts stronger (Simon et. al., 1997; Ortegren et
These structures, when imaged using electron microscopes, displayed distinct morphological invaginations and can be isolated without detergent extraction as caveolin rich vesicles (Westermann et al., 1999). These domains or microdomains in membranes are thought to be a functional platform for proteins to bind. A number of proteins that partition into cholesterol rich membranes have been identified. Interestingly, the mechanism by which these proteins sense or bind cholesterol or cholesterol-rich membranes varies from a few amino-acid peptides to entire domains dedicated to this purpose.

**Steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domains**

The START domain, which is a 210 amino-acid long module, is present in at least 15 different human proteins. In humans, 6 of these domain-containing proteins function by binding to membrane components like cholesterol (STARD1, SATRD3 and STARD5), 25-hydroxycholesterol (STARD5), phosphatidylcholine (STARD2, STARD10), phosphatidylethanolamine (STARD10) and ceramides (STARD11) (Alpy and Tomasetto, 2005). The remaining 9 human proteins with START domains have unknown functions. Apart from humans, these domains were discovered by sequence-based bioinformatics analyses in flies (4 START domains) and in nematodes (6 SATRT domains) (Soccio and Breslow, 2003). Interestingly, the START domain is more frequently found in plant genomes and commonly associated with homeodomain transcription factors (Schrick et. al., 2004), the significance of which is still unknown.
The function of all 15 human START domains has been identified in animal models using genetic screens. The STARD1 is associated with cholesterol movement in mitochondrial membranes, particularly from the outer to the inner leaflet. The STARTD2 domain performs a similar role, but for phosphatidylcholine (PC) in cell membranes. STARD4, STARD5 and STARD6 function in cholesterol synthesis, whereas STARD11 and STARD14 were shown to be involved in autoimmune disease and obesity disorders, respectively. The rest of the STAR-domain proteins (3, 5, 7, 8, 9, 10, 12, 13, and 15) were all found to be either up-regulated or down-regulated in different cancers types (Alpy and Tomasetto, 2005).

Multiple STAR domain structures have been published in recent years and reveal a very similar helix-grip fold in all constructs. The central groove, which has been proposed to be the ligand binding pocket, is mostly lined with β-sheets and one C-terminal α-helix, which acts as a cover. Truncating part of this C-terminal helix causes a complete eradication of the domain activity (Arakane et al., 1996; Feng et al., 2000). Models have been proposed where START domains interact directly with the membrane using the C-terminal sequence by sequestering specific lipid molecules in the central groove. The α-4 of the C-terminus caps the groove, thereby shielding hydrophobic lipids from water and making it safe to transport to another part of the membrane (Alpy and Tomasetto, 2005). Unfortunately, our knowledge regarding the mechanism of lipid specificity is inadequate and further investigation is necessary.
**Sterol-sensing domains (SSDs)**

The sterol-sensing domain protein (SSD) has five transmembrane domains and is responsible for sterol homeostasis and regulation. This protein synchronizes the expression of sterol biosynthesis genes and indirectly controls the rate-limiting step of related pathways by stabilizing the enzyme, 3-hydroxy-3-methylglutaryl (HMG)-CoA reductase (HMGR) (Theesfeld et al., 2011). Many of the cholesterol-lowering statin drugs in humans target this HMGR and knowledge about SSD domains could aid in designing drugs. In HMGR, the SSD domain is located within the N-terminal domain and functions by mediating the ubiquitin-dependent degradation of HMGR upon sensing a key intermediate’s concentration in the membrane. When the farnesyl pyrophosphate (FPP) concentration is high, the membrane-bound SSD domain binds INSIG, which in turn signals ER ubiquitine ligase gp 78 to destroy HMGR. Recent studies have identified that geranylgeranyl pyrophosphate (GGPP), an analog of FPP, is the signaling molecule for HMGR SSD (Theesfeld et al., 2011). Another enzyme of the same sterol biosynthesis pathway, 7-dehydrocholesterol reductase, also contains a sterol-sensing domain, but the functional significance of this domain in the activity of this protein not yet known.

SCAP, or SREBP cleavage activating protein, is a membrane protein with multiple transmembrane spans, which controls the expression of the SREBP protein. SREBP, a transcription factor, regulates the transcription of gene clusters in charge of the sterol biosynthesis pathway and LDL receptor production. When the sterol level becomes high, SCAP binds to cholesterol and actively traffics SREBP into the golgi
or ER with help of INSIG (or the ER localization protein) (Theesfeld et al., 2011; Epand, 2006). Under low sterol environments, SREBP becomes free from the INSIG complex and promotes the synthesis of key enzymes needed in the pathway. By itself, SREBP cannot sense the sterol concentration (Epand, 2006). Investigation with purified SCAP protein revealed a nanomolar affinity of SCAP for cholesterol, but also predicted some possible regulation factors, like phospholipid headgroups (Radhakrishnan et al., 2004). An additional patch of the SCAP protein comprised of only four amino acids (YIYF) on the cytoplasmic site was shown to be instrumental in INSIG binding. This small peptide is extremely conserved from humans to *C. elegans* and *Drosophila melanogaster* (Epand, 2006). Remarkably, many membrane-associated proteins not related to the sterol synthesis pathway also contain similar sequences as a part of the CRAC motif, which will be discussed below (Epand et. al., 2010; Epand, 2006).

**Cholesterol recognition/interaction amino acid consensus sequence (CRAC) motif**

Caveolin is a group of proteins identified as one of the main protein components of caveolae related proteins (Schlegel et al., 1999). For a long time, caveolin has been known to be a model protein performing cholesterol dependent membrane insertion with palmitoylated amino-acids responsible for this cholesterol dependence (Brown and London, 2000). However, mutating the palmitoylation sites or three Cys residues resulted in complete removal of palmitate incorporation without exhibiting any defect in the localization of the protein to cholesterol rich membranes (Dietzen et al, 1995). Using deletion mutants, Schlegel et al discovered that the
scaffolding domain of Caveolin-1 is necessary and sufficient for cholesterol rich membrane recognition and interaction (Schlegel et al., 1999). Further investigation identified that a small heptapeptide (VTKYWFYR) sequence inside the scaffolding domain is the cholesterol recognition/binding motif. Later, this observation was confirmed with synthetic peptides when part of the scaffold domain sequence (residues 82-101) was reported to bind to cholesterol in vitro (Wanaski et al., 2003).

This small variable length sequence is the CRAC sequence or CRAC motif (Li and Papadopoulos, 1998) with a consensus sequence of –L/V-(X)\(_{1-5}\)-Y-(X)\(_{1-5}\)-R/K-. The relaxed requirements of this motif pose challenges in identifying a functional CRAC motif just from sequence alone. For example, the consensus sequence –L/V-(X)\(_{5}\)-Y-(X)\(_{5}\)-R/K-, without including length variability, can represent >10\(^{13}\) different possible combinations (Epand, 2006) and occurs in genomes with a very high frequency (Palmer, 2004). The importance of each amino acid in the CRAC motif has been probed and the central tyrosine was found to be the key residue in the motif. This central tyrosine is absolutely needed for function and cannot be substituted with any other aromatic amino acids in the benzodiazepine receptor (Jamin et al., 2005; Epand, 2006). The CRAC motif from the human type 3 somatostatin receptor has the sequence \(^1\text{VICLYLLIVVK}^{12}\), which upon closer inspection reveals two CRAC-like motifs, one with \(^4\text{LCYLLIVVK}^{12}\) and the other with \(^1\text{VICLYLLIVVK}^{12}\). In molecular dynamics simulations, cholesterol bound this motif with a binding energy of -43 kJ/mol (Baier et al., 2011), but unexpectedly the cholesterol molecule did not interact with the central tyrosine (Y\(_6\)). Moreover, half
(two out of four) of the residues that are associated with cholesterol binding ($^{1}{\text{VICLYLLIVVK}}^{12}$ - residues in bold are the cholesterol interacting residues) are located in the degenerate regions of the consensus sequence.

However, another sequence from the same protein ($^{1}{\text{IVRYTKMK}}^{8}$) demonstrated a very different scenario of binding energetics. In this case, the central tyrosine not only interacted with cholesterol, but also provided more than half (-26 kJ/mol) of the total binding energy (-49 kJ/mol). Remarkably, more than half of the residues that associated with cholesterol, i.e. $^{1}{\text{IVRYTKMK}}^{8}$ (residues in bold are the cholesterol interacting residues), are located in the degenerate sequence (Baier et al., 2011; Fantini and Barrantes, 2013). In this particular case, the central tyrosine residue participated in π-stacking interactions with the planer rings of cholesterol. Theoretically, the tyrosine could also provide additional polar interactions with the hydroxyl group of the cholesterol molecule.

For transmembrane proteins, two additional criteria are put forward for a functional definition of the CRAC sequence. Firstly, the basic amino acid should stay close to the water interface and thereby the C-terminus of the motif will define, which leaflet (cytoplasmic or extracellular) of cholesterol the CRAC will bind. In other words, for a CRAC motif with the N-terminus towards the intracellular side and C-terminus pointing to the outside, the outer leaflet cholesterol is the binding partner. Secondly, no matter what the direction of the motif is, the degenerate amino-acids should be hydrophobic residues as they will always stay inside the hydrophobic core of the membrane.
Lipid raft dependence is evident in many cases of cellular entry for enveloped and non-enveloped viruses (Chazal and Gerlier, 2003; Epand, 2006). This could proceed through a loop in the HIV-1 gp120 protein, which is thought to anchor the protein in sphingomyelin-rich membranes, or through a CRAC motif in gp41. This CRAC motif has the sequence LWYIK, and mutational studies confirm the role of this sequence in membrane fusion of the intact virus. The LWYIK peptide alone was sufficient to induce cholesterol partitioning into cholesterol-rich microdomains, but additional research is needed to understand the mechanism.

**Inverted CRAC (C Arkansas) motif**

CARC motifs differ from the CRAC motif mainly in two respects. Firstly, it has an opposite orientation of the consensus sequence compared to the CRAC (i.e., -K/R-X1-5-(Y/F)-X1-5-(L/V)-) and secondly, the central tyrosine residue can be replaced by phenylalanine. Similar to the CRAC motif, the basic amino-acids help in anchoring the motif at the membrane-water interface and specify the interacting leaflet of the membrane (Fantini and Barrantes, 2013). However, the exchangeability of the tyrosine with the phenylalanine indicates that the potential electrostatic interaction between hydroxyl groups of tyrosine and cholesterol may not be required for binding. Perhaps interactions between cholesterol and the motif mostly happen through CH π-stacking between aromatic headgroups and the cholesterol molecule. A CARC motif from the same human type 3 somatostatin receptor with sequence

\[ ^{\text{1}} \text{RAGFIYTAAL}^{\text{12}} \]

was used in similar molecular dynamics simulations, and interestingly, this motif bound to cholesterol with a higher binding energy of -54
kJ/mol. The key residues in the motif that were actively associated with cholesterol are arginine, phenylalanine, tyrosine and leucine as marked in bold, \textbf{1RAGFIYTAAL}^{12}. More than 50% of the binding energy comes from interactions between cholesterol and these hydrophobic residues (-32.9 kJ/mol), while the leucine interacts with the isoctyl chain of the lipid molecules and accounts for -10 kJ/mol, a substantial amount for binding. Additional CARC domains have been identified in GPCRs, the AChR and other ion channels (Fantini and Barrantes, 2013).

CARC and CRAC domains can occur in the same protein and may potentially function in tandem by binding two cholesterol molecules instead of one, thereby increasing the stability and energetic contributions for membrane binding and insertion.

\textit{Tilted Peptide motif}

This type of cholesterol binding motif is sequence independent and functions by distorting the membrane upon insertion. The distortion induces a tilted conformation of \textasciitilde 45$^{0}$ of the peptide with respect to the membrane plane (Lins et al., 2008). With the HIV gp41 peptide, the total binding energy for cholesterol is comparable at -48.5 kJ/mol, even while lacking any aromatic residues that might contribute in $\pi$-stacking. Similar tilted peptides are also present in $\alpha$-synuclein and Alzheimer’s $\beta$-amyloid peptide and speculated to contain cholesterol-binding surfaces (Fatini et al., 2011; Fantini and Barrantes, 2013; Di Scala et al., 2013).
**Cholesterol consensus motif (CCM)**

CCM was first identified in the β-adrenergic receptor as a spatial cholesterol binding motif with a relatively loose definition (Hanson et al., 2008). Sequence analysis has led to the speculated presence of similar motifs in >80 membrane associated receptor proteins including the 5-Hydroxytryptamine (5HT1A) receptor and the oxytocin receptors (Hanson et al., 2008; Gimpl et al., 2010). This motif in the β-adrenergic receptor binds to cholesterol using two conserved aromatics, one non-aromatic hydrophobic, and one positively charged residue. To ascertain the mechanism and function of these motifs in the membrane receptors future investigation will be required.

**Membrane binding of VCC**

*Effect of cholesterol*

The direct membrane binding activity of VCC was first reported using phosphatidylcholine (PC) liposomes doped with cholesterol (Ikigai et al., 1996). As pore formation activities between erythrocytes and liposomes yield an identical population of oligomers, it is safe to assume that non-membrane components may not play any critical role in membrane binding or pore formation. Although the efficiency of membrane binding and pore formation drastically declines when a liposomal system is used, this could be due to the presence of additional receptors on the cell membranes not found on synthetic liposomal membranes (Zitzer et al., 1999). Ikiagi
et al. also pointed out that the pore formation process could be directly dependent on the cholesterol content of the membrane, a common theme for CDCs (Ikigai et al., 1996). Furthermore, other sterols like diosgenin or ergosterol can also induce pore formation as a cholesterol alternative, but with significantly reduced efficiency (Ikigai et al., 1996). This dependence was shown to be strict as an enantiomer of cholesterol, or epicholesterol, failed to promote any significant pore formation in the liposomal system (Zitzer et al., 2003). This observation signifies that the cholesterol dependence is not through indirect membrane fluidity, but rather through a direct interaction with cholesterol molecules or cholesterol-rich membrane microdomains. Also, this stereospecific mode of cholesterol interaction was more prominent for VCC than other classical cholesterol-dependent toxins or CDCs (SLO) tested (Zitzer et al., 2003). This indicates a different mechanism or motif for the cholesterol interaction in VCC. Many studies have confirmed this cholesterol-dependent pore formation using liposomal and other systems (Zitzer et al., 1999; Zitzer et al., 2000; Zitzer et al., 2001; Harris et al., 2002; Ikigai et al., 2006; Krasilnikov et al., 2007; Paul and Chattopadhyay, 2012). Additionally Zitzer et al. provided direct evidence that the pore formation activity of VCC is neutral to membrane fluidity, supporting the hypothesis that cholesterol regulates pore formation by directly interacting with the protein (Zitzer et al., 2000).

A detailed investigation of the cholesterol binding activity of VCC was performed by Ikigai et al (Ikigai et al., 2006), revealing that the 3βOH group and C-C Δ5 bond are indispensable for VCC binding. This observation also explained why
epicholesterol is unable to induce pore formation, as it has a 3αOH group and a single bond at the C5-C6 positions (Ikigai et al., 2006). Similarly, other sterols like diosgenin, campesterol, and ergosterol, which possess both structural motifs, can induce pore formation with reduced efficiency. It is worth mentioning that irrespective of the sterol used, or even with lipid membranes without any sterol present, some VCC monomer was always found in the liposome pellet. This points towards a non-specific membrane recognition process over a cholesterol interaction.

All of the aforementioned interaction studies were performed with cholesterol-rich membranes, which is not the same as free cholesterol, as the solubility of pure cholesterol is ~4.7 µM - too low for solution studies (Haberland and Reynolds, 1973). The first direct cholesterol-binding evidence came from Harris et al., where they used a cholesterol microcrystal suspension system to look at these interactions. VCC was able to bind and oligomerize on the cholesterol bilayer with much slower kinetics than cholesterol-rich membranes (Harris et al., 2002). To their amazement, oligomerization was predominantly observed only on the edges of the cholesterol bilayer, and this edge-only occurrence did not change even after incubating for 24 hours. Another interesting observation was the spontaneous discharge of VCC oligomers from cholesterol bilayers, which has never been detected in cholesterol-rich membrane systems. This clearly demonstrates the feeble nature of the interaction between the protein and cholesterol in the absence of a lipid membrane, specifying unknown structural functions of the lipid bilayer in the oligomerization process. The strange propensity of the toxin for the edges of the
cholesterol bilayer is possibly due to the availability of additional surface area at the edge, which definitely contains more than the cholesterol residing inside the bilayer (Harris et al., 2002). This agrees with the previous observation by Zitzer et al. that accessibility of cholesterol molecules determines VCC binding to the membrane and addition of non-sterol lipids augment the cholesterol interaction by changing the energetic state of cholesterol molecules (Zitzer et al., 2001; Harris et al., 2002). The 3βOH moiety of cholesterol, in the study by Harris et al., was not an essential requisite in contrary to results obtained by Ikigai et al. (Harris et al., 2002; Ikigai et al., 2006). Making things even more complicated, the incorporation of a short acetate ester in the 3β position did not have any detrimental effect on the binding, but the addition of a fatty acid side chain did abrogate the interaction. The addition of a second –OH group in addition to the 3βOH, and elimination of the C17 aliphatic chain of cholesterol had no significant consequence on the binding and oligomerization of the toxin. The inability of (2-trifluoroacetyl)naphthoxy cholesterol, which possesses a free 3βOH, but contains a bulky addition on the other side of the molecule, is an indirect implication of the importance of the overall chemistry and electrostatics of cholesterol in VCC binding and oligomerization (Harris et al., 2002).

VCC oligomerization and pore formation on symmetric bilayers exhibit a positive, but the non-linear relationship with the cholesterol content of the membrane. This is interesting as generally this type of non-linearity signifies the involvement of more than one cholesterol per VCC molecule. The membrane asymmetry also has an
influence on VCC oligomerization, as the rate of pore formation on asymmetric membranes is considerably lower (sometimes >80% lower) than symmetric counterparts. While keeping the cholesterol content of the cis leaflet constantly low, increasing the trans-leaflet cholesterol from 25% to 46% did not have any effect on pore formation. Therefore, the leaflet with the lowest cholesterol concentration limits the rate of channel formation. The amount of cholesterol on the cis leaflet appeared to be slightly more important than the cholesterol content of the trans-leaflet (Krasilnikov et al., 2007). However, the chemical composition of both the leaflets can essentially function as a rate limiting kinetic bottleneck for VCC pore formation. From the observations of VCC pore formation in the cholesterol bilayer and effects of cholesterol content in both leaflets on the same process, one can hypothesize that VCC may interact with a tail-to-tail cholesterol dimer, but future investigation will be needed to test such a theory (Harris et al., 2002; Krasilnikov et al., 2007).

There is, at least, one study questioning the cholesterol dependence of the VCC pore formation process (Menzl et al., 1996), but an overwhelming number of published results have actively refuted this study (Zitzer et al., 1999; Zitzer et al., 2000; Zitzer et al., 2001; Harris et al., 2002; Ikigai et al., 2006; Krasilnikov et al., 2007; Paul and Chattopadhyay, 2012). A few single point mutations have been identified in VCC resulting in a drastic loss of activity, which are speculated to be due to a loss of cholesterol-dependent membrane interaction, but detailed inquiries are necessary to confirm such claims (Paul and Chattopadhyay, 2012; Rai and Chattopadhyay, 2015). Still many questions remain, like if there is any motif, which
is responsible for cholesterol binding or recognition, or what is the effect of a cholesterol enantiomer on an asymmetric membrane?

Figure 1. 9 Effect of cholesterol on VCC’s pore forming activity. Activated VCC was incubated with brain-lipid liposomes with (left lane) and without (right lane) 30% cholesterol. The mixture was ran on a SDS-PAGE gel without boiling the sample. Adapted from Zitzer et al., 1999.

**Effect of lipids**

In addition to cholesterol, the chemical composition of the membrane is also a factor in VCC pore formation. The first evidence came from the observation that the dosage response for VCC pore formation varies with different membrane compositions. For example, liposomes reconstituted from bovine brain lipid extracts require at least 100-fold less VCC for pore formation than identical liposomes made from phosphatidylcholine (Zitzer et al., 1999). Asolectin, a crude extract from soyabean plants, is also a far superior candidate over PC (Zitzer et al., 2000). Although liposomes prepared from brain lipids and asolectin have distinct
characteristics in term of membrane fluidity, both of them have been used as model membrane components in VCC activity determination (Zitzer et al., 1999; Zitzer et al., 2000). The inferior performance of PC liposomes is perhaps the consequence of the inhibitory effect of the choline headgroup on channel formation by VCC (Zitzer et al., 1999). Sphingomyelin (SPM) and the phosphatidylethanolamine (PE) were proven to be superior candidates over PC, but no other glycerophospholipids tested showed any pore formation activity. The ceramides, whether free, attached to glucose (GlcCer) or attached to galactose (GalCer), were found to sensitize the membrane and enhance VCC pore formation activity better than any other lipids tested. Ceramide headgroups did not instigate any positive or negative influence on pore formation even when modified with a cyclopropyl group, benzyl group or when the $trans$ C4=C5 double bond was altered to a $cis$ double or triple bond (Zitzer et al., 2001). However, the amide-linked fatty chains of ceramide, when removed, caused a radical loss of membrane sensitization for VCC. Altering the acyl chains from the less bulky acyl-chain containing dilauroyl-PE, to bulkier dioleoyl-PE, resulted in higher membrane susceptibility by VCC. But an even bigger gain in membrane susceptibility was achieved when diphtanoyl-PE was used instead of dioleoyl-PE. As the diphtanoyl-PE contains several methyl branches, which confers a very large surface area compared to dioleoyl-PE, a direct correlation between the surface area of the acyl chain and membrane susceptibility can be postulated from the findings mentioned above (Zitzer et al., 1999; Zitzer et al., 2001).
The direct relationship between membrane lipid components and VCC pore sensitivity has been established in various reports. The absence of any known lipid binding motifs and lack of sufficient evidence for direct lipid binding in VCC points towards an indirect mechanism for regulating the sensitivity of the membrane. All observations together indicate that the presence of cone-shaped lipid (a smaller headgroup and a larger acyl surface) molecules competes with other cone shaped molecules like cholesterol when they are both in the membrane. The conical shape of lipids or cholesterol makes it harder for suitable headgroup coverage at the membrane-water interface. That in turn creates an energetically unfavorable situation for all cone shaped molecules involved. To avoid such a high energetic penalty, cholesterol molecules alternatively can take refuge in the corresponding cholesterol binding pockets or the hydrophobic surfaces of specific proteins. The spatial and energetic constraints rise sharply with the increase in the concentration of cone shaped molecules in the bilayer (Huang and Feigenson, 1999). Similarly, the presence of ceramides in the bilayer force the cholesterol molecules to attain an unfavorable state. In our case, when VCC monomers approach the membrane, cholesterol molecules find it favorable to interact with the cytolytin when more ceramide molecules are associated with the membrane. Thereby, lipid components modulate VCC pore formation by adjusting the energetic crosstalk in the coupled system of cone shaped lipids-cholesterol and not through any direct action on the cytolytin (Zitzer et al., 2001).
**Oligomerization and pore formation of VCC**

VCC belongs to the β-PFT family of pore forming toxins (Olson and Gouaux, 2005), which forms β-barrel pores with 1-2 nm internal diameter (Krasilnikov et al., 1992; Menzl et al., 1996; Zitzer et al., 1995; Zitzer et al., 1997; Zitzer et al., 1999; Pantano and Montecucco, 2006). The mechanism of pore formation and structural information regarding the membrane inserted pore was almost non-existent even after more than two decades of biochemical characterization of VCC. On the other hand, a large amount of mechanistic information for the anthrax PA channel (Collier and Young, 2003) and the detergent solubilized transmembrane pore structure of *S. aureus* α-hemolysin was available in the literature (Song et al., 1996). By using mutagenesis and protein engineering, Lohner et al. established that after primary membrane binding, and before the final membrane insertion, VCC forms a membrane-bound transient oligomer called the pre-pore (Lohner et al., 2009). This pre-pore state has been observed and proposed for other pore forming toxins, like the anthrax PA channel and *S. aureus* α-hemolysin (Miller et al., 1999; Walker et al., 1995). Lohner et al. argued that VCC, like its structural homologs, follow the same prototypical pathway of pore formation. Individual monomers bind to the cell surface and start a 2D search for other monomers to form a pre-insertion oligomer. With the formation of the transient oligomeric state, monomers undergo massive structural rearrangements and the pore-forming region breaks free from its hydrophobic shielding. All of the pore-forming transmembrane helices insert into the membrane at once to minimize any energetic penalty and form the mature membrane-spanning β-
barrel. As discussed above, a higher cholesterol content in the membrane and presence of conical shaped lipid molecules aids this process in VCC, but the precise molecular mechanism is still unknown.

Perhaps the most important parameter for the pore-formation process is the stoichiometry of the oligomeric pore. In archetypical pore forming toxins like *S. aureus* α-hemolysin and the anthrax PA channel, the pore has shown to be heptameric in nature (Song et al., 1996; Collier and Young, 2003), whereas the Staphylococcal bicomponent pore-forming toxins form octameric channels (Yamashita et al., 2011). However, another group of pore forming toxins, CDCs, display a variable stoichiometry of anything between 30-50 monomers for channel formation. In the case of VCC, studies have presented evidence for a heptameric assembly for the oligomeric pore (Harris et al., 2002; Olson and Gouaux, 2005; Pantano and Montecucco, 2006; He and Olson, 2010).

Another interesting observation reported by Zitzer et al. is that pore formation is sensitive to temperature and that any temperature below 10°C pauses the pore formation process. However, VCC still binds at these temperatures and data shows active membrane interactions at temperatures as low as 4°C (Zitzer et al., 2000; Zitzer et al., 1997; Rai and Chattopadhyay, 2015). Observations like these undoubtedly indicate the presence of a multistep mechanism for pore formation, i.e., membrane binding, pre-pore formation and final membrane insertion.
SECTION 5– GLYCANS AND LECTINS

Glycans: an overview

Diverse and complex oligosaccharides, or glycans, cover the surface of every cell. Generally, these arrays of sugars are attached to membrane-associated proteins or membrane lipids. Additionally, eukaryotes contain cytoplasmic and nuclear glycosylated macromolecules, which are not related to membrane glycans. These glycans are made up of monosaccharides stitched together through α- or β-glycosidic linkages and may be linear or branched in nature. Non-sugar components like acetyl groups or sialic acids can also be found in animal or plant glycans and are essential for various functions. It is worth mentioning that glycans hold the potential for much greater complexity and diversity when compared to the genome and proteome, and this consequentially enhances the diversity of different protein surfaces. Similar to gene and protein expression, glycan synthesis is highly regulated to sense and accordingly alter its expression profile giving rise to tissue-specific glycosylation. This is vital in multicellular organism’s development and survival. Moreover, these glycans control almost every important developmental pathway in multicellular organisms and have also been attributed to major roles in host-pathogen interactions.

Out of hundreds of potential biochemical saccharide monomers, only a very few are associated with glycans. These monosaccharides can be divided mainly into six categories – Pentoses, hexoses, hexosamines (with 2’ substitution), deoxyhexoses (6’-deoxy), uronic acids (6’ carboxyl groups) and sialic acids (nine carbon sugars).
Different types of glycans

The majority of all cellular proteins and a variety of lipids undergo glycosylation. Fascinatingly, the overall volume of glycan moieties encompasses as much as, or more volume, than the domains of proteins to which they are attached (Dwek and Opdenakker, 1998). Protein molecules are post-translationally modified to receive glycan molecules on the specific amino acid side chain. The three different types of protein glycosylation are N-linked glycosylation, O-linked glycosylation and glycosylphosphatidylinositol (GPI) anchor modification. Lipids also get glycosylated, but the extent of that glycosylation event is much more limited than their protein counterparts.

Figure 1.10 A, Conventional representations used for glycans. B, Symbols used for common monosaccharides. Although the symbols have their respective colors, they can be used in a color-independent manner. Adapted from Essentials of Glycobiology, 2nd edition (Varki et al., 2009).
**N-linked glycans**

These types of glycan molecules are attached to asparagine (Asn) residues through a glycosidic bond. Not all Asn residues can accept glycan moieties and the specific sequence Asn-X-Ser/Thr is needed for the transfer reaction to go forward. Based on this diversity, N-linked glycans can be categorized into three different classes with a common pentasaccharide core structure \((\text{Man}α1-6(\text{Man}α1-3)\text{Man}β1-4\text{GlcNAc}β1-4\text{GlcNAc}β1)\) attached to the Asn residue. The first class of N-linked glycans is known as the oligomannose type (or high-mannose type), which contains exclusively mannose residues linked to the core section. The second class comprises of complex type glycans, where after the core pentasaccharide residue, both branches receive N-acetylglucosamine (GlcNac) residues. The last type, or the hybrid type, display mannose residues on the \(\text{Man}α1-6\) branch and GlcNac on the other (Kornfeld and Kornfeld, 1985). Additional modifications like sialylation or fucosylation on the previously mentioned types of glycans also occur regularly in cells.

The ER, or endoplasmic reticulum, plays an instrumental and absolutely indispensable role in the glycan biosynthesis pathway. The initial steps of glycan synthesis start on the cytoplasmic side of the ER membrane, which contains membrane-associated enzymes (Schiller at al., 2012). Using a dolicholphosphate (Dol-P) group as a template, the \(ALG\) family genes (altered in glycosylation) sequentially add GlcNac and mannose residues to construct the aforementioned pentasaccharide core \((\text{Man}α1-6(\text{Man}α1-3)\text{Man}β1-4\text{GlcNAc}β1-4\text{GlcNAc}β1)\). After addition of a total of seven residues, the Dol-P-oligosaccharide complex is flipped to the inside face of
the ER. There, the addition of mannose residues continues to form a high mannose-type biantennary glycan with eleven carbohydrate residues attached. At this stage, glucose residues are added and a final fourteen-monoasaccharide complex (Glc$_3$Man$_9$GlcNAcP-P-Dol) is transferred from the Dol-P anchor to the recently synthesized polypeptide molecule. Following this step, the complex travels to the cis-golgi region and processing resumes. After few more steps, the protein-glycan complex relocates to the medial-golgi where a key fucosylation reaction takes place. Now, the glycan complex can undergo the addition of galactose and GlcNAc residues.

Figure 1.11 **Beginning steps of the N-linked glycan synthesis pathway.** Glc$_3$Man$_9$GlcNAcP-P-Dol is synthesized separately and attached to respective proteins. Further modification of the attached oligosaccharides continues as described in the text. Standard symbolic representations are used to depict glycan structures. Adapted from *Essentials of Glycobiology, 2nd edition* (Varki et al., 2009).
in the *trans*-golgi region and *trans*-golgi network (TGN) as needed. It is here where the glycan terminals are either sialated and transferred to the plasma membrane, or phosphorylated and transferred to endosomes. TGN also provides space for the conversion of the Man$_5$GlcNAc$_2$-Asn to hybrid and complex N-linked glycans (Kornfeld and Kornfeld, 1985; Schiller et al., 2012; Herscovics, 1999).

The final form of N-glycans varies slightly among yeasts, plants, mammals, and insects and act in various cellular functions. In humans, glycosylation defects have been ascribed to a hundred different genetic diseases and removal of complete N-glycan synthesis is fatal. Some of the major types of disorders are type I and type II congenital disorders of glycosylation, congenital dyserythropoiecic anemia type II, galactosemia, congenital muscular dystrophies, inclusion body myopathy, etc. (Kornfeld and Kornfeld, 1985; Schiller et al., 2012; Herscovics, 1999; Jaeken and Matthijs, 2007; Eklund et al., 2007).

Figure 1.12 A, Types of N-linked glycans and B, complex O-GalNAc glycans Adapted from *Essentials of Glycobiology*, 2nd edition (Varki et al., 2009).
**O-linked Glycans**

As the name suggests, O-linked glycans are attached to Ser or Thr residues of glycoproteins. First identified in mucin proteins in 1877, O-linked glycans since have been found widely distributed in all human mucins and mucin-related proteins. Primarily, these proteins are produced by mucous cells, but other epithelial cells have also been shown to secrete them.

O-linked glycosylation requires Thr or Ser residues, which are in abundance in the mucin family proteins, especially inside VNTRs (variable number of tandem repeats) (Muller et al., 1999). Interestingly, a single mucin protein may contain >100 sites for O-linked glycans to attach and this excessive density of glycans bestows a bottlebrush like morphology upon them (Brockhausen et al., 2009).

O-linked glycans, unlike N-linked glycans, contain multiple types of core carbohydrate arrangements. These different types were grouped into four major categories called extended core 1, 2, 3 and 4. The first core is composed of a GalNac residue bonded to a Gal through a β3-glycosidic bond. The second contains a trisaccharide branched core with Gal, GalNac, and Glucose moieties. The third and fourth types are made up of GlcNac and GalNac residues and form disaccharide single branch or trisaccharide double branches, respectively. Core 1, 2, and 3 can be found in human respiratory mucins, whereas the colonic mucins show specificity for core 4. Diverse types of oligosaccharide chains are found extending from these four
core sequences and are synthesized similarly after different cores have been produced.

The biosynthetic pathways for O-linked glycans also vary considerably from N-linked glycans. Firstly, the single sugars are attached sequentially on Ser/Thr-containing polypeptides, unlike N-linked glycans where the oligosaccharide is primarily made on a Dol-P platform. Secondly, these types of glycans are surprisingly low in mannose content and are mostly comprised of Gal, GlcNac and GalNac residues along with some sialic acids and fucose moieties. In all the core structures, a GalNac residue is initially attached on the Ser/Thr-containing proteins, followed by other residues like Gal (core 1), Gal and GlcNac (core 2), GlcNac (core 3 and 4) (Brockhausen et al., 2009; Tabak, 1995).

More complex O-glycans are synthesized on the previously assembled core structures. Some elongation reactions also construct oligosaccharides, like GlcNAcβ1-3Galβ1-4(poly-N-acetyllactosamine), GalNAcβ1-4GlcNac-(LacdiNac), and Galβ1-3GlcNac sequences. One interesting fact is that ABO blood-group-defining glycans (and relatives) are commonly found on human mucin proteins and other glycoconjugates. Fascinatingly, some enzymes that transfer sialic acid residues or sulfate ions onto the terminal sugar molecules of glycans are common between both types of glycans, but generally do exhibit bias in specificity for N-linked glycan substrates.
Although the precise roles of glycans are not well understood, genetic studies have shown they affect various processing mechanisms like protein folding, cell adhesion, cell differentiation, cytokine production, lymphocyte activation, and sperm-egg interactions. Not surprisingly, cancer tissues exhibit differential expression of some O-linked glycans, although the precise mechanism of these molecules in cancer is not yet clear (Brockhausen et al., 2009, Vasudevan and Haltiwanger, 2014; Brockhausen, 1999; Fukuda, 2002; Hollingsworth, 2004; Van den Steen et al., 1998).

**Glycosylphosphatidylinositol (GPI) anchors**

A number of proteins associated with peripheral membranes, including transmembrane proteins, contain GPI molecule attached to their C-termini. In recent years, GPI anchors have been established to be ubiquitous with hundreds of proteins identified to contain GPI anchors. Every known GPI anchor possesses a common core structure with an ethanolamine-PO₄-6Man α1–2Manα1–6Manα1–4GlcNa1–6myo-inositol-1-PO₄ moiety, with varieties containing a substitution of the sugar hydroxyl groups. The main function of these anchors is to attach proteins to the outer or inner membrane leaflets, but additional roles in cell adhesion (*S. cerevisiae*), cell-wall synthesis (*S. cerevisiae* and *Candida albicans*), olfaction (*C. elegans*) and plant reproduction (*Arabidopsis thaliana*) have been demonstrated. Additionally, in mammals, many studies have identified roles for these molecules like an immune regulator, neural receptor, LPS receptor and cell-cell adhesion mediator (Ferguson et al., 2009; Maeda et al., 2006; Paulick et al., 2008; Orlean et al., 2007).
**Glycosphingolipids (GSL)**

Glycosphingolipids are cell-surface glycolipid molecules found in almost every form of life. These glycosylated lipids are predominantly found in the outer leaflet of the membrane and function by mediating cell-cell adhesion and membrane protein regulation (Schnaar et al., 2009).

**Lectins: a brief history**

Lectins were first identified by biochemists working on plant seed components. The toxic effects of castor bean extracts (*Ricinus communis*) were first recognized by Dixson et al. (1886), but it was P. Herman Stillmark who identified the toxic substance to be a hemagglutinin. Using salt extraction, protein precipitation, and dialysis, Stillmark isolated the fraction responsible for toxicity and hemagglutination and named it ricin. Later, after almost 70 years, Stillmark’s fractions were identified to be a mixture of two proteins – a toxin with weak agglutinating activity (ricin) and an agglutinin (Singh and Sarathi, 2012). Following Stillmark’s work, a number of plant seed extracts were described to have hemagglutination activity, but the mechanism of this reaction was unknown. A major breakthrough happened when Summers et al. (1936) isolated the concanavalin A protein and demonstrated the sugar binding specificity of the molecule. William C. Boyd and Karl O. Renkonen individually discovered the ability of many plant agglutinins to specifically distinguish between different blood types, leading to the designation lectins (from
Latin “legere” meaning to choose) for these groups of molecules (Boyd and Shapleigh, 1954). These findings were instrumental to the lectin field and the presence of lectin genes was later identified outside the plant kingdom and virtually in every form of life. Lectins are thought to be involved in many diverse pathways making them one of the most important groups of proteins in eukaryotes, especially in the animal kingdom.

The discovery of the mitogenic properties of lectins not only impacted the lectin field, but also paved the way for many revolutionary findings in immunology, including the discovery of interleukin-2. Significant roles, like the regulation of protein biosynthesis and stability, the triggering of innate immunity responses, the modulation of cell-cell connections, and cell trafficking through the circulatory system have been attributed to individual lectins or proteins containing lectin domains (Singh and Sarathi, 2012). Concurrent research demonstrated the presence of lectin molecules in all edible plant sources, and many foreign lectins were found to be associated with human health problems. Another area of lectin biology came to light when lectins were demonstrated to play key functions in amoebic, bacterial and viral infections. It is worth mentioning that since the predominant lectin identification method uses mammalian cell surface carbohydrates, already identified lectins show a skewed preference for animal glycans. For the same reason, lectins that do not bind to the mammalian cell surface glycans are underrepresented and new methods of detection need to be developed.
Molecular aspects of carbohydrate binding

Flexibility

Oligosaccharides are exceptionally flexible molecules that can adopt multiple different conformations (Carver, 1993) due to the high rotational degrees of freedom of glycosidic bonds. For a lectin to bind to an oligosaccharide with high specificity and affinity, it has to either screen the most favorable conformer or induce the best-fit state in a heterogeneous population. This is a challenging task as the most populated conformation is often energetically distant from the ideal binding conformation (Imberty and Perez, 2000; Sharon and Lis, 2007). To overcome this problem, lectins must maintain a delicate balance between entropic contributions, which come from flexibility, and enthalpic contributions, which arise from bonding and hydrophobic interactions. Lectins can theoretically achieve this by reducing the entropic change without disturbing enthalpy, by intensifying the enthalpic stability without increasing entropy, or through some combination of both (Imberty and Perez, 2000). Crystal and NMR structures along with other experiments suggest a sharp reduction in the entropy and a major gain in enthalpic energy of the oligosaccharide upon binding. Presumably, solvation and desolvation of the carbohydrate molecules upon binding provide substantial impacts to the binding reaction through both entropic and enthalpic contributions. Some lectins avoid such a scenario by choosing to bind the lowest energy state of the ligand, as in the case of the Griffonia simplicifolia lectin-1B₄ binding to Galα3Gal (Tempel et al., 2002).
The N-linked glycan pentasaccharide core Manα6(Manα3)Manβ4-GlcNAcβ4GlcNAc is another example of an oligosaccharide whose conformational dynamics have been studied. The presence of two possible rotamers of the α6 mannose along the C5-C6 bond (to the previous mannose) makes it exceptionally heterogeneous in nature. Interestingly, the addition of a β 1-4 linked N-acetylglucosamine to this mannose introduces bias for one of those two rotamers (Sharon and Lis, 2007) and thus reducing the heterogeneity of the system. Flexibility in this glycan is of particular interest because two lectin domains described in this thesis exhibit nanomolar affinity to this particular pentasaccharide.

**Multivalency**

Multivalency of the lectin-carbohydrate binding reaction transforms millimolar affinities for monosaccharides to nano-micromolar affinities of polysaccharides through cooperativity. This can be achieved through introducing avidity to the system by binding to multiple substrates or by creating clustered lectin complexes. In the case of cholera toxin B (CTB), as discussed previously, the pentameric association of the B subunit boosts the affinity for substrates from 40 nM to 40 pM compared to the monomeric form. Moreover, multivalency can also modify the specificity of binding. For example, concanavalin A displays only a 4-fold higher affinity for Me-α-Mannose (MMA) than Me-α-glucose (MAG) monomers. However, when multivalent derivatives of the same sugars are used, the specificity for MMA becomes >150 fold higher compared to MAG (Mortell et al., 1996).
Energetics of binding

The binding of carbohydrate molecules to lectins starts with the presentation of a completely solvated sugar to a solvated binding site. This results in concurrent desolvation of both the molecules, aiding in specific interactions between the protein and ligand. The carbohydrate molecule must cross the entropic barrier at this stage for successful binding to occur. The reaction proceeds forward depending on the difference between the energetic contribution of protein-water and sugar-water complexes compared to that of protein-sugar binding (Sharon and Lis, 2007). It is possible that the release of higher energy water molecules from the binding site provide an additional enthalpic and entropic impetus for binding (Lemieux, 1996). This, along with numerous hydrogen bonds, van der Waals interactions, and hydrophobic interactions that will be formed between the protein and sugar (Sharon and Lis, 2007). Experimental data has revealed that carbohydrate molecules binding to lectins are an enthalpy-driven process, however, multivalent binding benefits from large entropic contributions (Dam et al., 2000). As discussed previously, protein-carbohydrate binding is a constant battle between the unfavorable entropy from lost rotational and torsional degrees of freedom and favorable enthalpy from the expulsion of high energy water molecules and protein-ligand interactions (Sharon and Lis, 2007).
Old classification of lectins

Based on a very broad carbohydrate specificity, traditionally, lectins were divided into two classes – monosaccharide/oligosaccharide binding lectins and only oligosaccharide binding lectins (Wu et al., 2001).

Monosaccharide binding lectins

The primary specificity of lectins is defined by affinities for five monosaccharides - mannose, galactose/N-acetylgalactosamine, N-acetylglucosamine, fucose and neuraminic acid. There are some lectins that possess an affinity for more than one monosaccharide, but other monosaccharides binding to lectins are very unusual. Galactose-binding lectins are very abundantly present in different organisms, whereas mannose binding lectins are limited to mainly to vertebrates. One interesting observation regarding neuraminic acid (also known as sialic acid) binding lectins is that they bind to sialated oligosaccharides very tightly, but are unable to bind free sialic acids.

The mannose-specific lectins and Man/Glc grouped lectins are thought to have dual specificity for mannose and glucose molecules, as seen in the case of concanavalin A. Presently, the majority of identified members do not exhibit dual specificity and are specialized for mannose only. Some of the well-known lectins belonging to this group include mannose binding lectins (MBL) of animals, the *Pisum sativum* lectin, the *Crocus sativis* lectin, the lectin from *Allium sativum*, and the lectin from *E. coli* type 1 fimbriae. One interesting example of Man/Glc lectins are the
Leguminosae lectins, which show specificity for mannose and trimannoside Manα3(Manα6)Man molecules, but with a 100-fold stronger affinity for the later (Gupta et al., 1996).

Galactose (Gal) specific lectins generally show affinities for both galactose and N-acetylgalactosamine. However, many exceptions exist, including the PNA lectin, which does not bind to N-acetylgalactosamine and the DBL lectin, which binds to N-acetylgalactosamine, but not galactose. Perhaps one of the most famous galactose-specific lectin is the Galactin lectin, which is present in animals like rats, chickens and humans. All members of this lectin family bind to galactose or galactose-rich oligosaccharides in vitro and exhibit almost similar affinities for Gal and Galβ3GalNAc (Leffler & Barondes, 1986; Sparrow et al., 1987). Galactins are speculated to be involved in many molecular signaling events by interacting with various membrane proteins, CD-45, and phosphatases (Perillo et al., 1998).

Other groups of lectins are smaller in terms of the number of members identified. One of these smaller groups is fucose specific lectins, which were identified as agglutinins that precipitate blood group O erythrocytes. Among other groups, the N-acetylglucosamine binding lectins are generally found in cereals with wheat germ agglutinin (WGA) as the most well-known member of this family, interacting with both N-acetylglucosamine and N-acetylneuraminic acid. The neuraminic acid binding lectins prefer binding to sialic acid moieties attached to galactose or lactose molecules.
Oligosaccharide binding lectins

This lectin group can bind only to oligosaccharides, being unable to interact with monosaccharides. The majority of these lectins show a preference for terminal sialic-acid-containing oligosaccharides and can be found in pathogens like *H. pylori*, *Mycoplasma pneumonia*, and *Neisseria gonorrhoea*. Interestingly, these lectins perform important functions in the life cycle of the influenza virus, which is responsible for major health concerns across the globe. Some oligosaccharide-preferring lectins also show specificity to terminal galactose-containing oligosaccharides and are mostly located in bacteria. The precise function of these lectins is not well established, but they are proposed to be an integral part of the bacterial infection process.

Specificities for other terminal-residue-containing oligosaccharides are a little uncommon, but examples can be found including terminal N-acetylglucosamine, glucose, mannose or fucose-dependent binding. Some of these lectins, like calnexin and calreticulin, which is involved in the glycosylation pathway of the endoplasmic reticulum (Parodi et al., 2000), are found to play a significant role in key cellular processes.
Modern classification of lectins

Lectins are classified into three main categories – plant lectins, animal lectins and lectins from other sources. The plant and the animal lectins are individually subcategorized into nine subgroups based on their phylogenetic origin. To avoid such a complicated classification system, recently a simpler system has been proposed based on structural and evolutionary similarity. In this way, lectins can be divided into a total of seven groups and are as follows – R-type, L-type, P-type, C-type, I-type, the galectin family and microbial lectins.

R-type lectins

R-lectins are members of a superfamily of lectin, which are related to the ricin lectin from *R. communis*. Ricin is 60 kD protein with two separate subunits - the first is the cytotoxic A chain and the second is the lectin. The lectin domain is speculated to result from a gene duplication event and contain two similar sugar-binding domains. The sugar-binding sites are formed by close interactions between β-sheets in the β-trefoil domain (trfoil from Latin *trifolium* meaning three-leaved plant) and exhibit millimolar affinities for galactose and N-acetylglucosamine. However, on cellular surfaces, these domains display higher nanomolar binding, perhaps due to multivalency and avidity effects. The carbohydrate-binding module (CBM) of ricin has been identified not only in various plant and animal glycosyltransferases, but also in fungal and bacterial proteins. Interestingly, among the three symmetrical lobes of the ricin lectin domain, generally, not all of them are active against carbohydrates.
This activity is located within pockets containing \((QXW)_n\) motifs. Structural analyzes have shown the importance of aromatic residues within the binding cleft, which stacks against Gal/GalNac moieties on cell surface glycans.

Two outlaying examples of plant lectins belonging to this family include lectins from *Arbus precatorius* (SSA) and *Sambucus sieboldiana* (SNA). Even though these domains contain a conserved R-type carbohydrate-binding motif, they show specificity towards \(\alpha2-6\) linked sialated oligosaccharides. Several other cases of R-type lectins have been identified in ricin-like ribosome-inactivating proteins from various plant seeds and vegetative tissues (Peumans et al., 2001), but their functions are yet to be uncovered.

R-type lectins were first thought to be limited to plants. Over the years, members of this family have been identified in several well-known animal protein families, like the mannose receptor family and the \(\alpha\)-N-acetylgalactosaminetransferase (ppGalNacTs) family. The mannose receptor (MR) family proteins contain multiple lectin domains in tandem that are involved in controlling innate immune responses in animal systems. These MR family R-lectins exhibit millimolar binding for sulfated GalNac residues and are thought to be responsible for interacting with glycoprotein hormones or leukocyte membranes. ppGalNacTs, on the other hand, are involved in protein glycosylation pathways, and have at least 24 members in humans alone. All of these enzymes include C-terminal R-type lectin domains of \(\sim130\) amino acids, which are vital for the catalytic transfer
of sugar residues in some cases. Unfortunately, other proteins containing R-type lectin domains, like ENDO-180 and phospholipase A2 receptors, are not been well characterized. Sequence-based studies show that almost all proteins containing R-type related domains are involved in either O-linked biosynthesis or glycan metabolism. A handful of the proteins containing this domain has also been implicated in additional metabolic pathways like sphingolipid, glycerosphingolipid and nucleotide sugar synthesis pathways.

In the microbial world, R-type lectins are well represented. From *Streptomyces* to *Vibrio*, these lectin domains are found in proteins essential bacterial survival and growth. The lectin domain in endo-β1-4xylanase is one example of an R-type lectin that exhibits specificity for xylan, galactose, lactose and arabinose oligosaccharides. It is worth mentioning that in these lectins all three potential binding sites are active and assumed to be important for the function of the corresponding proteins.

Several of these domains are also present in various exotoxins produced by pathogenic *Vibrio* species and are thought to function by targeting cell surface carbohydrates, helping toxin binding to the host cell. In total, a massive ~3,500 R-type lectin domain-related genes are recognized in bacteria, the majority of which are proposed to be involved in host-pathogen and host-symbiont interactions.
**L-type Lectins**

The L-type lectins, or legume lectins, are a family of structurally related lectins found in hundreds of leguminous plants (Sharon and Lis, 1990; Sharon and Lis, 2007). This is extensively studied group of lectins are known for their structural stability. The L lectins are multimeric in nature with two to four subunits, each carrying an independent carbohydrate-binding site and forming homomeric or heteromeric complexes. Generally, each of the ~25 kDa subunits of requires calcium and manganese ions for binding to carbohydrates. Structurally these subunits are made from a single polypeptide, but are often cleaved into two pieces resulting in a small C-terminal part. Interestingly, the carbohydrate-binding site is only composed of four flexible loop regions and no well-structured part forms the binding pocket. These loops contain four extremely conserved and absolutely necessary amino acids for function. These amino acids include one aspartate and one alanine from loop A, one glycine or arginine from loop B, and one asparagine from loop C. The aspartate mediates a key hydrogen-bond with the ligand and contains an unusual *cis*-peptide bond indispensable for proper alignment of the side chain with the ligand. The high conformational energy of this rare *cis*-peptide bond is alleviated by the presence of the metal ions, which essentially stabilize the bond playing a key role in ligand binding. The glycine or arginine residue from loop B and the asparagine residue from loop C initiates hydrogen bonding with the sugar molecule through the backbone amide or side chain. The D loop does not participate in primary binding, however, the sequence variability and size of the loop contribute to substrate specificity.
The β-sheet and β-bend loops connecting the strands form a β-jelly roll or Swiss role topology (Srinivasan et al., 1996). The metal ions, mentioned previously, function by stabilizing the structural scaffold needed for binding. The presence of a calcium ion is imperative and cannot be substituted with any other metal ions. On the other hand, manganese is also essential for binding, but can be replaced with other transition metals like zinc, nickel, and cobalt. These two metal ions are positioned in the immediate vicinity (~9-13 Å) of the binding site and are 4.25 Å apart from each other as shown in the crystal structure. Removal of the metal ions from the lectin leads to an altered 3D structure and results in complete loss of ligand-binding activity, which can be reversed upon re-addition of the ions (Bouckaert et al., 1995; Bouckaert et al., 2000; Lescar et al., 2002).

Many of these lectins oligomerize to enhance their activity. The structured β-sheets are responsible for the multimerization of these lectins and mostly form dimeric or trimeric assemblies. Variability in oligomer formation, secretion, glycosylation, and proteolytic modifications make these lectins functionally diverse. In concanavalin A, a removal of the central peptide sequence and attachment of the original N and C-termini creates the final protein through a process called circularly permuted protein formation. This gives rise to a circular sequence homology with other legume lectins, where the homology can be found only after aligning residue 119 with the amino terminus of other lectins. Another unexpected sequence homology of these legumes exists between mannose-specific animal lectins.
containing the conserved L-lectin binding site (Roche and Monsigny, 2001; Velloso et al., 2002).

It is worth mentioning that L-lectins are present in both seeds and vegetative bodies, and both lectins are synthesized by two distinct genes with 90% sequence identity. However, the precise role of these lectins in seeds or in the vegetative body is unknown even after many decades of research. One possible function of these lectins is that they act as carbohydrate storage proteins. Another possible role could involve aiding in the nitrogen fixation process by interacting with the *Rhizobia* cells. However, this hypothesis has been criticized by an overwhelming majority of lectin biologists as the L-lectins do not exhibit a preference for plant glycan-based signaling molecules. A role in plant defense is another proposed function of these lectins as many of them are extremely toxic to animals. Additionally, they are thought to play a part in the innate immune response, but further research is necessary to confirm other hypotheses regarding the function of L-lectins.

It is important to note here that these types of lectins are not only limited to the plant world and many examples can be found in the microbial world. Recently studies have identified lectins with jelly-roll motifs in *Clostridium* neurotoxins, *Pseudomonas aeruginosa* exotoxin A, *V. cholerae* neuraminidase, and *Macrobdella decora* trans-sialidase enzyme – all of which are actively related to infectivity and pathogenicity of the corresponding microbes.
**P-type Lectins**

The P-type lectins were first identified in animal mannose-6-phosphate (Man6P) receptors (MPRs), which are transmembrane glycoproteins (Dahms and Hancock, 2002). Generally functioning as a dimer or tetramer, the MPRs can be divided into two groups. The first group is made of smaller 45 kDa subunits, which require Ca\(^{+2}\) for lectin binding (CD-MPR), whereas the second group contains large 300 kDa cation-independent subunits (CI-MPR). Both groups bind specifically to oligosaccharides with terminal Man6P residues and the large subunit binds other ligands with GlcNac-P-Man phosphodiester bonds, methylphosphomannose, and retionic acid.

The smaller subunits are extremely conserved among animals. For example, the human and the bovine subunits contain absolutely identical intracellular regions and share a staggering >90% identity in sequence. Although the CD-MPR was first purified and characterized as a Ca\(^{+2}\) dependent lectin, recent studies point out that the presence of calcium is not absolute and can be replaced by manganese or magnesium. Interestingly, the lectin displayed higher binding affinity when manganese ions were used. Overall, the CD-MPR binds to cell surface glycans containing two Man6P residues with high-nanomolar to low-micromolar affinity and binding to any other ligands have not been reported.

The CI-MPR is a very large protein with 2,451 amino acids and contains a very small 23-residue transmembrane region. The extracellular domain is comprised
of 15-homologous continuous repeats with ~145 amino acids each and this unique domain structure is instrumental in ligand recognition. Proteins with glycan units containing two Man6P residues stack in the binding cleft formed between repeat 3 and 9 or sometimes between repeat 11, 12 and 13. This way, binding to two mannose-6-phosphate residues per monomer improves CI-MPR’s affinity 10-100-fold (to the low-nanomolar range).

In recent years, many proteins, like cellular repressor of E1A-stimulated genes (CREG), thyroglobulin and leukemia inhibitory factor (LIF), have been predicted to be binding partners of P-type lectins. Moreover, a number of important viral glycoproteins containing Man6P motifs have been identified, which could be crucial for herpes simplex virus and varicella zoster virus infections.

**C-type Lectins**

C-type lectins have been identified in well over 1,000 proteins and show a compulsory requirement for calcium ions. Interestingly, the C-type lectins lack a strict sequence requirement. Theoretically predicted to be present in ~10^{13} different proteins, the C-type lectin is one of the most diverse domains in the protein world maintaining structural conservation despite dissimilar sequences (Vakri et al., 2009). Generally, C-type lectin domains (CTLDs) are small, with ~100-130 amino acids containing multiple α-helical and β-strand regions. The calcium-binding site utilizes glutamic or aspartic acid residues, which in turn form direct interactions with individual carbohydrate molecules. Among the four calcium binding sites, the second
binding site is perhaps the most dominant and helps in the formation of a three-way complex between ligand, lectin and metal ions. These weak sites are located in the loop regions and contain weakly conserved sequences like EPN and WND. Upon calcium binding, these loops undergo a large conformational change necessary to stabilizing the carbohydrate binding site.

CTLDs are found in various different proteins. In humans alone, more than 100 proteins are predicted to contain these domains. Some proteins carry only single CTLDs, but others contain multiple numbers, like in the macrophage mannose receptor (MMR). According to their function, CTLD-containing proteins are divided into at least seventeen different groups. However, some proteins with CTLDs have been identified lacking the ability to bind sugars, like some natural killer cell receptors (Drickamer, 1999; Kogelberg and Feizi, 2001).

CTLDs have been speculated to play important roles in host-pathogen interactions, as these domains are widely spread not only in metazoans, but also across the microbial world. In a study in *C. elegans*, it was seen that host CTLD genes take part in the response against different pathogens in concert with the many other immune proteins. Some domains which have reported roles in pathogen sensing include MMR lectins from *Mycobacterium tuberculosis*, *C. albicans*, HIV; MBL collectins from *S. aureus*, *S. pneumoniae*, *Aspergillus fumigates*; type-II receptors from dengue, *H. pylori*, *M. tuberculosis*, *Leishmania*, *S. mansoni*, *C. albicans*; and NK receptors from *Pneumocystis* and *C. albicans* (Cambi et al., 2005).
Therefore, CTLDs do not just act as cell-cell adhesion or cell signaling molecules, but instead, they form an important aspect of innate immunity by recognizing and identifying pathogenic microbes. Although the mechanisms of these observations are still unknown, some complex signaling pathways involving Toll-like receptors and lipid-raft platforms have been proposed.

**I-type Lectins**

I-type lectins are sugar-binding proteins from the immunoglobulin superfamily (IgSF). The immunoglobulin-like (Ig) fold was identified in immunoglobulins and consists of ~100 amino acids containing only β-strands. Alternatively, another term, siglec (sialic acid-binding, immunoglobulin-like lectin), was proposed for the core members of this superfamily. The siglec family is comprised of type I membrane glycoproteins that show specificity towards sialic acid residues (Munday et al., 1999) and display sequence similarity with the V-domain of IgG (Sharon and Lis, 2007). The rest of the protein contains tandem repeats of C2-set Ig domains with varying numbers. All C2 segments share 50-80% sequence identity with the Siglec-3 protein and together they form the CD33-related siglec group. Additionally, an unusual disulfide bond between the V-set domain and the C-set domain promotes tight packing in the molecule, which is necessary for its cellular function or proper folding of the protein. Another interesting characteristic of siglecs is the presence of ITIM (immunoreceptor tyrosine-based inhibition motif) sequences on the cytoplasmic side of the membrane, which functions in the regulation of these receptors. Other than the
V-set IgG domain, all additional domains are generally not involved in the carbohydrate binding activity of the protein and removal of these domains does not affect the lectin activity (Nath and Crocker, 1995).

Siglecs exhibit an absolute necessity for sialic acid residues for binding to glycan targets like many other sia-specific lectins, such as selectins. The major difference between other sia-specific lectins and siglecs is the mechanism of binding. Siglecs, unlike the others, do not only interact with the carboxyl terminus, instead, they physically connect with the glycerol group, N-acyl group and the hydroxyl group in the sialic acid. Biophysical methods were employed to measure the binding affinities of siglec-glycan interactions, which fell in the micromolar range. This is low, but not unexpected, as potential multimeric assemblies in the membrane can enhance the affinities >10-100 fold. Moreover, the excess presence of sialated glycoproteins on the membrane may also augment the binding kinetics of siglec proteins. Interestingly, all of these proteins are so specific for sialated glycoproteins that cells can change the expression profile of cis sialated glycoproteins for regulation (Powell and Varki, 1995).

Many classical immunoglobulins have been found to have carbohydrate binding properties. It is possible that the I-type lectins initially developed from ancient immunoglobulins by divergent evolution of the binding site (Langman and Cohn, 1987) to fulfill the demand of carbohydrate-based signaling in multicellular organisms (Powell and Varki, 1995). However, some evidence also suggests that the
classical immunoglobulins and IgSFs are outcomes of closely related parallel evolutionary pathways (William et al., 1989; Hunkapiller and Hood, 1989; Langman and Cohn, 1987; Powell and Varki, 1995).

**Galectins**

Galectins are β-galactose-specific lectins found mainly in the animal kingdom. These domains are small in size (130 amino acids) and generally contain one sugar-binding site per monomer. According to domain arrangements, galectins can be divided into three categories – prototypical galectins, chimeric galectins and tandem repeat galectins. Almost all members of this family bind to galactose-containing sugar moieties like disaccharides or trisaccharides with micromolar to millimolar affinities in vitro, whereas cell surface glycoconjugate binding occurs at much greater affinity. A surprising diversity of cell surface glycan binding partners has also been identified with different galectins. For example, galectin-1 interacts with the terminal β-gal-containing N-acetyllactosamine complexes while galectin-3 shows specificity towards repeating units of [-3Galβ1-4GlcNAcβ1-]. Structural studies demonstrate the importance of the C4, C5 hydroxyl groups of galactose and the C3 hydroxyl group of N-acetylglucosamine in the binding process. Additionally, these exclusively β-strand-containing proteins show some resemblance with β-sandwich or jelly-roll motifs, and in some cases have been predicted to have disordered linking peptide regions (Leffler, 2001; Barondes et al., 1994).
The biological functions of galectins have been well characterized and mostly fall in the realm of cell-cell or cell-matrix adhesion. Some of the important roles galectins perform include regulation of immune responses and inflammation, control of T-cell apoptosis, management of internal organ development, and induction of the selective loss of CD4\(^+\) and CD8\(^+\) cells. Furthermore, some reports have also predicted roles in innate immunity and tumor survival (Barondes et al., 1994; Leffler, 2001; Liu et al., 2002; Hsu et al., 2006; Lie and Rabinovich, 2005).
Figure 1. 13 **Structural motifs found in the lectin family.**
Microbial lectins

In recent years, a number of microbial lectins with significant functions in the microbial lifecycle have been characterized biochemically.

Viral lectins and hemagglutinin

Viral lectins are highly diverse, not sharing major structural features among them. The most systematically examined viral lectin is influenza hemagglutinin, which is a heterodimer of two smaller proteins (Sauter et al., 1992; Wiley and Skehel, 1987). This hemagglutinin binds to Neu5Acα2-6Gal-containing glycans in the upper respiratory track of infected humans (Tharakaraman et al., 2013). Another example of a viral carbohydrate binding protein is the murine polyoma virus protein 1 (VP1) containing presumptive jelly-roll motifs. The herpes simplex glycoproteins, HIV gp120, and Dengue envelop proteins demonstrate specificity for sulfated heparans on their consecutive receptors.

Bacterial lectins

Lectins found in bacteria are generally part of larger proteins like adhesins, fimbriae, and other virulence factors (Sharon and Lis, 2007). Some examples of soluble lectins that have been biochemically characterized include Pseudomonas aeruginosa PA-IL and PA-IIL (Gilboa-Garber et al., 1997) and Nostoc ellipsosporum cyanovirirn-N (Bewley et al., 1998). Membrane-bound or cell-surface lectins are more well-known and well-investigated and contain fimbriae (from Actinomyces,
Bordetella), P fimbriae (from E. coli), type 1 fimbriae (from E. coli), type IV pili (from Pseudomonas aeruginosa) and K99 fimbriae (from E. coli) (Choudhury et al., 1999; Sauer et al., 2000; Forest and Tainer, 1997; Sharon and Lis, 2007). Additionally many bacterial virulence factors and toxins also carry lectin genes for host cell binding and play an indispensable role in bacterial pathogenesis.

**Lectins and bacteria**

In the 1950s, two scientists, Duguid and Brinton, first established that many bacteria are able to influence erythrocytes the same way as pure hemagglutinins do. Further research proved that these effects are the result of bacterial hemagglutinins or lectins, which follow similar structural-functional aspects as the eukaryotic ones (Duguid and Old, 1980; Brinton, 1965; Sharon, 1986). The bacterial lectins were primarily divided into two classes, mannose sensitive and mannose resistant (Sharon, 1987), but recent classification has created multiple groups out of the mannose resistant class because a varying specificity to different carbohydrates. Previous studies identified a number of bacterial cell surface lectins and recent discoveries recognize much more lectin domains as a part of bacterial cytoplasmic or secreted proteins. The functions of different kinds of bacterial proteins containing lectin domains can be broadly (and loosely) categorized into two groups – lectins involved in adhesion to the host and lectins involved in target recognition of bacterial toxins.
Lectins are unquestionably required for host recognition and binding by bacterial cells. Primarily, bacterial lectins were thought to function only in the cell adhesion process, and one of the first molecules that were characterized was *E. coli* fimbriae (Sharon, 1987). All strains of *E. coli*, and many other enteric pathogens, express fimbriae proteins dispersed throughout the bacterial membrane. These molecules are extremely hydrophobic and capable of agglutinating yeast cells and mammalian red blood cells through a mannose-dependent pathway. It has been speculated that the carbohydrate binding activity resides in a tip of the fimbriae, which is most likely to interact with a eukaryotic cell (Sharon, 1987). Oligosaccharide specificity of many different types of fimbriae has been studied with great detail and the top three binders are Manα6[Manα3]Manα6[Manα3]ManαOMe, Manα6[Manα3]Manα6[Manα2Manα3]ManαOMe and Manα3Manβ4GlcNAc - all with high mannose content. These carbohydrate-binding events are perhaps the first true specific contact between a bacterial cell and the host. Again and again, many reports demonstrate the ability of simple sugars to minimize infection in animal urinary models not only in *E. coli*, but also in infections caused by *Pseudomonas mirabilis*, *Klebsiella pneumoniae*, and *Shigella flexneri*. Recent studies have identified multiple fimbriae-associated and fimbriae-independent carbohydrate binding adhesin molecules, like AfaE, antigen-43, AIDA-I, TibA, and intimin (Ohlsen et al., 2009). Interestingly, most of the gram-positive adhesins studied bind to protein receptors instead of sugar or glycan partners, as seen in many gram-negative species. For their potential role in early phases of infection, many anti-adhesin
strategies have been investigated to inhibit bacteria-host tissue infections. Unfortunately, as bacteria use multiple types of adhesins to attach to the host cells at the same time, multiple binding events need to be targeted for the development of an efficient antibiotic.

Apart from the adhesin molecules, several bacterial toxins also carry lectin domains and interact with cell surface glycans. As many of these toxins act as oligomers, the membrane-binding step represents the slowest and perhaps most important phase in the toxin’s function. Some examples of potent bacterial toxins with carbohydrate or glycan binding activities include botulinum toxin from *Clostridium botulinum*, Cry toxins from *Bacillus thuringensis*, Shiga toxin from *Shigella dysenteriae*, toxins from *Bordetella pertussis*, Cholera toxin from *V. cholerae*, VCC from *V. cholerae*, and hemolysin from *V. vulnificus*.

Botulinum toxin from *C. botulinum* (BoNT) contains three subunits with the C-terminal portion of the protein interacting with glycosphingolipids on the membrane of neuronal cells. BoNT uses a conserved motif of H-(X)\(_n\)-SXWY-(X)\(_n\)-G to bind to gangliosides of the Glb series, G\(_{D2}\), G\(_{T1b}\), G\(_{D1b}\), and G\(_{Q1b}\) specifically (Singh et al., 2000; Kitamura et al., 2005). Similar ganglioside binding partners can be found in additional toxins like tetanus toxin (Kitamura et al., 2005) and cholera toxin (CT) (Holmgren et al., 1975). Another toxin family which works in a similar way is the shiga and shiga-related toxin family. In this family, toxins target and attach to cell surfaces through G\(_{b3}\) ganglioside binding by the 1B subunit and get internalized (Sandvig, 2001; Kovbasnjuk et al., 2001). Each of the previously mentioned toxins
belongs to the AB type toxin family and contains two subunits, one for target binding and the other for toxicity. Interestingly, for all of these toxins the sugar-binding steps are found to be essential for the activity of the toxins.

**Lectin domains in V. cholerae**

*Vibrio cholerae* carry a number of lectin genes in its two chromosomes. Several proteins that contain these domains are of fundamental importance in *V. cholerae*’s infectivity and pathogenesis. Some of these critical lectin-containing proteins include *V. cholerae* mannose/glucose specific hemagglutinin (MSHA), neuraminidase, cholera toxin, *Vibrio cholerae* cytolysin (VCC), biofilm-associated protein (Bap1), and rugosity and biofilm structure modulator C (RbmC).

**V. cholerae mannose/glucose specific hemagglutinin (MSHA)**

MSHA proteins constitute type-4 pili structures of bacterial cells. *V. cholerae* MSHA displays a strong preference for mannose, glucose, N-acetylglucosamine, and glucosamine molecules, but exhibits no specificity for fucose (Sasmal et al., 1997; Finkelstein et al., 1987). The MSHA pilus functions by aiding the attachment of bacterial cells to host tissue or cell surfaces. Additionally, type 4 pili are involved in lateral DNA transfer and bacteriophage infection, both of which are instrumental in diversifying the gene pool of *V. cholerae* (Marsh and Taylor, 1999). Sequence homologs of this protein in *Pseudoalteromonas tunicate*
were established to be an initiator and promoter of cell surface attachment and play additional roles in environment sensing (Dalisay et al., 2006).

**Neuraminidase**

The neuraminidase enzyme is part of the mucinase complex of *V. cholerae* and works by processing sialated GM1 residues on the cell surface. The main cause of the characteristic diarrheal symptoms is classical cholera toxin (CT), which binds to neuraminidase treated GM1 residues as receptors. Neuraminidase itself contains two L-type lectin domains and both are relatives of L-type coral tree lectins. These exhibit strong structural similarity (1.8Å rmsd) despite very little sequence similarity (23%). As discussed previously, L-type lectins require closely fit Ca$^{+2}$ ions as a structural component of the domains. Surprisingly, neither the metal ion nor the loops coordinating the metal ion is present in *V. cholerae* neuraminidase. The function of these domains is poorly understood (Crenell et al., 1994; Moustafa et al., 2004).

**Cholera toxin (CT)**

Cholera toxin (CT) is the main causative agent of cholera symptoms. CT works as an ABs type toxin using pentameric B subunits to attach to GM1 gangliosides on epithelial cell membranes. Once the complex is bound to the cell membrane, the entire complex gets internalized, where after proteolytic processing, the A subunit is released into the cytoplasm (Nedelkoska and Benjamins, 1998; Rodighiero et al., 2001; Uwiera et al., 1992). The CT toxin structure and function have been reviewed elsewhere in this chapter in detail.
**Vibrio cholera cytolysin (VCC)**

VCC contains two lectins domains, which are not related to each other. The first domain is a β-trefoil lectin domain and the second one is a β-prism lectin domain. The β-trefoil lectin domain (VCC\(_{β-	ext{trefoil}}\)) exhibits sequence similarity with R-type lectins, especially within the B subunit of ricin. Ricin contains two lectin domains with only one of them binding to carbohydrates. This domain is widely present in hemolysins from various *Vibrio* bacteria like *V. vulnificus*, *V. cholerae*, *V. anguillarum*, *V. fluvialis*, *V. coralliilyticus*, *V. mimicus* and *V. campbelli*, but only a handful of these proteins have been biochemically characterized (Kaus et al., 2014). A small tripeptide sequence containing QXW is the functional motif for carbohydrate binding, but several proteins, including VCC, are known to carry a non-functional QXW motif (Rutenber et al., 1987; Loris, 2002). Unfortunately, very little is known about the functional aspects of the VCC\(_{β-	ext{trefoil}}\) domain and possibly it is present in the PFT as an evolutionary scaffold. *V. vulnificus* hemolysin has a domain related to the VCC \(β\text{-trefoil}\) and functional studies have been performed on it. Structurally, the VVH \(β\text{-trefoil}\) domain possesses a pseudo three-fold axis of symmetry and binds to N-acetyl-D-galactosamine (GalNAc) and N-acetyl-D-lactosamine (LacNAc) with micromolar affinities (Kaus et al., 2014). However, under *in vivo* conditions this binding may occur at nanomolar or sub-nanomolar ranges through ligand clustering, avidity effects and the presence of multiple binding sites. In contrast, the VCC\(_{β\text{-trefoil}}\) domain when tested with glycan screening or individual monosaccharides and disaccharides did not exhibit any binding even at millimolar concentrations. Upon closer inspection of the
structures of the VVHβ-trefoil domain, Kaus et al. discovered an eight amino acid loop with sequence YGENGDKT making key hydrogen bonds and van der Waals interactions with the ligands in VVH. The Ricinβ-trefoil domain, which is active against Galβ1–4GlcNAc or GalNAcβ1–4GlcNAc moieties, also have a variant of this loop with sequence RDGRF present (Cummings and Etzler, 2009). This loop surprisingly is missing in VCC and could be the reason behind the non-functional nature of this domain. Additionally, the VVHβ-trefoil domain forms a heptameric assembly spontaneously under crystallization conditions, which may suggest that these domains may act as a structural scaffold for oligomerization in VVH or VCC.

The C-terminal sequence of VCC is a small 15-kDa β-prism lectin domain (VCCβ-prism), which is a sequence and a structural homolog of the jacalin lectin family. The VCCβ-prism domain is not as broadly present as the VCCβ-trefoil domain in the Vibrio family. Some examples of Vibrio species containing the β-prism lectin are V. mimicus, V. campbelli, V. fluvialis, V. furnissii, and V. anguillarum. In VCC, the removal of the β-prism domain results in a 99% loss in activity and the domain itself associates with β-galactosyl residues and erythrocyte membranes with a 10^{-7} M affinity (Olson et al., 2005; Saha and Banerjee, 1997; Ganguly et al., 2014). When the VCC monomer structure was solved, Olson et al. saw a β-octoglucoside (β-OG) molecule, a detergent with a glucose head group, bound to the VCCβ-prism domain. Additionally, they identified key hydrogen bonds between the β-OG and the protein through the D617 sidechain. In our hands, mutating the D617 residue causes a >100-fold loss of hemolysis activity of VCC and a 200-fold loss in sugar binding
affinity (Levan et al., 2013). Also, the structure of the VCC$_{\beta\text{-prism}}$ bound to methyl-\(\alpha\)-mannose (MMA) supported conclusions from the first structure of the VCC monomer bound to \(\beta\text{-OG}\) and reiterated the importance of D617 in sugar binding (Levan et al., 2013). Moreover, we demonstrated that the VCC$_{\beta\text{-prism}}$ domain interacts with an N-linked heptasaccharide core (NGA2) with nanomolar affinity and thereby helps the toxin to target the cell surface. Structurally, this domain is speculated to perform multiple functions. Firstly, in the monomer structure, this domain prevents the pre-stem region from falling down and inserting into the membrane, thus preventing immature membrane insertion. Secondly, some studies speculate a role of this C-terminal domain in the pre-pore to pore transition (Mazumdar et al., 2011). However, truncation studies establish that a \(\beta\text{-prism}\)-truncated version can form a similar heptameric pore as the wild-type protein itself (He and Olson, 2009). One hypothesis put forward is that the VCC$_{\beta\text{-prism}}$ acts as a structural scaffold that holds the whole structure together, and in the absence of this domain the other lectin domain crumples on the cytolysin domain resulting in a complete collapse of the structure (Rai and Chattopadhyay, 2015). Several single mutations have been reported, which cause a few fold loss of activity, but the precise mechanism of glycan binding is still not known. Other than the NGA2 glycan, glycoproteins have been used as a model for cell surface binding. Asialofetuin, a glycoconjugate in calf serum, shows an astonishing lower nanomolar affinity with the VCC$_{\beta\text{-prism}}$ domain (Rai et al., 2012). Interestingly, the sialylated version of asialofetuin, fetuin, demonstrates a \(~30\)-fold lower affinity for the VCC$_{\beta\text{-prism}}$ domain. This is interesting, as cholera toxin also
requires neuraminidase to asialate the receptor for binding. It is possible that VCC also has similar requirements, and perhaps the same neuraminidase prepares the cell surface receptors for VCC.

**Biofilm associated protein (Bap1) and Rugosity and biofilm structure modulator C (RbmC)**

The proteins Bap1 and RbmC play instrumental roles in *V. cholerae* biofilm formation. Both proteins, along with a third called RbmA, are secreted coating the bacterial cells (Teschler et al., 2015; Absalon et al., 2012). Bap1 and RbmC both contain β-prism lectin domains, with one in Bap1 and two in RbmC. All of these lectins domains are sequence homologs of the VCC β-prism domain. Functionally, Bap1 and RbmC form the matrix of the biofilm and are dynamic enough to accommodate growing cells and combat against environmental changes inside the host epithelia. Overall, these proteins help bacterial cells adhere to the surface of the human epithelial lining and to the biofilm itself (Berk et al., 2012; Giglio et al., 2013; Absalon et al., 2011; Fong and Yidiz, 2015). As expected, deletion of the two genes causes reduced biofilm formation, and when absent together, the bacteria completely lose the capability to form biofilms (Fong and Yidiz, 2007).
CHAPTER 2
STRUCTURE OF THE VCC HEPTAMER
**Introduction**

*Vibrio Cholerae* cytolysin (VCC) is a pore-forming toxin produced by the endemic gastroenteric pathogen *Vibrio cholerae*. Similar to other pore-forming toxins, VCC has been shown to lyse various types of human cell lines including intestinal and immune cells (Zitzer et al., 1997; Valeva et al., 2008; Debellis et al., 2009). Additionally, VCC is able to cause gastrointestinal fluid accumulation in rabbit models and is involved in enterotoxicity mediated by *V. cholera* infection. On the cellular level, this toxin causes inhibition of protein synthesis, cell vacuolization and cell apoptosis in model human cell lines (Rivas et al., 2015; Saka et al., 2007).

VCC, like other PFTs, is secreted as a water-soluble monomer that targets specific cell surfaces and forms a membrane spanning β-barrel pore (Olson and Gouaux, 2005). This interesting mechanism of action creates some big obstacles that the toxin must overcome in order to be pathogenic. The first one of these is how to hide hydrophobic residues, which are instrumental in forming the membrane-spanning pore, from the aqueous solvent phase in the water-soluble stage. The second obstacle is to find a suitable membrane and reorganize hidden hydrophobic residues to insert into the membrane. Although the membrane spanning region of VCC and other PFTs are relatively small compared to the rest of the protein, the hydrophobic mismatch could cause devastating consequences for the protein if not taken care of properly.
Based on the type of secondary structure that penetrates the membrane, PFTs are classified into \( \alpha \)-PFTs or \( \beta \)-PFTs. As mentioned earlier, VCC belongs to the \( \beta \)-PFTs as it forms a \( \beta \)-barrel membrane pore. Although low in sequence similarity and identity, the *Staphylococcus aureus* PFTs are predicted to be structural homologs of VCC (Olson and Gouaux, 2003). Interestingly, several structures of the *Staphylococcus* PFTs are available in both water soluble (LukF and LukS) and oligomeric pore states (\( \alpha \)-hemolysin). The water-soluble structure of the VCC monomer was also solved, but very little is currently known regarding the membrane spanning oligomer (Olson and Gouaux, 2005).

Another structural homolog of VCC is the *Bacillus anthracis* protective antigen (PA) channel. This PFT is instrumental in anthrax infection and functions by translocating two additional enzymatic toxins into the cytosol of the cell. The two translocating toxins, edema factor and lethal factor, are the main causative agents of cytotoxicity, but are absolutely ineffective in the absence of the PA pore. One very striking similarity between the PA channel and VCC is the requirement of proteolytic activation, which is not often a necessity for other PFTs (Petosa et al., 1997). Secreted by *Aeromonas hydrophilla*, Aerolysin is another PFT, which displays overall structural similarity to VCC, even though it posesses very low sequence identity (Fivaz et al., 2001).

The cholesterol-dependent cytolysins (CDCs) are another large group of pore-forming toxins, which belong to the same \( \beta \)-PFT family. Until today, at least 22 CDCs have been identified, which are characterized by their absolute requirement for
cholesterol on the membrane and for the massive pores they form constituting 20-100 monomeric toxins. CDCs are generally distinct from VCC in several ways including the size of the pore or and diverse sequence similarity. The main similarity between these two β-PFTs is that VCC has also been demonstrated to prefer cholesterol-rich membranes for pore formation.

The VCC monomer structure was solved to 2.3 Å and bears striking structural similarities to the *Staphylococcus* β-PFTs. Structurally, the VCC protoxin consists of four major domains named prodomain, cytolysin domain, β-trefoil lectin domain and the β-prism lectin domain. The VCC prodomain is a small ~15kDa domain, which needs to be proteolytically removed for activating the toxin. Although, the precise function of the domain is not clear, the prodomain is postulated to help in the toxin’s folding and stability. The cytolysin domain, on the other hand, is the central core of the protein and also contains the prestem region, which forms the membrane-spanning pore. Additionally, this domain may also interact with the membrane for proper targeting and may be responsible for the cholesterol preference of VCC. Interestingly, VCC carries two lectin domains in its C-terminal region and both of them have pseudo three-fold symmetry in their structure. The β-trefoil lectin is related to the R-type lectin family, which includes the ricin lectin, whereas the β-prism lectin shows similarity to the jacalin lectins. Although the β-prism lectin is assumed to bind cell surface carbohydrate molecules, the function of the β-trefoil lectin is not well understood.
Several studies have looked at the mechanism of pore formation by different β-PFTs and multiple models have been proposed. Experiments support one of the models, where each monomer binds to the membrane independently, the monomers oligomerize on the membrane to form a pre-pore stage, and finally the mature pore is formed through a membrane insertion process. Again, several aspects of this process are not completely understood and further studies are needed to understand the mechanism with more precision.

VCC has been the subject of a number of studies, which have reported various aspects of the membrane interaction, membrane targeting, cholesterol preference, and mechanism of pore formation. However, other than a very low resolution (He and Olson, 2010) cryo-electron microscopic structure, no additional useful information is available regarding the structural aspects of the pore. In previous studies, the final heptameric assembly and overall domain rearrangements that occur during the aqueous to transmembrane forms have been shown, but the resolution of these studies was insufficient to show a clear picture (He and Olson, 2010). In this study, we solved the structure of the detergent-solubilized VCC oligomer to 2.9 Å revealing structural features of the VCC heptamer. Firstly, there is a massive rearrangement of the β-prism lectin domain to a completely different face of the protein, enabling the otherwise blocked prestem region to insert into the membrane. Secondly, we also showed the presence of charged amino acids inside the pore and a tryptophan ring, which constricts the pore diameter to only 8 Å, both of which resemble the anthrax
PA channel. These observations could potentially uncover unknown aspects of VCC and may help us understand the function of the toxin in the pathogenicity of cholera.

**Materials and methods**

**Assembly of the VCC oligomer**

Monomeric VCC was expressed and purified as described previously (Olson and Gouaux, 2005). Trypsin-cleaved toxin was added to freshly prepared soybean asolectin liposomes (1:2 protein:lipid ratio) containing 20% cholesterol and incubated for 30 min at room temperature. Following centrifugation for 30 min at 40 K rpm in a Beckman TL-100 rotor tabletop ultracentrifuge, the oligomeric toxin was solubilized in a buffer containing 20 mM Tris pH 7.6, 150 mM NaCl, 1 mM EDTA, and 40 mM C_{10}E_{6}. After a second centrifugation step, the solubilized toxin was purified over a Superose 6 column in a buffer containing 20 mM Tris pH 7.6, 150 mM NaCl, 1 mM EDTA, and 1 mM C_{10}E_{6}. Peak fractions were concentrated to 10 mg/mL before crystallization using a 100-kDa cutoff membrane (Millipore Ultrafree).

**Crystallization and data collection**

Oligomeric toxin was crystallized by the hanging drop method. One microliter of the concentrated oligomer was added to an equal volume of buffer containing 20 mM HEPES pH 7.6, 10% PEG 2000, and 20 mM Co(NH_{3})_{6}Cl_{3}. Crystals were cryoprotected in 20% glycerol before freezing in liquid nitrogen. For heavy atom derivatization, crystals were soaked overnight in a buffer containing 20 mM HEPES pH 7.6, 10% PEG 2000, 10% glycerol, and 1 mM Ta_{6}Br_{12}^{2+} clusters (Jena
Bioscience). All native and heavy-atom data were collected at beamline 12-2 at the Stanford Synchrotron Radiation Laboratory (SSRL), and X-ray data were processed by HKL2000 (Otwinowksi and Minor, 1997). MAD data were collected using the inverse-beam method at the peak, infection point, and high remote energies based on tantalum fluorescence scans.

**Structure solving and refinement**

The program SHELX (Sheldrick, 2008) located cluster sites and provided preliminary phases (using data to 6.5 Å) resulting in electron density maps with a clear molecular outline for two VCC heptamer molecules in the asymmetric unit. Phases were extended to 4.0 Å by solvent flattening and 14-fold NCS averaging using the program PARROT. Real-space 14-fold NCS averaging by COOT (Emsley and Cowtan, 2004) allowed manual docking of the three main domains of the 2.3-Å VCC monomer structure [Protein Data Bank (PDB) ID code 1XEZ]. The resulting 42 domains were subjected to rigid-body refinement [REFMAC (Murshudov et al., 1997)], first in the 4-Å data and then in a native dataset diffracting to 2.9 Å (resulting in an R factor of 44%). Missing regions, including the transmembrane channel region and several connecting loops, were built into 14-fold NCS averaged and B-factor sharpened $2Fo - Fc$ electron density maps. Alternating refinement and model-rebuilding steps were carried out using the PHENIX suite (Adams et al., 2002) and COOT. NCS restraints (coordinate sigma = 0.05 and B-factor weight = 10) were used in refinement, with residues 589–716 restrained in a second NCS group for chains A and L (due to crystal-contact induced loop rearrangements). Model validation was
performed using Molprobity (Davis et al., 2007). Movies were created using the Gerstein morph script (Adams et al., 2002; Davis et al., 2007) and CNS (Brunger et al., 1998).

**Results**

**VCC heptamer structure**

The VCC heptamer forms a ring-like structure approximately 140 Å high perpendicular to the membrane plane with a widest outer diameter of 135 Å. The channel pore, which runs along the 7-fold symmetry axis of the heptamer, is separated into an upper large vestibule, formed by β-prism and cytolysin domains, and a 14-strand transmembrane β-barrel formed by the stem domains. The β-barrel has a backbone-to-backbone diameter of approximately 25 Å and is ringed by a single layer of aromatic residues (composed of F288 and Y313) commonly observed in transmembrane β-barrel proteins near the lipid/solvent interface (Gouaux, 1998). However, VCC is missing the second ring on the opposite side of the barrel predicted to occur in many β-PFTs [primarily phenylalanines (Iacovache et al., 2006)]. At its narrowest point, the β-barrel has an approximately 8-Å-wide constriction [calculated by the program HOLE (Smart et al., 1993)] formed by a heptad of tryptophan residues (W318) conserved throughout many *Vibrio* species (except *Vibrio vulnificus*), but not *Aeromonas* species (which also lack the β-prism lectin). The upper vestibule surface is primarily acidic, whereas the β-barrel region contains alternating bands of basic and acidic amino acids with two rings of lysine residues near the intracellular opening of the stem (K304 and K306). This is in contrast to staph α-HL,
which is more neutral in character and wider within the channel (approximately 10-Å constriction formed by M113 and K147). Consistent with these observations, electrophysiological measurements of single channels in planar lipid-bilayers indicate that VCC is moderately anion-selective and has a 4-fold lower conductivity than α-HL (Menzl et al., 1996). Previous modeling studies of the VCC pore predicted an excess of positive charge near the intracellular opening (Pantano and Montecucco, 2006), which may also help to explain the low permeability of Ca\(^{2+}\) ions through the channel (Zitzer et al., 1997). The VCC outer vestibule is more open than in α-HL, but repulsion of ions due to charges within the barrel could underlie the lower conductivity of VCC, similar to the 10-fold drop in conductance seen between ScrY and LamB, two glycoporins with nearly identical pore geometries, but different electrostatic profiles (Ranatunga et al., 1999).

The ring of β-trefoil lectin domains sits atop the cytolysin domains and heightens the upper vestibule by 30 Å. An extended loop connects the β-trefoil lectin to the second β-prism lectin in a VCC protomer. This lectin domain interacts with the outer surface of the cytolysin domain, burying approximately 600 Å\(^2\) of accessible surface area. The cytolytic core region of the VCC protomer shares an overall topology with α-HL (RMSD 2.9 Å for 1453 of 2051 Ca residues, 11.9% sequence identity), with important differences discussed below.
Figure 2.1 Structure of the VCC heptamer. A, Ribbon representation of the assembled heptamer. Core cytolysin domain (including rim region), blue; β-trefoil lectin, purple; β-prism lectin, gold; β-barrel stem, green. Side chains of aromatic residues near the putative membrane-solvent interface are shown in red. The approximate outline of the membrane is in gray. B, Top view of the heptamer. C, Surface representation of the heptamer sliced in half along the sevenfold axis and colored by electrostatic potential. Figure generated using APBS and Chimera. D, Outline of the central vestibule/channel of the VCC heptamer generated using HOLE. The sevenfold axis is shown as a yellow bar. E, Graph showing the inner pore diameter along the sevenfold symmetry axis for VCC (purple solid line) and α-HL (dotted black line).
Figure 2.2 Comparison of the VCC protoxin structure and a protomer from the VCC heptamer. A, The VCC water-soluble monomer structure with bound glucoside (PDB ID code 1XEZ). Domains are colored as in Figure 1 with the prodomain in red and W318 shown as green spheres. B, In the assembled form, the stem domain is completely unfurled and the β-prism lectin domain moves to the opposite side of the cytolysin domain. The cradle loop has rearranged, contacting the neighboring protomer. The sugar headgroup seen in A is modeled into the β-prism lectin-binding site. C, Schematic of the putative backbone hydrogen-bonding pattern in the prestem. Hydrogen bonds (using a 3.2-Å cutoff) are shown as black dashed lines. D, The shifted hydrogen-bonding pattern of the assembled stem loop. Amino-acid side chains facing the membrane are marked with black dots, and the aromatic residues near the membrane/solvent interface are marked with gold asterisks.
Figure 2.3 **Key residues at the interprotomer interface (two protomers shown).**

A, The cradle loop (yellow) knits together multiple domains through hydrogen bonding (dashed lines) and van der Waals (dotted surface) interactions. B, An insertion containing three consecutive aspartate residues not present in staph toxins (red) participates in multiple salt-bridge interactions. Additional putative hydrogen bonds involving R330 form links between the protomers.
Structural rearrangements during assembly

Superimposing the cytolytic domains (residues 136–278 and 325–459) of the protoxin monomer and a protomer from the VCC oligomer reveals five major rearrangements that occur along the pathway between the protoxin and assembled states. Firstly, the amino-terminal prodomain (amino acids 1–105) is absent from the heptameric assembly after liberation by proteolytic cleavage. Secondly, the β-prism lectin domain swings around the cytolytic domain to a location 180° opposite its starting place. This new position on the exterior of the ring (forming the “spikes” seen in the low-resolution EM structure) partially overlaps the previous location of the prodomain in the water-soluble protoxin and involves a different surface of the β-prism domain than utilized in interactions with the prestem. Our water-soluble monomer structure exhibits electron density for a bound glucoside moiety consistent with reports that this domain interacts with carbohydrate receptors on cell membranes. It appears that the β-prism rearrangement could occur while still bound to a carbohydrate receptor, with the final location of the site facing downward toward the cell membrane. The third transition involves a 35° rigid-body rotation of the β-trefoil lectin domain around the loop connecting it to the cytolytic domain. The short helical turn that precedes the β-trefoil linker is anchored within the cytolysin domain by a phenylalanine residue (F455) that is necessary for oligomerization in the related Vibrio vulnificus hemolysin (Kashimoto, 2009). The fourth transition is a movement of the loop that cradles the prestem in the water-soluble structure (residues 191–203) through hydrophobic side-chain interactions (notably L192, Y194, L307,
and A309) and backbone hydrogen bonds involving G291. Superposition of the water-soluble monomer on top of the heptameric structure indicates this loop would sterically bump with each neighboring protomer if a rearrangement did not occur.

Reordering of the cradle loop may destabilize the interactions holding the stem in the water-soluble position and initiate unfolding of the stem loop, which constitutes the fifth major transition. Rearrangement of the prestem loop from the water-soluble to assembled state requires a significant breaking and reforming of numerous polar and nonpolar interactions. The prestem in the water-soluble monomer consists of a 10-residue-long antiparallel β-sheet with an intervening 18-residue loop with \(3_{10}\) helical characteristics. Analysis of hydrogen-bonding patterns identifies approximately 18 bonds between backbone atoms within the prestem that are broken during pore assembly. Upon formation of the β-barrel stem, each of the seven stem loops transforms into 19-residue-long antiparallel β-sheets held together by 20 newly formed backbone hydrogen bonds. Each protomer loop additionally forms 21 main chain hydrogen bonds with each of its two neighboring protomer stem loops. The net gain in hydrogen-bonding interactions explains the irreversibility of oligomer assembly, and the rearrangement of bonds likely constitutes an energy barrier that must be overcome to initiate stem unfolding. This transition represents the rate-limiting step in \(α\)-HL assembly, with a measured \(t_{1/2}\) of 8 min on rabbit erythrocyte membranes (Walker et al., 1995). In VCC, the rate-limiting step for pore formation is highly dependent on the membrane cholesterol content, an additional requirement necessary for the insertion of the stem (Krasilnikov, 2007). Hydrophobic residues that
contact the inner leaflet of the target membrane in the assembled stem are mostly packed against hydrophobic residues on the cytolytic core in the water-soluble monomer, shielding them from water.

Most of the interactions between protomers in the heptamer are localized within the interface between cytolysin domains and stem loops and, with the exception of stem hydrogen bonds, are distinct from interactions that bridge α-HL protomers. The VCC cradle loop, which is absent in α-HL, forms multiple interactions between the cytolytic and lectin domains of neighboring protomers and is in a position to coordinate assembly-related conformational movements between domains. This loop may play a functionally analogous role to the amino latch in the staph toxins, which prevents premature oligomerization of monomers in solution and may form cooperative interactions with the prestem (Jayasinghe et al., 2006). In α-HL, a key histidine residue (H35) forms important interprotomer contacts, and mutations to this residue arrest assembly at the prepore state (Panchal and Bayley, 1995). In VCC, the H35 position is replaced by a unique loop structure containing three consecutive aspartate residues that form salt bridges within and between protomers. Together, interactions between each pair of protomers bury 2,854 Å² of accessible surface area with a total of 19,978 Å² in the entire heptamer.

**VCC membrane interactions**

The functional role of cholesterol in the assembly of VCC and other β-PFTs is still an area of intense investigation and may vary between different toxins. For some proteins such as aerolysin and anthrax toxin, cholesterol may serve to cluster
membrane receptors and facilitate productive oligomerization of bound monomers (Abrami and van Der Goot, 1999; Abrami et al., 2003). In contrast, cholesterol is a receptor for the CDC perfringolysin O, which contains a two-residue motif (T490–L491 in YTTL sequence) within a membrane-interacting loop responsible for cholesterol recognition (Farrand et al., 2010). Inspection of the membrane-proximal rim domain of VCC reveals an identical motif (T237–L238 in TTLY sequence) within a loop facing the membrane surface, which could similarly mediate interactions with cholesterol (the L238 side chain is disordered in our maps). A second possible site (A360–L361) is located in a comparable orientation on an adjacent loop. We do not observe any bound lipid or cholesterol moieties in our crystal structure even though the toxin was solubilized from membranes containing 20% cholesterol. It is possible that such interactions are weak, disrupted by detergent, or nonspecific in nature; or that lipids and cholesterol only play a role in earlier stages of assembly. Additionally, the lipid-headgroup binding pocket observed within the LukF (Olson et al., 1999) and α-HL (Galdiero and Gouaux, 2004) rim domains is absent in VCC (neither staph toxin rim contains a Thr–Leu motif). A superposition of the VCC and α-HL β-barrel stem domains indicates that the hydrophobic and presumably membrane buried region of the VCC stem is approximately 3–5 Å longer than the α-HL stem, possibly due to complementarity with thicker membranes, such as found in cholesterol-rich “lipid-raft” subdomains (Lingwood and Simons, 2010). It remains to be seen whether a longer stem is a consequence of the toxin evolving
toward increased stability within thicker regions of the membrane and to what extent cholesterol or lipid-binding motifs are responsible for membrane specificity.

Another distinction between the VCC heptamer and α-HL is the significantly longer loops within the VCC membrane-proximal rim domain. These loops are 10–15 Å longer in VCC, adopt a conformation nearly identical to the VCC water-soluble monomeric state, and would presumably extend much deeper into the membrane bilayer than the loops in α-HL. Aside from two valine residues (V422 and V423) on the tip of the longest loop and the two previously mentioned leucine residues (L238 and L361), the amino-acid composition within the rim domain loops is richly aromatic and contains a balance of tyrosine, tryptophan, and histidine residues: side chains less likely to penetrate deep into the bilayer. We note that our VCC EM structure (He and Olson, 2010) is consistent with the loop conformation seen in the crystal structure, which is not surprising because both were determined from solubilized heptamer in micelles rather than in a lipid bilayer environment. It is possible that these loops adopt a splayed conformation when the heptamer is sitting on the membrane and that detergent solubilization of the complex may have allowed them to relax to their preferred water-soluble conformation. Without significant reorientation, it is also possible that the loops distort the local lipid bilayer structure in a way that might facilitate insertion of the stem domain.

**The VCC pore**

Analysis of alignments of the predicted membrane-spanning loops of β-PFTs reveals common features within the family of β-barrel membrane proteins. Many
channels are lined with charged amino acids and have an approximately 5-residue loop (or turn) containing hydrophobic and/or aromatic residues at the \textit{trans} (opposite the side of entry) end of the pore. A “rivet” model for membrane insertion has been proposed, where this hydrophobic loop at the end of the stem hairpin folds back into the membrane and anchors the channel (Iacovache et al., 2006), a motif also observed in some bacterial outer membrane proteins. The α-HL toxin is an exception, with a neutral pore surface and two charged aspartate residues in the \textit{trans}-loop (with sequence DDTGK). VCC combines features from both groups, having a highly charged lumen and a loop containing a single aspartate residue (with sequence SGDG). Instead of fluting outward on the \textit{trans} side as α-HL does, the tip of the VCC channel backbone curves slightly inward.

Another feature linking VCC with a nonhomologous β-PFT is the W318 aromatic ring positioned at the narrowest aperture of the pore. Within the anthrax PA channel lies a “φ-clamp,” a heptameric ring of phenylalanine residues (F427) that may also outline a pore constriction (Krantz et al., 2005; Katayama et al., 2010). This hydrophobic belt forms a seal around translocating polypeptides and facilitates unidirectional transport through the pore (Krantz et al., 2005). Mutations at position F427 in PA disrupt polypeptide translocation and in some cases also inhibit the prepore to pore transition (Sun et al., 2008). An interprotomer salt bridge lying directly above the φ-clamp (formed by K397 and D426) may help position F427 and neutralize charges near the constriction that interact strongly with a translocating protein (Melnyk and Collier, 2006). Quite remarkably, two charged residues in VCC
reside directly above the W318 ring (K283 and E281) within close enough proximity to form a salt bridge (less than 4 Å), and VCC has an acidic central vestibule similar to anthrax toxin. At this time, we are unaware of any evidence that VCC serves as a translocation channel, although electrophysiological experiments indicate that synthetic peptides can fit within the α-HL pore (Movileanu et al., 2005) and presumably VCC as well. Certainly these intriguing similarities between VCC and anthrax toxin require further inquiry to elucidate whether analogous features exist in other β-PFTs and to determine to what extent they participate in the assembly mechanism.

**Discussion**

Previous studies have placed VCC as an archetypical β-PFT, which was thought only to target and lyse cells by forming transmembrane pores. Several studies have reported VCC’s ability to instigate phagocytosis or apoptosis in different human cell lines.

The VCC heptamer pore structure has revealed several features of this β-PFTs that were previously unknown. The heptamer forms a ring-like structure with dimensions of 140 Å x 135 Å. The seven-fold axis of symmetry runs right through the middle of the pore and the 14-strand β-barrel transmembrane pore is formed around the axis. The vestibules, which were seen by cryo-EM images are formed by the β-prism lectin domain, which moves almost 180° around the cytolysin domain. The most interesting observation was that even though the outer diameter of the barrel is
25-Å wide, the inner pore is constricted to only 8-Å diameter at its narrowest spot. Surprisingly, this constriction is formed by a tryptophan (W318) residue, which is conserved in *Vibrio* spp. Aromatic residues like phenylalanine are generally present inside an peptide translocating channels. In the related α-hemolysin, the narrowest point of the pore has a 10-Å constriction formed by two non-aromatic residues, a methionine, and a lysine. The similar aromatic constriction in the anthrax PA channel was formed by a phenylalanine residue and functional studies have established that this residue is involved in translocating two other toxin molecules across the membrane. In addition to the heptad phenylalanine clamp, a nearby interprotomer salt bridge also plays a critical role in the translocation process. Upon close inspection, we discovered two residues, K283 and E281 that form a salt bridge proximal to the tryptophan clamp. Interestingly, the vestibule region of VCC resembles the acidic vestibule of the anthrax PA channel more than the neutral vestibule of the α-hemolysin. The phenylalanine clamp in the anthrax PA channel is absolutely necessary for the translocation process to occur (Sun et al., 2008). Unfortunately, to the best of our knowledge, no translocating partner has been identified for VCC and mutating the W318 to alanine, arginine, lysine and phenylalanine did not lead to a significant difference in the rabbit erythrocyte lysis activity (data not shown). However, the presence of such remarkable similarities does point towards possible additional functions for VCC. Although purified VCC alone is capable of lysing cells and the W318 residue does not seem to affect the hemolytic activity or the folding of the protein, it is possible that unknown toxic factors use VCC pore as a translocation
device to reach the cytoplasm of the host cells. Further research is necessary to elaborate the function of this highly conserved residue and to identify any translocating partners if they exist.

VCC oligomeric pores were shown to have a substantially lower conductance of Ca\textsuperscript{2+} ions compared to other prototypical PFTs, and in one particular study, VCC displayed 4-fold less conductivity than the α-hemolysin (Menzl et al., 1996). The inner pore diameter of α-hemolysin is ~2 Å greater than that of VCC (Song et al., 1996), but that alone cannot explain the ~75% lower conductance of the channel. Our structure reveals that the extracellular part of the pore contains a large number of positively charged residues and the density of these acidic patches are exceptionally high in the intracellular opening of the pore. These positively charged patches may have repulsive effects towards positively charged Ca\textsuperscript{2+} ions, and this along with the smaller pore diameter, may explain why the conductance of VCC is lower than α-hemolysin.

The large domain rearrangements that occur during the assembly process were expected based on the earlier EM study (He and Olson, 2010). Firstly, the β-prism domain swings around the core of the toxin and forms an interaction where the prodomain was located in the monomer water-soluble structure. This explains why proteolytic removal of the prodomain is necessary for the function of VCC, because if the prodomain is present in this position, further steps of the assembly process would be blocked. Once the β-prism domain moves almost 180\textdegree to the other side of the toxin, the so-called cradle loop becomes free to reorder itself. This loop has been
identified in related toxins like α-hemolysin and also shown to play a key role in unfolding of the membrane penetrating sequence. Similarly, in VCC the cradle loop reorientation causes destabilization of the prestem cytolysin interactions, which in turn initiates the structural rearrangement of the prestem region. This pushes the prestem to partially unfold and insert itself into the membrane. As the prestem region contains a large number of hydrophobic residues, once the destabilization occurs, the prestem spontaneously enters the membrane. Some studies have identified that in CDCs the tip of the β-hairpin loops play a crucial role in the insertion as well as in the anchoring of the pore. It is possible that the β-hairpin tips in VCC also carry out a comparable role in this process. The membrane insertion reaction is highly favorable from a thermodynamic viewpoint. The pre-stem to membrane-inserted stem only form 20-hydrogen bonds between the two β-sheets from the same monomer (compared to 18 in the monomer structure), but also make 21-new hydrogen-bonds with the β-sheet from the neighboring promoter. Additionally, the large amount of hydrophobic interactions with non-polar amino acids with membrane lipids makes the reaction irreversible in nature. Together these hydrogen bonds and the hydrophobic interaction provides the sufficient free energy to the pore to form spontaneously under physiological temperatures. There is some supplementary reordering that also occurs during pore formation, such as a 35° rigid body rotation of the β-trefoil lectin, but these rearrangements are minor and perhaps are not directly involved in the pore formation process.
VCC exhibits a preference for cholesterol-rich membranes for pore formation, but cholesterol is not absolutely necessary for the process. A thorough inspection of the membrane proximal rim region uncovered multiple TL or TL like motifs, which are responsible for cholesterol recognition in CDCs like perfringolysin O (PFO), SLO and PLO (Farrand et al., 2010). Two sites in VCC - T237-L238 and A360-L361 are located in a position close to the membrane where a physical interaction can occur. These motifs are not present in Staph PFTs, which do not have any direct affinity for cholesterol-rich membranes. These motifs are found in various membrane-associated proteins that interact with cholesterol, like the benpyrinoid receptor. VCC contains at least three CRAC or CARC motifs in the membrane-proximal rim region, but the degenerative nature of this motif makes it hard for us to confirm without experimental evidence. It is possible that the interaction between these loops and the membrane components disrupt the regular membrane morphology, further helping the pre-stem region to insert in the membrane. Moreover, the α-hemolysin stem is almost 4 Å shorter than in VCC, which could be due to the greater thickness of the cholesterol-rich membranes or lipid rafts.

Our study on the VCC heptamer structure provides detailed insights into the mechanism of pore formation and the characteristics of the channel itself. Many questions remain, such as the exact molecular mechanism of pore formation, how monomers bind and target the membrane, the role of individual domains in the process, how the pre-stem region inserts into the membrane, and if the insertion process is individual or sequential. Further experimentation will be necessary to
completely understand the mechanism of this PFT in the pathogenesis of cholera, but the results of this study, including the presence of putative cholesterol binding sites, the aromatic ring inside the pore, and the large domains rearrangements, bring us closer to this understanding.
CHAPTER 3
MEMBRANE BINDING OF VCC
**Introduction**

*Vibrio cholerae* cytolysin (VCC) is an important accessory toxin produced by *Vibrio cholerae*. The role of this toxin in the virulence and pathogenicity of cholera has been demonstrated using strains lacking classical cholera toxin (CT) (Honda and Finkelstein, 1979; Ichinose et al., 1987; Coelho et al., 2000; Mitra et al., 2000; Saka et al., 2008). VCC is a PFT that is not only cytolytic, but also cytotoxic to human cell lines (De and Olson, 2011). Both the cytolytic and cytotoxic effects of VCC are attributed to its transmembrane pore formation in the target membrane. In the cholera infection, VCC promotes intestinal colonization of the bacteria by attacking host immune cells like neutrophils as a defense measure (Olivier et al., 2007; Queen and Stachell, 2012). Additionally, lysing host cells may result in release of essential nutrient molecules further aiding bacterial growth (Geny and Popoff, 2006; Geny and Popoff, 2006; Los et al., 2013). As accessory toxins play vital roles in the pathogenesis of cholera and act as important virulence factors involved in the intestinal colonization process, it is imperative that we understand the mechanism of action of these toxins.

VCC belongs to a prototypical family of pore forming toxins that is characterized by the presence of β-barrel transmembrane regions (Heuck et al., 2001) and includes other PFTs like *Staphylococcus aureus* α-hemolysin. Crystal structures are available for many of these PFTs in both water soluble and membrane-inserted forms, including VCC, *Staphylococcus aureus* α-hemolysin, *Staphylococcus aureus* γ-hemolysin, and *Clostridium perfringens* necrotic enteritis B-like toxin (NetB).
Apart from these ones, *S. aureus* LukF-PV, LukF, leucotoxin S are the other PFTs with known crystal structures. (Olson and Gouaux, 2005; De and Olson, 2011; Song et al., 2006; Olson et al., 1999; Pedelacq et al., 1999; Guillet et al., 2004; Yamashita et al., 2011; Foletti et al., 2013; Savva et al., 2013). Overall, these toxins are secreted as a water-soluble monomers and after targeting a membrane, form a membrane-spanning channel that leads to toxicity. Widely accepted models indicate that individual monomers bind to the membrane independently of each other and assemble into the final oligomeric form through a non-lytic prepore stage (Walker et al., 1995; Lohner et al., 2009). However, despite the similarities between them, PFTs in this family vary significantly, particularly in the way they interact with the membrane to target specific cell surfaces.

PFTs recognize and bind to the host cell membrane using various membrane-associated receptors that include protein receptors, cell surface glycans and membrane lipids and cholesterol as receptors. For instance, Staph α-hemolysin utilizes a membrane-bound metalloprotease ADAM-10 as a receptor and thereby targets specific cell membranes (DuMont and Torres, 2014), and intermedialysin (a member of the cholesterol-dependent cytolysin (CDC) family) uses CD59 as its membrane receptor to recognize membranes (Wickham et al., 2011). Additionally, Staph α-hemolysin has been predicted to bind lipid headgroups in the membrane. CDCs, on the other hand, are known for their absolute requirement of cholesterol for membrane binding as well as pore formation. Moreover, several toxins like the CT exploit the cell surface carbohydrate molecules as receptors (*Fukuta et al., 1988*).
To the best of our knowledge, VCC does not have an identified protein-binding partner on the cell surface. In addition, unlike Staph α-hemolysin, a conserved lipid binding pocket is also absent in VCC. Interestingly, VCC contains two lectin domains, a β-trefoil lectin domain and a β-prism lectin domain, both of which can potentially interact with cell surface carbohydrate molecules. Detailed investigation of these lectin domains have revealed that the β-trefoil lectin domain is inactive in terms of sugar binding, while the β-prism lectin domain shows specificity for N-linked heptasaccharide cores on the cell surface (Levan et al., 2012). Interestingly, N-linked glycans are present in almost all eukaryotic cells and perhaps act as receptors for binding to cell surfaces.

Multiple investigations have provided evidence that various membrane components like cholesterol, sphingomyelin, and ceramide drastically enhance the activity of VCC (Zitzer et al., 1999; Zitzer et al., 2000; Zitzer et al., 2001; Harris et al., 2002; Ikagi et al., 2006; Krasilinikov et al., 2007). Moreover, VCC is able to form fully mature pores on pure cholesterol bilayer, but fails to show any preference for epicholesterol, an enantiomer of cholesterol (Harris et al., 2002; Zitzer at al., 2003). These observations indicate that the preference for cholesterol is not through some indirect membrane fluidity issue, but rather through direct physical interactions with the cholesterol or cholesterol-rich membranes. However, the mechanism by which VCC interacts with the cholesterol-rich membranes is not well understood. Closer inspection of the VCC heptameric structure revealed the presence of three rim-region loops that are located in positions potentially close to the membrane. Thus, these
loops may likely contain binding sites for cholesterol or other membrane components. The amino acid sequences of these loops contain several polar amino acids, like glutamines and histidines, and the presence of these residues may prevent these loops from being able to completely penetrate the membrane. However, these residues may cause an alteration in the lipid bilayer or reordering of the local structure of the loops.

In toxins like perfringolysin O (PFO) and *Bacillus thuringiensis* Cry4a, mutational approaches were employed to understand the mechanism by which these toxins interact with membranes (Farrand et al., 2010; Howlader et al., 2010). Similarly, in this study, we conducted a thorough and systematic alanine-scanning mutagenesis of the membrane-proximal rim domain loops to examine the possibility that these loops are involved in the membrane interaction. A detailed examination of the VCC sequence reveals the presence of multiple potential cholesterol-interacting motifs or motif-like sites in the rim region loops. The TL (threonine-leucine) motif, which was identified in PFO as responsible for cholesterol recognition and binding, is present in these three loops at least once and is in a comparable position in PFO (Farrand et al., 2010). A second degenerate TL motif was also located in the rim region, containing the sequence AL (alanine-leucine). Another type of cholesterol interacting motif, which is present in VCC rim region is the CRAC (-L/V-(X)1-5-Y-(X)1-5-R/K-) or inverse CARC (-K/R-(X)1-5-Y/F-(X)1-5-L/V-) type motif, found in benzodiazepine receptors (Li and Papadopoulos, 1998; Epand, 2006; Epand, 2010). Other types of cholesterol binding motifs like SSD, STAR domains, and sequences found in CDCs are absent in VCC.
In this study, we identified ten rim-domain residues, which when mutated to alanine resulted in >90% loss of hemolytic activity against rabbit red blood cells. We confirmed that this loss of activity is attributed to a true loss of function and is not caused by aggregation or global misfolding of the protein. We explore the relationship between the rim region and the β-prism lectin domain and confirmed that these interactions are independent of each other and additive in nature. Additionally, we establish that the reduction in activity is not unique to rabbit blood cells, since comparable differences were observed utilizing a liposome-based system lacking protein or glycan receptors in the lipid bilayer. We also used oligomerization assays and surface plasmon resonance (SPR) experiments to uncover the effect of these mutations on the interaction with cholesterol-rich membranes. In short, this report provides comprehensive functional insights into how VCC utilizes membrane and glycan interactions to target and bind cell membranes.

**Materials and methods**

**Construction of VCC mutants**

Site-directed mutations were introduced into the full-length VCC gene (hlyA) from *V. cholera* O1 El Tor strain 8731, previously cloned into the pHis-Parallel2 vector (Sheffield et al., 1999) using an overlapping primer protocol (Braman et al., 1996). Briefly, complementary 5’ and 3’ primers 39 to 45 bases long were designed with the desired mutation introduced at the center of the primer. The entire vector was
amplified with these primers in a PCR reaction using *PfuUltra* DNA polymerase (Agilent) and the parental strands removed by digestion with DpnI. DNA was transformed into competent NEB5α cells (New England Biolabs, Inc.) and individual colonies grown, miniprepped, and sequenced to confirm the introduction of the mutation.

**Protein expression and purification**

VCC constructs (WT and mutants) were transformed into SHuffle Express T7 (NEB) competent cells and cultured overnight at 37 °C in 20 ml Luria Broth (LB) supplemented with 100 μg/ml ampicillin (Amp). The overnight culture was added to 1 L LB/Amp and incubated with constant shaking at 37 °C until the OD$_{600}$ reached approximately 0.6. At this point, 1 mM isopropylthio-β-galactoside (IPTG) was added to induce VCC expression and the culture incubated with shaking for another 3 hours at 30 °C. Following expression, cells were pelleted in a swinging bucket centrifuge at 3,900 x g for 30 minutes and the cell pellet resuspended in 10 ml TBS buffer (20 mM TRIS pH 7.5, 150 mM NaCl) supplemented with 1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), and 10 mM imidazole. Following passage 3 times through an EmulsiFlex C5 High Pressure Homogenizer (Avestin, Inc.), the lysate was spun at 40,000 x g for 30 minutes at 4 °C in a Lynx 6000 centrifuge (F20-12x50 LEX rotor, Thermo Scientific). Following centrifugation, the supernatant was collected and loaded onto a Ni-chelating column and washed with TBS and TBS+40 mM imidazole. Purified VCC was eluted with TBS containing 250 mM imidazole. The protein was further purified over an S200 10/300 size exclusion
column (GE Healthcare) in TBS buffer and the purity estimated by running on an SDS-PAGE gel. Mutants that failed to produce a monodisperse elution profile on the size exclusion column were excluded from further analysis.

Isolation of primary human polymorphonuclear leukocytes and culture of primary human immune cells lines

All human immune cell lines used in this study (THP-1, Jurkat-R5, HUT) were maintained in Roswell Park Memorial Institute (RPMI) media supplemented with 10% fetal bovine serum (FBS) at 37° C with 5% CO₂. Human blood samples were obtained from normal healthy donors from the New York Blood Center the morning of each experiment. The New York Blood Center obtains written informed consent from all participants and all identifiers are stripped from material prior to shipment. Polymorphonuclear leukocytes (PMNs) were isolated from the peripheral blood of two independent donors according to previously described methods (DuMont et al., 2013). Briefly, packed leukocytes were resuspended in Hank’s Balanced Salt Solution and overlayed onto Ficoll-Paque PLUS (GE Healthcare) followed by centrifugation at 2,000 RPM for 30 minutes at room temperature. After centrifugation the peripheral blood mononuclear cell layer was removed as well as any remaining Ficoll. The lower fraction containing red blood cells and PMNs was resuspended in sterile endotoxin-free 0.9% NaCl + 3% dextran and placed at 25° C for 25 minutes to allow for sedimentation of red blood cells. The upper layer, containing PMNs was removed and residual contaminating red blood cells were lysed in 9 ml of sterile ACK lysing solution (Life Technologies) for 2-3 minutes. Cells were resuspended in RPMI +
10% FBS, filtered through a sterile 70 μm filter, and used immediately in VCC cytotoxicity assays.

**VCC cytotoxicity assays**

Human immune cell lines (THP-1 – monocyte, Jurkat-R5 - T cell, and HUT – T cell) or primary human PMNs were seeded into a 96-well plate at 100,000 cells per well in 90 μl of RPMI + 10% FBS. A 2-fold serial dilution of VCC in phosphate buffered saline (PBS) was added to cells for 1 hour at 37° C and viability was assessed after addition of 10 μl CellTiter colorimetric viability indicator (Promega) followed by incubation at 37° C according to the manufacturer’s instructions. All measurements were made on an EnVision 2103 Plate reader (Perkin-Elmer).

**Hemolysis assays**

Purified VCC was activated by proteolytic cleavage with a 1:350 (wt/wt) ratio of α-chymotrypsin for 30 minutes at room temperature. Mutants that failed to show a clean proteolysis profile were excluded from further analysis. A serial dilution of activated VCC was added to individual wells of a 96-well clear bottom plate containing a ~0.5% solution (normalized to 1 OD<sub>595</sub>) of defibrinated rabbit blood in blood dilution buffer (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl, 1 mg/ml BSA, pH 7.4). Absorbance data were collected at room temperature in an iMark microplate reader (Bio-Rad Laboratories, Inc.) at 595 nm every 15 seconds for 60 minutes with 1 second of shaking between measurements. A measure of percent lysis after 30 minutes was calculated from the data at the two time points using the following
equation: \( \frac{A_0 - A_{30}}{A_0} \times 100 = \% \text{ lysis (raw)} \), where \( A_0 \) and \( A_{30} \) are the absorbances at time = 0 min and time = 30 min, respectively at 595 nm. The raw percent hemolysis value at time \( i \) was normalized to a maximum value of 100\% using this equation:

\[ \frac{100}{\text{max \% lysis}} \times \text{lysis}_i = \% \text{ lysis}. \]

The average percent lysis at each concentration of VCC was plotted and fit to a sigmoidal function for calculating HD_{50} values (concentration of VCC required to achieve 50\% hemolysis after 30 minutes) using KaleidaGraph v. 4.1.3 (Synergy software). Data points were collected in triplicate and plotted as the mean with error bars denoting the SEM.

**Isothermal titration calorimetry (ITC)**

Purified VCC was dialyzed for 24 hours against a buffer containing 20 mM TRIS pH 7.5 and 150 mM NaCl using SnakeSkin Dialysis Tubing (Pierce) with a 10 kDa molecular weight cutoff. VCC sugar-binding activity was assayed by measuring binding to asialofetuin, a glycoprotein previously shown to bind to VCC with approximately 1 \( \mu \)M affinity (Saha and Banerjee, 1997). A 120 \( \mu \)M asialofetuin solution was made by dissolving lyophilized asialofetuin (Sigma) into the dialysate buffer. ITC data were collected using a MicroCal VP-ITC (Microcal, LLC). The sample chamber (1.44 ml) was loaded with dialyzed VCC (23 \( \mu \)M) and titrated with 30 injections of 9.3 \( \mu \)l asialofetuin (110 \( \mu \)M). Blank-subtracted ITC data were processed using Nitpic v. 1.1.2 (Keller et al., 2012), fit to a two-site binding model using SEDPHAT v. 12.1b (Houtman et al., 2007), and displayed with GUSSI v. 1.0.8. (http://biophysics.swmed.edu/MBR/software.html)
Analytical ultracentrifugation

Purified VCC\textsubscript{WT} protein and mutants were concentrated to 1 mg/ml and dialyzed against TBS buffer (20 mM TRIS pH 7.5 and 150 mM NaCl). Samples were loaded into double sector synthetic boundary cells equipped with sapphire windows allowing matching of the sample and reference menisci. Sedimentation velocity analysis was conducted using interference optics with a Beckman-Coulter XL-I analytical ultracentrifuge. The rotor was equilibrated under vacuum at 20 °C and after a period of ~1 hour the rotor was accelerated to 40,000 RPM. Interference scans were acquired at 60-second intervals for approximately 7 hours. Physical constants were calculated for the protein (molecular weight and partial specific volume) and buffer (density and viscosity) using the program Sednterp (Laue et al., 1992). Ultracentrifugation data were fit using the g(s*) method as implemented in the DCDT+ software package (v. 2.4.0) by John Philo (Stafford, 1992; Philo, 2006).

Liposome preparation

To prepare liposomes, 50 mg of dried lipid was dissolved in 2-methyl-2-propanol at 45 °C for soy asolectin (Sigma Aldrich) or 55 °C for total porcine brain extract lipids (Avanti Polar Lipids) in a 50 ml round bottom flask. Cholesterol (10 mg/ml in ethanol) was added at a 30% molar ratio and the lipids dried under a nitrogen flow. The resulting film was vacuum dried overnight in the dark and resuspended in 10 ml TBS buffer (20 mM Tris, 150mM NaCl, pH 7.5) to a final concentration of 5 mg/ml. Following resuspension, the mixtures were subjected to 20 freeze-thaw cycles in
alternating liquid nitrogen and 45 °C/55 °C water baths. The resulting unilamellar liposome solutions were passed 11 times through a 100 nm membrane using a mini-extruder (Avanti Polar Lipids).

**SDS-PAGE gel-based oligomerization assays**

Toxins were activated by α-chymotrypsin proteolysis (as described previously) and incubated with liposome solutions (1:1 lipid:protein by weight) at room temperature for 1 hour. Samples were run on a 5% polyacrylamide gel and stained with coomassie dye. After destaining, gels were scanned using a Typhoon TRIO scanner (GE Healthcare Life Sciences) and images quantified by densitometric analysis using the ImageQuant TL software package (version 7.0). For competition gels, all bands were normalized to the VCC_{WT} band after subtracting the background. Similarly, for assembly kinetics gels, bands were normalized using the overnight VCC_{WT} assembly band after background correction. All data were plotted using KaleidaGraph 4.1.3.

**Surface plasmon resonance**

All SPR data were collected on a Biacore T200 system using a L1 lipophilic chip (GE Healthcare Life Sciences). The chip surface was pre-conditioned with 10 μl of 20 mM CHAPS (solubilized in HBS-N buffer containing 10 mM HEPES, pH 7.4 and 150 mM NaCl) followed by 10 μl of 50 mM NaOH. Asolectin and brain lipid liposomes were diluted to 0.5 mM stocks and injected at a flow rate of 10 μl/min for 15 minutes. Typical loading densities were 3,000 and 7,000 RUs for asolectin and
brain lipid liposomes, respectively. The surface was washed with 50 mM NaOH for one minute and blocked with 0.1 mg/ml BSA for one minute. Liposome surfaces were stable for the duration of the experiment using this procedure. VCC\textsubscript{WT} and mutant proteins were dialyzed into HBS-N buffer and cut with α-chymotrypsin before analysis. To test toxin activity, eight consecutive 5-minute injections (50 μg/ml) were flowed over the chip with 2.5-minute wash phases in between. To regenerate the chip, three-minute injections of 20 mM CHAPS were followed by one-minute injections of 50mM NaOH until no further loss in signal was detected.

**Results**

*Activity of VCC against primary and derived cell lines*

Previous studies have demonstrated that VCC is active against a wide range of cell types including erythrocytes (Honda et al., 1979), intestinal cells (Zitzer et al., 1995), neutrophils (Valeva et al., 2008), and other nucleated cell lines (Zitzer et al., 1995). Related toxins from *S. aureus* display tropism due to the presence of specific receptors on different cell types (DuMont and Torres, 2014). In order to assess whether VCC shows similar differences in activity across different human cell lines, we tested purified VCC against several human primary and derived cell lines including HUT and Jurkat human T cell lines, the THP-1 human monocyte-like cell line, and primary human neutrophils. Wild-type VCC (VCC\textsubscript{WT}) was titrated in a dose-response experiment (15 pM to 120 nM) against roughly 100,000 cells and toxin-
dependent killing was measured after 1.5 hours. Cell killing was confirmed by measuring ethidium bromide uptake into treated cells, which began at approximately 30 minutes and plateaued at around 60 minutes. All four cells types displayed a similar susceptibility within a narrow window, with the half-point of toxin activity ranging between 200 and 800 pM. These results suggest that VCC does not display selectivity for any one of these four immune cell types.

It is known that VCC prefers rabbit erythrocytes over human, sheep, and chicken (Saha and Banerjee, 1997), perhaps due to the overabundance of complex type N-glycans on the cell surface (Sutton-Smith et al., 2007). To ascertain whether the efficacy of VCC towards rabbit erythrocytes is comparable to the human immune cell types tested, we tested VCC against a similar titer of washed rabbit erythrocytes (~100,000 cells) in a microtiter-plate lysis assay using optical density at 595 nM as a measure of cell lysis. The concentration of half lysis (HD$_{50}$) by wild-type activated VCC was approximately 100 pM. This number is not largely different than the efficacy of VCC towards the human immune cells we tested, as would be expected if a receptor were not involved. These results suggest that rabbit erythrocytes are a suitable model for understanding VCC/membrane interactions and that VCC does not exhibit tropic behavior indicative of a selective receptor on neutrophils or other human-cell lines tested.
Figure 3.1 **VCC/Membrane interactions.** A, Schematic representation of steps in VCC pore-formation. Water-soluble monomers bind to cell membranes (1) utilizing both complex N-glycan binding by the β-prism lectin domain (yellow) and direct membrane-interactions with the rim domain (blue). Monomers diffuse across the cell membrane (2) forming pre-pore oligomeric intermediates before unfolding amphipathic loops (3) that form the transmembrane β-barrel channel pore. B, Cartoon representation of the fully assembled heptameric VCC pore structure determined by X-ray crystallography. The transmembrane β-barrel (stem) is surrounded by putative membrane-contacting loops in the rim (blue). A ring of aromatic amino acids denotes the predicted membrane-interface shown in stick representation (orange). A second inactive lectin with a β-trefoil fold is shown in magenta. C, Close-up of the three loops that comprise the membrane-contacting rim domain. D, Sequence of the loops subjected to alanine-scanning mutagenesis. Residues in green were already alanine or glycine and not mutated while red residues failed quality control specifications and were not tested further. Black underlined residues passed quality control and were included.
Figure 3.2 Cytotoxicity and hemolysis data. A, VCC$_{WT}$ exhibits similar levels of potency against four human cell-types. Error bars indicate the standard error of the mean (SEM) calculated from three repetitions (six for neutrophils representing two donors). B, VCC$_{WT}$ is equally effective against rabbit blood. Hemolysis curves for VCC$_{WT}$ (red) and several rim-domain mutants are displayed with overlaid sigmoidal fits. Error bars represent the SEM from three repetitions.
Construction and quality control of VCC rim-domain alanine-scanning mutations

Based on the oligomeric structures of VCC and related pore-forming toxins, we expected significant direct steric interactions between the rim-domain loops and the outer leaflet of the plasma membrane. We employed a complete systematic alanine-scanning mutagenesis of the three primary rim-domain loops predicted to be membrane proximal in order to uncover specific interactions necessary for lytic activity of the VCC toxin. If a receptor-binding site exists within the rim domain, or if non-specific interactions are necessary for monomer binding to the membrane, we might uncover these hot spots through a mutagenesis strategy. In all, 30 individual mutants were made by site-directed mutagenesis encompassing residues 234-243, 356-365, and 417-430.

The transition of VCC from water-soluble monomer to oligomeric transmembrane channel involves several steps, including glycan-receptor binding, rim-domain interactions with the membrane, oligomerization into a pre-pore, and channel insertion. In order to guard against unanticipated factors that might affect other steps in the process, we employed a four-tier quality control procedure to filter out mutants that displayed aberrant behavior. First, mutants were purified over a size-exclusion column and only passed if they displayed a monodisperse elution profile (to address possible solution aggregation). Second, only mutants that could be cleanly cleaved by α-chymotrypsin (necessary to activate the pro-toxin by removing the prodomain) were included with the assumption that misfolded or misbehaved protein would be more susceptible to spurious cutting by the protease. Third, a subset of
mutants with varying losses in activity was subjected to analytical ultracentrifugation to confirm their monomeric state in solution. And fourth, the glycan-binding activity of mutants with varying losses in activity was measured against a model glycoprotein (asialofetuin) using isothermal titration calorimetry (ITC). In all, out of 30 mutations, 25 expressed and purified mutants passed our specifications and were included in our analysis.

Figure 3.3 Analytical ultracentrifugation of selected VCC mutants. VCC\textsubscript{WT} and seven representative mutants were subjected to sedimentation velocity analytical ultracentrifugation to determine if introduced mutations have any effect on the oligomerization state of the protein. All proteins displayed a similar profile when analyzed using a g(s*) analysis suggesting that the mutations did not lead to any observable change in the hydrodynamic properties of the protein.
Hemolysis assays of VCC rim-domain alanine-scanning mutants

To assess the effects of rim-domain mutants on toxin-mediated cell lysis, purified and activated VCC\textsubscript{WT} and mutant proteins were added to washed rabbit blood in microtiter plates and cell lysis was monitored spectrophotometrically. The degree of cell-lysis for VCC\textsubscript{WT} and mutants was plotted against toxin concentration after 30 minutes at room temperature. The concentration resulting in 50% lysis (HD\textsubscript{50}) was calculated based on a sigmoidal fit of the dose-response data and compared to the HD\textsubscript{50} of VCC\textsubscript{WT}. For all mutations, the shape and slope of the hemolysis curve was not significantly affected, merely the position of the half-point of lysis.

Of the impaired mutations, ten displayed a greater than 10-fold loss in hemolytic activity (loss of >90.0% activity) compared to VCC\textsubscript{WT} (L238A, F242A, L361A, W362A, H419A, Y420A, Y421A, V423A, H426A, and Q427A). Two neighboring mutants, Y420A and Y421A, exhibited the largest loss in activity with 225-fold (loss of 99.5% activity) and 684-fold (loss of 99.9% activity) losses, respectively. The six least active mutants involve the removal of hydrophobic and aromatic side chains that form a band around the lower tip of the rim-domain. The most severe mutations are spread across the three rim-domain loops with two coming from loop 1 (L238A and F242A), one from loop 2 (W362), and three from loop 3 (Y420A, Y421A, H426A). A second histidine mutation (H419A from loop 3) had a more modest effect, with a roughly 10-fold loss in hemolytic activity. The six aforementioned mutations represent most of the aromatic residues below the stem F/Y aromatic ring, which likely demarcates the lipid-water membrane interface. One
additional tyrosine mutation (Y235A) near the membrane interface was relatively unaffected in activity (~4-fold) and another tyrosine mutation (Y417A) did not pass quality control. Three hydrophobic residues below the F/Y line also had modest effects on hemolytic activity, including L361A (37-fold loss) and V423A (10-fold loss). Lastly, one polar residue located at the predicted membrane interface (Q427A) resulted in a 19-fold loss, although this buried residue contributes no exposed surface area on the rim-domain based on VCC crystal structures.

**Additional mutagenesis of rim domain residues**

Several sequence-specific motifs have been identified in membrane-active toxins that facilitate membrane interactions via cholesterol binding. We performed additional targeted mutagenesis beyond our initial scan to investigate these motifs in VCC and used the same quality control procedure to ensure that mutations did not have unintended effects. The first we considered is the threonine-leucine (TL) sequence first identified in perfringolysin O (PFO). VCC has a single TL motif within loop 1 (T237-L238) and a second similar AL motif within loop 2 (A360-L361). Mutation of L238A and L361A led to a 70-fold and 37-fold loss in activity, respectively. This degree of this loss is not as high as in PFO, where mutation of the leucine to alanine led to a 100-fold loss in activity (Farrand et al., 2010). Mutating both leucine residues in VCC (L238A and L361A) at the same time had a much more dramatic effect, with a total loss of over 1,000-fold activity. This suggests that the two putative motifs may compensate for each other’s loss of function. Mutation of L238 to isoleucine and valine resulted in a 3-fold and 7-fold loss, respectively, while
in the case of PFO, mutation of the leucine to isoleucine and valine led to a loss of 4.4 fold and 5.6-fold, respectively. These data suggest that leucine residues play an important role in membrane interactions, possibly via cholesterol binding. However, the fact that VCC has other residues that when mutated to alanine exhibit an even greater loss in activity (such as Y420 and Y421) demonstrates that a single TL motif alone is not responsible for membrane recognition as is seen in PFO.

A second set of motifs we investigated included the CRAC and inverse CARC sites. Mutagenesis and modeling studies suggest that in these motifs cholesterol stacks against the key central aromatic residue through CH-π stacking interactions (Fantini and Barrantes, 2013). Inspection of the VCC rim-domain indicates a number of overlapping putative CRAC and CARC sequences, although only one sequence contains a central aromatic residue that led to a significant loss in hemolytic activity when mutated to alanine. This putative CARC motif, containing the sequence \(418\text{KHYYVV}^{423}\), included the two residues most sensitive to mutagenesis in the VCC rim-domain. If this sequence were a functional CARC motif, mutation of the central tyrosine to phenylalanine would be tolerated, whereas mutation to other amino acids would not. Because there are two tyrosine residues in the center of the sequence, we mutated each independently to phenylalanine and proline while also creating double phenylalanine and proline mutants. Interestingly, the activity of Y420F actually improved upon mutation, whereas the activity of Y421F decreased 25-fold. The Y420F/Y421F double mutant displayed a slightly more substantial loss than the Y421F mutation alone (>97.6% loss in activity). Further, mutation of either residue to
proline almost completely abolished activity (to less than 0.1% VCC<sub>WT</sub> activity) and mutation of both tyrosines to proline resulted in protein that passed quality control, but had no measurable hemolytic activity. These results are consistent with VCC containing a CARC motif within the rim-domain. CRAC and CARC motifs typically reside on α-helical segments, but the inherent flexibility in the membrane-interacting loops may allow a productive cholesterol-interacting conformation to occur.

**The interplay between sugar-binding mutants and rim-domain mutants**

In addition to direct rim domain interactions, lectin-glycan interactions play an important role in membrane targeting by VCC (Levan et al., 2013; Saha and Banerjee, 1997). To ensure that rim-domain mutants do not have unintended effects on the β-prism domain lectin activity, we measured binding of VCC<sub>WT</sub> and a selection of single-alanine mutants to the glycoprotein asialofetuin by isothermal titration calorimetry (ITC). Previous studies indicate that asialofetuin contains glycans that mimic membrane sugar-receptors and is an indicator of glycan-binding activity (Rai et al., 2013). Our ITC experiments confirm that unaffected (S357A), moderately affected (H419A and H426A), and severely affected (Y420A and Y421A) rim-domain mutations all retain nearly wild-type levels of asialofetuin binding activity. Mutation of a key residue necessary for glycan binding (D617K) or complete removal of the VCC β-prism domain completely abolished measurable binding of asialofetuin (data not shown).

To investigate the extent to which rim-domain mutations and glycan-binding
mutations collectively contribute to hemolytic activity, we created double mutations of selected alanine-scanning mutations on a D617K background. D617K alone leads to a roughly 180-fold loss in hemolytic activity against rabbit blood. Introduction of the F242A and H426A mutations on the D617K background led to a larger loss in activity than any of the single mutations alone. Introduction of L238A, W362A, Y420A, and Y421A (all on the D617K background) led to mutant protein with little or no measurable lytic activity up to a concentration of 5 μM. These results demonstrate that glycan-binding and rim-domain interactions are independent and additive, each contributing to toxin activity. Mutations within the rim domain are made worse with the removal of glycan-interactions to the point where four of the double mutants tested are essentially dead.

*The effect of rim-domain mutations on activity against liposomes*

To further explore the mechanism by which rim-domain mutations affect hemolytic activity, we moved into a liposome system in order to control individual membrane components. Previous studies show that VCC maintains a fairly high activity of pore formation in liposomes made of asolectin lipids (a crude extract of soy lipids) supplemented with a 30% molar ratio of cholesterol (Zitzer et al., 2000). Asolectin liposomes are not expected to contain complex N-glycans targeted by VCC or specific protein receptors, thereby simplifying interpretation of oligomerization results. If the rim-domain mutants affect binding to protein receptors or glycan receptors, then we should see no difference in the pore-forming activity of VCCWT.
Figure 3.4 ITC experiments measuring VCC binding to asialofetuin. Mutations introduced into the rim domain do not negatively affect the sugar-binding ability of the VCC β-prism domain. The glycoprotein asialofetuin was used as a substrate for assaying the activity of the VCC lectin. Data fit best when 2 glycans per asialofetuin were included in the fit, although some glycosylation heterogeneity is expected in the protein. Five representative mutants were tested, including Y421A, which exhibited the greatest loss of activity in hemolysis assays. Representative data from one of three experiments are shown. Error bars represent the SEM of three repetitions.
Figure 3.5 **Alanine-scanning mutagenesis.** A, Mutagenesis data from 25-alanine mutations are displayed as the ratio of HD$_{50}$ values for each mutant compared to VCC$_{WT}$. The HD$_{50}$ is defined as the concentration of VCC required to elicit 50% lysis of rabbit blood in 30 minutes at 25 °C. Ten mutants exhibited a greater than 10-fold decrease in lytic activity, two mutants were near 100-fold (F242A, W362A) and two mutants (Y420A, Y421A) lost more than 100-fold activity. Loop 3 was the most susceptible overall to activity loss upon mutagenesis. Numerical data are presented in Table 1. B, Surface representation of the VCC rim domain with the mutagenesis boundary outlined with a thick black line. Residues are color-coded by loss in activity; green for less than 10-fold, yellow for 10 to 40-fold, and red for greater than 40-fold. Residues that failed to meet quality control specifications are shown in white. The Q427A mutant is completely buried in the structure and therefore does not contribute to the protein surface.
and mutant proteins on soybean liposomes. This system also allows us to directly test the influence of cholesterol on membrane activity.

We tested the pore-forming activity of a subset of rim-domain alanine mutations to assess whether their loss in activity is specific to cell membranes or a more general phenomenon of lipid membrane interactions. Rim-domain mutations that exhibited deficient activity against rabbit blood also displayed reduced pore-forming activity on asolectin liposomes when assessed on SDS-PAGE gels. It is important to note that these assays are conducted at much higher protein concentrations (μM) than cell assays (pM to nM) and so the degree of loss in activity may not be as severe as that measured on cells. Still, the mutants exhibited substantially less pore-forming activity on liposomes with and without added cholesterol, with VCCWT also exhibiting less activity against liposomes without cholesterol. This suggests that rim domain membrane binding is enhanced by cholesterol, although cholesterol is not absolutely necessary for toxin activity at high enough concentrations.

The loss in activity of rim-domain mutants could be caused by a reduced ability of the monomer to productively associate with the cell membrane, but could also be attributed to a disruption in later stages of oligomerization and pore-formation. To address the possibility that mutations are merely affecting the kinetics of pore-formation, we performed a time course assay to look at the magnitude and rate of SDS-resistant oligomer formation. Fully assembled β-barrel channels remain heptameric in sodium dodecyl-sulfate (SDS) as long as the samples are not boiled.
This allows for visualization of the degree of pore-formation on SDS-PAGE gels and quantification of the extent of VCC oligomerization can be accomplished by gel densitometry of assembled species. Focusing on the two most significant mutations (Y420A and Y421A), we see that the magnitude of pore-formation, rather than the rate, is most affected by the mutations. Even after an overnight incubation, the mutants do not reach the amount of VCC\textsubscript{WT} oligomer seen on the gel suggesting that the concentration of monomer on the membrane is not sufficient to promote oligomerization.

If the mutations are affecting the mechanism of oligomerization or pore insertion, we would predict that formation of heterooligomers between mutants and VCC\textsubscript{WT} would result in less channel formation. If the mutations merely caused a decrease in the ability to bind to the cell membrane, the mutants should not interfere with the VCC\textsubscript{WT} protein’s ability to bind and assemble on liposomes. We observed that adding up to one molar equivalent of the mutant had no effect on the ability of VCC\textsubscript{WT} to form SDS-resistant oligomers and did not reduce the activity of VCC\textsubscript{WT} against rabbit blood in competition assays, even when added at a ~100-fold excess of the HD\textsubscript{50} of VCC\textsubscript{WT}. These data suggest that rim-domain mutants primarily affect the ability of VCC to adhere to the lipid-membrane, rather than subsequent steps in toxin oligomerization and channel-insertion.
Figure 3.6 **L/Y mutants and glycan double mutants.** A, Additional rim-domain mutations were made to investigate the possibility of TL, CRAC, or CARC-type cholesterol-binding motifs. The proline double mutant did not result in any measurable cell lysis. Numerical data are presented in Table 3.1. B, Rim-domain mutations were combined with a mutation in the β-prism domain (D617K) that blocks glycan-binding. Most double mutants, with the exception of F242A/D617K and H426/D617K, resulted in a nearly complete loss of hemolytic activity up to a 5 μM concentration. Error bars represent the SEM of three repetitions.
Figure 3.7 **Liposome gel-based assays.** A, SDS-PAGE assembly gel showing VCC\textsubscript{WT} and a panel of additional glycan and rim-domain mutations assembled on asolectin liposomes +/- 30\% cholesterol. Monomer and oligomer bands are designated by black arrows. All toxins are cut by chymotrypsin prior to assembly and samples are not boiled before loading on the gel. B, Time-course of VCC assembly. VCC assembly reactions were stopped at various time points and only the oligomer band is shown. Bands were quantified and plotted normalized to the maximum amount of VCC\textsubscript{WT} oligomer. Error bars denote the SEM of two experiments. C, Competition assay showing that mutant toxins do not interfere with VCC\textsubscript{WT} activity. Numbers in boxes and under the bar graph indicate the molar ratio of VCC\textsubscript{WT} : mutant protein. D, Competition assay performed on rabbit blood. Curves result from a mixture of VCC\textsubscript{WT} protein (concentration on the x-axis) and a 10 nM constant concentration (100-times the VCC\textsubscript{WT} HD\textsubscript{50}) of each designated mutant protein. Mutant proteins did not reduce the potency of VCC\textsubscript{WT} against cells. The \(\beta\)-prism lectin domain from VCC was used as a control to show that cell-surface glycan-sites are not saturated at this concentration.
**Surface plasmon resonance (SPR) studies of membrane binding by VCC\textsubscript{WT} and alanine-scanning mutants**

To further investigate the effects of rim-domain mutations on membrane interactions, we used SPR to look at VCC binding and assembly on liposomes. Liposomes were first captured on an SPR chip containing a lipophilic surface and blocked using BSA. Analysis of SPR sensorgrams allowed us to quantify the relative amount of monomer bound to the membrane during the binding phase and the resulting amount of oligomer assembled on the liposome following the wash step.

To first ensure that asolectin liposomes do not interact with the glycan-binding domain of VCC, we tested the D617K mutant on the chip. We saw a similar amount of both membrane binding and oligomerization of the D617K mutant as VCC\textsubscript{WT} on asolectin liposomes containing 30% cholesterol. A number of the most significant VCC rim-domain mutations, including L238A, F242A, Y420A, and Y421A all exhibited less signal during the binding phase and consequently less oligomer formation measured following the wash phase. Remarkably, the difference between VCC\textsubscript{WT} and the mutants in both membrane association and oligomer formation nearly disappeared when cholesterol was excluded from asolectin liposomes. To see if this effect is specific to soy lipids, we also tested several mutants on liposomes made with total porcine brain lipids, which contain residual cholesterol (20). In brain lipids supplemented with 30% cholesterol, we saw an even more dramatic difference between VCC\textsubscript{WT} and rim-domain mutant toxin (liposome capture was more efficient as well) in both membrane binding and oligomerization. Removal
of cholesterol in this case did result in less efficient binding and oligomerization by VCC\textsubscript{WT}, but not to the same degree as observed with cholesterol-free asolectin lipids. The activity that VCC\textsubscript{WT} retains may be attributable to the residual cholesterol in crude brain lipid preparations, or additional lipid interactions in brain vs. soy lipids that rim-domain mutations disrupt, such as those involving ceramides or sphingolipids.

**Discussion**

VCC exploits and integrates two different types of membrane interactions in membrane binding, ensuring high-affinity binding to the cell surface through additive effects. These two types of interactions happen through β-prism lectin domain, which bind to the cell surface glycans, and through the membrane proximal rim region of the cytolysin domain, which makes direct contact with the lipid bilayer. The glycan interaction via the β-prism lectin domain is particularly strong and responsible for a nM affinity of VCC for the cell surface. However, this nM binding is neither entirely sufficient, nor absolutely necessary, for VCC to interact with the bilayer or to form the channel on the membrane. The β-prism domain-truncated VCC monomers are capable of forming heptameric complexes on the membrane and display reduced hemolytic activity. The second type of interaction happens through the rim region directly with lipid membrane components. Previous studies have identified sphingomyelin, ceramide, and cholesterol as enhancers of VCC’s pore-forming
Figure 3.8 **SPR analysis of rim-domain mutations.** A, Activity of VCC\textsubscript{WT} and rim-domain mutants on asolectin liposomes containing 30\% cholesterol. Scatter plot points indicate the residual number of resonance units (RUs) measured on liposomes after each injection and wash. The total mass of toxin flowed over the chip is designated on the x-axis. RUs on the y-axis are normalized to the total amount of liposomes captured in each experiment. The bar graph (inset) shows the highest RU signal measured during the binding phase of the first injection point in each series. B, Same analysis shown for asolectin liposomes without any added cholesterol. C, and D, show similar data for brain lipid liposomes +/- 30\% added cholesterol.
Table 3.1 **HD$_{50}$ values for all mutants described in this study.** Hemolysis data were fit using a sigmoidal function to determine the amount of protein required to elicit 50% lysis (HD$_{50}$) for wild type and 25 mutants (left three columns). For comparison, the ratio of the mutant vs. wild type HD$_{50}$ is displayed along with the percent wild type activity exhibited by each mutant. All values are significant to $p \leq 0.01$ except for those marked with an asterisk based on a single factor analysis of variance (ANOVA) test (n=3). Hemolysis data for additional mutations involving L238A, Y420A, Y421A, and glycan/rim domain double mutants are shown in the right three columns. Mutants without measurable activity are denoted with NA and mutants with less than 10% wild-type activity are highlighted in bold type.

<table>
<thead>
<tr>
<th>VCC Mutant (alanine scan)</th>
<th>HD$<em>{50}^{\text{Mutant}}$ /HD$</em>{50}^{\text{WT}}$</th>
<th>% Wild Type Activity</th>
<th>VCC Mutant (L238/Y420/Y421)</th>
<th>HD$<em>{50}^{\text{Mutant}}$ /HD$</em>{50}^{\text{WT}}$</th>
<th>% Wild Type Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>1.0 ± 0.1</td>
<td>100.0</td>
<td>L238A+L361A</td>
<td>1152.8 ± 63.9</td>
<td>0.2</td>
</tr>
<tr>
<td>S234A</td>
<td>1.0 ± 0.06</td>
<td>100.0</td>
<td>L238I</td>
<td>2.82 ± 0.3</td>
<td>35.4</td>
</tr>
<tr>
<td>Y235A</td>
<td>4.12 ± 0.8</td>
<td>24.2</td>
<td>L238V</td>
<td>6.9 ± 0.9</td>
<td>14.5</td>
</tr>
<tr>
<td>T236A*</td>
<td>0.8 ± 0.07</td>
<td>120.8</td>
<td>L361I</td>
<td>1.7 ± 0.2</td>
<td>58.7</td>
</tr>
<tr>
<td>L238A</td>
<td>70.3 ± 4.1</td>
<td>1.4</td>
<td>Y420F</td>
<td>0.07 ± 0.02</td>
<td>1366.9</td>
</tr>
<tr>
<td>D239A</td>
<td>6.7 ± 0.7</td>
<td>14.9</td>
<td>Y421F</td>
<td>24.9 ± 2.9</td>
<td>4.0</td>
</tr>
<tr>
<td>F242A</td>
<td>96.9 ± 19.9</td>
<td>1.03</td>
<td>Y420F+Y421F</td>
<td>41.7 ± 3.4</td>
<td>2.4</td>
</tr>
<tr>
<td>R243A*</td>
<td>1.0 ± 0.2</td>
<td>101.5</td>
<td>Y420P</td>
<td>1001.2 ± 132.7</td>
<td>0.1</td>
</tr>
<tr>
<td>R356A*</td>
<td>1.2 ± 0.3</td>
<td>82.1</td>
<td>Y421P</td>
<td>6377.1 ± 2602.5</td>
<td>0.02</td>
</tr>
<tr>
<td>S357A</td>
<td>0.5 ± 0.06</td>
<td>186.4</td>
<td>Y420P+Y421P</td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>T358A</td>
<td>1.7 ± 0.1</td>
<td>57.9</td>
<td>VCC Mutant (glycan/rim double mutant)</td>
<td>HD$<em>{50}^{\text{Mutant}}$ /HD$</em>{50}^{\text{WT}}$</td>
<td>% Wild Type Activity</td>
</tr>
<tr>
<td>D359A</td>
<td>2.4 ± 0.002</td>
<td>41.6</td>
<td>D617K</td>
<td>181.3 ± 19.6</td>
<td>0.6</td>
</tr>
<tr>
<td>L361A</td>
<td>37.2 ± 2.4</td>
<td>1.0</td>
<td>L238AD617K</td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>W362A</td>
<td>99.5 ± 17.9</td>
<td>1.0</td>
<td>F242AD617K</td>
<td>345.3 ± 44.2</td>
<td>0.3</td>
</tr>
<tr>
<td>V363A*</td>
<td>1.1 ± 0.2</td>
<td>90.1</td>
<td>W362AD617K</td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>N364A</td>
<td>4.2 ± 0.6</td>
<td>23.9</td>
<td>Y420AD617K</td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>T365A*</td>
<td>0.8 ± 0.2</td>
<td>120.1</td>
<td>Y421AD617K</td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>H419A</td>
<td>10.2 ± 1.8</td>
<td>9.8</td>
<td>H426AD617K</td>
<td>460.1 ± 76.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Y420A</td>
<td>225.2 ± 25.5</td>
<td>0.4</td>
<td></td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>Y421A</td>
<td>684.0 ± 59.7</td>
<td>0.1</td>
<td></td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>V422A</td>
<td>5.6 ± 1.0</td>
<td>17.8</td>
<td></td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>V423A</td>
<td>10.3 ± 3.1</td>
<td>9.7</td>
<td></td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>H426A</td>
<td>44.4 ± 7.2</td>
<td>2.3</td>
<td></td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>Q427A</td>
<td>18.8 ± 1.0</td>
<td>5.3</td>
<td></td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>S428A</td>
<td>1.8 ± 0.3</td>
<td>55.1</td>
<td></td>
<td>NA</td>
<td>--</td>
</tr>
<tr>
<td>H430A*</td>
<td>1.1 ± 0.1</td>
<td>92.2</td>
<td></td>
<td>NA</td>
<td>--</td>
</tr>
</tbody>
</table>
activity, but VCC can form pores even in the absence of these above-mentioned components, although with considerably less potency. The concentration-dependence of VCC pore formation on these membrane constituents indicates some sort of interaction between the toxin and these molecules.

The rim-domain of VCC can also theoretically interact with membrane-associated proteins as receptors. Protein receptors are not uncommon in pore-forming toxins and a number of them have been identified including ADAM10 for the *S. aureus* α-hemolysin, CCR5/CXCR1/CXCR2/CCR2/C5aR C5L2 for the *S. aureus* leukocidins (HlgACB or LukED), CD59 for the *S. intermedius* intermedilysin, and furin for the *Bacillus anthracis* PA channel (Peraro and van der Goot, 2016). The presence of protein receptors generally cause tropism towards specific cells types that contain that specific receptor. To determine if VCC utilizes a protein receptor, we tested the toxin against multiple human cell lines including neutrophils, which have been speculated to be the target of this PFT. Interestingly, no substantial difference in the cytolytic activity is noted between the neutrophils and other cell lines like THP1, HUT and Jurkat-R5. Moreover, the potency of VCC for killing human neutrophils is comparable to that of rabbit blood cells, indicating the unlikely possibility that there exists a protein receptor for VCC. Perhaps VCC compensates for not having a protein receptor by using the glycan-based high-affinity membrane binding, a mechanism that is missing in most PFTs with protein receptors, like the Staph α-hemolysin, the leukocidins, or the anthrax PA channel (Peraro and van der Goot, 2016). This subtle
difference in the targeting mechanism allows VCC to target and lyse a broader group of cells, as glycans are predominantly present in almost all eukaryotic cells.

Our systematic and comprehensive alanine scanning mutagenesis study revealed, at least ten amino-acid residues, which upon mutation to alanine, caused a drastic loss (>90%) of hemolytic activity. Interestingly, this list of ten residues includes four aromatic residues, three non-aromatic hydrophobic residues, and three polar residues. As not all of these ten residues are hydrophobic in nature, we can safely predict that something other than just a simple loss in lipid-mediated hydrophobic interactions are responsible for this loss of activity. Moreover, these mutations displayed an analogous loss of activity in the liposomal system, where membrane proteins and glycans are absent. This provides additional support towards the notion that VCC does not utilize membrane-associated proteins as receptors. Furthermore, our experiments with double rim/glycan mutants further support the plausibility of glycan interactions in VCC’s activity. The double mutants are significantly less active compared to the rim domain mutants, which is indicative of an independent and additive mechanism. This conclusion directly conflicts with assertions in the literature that the VCC β-prism lectin domain is a regulatory switch and that the β-prism lectin domain participates in rim-membrane interactions. Our results clearly show that this is not the case, and the toxin can form pores even without a functional glycan binding activity (VCC_{truncated β-prism domain}). In addition, in the absence of rim domain interactions (rim mutants), the removal of glycan binding adversely affects the hemolytic activity of VCC. We argue that these two interactions
together make the VCC membrane interaction stronger than the two individual binding events (pointing towards an avidity effect). Impeding any one of the interactions causes a decline in the membrane-bound population of the toxin, thereby making the toxin a less potent pore former.

Close inspection of the data uncovers that many of the mutants with a >90% loss of activity are positioned inside the putative cholesterol binding motifs in VCC. The first motifs that we identified are TL motifs and TL-like motifs (AL sequence). VCC TL or TL-like motifs possess similar molecular characteristics that define these motifs in non-related β-PFTs like CDCs. Mutation of the leucine to alanine triggers a 37-70-fold loss of activity in VCC, which is equivalent to what was seen in PFO TL motifs. However, mutations of the leucine residues to isoleucine and valine led to only a 4-6-fold loss of activity, mirroring what Farrand et al. saw in their system. However, TL motifs do not constitute the entire cholesterol interacting surface of VCC, as there are eight other residues that exhibited a >90% activity loss, and those residues do not reside in TL or TL-like motifs.

The second set of cholesterol-interacting motif that we identified includes CRAC and inverse CARC type motifs. The third rim domain loop of VCC holds a CRAC like sequence $^{418}$KHYYVV$^{423}$, which contains four of the ten most sensitive alanine mutagenesis residues, including the top two mutants (Y420 and Y421). Interestingly, by definition, in the CRAC motif the central tyrosine residues helps stabilize cholesterol binding by stacking with sterol rings makeing up the most vital residues in the motif. Also, in CARC motifs, replacement of the tyrosine by
phenylalanine results in a lower loss of activity when compared to alanine mutations. Changing the tyrosine to proline completely disrupts the planar surface for cholesterol binding, resulting in a total loss of cholesterol interactions. In our experiments, the Y420A and Y421A mutations showed a drastic 200-685-fold loss in activity. However, replacement of the tyrosine residues with phenylalanine was still disruptive, but surprisingly this mutation was tolerated well than the alanine mutations. Strikingly, substitution of the tyrosine residues with proline caused a drastic loss of activity (1,000 and 6,000-fold) and the double-proline mutant was completely deficient of any measurable activity. It is possible that this motif acts as a third cholesterol binding motif and work in concert with other TL and TL-like motifs to recognize and bind cholesterol in the membrane.

There are a couple of studies that reported single mutations in the rim region leading to a loss in activity (L238, W362, and Y421) (Rai and Chattopadhyay, 2015). They also stated that some of the single mutants are unable to bind to the membrane in addition to forming abortive oligomers on the membrane. We identified a total of ten mutations, including these three, and confirmed using a liposomal assay that all of can form pores in the membrane. Additionally, we also did not detect any abortive oligomerization during our SPR experiments. Instead, we observed that the degree of channel formation is directly proportional to the amount of protein that binds to the membrane. Based on these results, we argue that these mutations affect the membrane-binding ability of the toxin and a decline in the monomer concentration on the bilayer is the major cause of the overall loss of activity. This argument is further
supported by our hemolysis and liposomal oligomerization assays, where the addition of a higher amount of toxin compensates for the loss in pore formation. These results also allow us to propose that the mutants are flawed mainly in the membrane-binding phase and that a similar amount of pore formation can be achieved by increasing the number of monomers bound to the membrane.

In summary, the VCC rim domain plays an essential role in membrane binding events along with the β-prism lectin domain. Together these independent interactions add up, perhaps through avidity effects, to a total pM HD$_{50}$ value. Through our systematic investigation, we identified the membrane-proximal rim domain residues that are involved in direct membrane interactions and confirmed that the mutations abrogate the primary membrane-binding step of the toxin. We also demonstrate that the loss of activity is easily compensated by rising the concentration of mutant toxin, thereby increasing the number of membrane-bound monomers. Although our SPR and gel electrophoresis-based assays show a clear and direct correlation between the cholesterol content and the effect of these rim domain mutants on VCC’s pore formation activity, further experimentation will be needed to understand the mechanism of the rim domain motifs in membrane binding. Moreover, the effect of different lipid compositions and epicholesterol on these mutants might provide insights into the precise chemical nature of cholesterol interactions by these motifs. In conclusion, our study provides a framework for designing drug molecules to prevent membrane interactions, not only in VCC, but also in other archetypical β-pore forming toxins.
CHAPTER 4

GLYCAN BINDING BY VCC
**Introduction**

*Vibrio cholerae* is a human pathogen that infects an estimated 1.4 to 4.3 million individuals every year and causes an acute gastrointestinal disease characterized by severe watery diarrhea. Throughout documented human history, *V. cholerae* has ravaged human civilization by infecting millions of people worldwide. Cholera is mainly prevalent in areas with a scarcity of clean water, such as places struck by massive natural disasters or savaged by civil unrest. Two major factors that potentially lead to the epidemic nature of this disease include the short incubation period of 2 hours to a few days, and the recurrent entry of the pathogen through polluted drinking water. Though the acid sensitivity of the bacteria means only a small handful of severe cases occur amongst all exposed individuals (Almagro-Moreno et al., 2015), over the centuries *V. cholerae* has managed to cause six separate cholera pandemics. Surprisingly, even though cholera is a treatable disease with simple remedies, the seventh pandemic has been ongoing since 1961, causing more than 100,000 deaths per year (Sack et al., 2004; Sack et al., 2006; Fournier and Quilici, 2007; Griffith et al., 2006). Recently, the distribution pattern of the disease has been altered by various natural forces and currently almost 40% of the cases are reported in the Americas, which used to contribute only 1% of total cases even a decade ago.

*V. cholerae* is normally found in an estuary ecology, mainly as biofilms or individual cells (Binsztein et al., 2004; Alam et al., 2006; Worden et al., 2006), where they can attach themselves to various marine life forms like cyanobacteria (*Anabaena*...
variabilis), phytoplakton (Skeletonema costatum), algae (Rhizoclonium fontanum), hyacinth (Eichornia crassipes), eastern oysters (Crassostrea virginica), aquatic arthropod (Gerris spinolae), and Atlantic blue crabs (Callinectes sapidus) (Islam et al., 1989; Martin & Bianchi, 1980; Islam et al., 1989; Hood et al., 1981; Spira et al., 1981; Shukla et al., 1995; Huq et al., 1986). The surface attachment of the V. cholerae cells depend on various environmental factors like temperature and phytoplankton bloom condition (Kierek and Watnick, 2003; Worden et al., 2006). The communal reservoirs not only provide better protection against the environmental factors and protozoan predators, but also present an opportunity for genetic exchange (Worden et al., 2006; Blokesch and Schoolnik, 2007). The bacteria shift their lifecycle upon entering the human body through tainted water or food materials and enter the human intestine after surviving the acidic environment of the stomach. In the intestine, individual bacteria cross the mucin layer and attach themselves to epithelial cell surfaces, where depending on the environment and nutrient availability, they begin reproducing and forming biofilms (Schild et al., 2008; Nelson et al., 2009; Klose and Mekalanos, 1998; Millet et al., 2014). Upon reaching a certain threshold, colonies begin expressing cholera enterotoxin, causing massive diarrhea, through which the pathogen escapes from the human body. During both phases of the cholera life cycle, the bacteria utilize carbohydrate-binding domain containing proteins for targeting and attaching to host cell surface glycans (Almagro-Moreno et al., 2015; Bhowmick et al., 2008; Absalon et al., 2011).
Lectins are small protein domains with carbohydrate binding specificity, which acquire their name from their ability to identify different blood groups. These domains are generally non-immunogenic in nature and function by decoding oligosaccharide arrays on the cell surface through stereo-specifically binding to carbohydrate molecules. Present in almost all life-forms including plants, animals and bacteria, lectins perform a variety of functions in cell-cell adhesion and signaling, cell differentiation, immune responses, mitogenic stimulation, embryonic growth, and host-pathogen interactions (Sharon and Lis, 1989; Lis and Sharon, 1998; Vijayan and Chandra, 1999; Rini, 1999; Drickamer, 1999). To recognize and bind to an enormous diversity of glycan combinations, lectin domains are also extremely diverse in nature. In addition, these domains are also used in vitro for glycoconjugate isolation, blood cell agglutination, and protein purification; potential applications as an antiviral, antifungal or antibacterial agent are also present.

Plant lectins are the most well-studied of all lectins and can be divided into five different categories based on their carbohydrate binding fold – L-type lectins (Legume), Hevein domain lectins (cereal lectin), β-prism fold I lectins (jacalin type), β-prism fold II lectins (monocot mannose-specific lectin) and β-trefoil fold (R-type) lectins (Sharma et al., 2007). Plant lectin domains are assumed to only perform two functions, defense against animal predators, parasitic fungi, and other invasive species and storage of carbohydrate molecules. However, knockout studies have clearly established that the functions of these lectins are not limited to protection nor storage, rather they are essential for developmental pathways. Consequently, the
absence of these lectins can cause impairment of embryogenesis and embryonic development. In bacteria, the absence of plant or animal like complex glycosylation profiles reduces the ability of these proteins to perform cytosolic or cellular biochemical tasks. Instead, lectin domains in bacteria are specialized in targeting and binding to host cell surface glycans, a vital step in host-pathogen interaction (Esko and Sharon, 2009; Lannoo and Van Damme, 2014).

The *V. cholerae* genome possesses multiple genes with lectin domains, especially domains that are structurally similar to the β-prism, β-trefoil (Olson and Gouaux, 2005; Absalon et al., 2011) and putative microbial lectins (Wong et al., 2012). Some examples of *V. cholerae* proteins that are involved in carbohydrate binding include the N-acetylglucosamine binding protein or GbpA (Absalon et al., 2011), the rugosity and biofilm modulator RbmC (Teschler et al., 2015), the biofilm-associated protein or Bap1 (Duperthuy et al., 2013), *Vibrio cholerae* cytolysin or VCC (Olson and Gouaux, 2005) and cholera toxin or CT (Sanchez and Holmgren, 2011). GbpA or N-acetylglucosamine (GlcNAc) binding protein is an adhesion factor of *V. cholerae*, which binds to GlcNAc containing oligosaccharides like chitin or mucin, permitting adhesion to planktons and human intestinal cells (Kirn et al., 2005; Zampini et al., 2005). Presumably, GbpA crosslinks *V. cholerae* cells to the mucin layer and to the bacterial surface. This crosslinking function is essential for plankton attachment as well as for successful intestinal colonization, therefore being crucial for the pathogenesis of cholera (Jude et al., 2009; Bhowmick et al., 2008; Wong et al., 2012).
The biofilm matrix associated proteins RbmC and Bap1 also contain lectin-like domains, which are instrumental in their function (Figure 4.1). With an overall 54% identity and 70% similarity between the full-length proteins, and 35-40% identity between the three lectin domains, it has been proposed that they act in a redundant manner (Hobley et al., 2015). Bap1 was identified as a protein associated with biofilm and antimicrobial resistance (Duperthyu et al., 2013). Recently, Bap1 has also been shown to be responsible for maintaining the mechanical flexibility of the pellicle part of the biofilm and preserving the hydrophobicity of the liquid-air edge (Hollenbeck et al., 2014). RbmC, on the other hand, covers the cluster of cells and forms a flexible cellular envelop that helps withstand the dynamic environment. Overall, Bap1 helps the bacteria adhere to the surface, and along with RbmC, the two provide a dynamic structure necessary for the survival of the bacterial colony (Berk et al., 2012). The role of the three homologs of β-prism lectin domains in these proteins is presumably to attach to cell surfaces. However, experimental deletion of the C-terminal lectin in RbmC does not affect the biofilm formation process (Absalon et al., 2011), perhaps due to the redundant function of these domains and presence of two proteins with similar functions. Moreover, removal of both proteins leads to a complete phenotypic loss of biofilm formation in V. cholerae cells.

VCC is a lectin-domain containing pore-forming toxin produced by V. cholerae. VCC is secreted as a water-soluble monomer that forms a pore in host cells upon targeting cell membranes (Olson and Gouaux, 2005). VCC is thought to be responsible for defending the bacteria against the host innate immune responses by
killing immune cells like neutrophils (Satchell et al., 2012), which in turn also provides nutrients for bacterial growth. VCC contains two kinds of lectin binding domains – a β-trefoil lectin domain and a β-prism lectin domain. The β-trefoil domain is inactive and perhaps performs only a structural function, whereas the β-prism lectin domain is active and binds to cell surface carbohydrate molecules. (Saha et al., 1997; Zhang et al., 1999). Previously, the VCC β-prism domain was thought to bind terminal β-galactosyl residues (Saha et al., 1997), but recent studies have established the N-linked glycan heptasaccharide core as its specific binding partner (Levan et al., 2013). This binding event involving N-linked glycans helps the β-prism lectin domain and the toxin itself bind to specific membranes, which explains the >90% loss of hemolytic activity in VCC when the β-prism lectin domain is removed (Olson and Gouaux, 2005; Levan et al, 2013). Certain studies have proposed additional roles for the β-prism lectin domain in oligomerization and pore formation (Rai and Chattopadhyay, 2013; Dutta et al., 2010), but experiments with β-prism truncated VCC have shown otherwise.

Similar to the β-trefoil domain, the β-prism lectin domain has a pseudo-three-fold symmetry (Olson and Gouaux, 2005) that can be found in related toxins from other Vibrio species. This lectin domain is also quite widespread outside of the Vibrio family, especially in the plant and animal kingdoms. Sharma et al. suggested the presence of at least 194 non-redundant β-prism lectin domains in plants with medium to high similarity in sequence (28-64%). A general consensus sequence of G…GXXXD was proposed as the carbohydrate-binding motif for β-prism lectins, but
due to the extreme diversity within these bacterial domains the working definition appears to be weak (Sharma et al., 2007).

In our study, we investigated three different *V. cholerae* β-prism lectin domains, two from the RbmC protein and one from VCC. Generally, lectins with a β-prism fold show millimolar affinity to monosaccharides, but exhibit nanomolar or micromolar affinities for cell surface glycans. Chip-based glycan screening confirmed that all three β-prism lectins in *V. cholerae* display similar specificity for cell surface oligosaccharides. We subjected all three *V. cholerae* β-prism lectins to systematic binding studies and characterized the binding affinities for monosaccharides and N-linked glycans. Interestingly, the carbohydrate binding affinity is stronger for the RbmC β-prism lectins than the VCC lectin, suggesting the need for stronger binding to the host cell surface in the biofilm matrix. We also attempted crystallization of the *V. cholerae* β-prism lectin domains with multiple N-linked glycan fragments like NGA2 heptasaccharide, pentasaccharide core, trimannose, and LanNAc molecule to understand the molecular mechanism of oligosaccharide recognition by these β-prism lectin domains. Out of all the trials, we only successfully solved the structure of the RbmC 2nd lectin domain bound to trimannose. Using mutagenesis and *in vitro* binding assays, we identified specific residues in all three lectins responsible for carbohydrate binding. An understanding of this interaction may allow us to inhibit the progress of infection by *V. cholerae* and may also lead to the development of inhibitors for various viral or pathogenic bacterial infections.
Figure 4.1 Comparison between the β-prism domains of VCC, Bap1, and RbmC. A, The domain arrangements in the VCC, Bap1, and RbmC full length proteins. VCC and Bap1 contain one β-prism domain (colored in gold) each, whereas the RbmC has two such domains. The β-prism domain in Bap1 has a long insertion in the sequence. B, The crystal structure of the VCC monomer rendered from PDB ID 1XEZ. The different domains are colored as follows; pro-domain (red), cytolyisin domain (gray), rim region (marine), β-trefoil lectin (magenta), and β-prism lectin (gold). C, Sequence alignment of all four β-prism domains from *Vibrio cholerae*. The domains have 35-40% sequence identity between them. The major difference between these domains is a five amino acid sequence, which is only present in the RbmC 2nd lectin domain and not in the other three. The Bap1 β-prism domain also differs from the other three domains by the presence of a large insertion with an unknown function.
Materials and Methods

Construction of RbmC lectin constructs

Full-length RbmC was cloned into the pET28 vector from the Vibrio cholerae El Tor str. N16961 (Tax ID 243277) genome. The individual β-prism lectin domains (RbmC 1st lectin from residue 506 to 638 and RbmC 2nd lectin from residues 823 to 955) were cloned out of the full-length protein and cloned as a his-GFP fusion protein product using the pNGFPBC vector. All constructs were sequenced and checked for additional mutations before use in our study. For expression purposes, the recombinant plasmids were transformed into the T7Express cell line (NEB inc.) and colonies were picked after overnight incubation.

Site-directed mutations were introduced into the lectin domains using a previously described method (Braman et al., 1996). Briefly, overlapping 5’ and 3’ primers varying between 39-40 bases in length were designed with the desired mutation in the center. Using Pfu Ultra DNA polymerase (Agilent Technologies), the entire vector was amplified for 16 cycles. Following amplification, the template strands were digested using DpnI (NEB Inc.), and PCR product was transformed into NEB5α cells (NEB inc.). The resulting colonies were grown overnight, minipreped and sequenced to confirm the presence of the desired mutations.

Construction of VCC lectin mutants

A full length VCC gene (hlyA, derived from Vibrio cholerae O1 El Tor strain 8731) previously cloned into a pHis-parallel vector (De and Olson, 2011; Olson and
Gouaux, 2005), and the VCC β-prism lectin domain cloned into a pET 28 vector (Levan et al., 2012), were subjected to site-directed mutagenesis to introduce mutations into the VCC lectin domain using the same procedure described above (De et al., 2015)

**Glycan-chip assay**

RbmC lectin domains were purified as GFP-fusion proteins and digested with thrombin to remove the GFP fusion partner. The lectin domains were labeled using primary amine chemistry with Alexa Fluor 488 dye carboxylic acid, TFP ester, bis-(triethylammonium salt) (Molecular Probes Alexa Fluor 488 Labeling kit, Invitrogen). The excess dyes were removed using a desalting column (provided by the manufacturer) before running the sample on an SDS-gel. Dye incorporation was confirmed by scanning the gel using a Typhoon FLA 9400 (GE Healthcare) imager. The efficiency of labeling was determined using the absorbance of the dye at 494 nm and the labeled proteins were sent to the Consortium for Functional Glycomics (Core D) for analysis using their mammalian glycan screen version 5.0. In brief, glycan chips were incubated with 180 µg/ml of the labeled lectin domains for 1 hour, washed three times to avoid non-specific binding, dried under nitrogen and imaged using a Perkin Elmer MicroScanarray XL4000 scanner. The results are represented as the average response units of six replicates and the standard deviation reported after removing the highest and the lowest data points. The entire dataset is freely available through the CFG website (www.functionalglycomics.org).
**Fluorescence binding assay**

The VCC β-prism lectin domain was purified using an S200 10/300 size exclusion column (GE Healthcare) that was pre-equilibrated with Phosphate buffered saline or PBS (20mM Phosphate pH 7.4, 150mM NaCl). Protein fractions were pooled together and dialyzed against the same PBS buffer. Fluorescence assays were performed using a Fluoromax-2 spectrophotometer (Horiba Jobin-Yvon, Edison, NJ) with the sample chamber temperature maintained at 15°C by a circulating bath temperature controller (RTE model 111, NESLAB Instruments, Inc.). Glycans were solubilized in PBS buffer before being titrated into the protein solution. All samples were incubated in a 6 mm x 6 mm quartz cuvettes (Starna Cell, Inc.) and were continuously stirred throughout the span of the experiment. Intrinsic protein fluorescence was monitored after exciting the protein samples at 295 nm with a band-pass of 4 x 4 nm. Upon addition of the solubilized glycans, increase in the intrinsic fluorescence was measured and the fractional increase in the fluorescence was calculated and normalized to the highest point. All data were procured in triplicates and the results were plotted in Origin 8.0 and fitted using a RandoA function. The standard error of mean (S.E.M.) was calculated from all three replicates.

**Expression and purification of constructs**

For the RbmC 2nd β-prism lectin domain, cultures were expressed at 30°C for 4 hours after inducing at an O.D. of ~0.6. Cells were pelleted by centrifugation in a LYNX 6000 Sorvall centrifuge at 3900 x g for 30 minutes and lysed by running
through a high-pressure homogenizer (Avestin Emulsiflex-C5). The lysate was spun at 40,000 x g for 30 minutes at 4 °C using an F-20-12x50 LEX rotor (Thermo Scientific) and the supernatant was loaded onto a ToyoPearl Nickel column (ToyoPearl). After washing the column with TBS + 40 mM imidazole, the purified proteins were eluted with TBS + 250 mM imidazole. The purified GFP-fusion protein was incubated with a 1:100 thrombin:protein (wt:wt) ratio to remove the GFP, and was ran on an SDS gel to confirm complete cleavage. For further purification, an S6 10/300 (GE Healthcare) size exclusion column was used and the fractions pooled together after running them on an SDS gel.

VCC full-length and β-prism lectin domain constructs were expressed at 30 °C for 3 hours and 37 °C for 8 hours, respectively, after being induced at an O.D. of ~0.6. Cells were pelleted and lysed as described above and loaded onto a Ni-chelating column (GE Healthcare). The purified proteins were eluted with 250 mM imidazole followed by running through an S200 16/60 column for further purification. The peak fractions were pooled after confirming the presence of the correct size bands on an SDS gel.

**Hemolysis assays**

For hemolysis assays, purified VCC was cleaved using α-chymotrypsin (at 1:350 wt/wt enzyme:protein) for 30 minutes at room temperature and serially diluted to varying concentrations. Activated VCC dilutions were added to the wells of 96-well clear bottom plates containing rabbit blood cells diluted in a blood dilution
buffer (20 mM NaH2PO4, 150 mM NaCl, 1mg/ml BSA, pH 7.4). The optical density of the solutions was monitored at 595 nm every 15-seconds at room temperature in a 96-well plate reader (iMark, Biorad). The raw data were converted into % lysis and HD50 values are represented as described previously (De et al., 2015) using KaleidaGraph v. 4.1.3 (Synergy software).

**Isothermal calorimetry**

Protein samples were purified and dialyzed against TBS (20mM Tris pH 7.5, 150mM NaCl) solution overnight using a snake skin dialysis membrane (Pierce). Ligand solutions were serially injected (5 mM for TriMan, LacMan, and NacMan; 120 μM for Asialofetuin) into 1.44 ml of the dialyzed protein (23 μM for VCC, 100 μM for β-prism Lectin domain, 48 μM for RbmC 2nd β-prism lectin domain) solution and thermograms were collected using a MicroCal VP-ITC (Microcal, LLC). Blank-subtracted ITC raw data were processed using NITPIC v 1.1.2, fitted using SEDPHAT v. 12. 1b and image processed using Gussi v. 1.0.8. The error of fitting was determined using SEDPHAT.

**Crystallization**

The RbmC 2nd β-prism lectin domain was purified using a ToyoPearl Nickel column as a GFP-fusion protein and was further purified using an S6 10/300 GE (GE Healthcare) size exclusion column. This was followed by digestion with thrombin to remove the His tag and GFP. The proteins fractions were pooled together and concentrated using a 3,000 kDa molecular cutoff concentrator (Vivaspin, GE
Healthcare) to 4.5 mg/ml. For crystallization, aliquots of the RbmC 2\textsuperscript{nd} β-prism lectin domain were incubated with a 1:10 molar excess of trimannose. Crystals were grown using the vapor diffusion method (RbmC 2\textsuperscript{nd} β-prism lectin domain + trimannose) with a reservoir solution containing 2M Ammonium sulfate, 0.1M Sodium acetate pH 4.6. Crystals were transferred into a reservoir solution supplemented with 20% glycerol and 3 mM trimannose and flash frozen in liquid nitrogen. X-ray diffraction data were collected using an Oxford diffraction X-ray source (Oxford Instruments) and indexed using CrysAlias Pro (Oxford Instruments). Molecular replacement was used to solve the structure using the VCC β-prism lectin domain (from PDB ID 1XEZ) as a template in Phaser (CCP4 Program Suite). Difference maps calculated by COOT showed unambiguous density for the trimannose molecule allowing us to build in the ligand. The resulting structure was refined using phenix.refine (Adams et al., 2002).
Figure 4.2 Glycan screening of the RbmC 1st and 2nd lectin domains identified several glycans with strong binding to RbmC β-prism domains. Purified RbmC labeled with Alexa Fluor 488 was screened against an array of 611 mammalian glycans by the Consortium for Functional Glycomics. Panel A shows the glycan screening results for the RbmC 1st lectin domain and panel B shows the results for the 2nd lectin domain. The average fluorescence of RbmC binding to each glycan is shown by red bars and the standard deviation from four replicates (after the highest and lowest replicates were dropped) is depicted by the black bars. The top three hits for each domain are illustrated in schematic form. Panels C and D show the positions of the lysine residues in the RbmC 1st and 2nd lectin domains. For the RbmC 1st lectin domain, a homology model is used to render the structure as no crystal structure is available. The RbmC 2nd lectin domain is rendered from the crystal structure of the domain with trimannose residue bound (this study). Symbols used to depict the corresponding sugars are shown in Panel E.
Results

Comparison of the RbmC 1st lectin, RbmC 2nd lectin and VCC β-prism lectin

We investigated three β-prism lectin domains from *V. cholerae* to understand the mechanism of glycan binding by these lectins. The sequence identity between the VCC β-prism lectin and the RbmC 1st lectin is 35.3%, whereas between the VCC β-prism lectin and the RbmC 2nd lectin it is 38.0%. Additionally, the two lectin domains from RbmC share a 40.6% sequence identity between them. Therefore, the two lectins from RbmC are more closely related to each other than they are to the VCC β-prism lectin. For example, the RbmC lectins have pI values of 9.16 and 8.94, respectively, for the 1st and 2nd lectins, but the VCC β-prism lectin has a much lower pI of 5.63. Although the general charged amino acid content for all the three lectins is very similar, the RbmC lectins differ from the VCC β-prism lectin in aromatic amino acid content and in overall hydrophobicity. The abundance of aromatic amino acids theoretically can bolster sugar-binding activity by introducing new stacking interactions with the ribose rings. The RbmC 1st and 2nd lectins carry 16 and 15 aromatic residues (Phe, Trp and Tyr), respectively, compared to 11 for the VCC β-prism lectin. Similarly, a GRAVY analysis (Kyte and Doolittle, 1982) illustrates a significantly higher average hydrophobicity for the RbmC lectins (-0.017 and -0.066 for the 1st and 2nd lectins respectively) compared to -0.266 for the VCC β-prism lectin. It is possible that the greater hydrophobicity is related to the RbmC protein
function in the biofilm matrix, with a plethora of non-polar molecules such as the sugar matrix and cellular debris.

**Comparison of glycan screens for RbmC 1st lectin, RbmC 2nd lectin and VCC β-prism lectin**

We identified the target glycans of the RbmC lectins using a glycan screen from the Consortium for Functional Glycomics. In doing so, we labeled the lectin domains of RbmC protein with Alexa Fluor 488 and monitored their binding against the glycan chip. The RbmC 2nd lectin domain resulted in slightly higher RFUs compared to the RbmC 1st lectin domain (Figure 4.2), but both lectins displayed an almost identical preference for glycans with a high binding affinity (above 400 RFUs). Most of the glycans with the highest number of RFUs contained either a part of the N-linked core heptasaccharide (NGA2) or some higher glycosylated version of the core heptasaccharide. This is not unexpected, as the VCC β-prism lectin domain previously displayed a similar propensity for N-glycan core heptasaccharides. Also, this observation is not surprising as all the lectin domains are β-prism lectins, which are sequence and structural homologs (paralogs to be specific) with 35-40% identity.

All three top hits for the RbmC 1st lectin domain are close chemical relatives of the glycan NeuAc α2-6 Gal β1-4 GlcNAc β1-2 Man α1-3(NeuAc α2-6 Gal β1-4 GlcNAc β1-2 Man α1-6) Man β1-4 GlcNAc β1-4 GlcNAc, or the A2 glycan. As one arm of this glycan is missing in one of the top binders and another one fucosylated in the GlcNAc tail, we can assume that the lectin is able to bind with only
one branch present, and the conservative tail modification does not interfere. All of
the top three glycans are also sialylated, which is interesting as some other V.
cholerae carbohydrate interacting proteins have shown to require asialated version of
lectins for binding, like cholera toxin or VCC (Galen et al., 1992; Saha and Banerjee,
1997). Additionally, the RbmC 1\textsuperscript{st} lectin bound to at least six more sialated glycans in
the screen along with sialic acid only. Moreover, it showed specificity for some
smaller oligosaccharides like GlcNAcβ1-4GlcNAc (RFU 519±70) and an array of
GlcNAc molecules attached to Gal molecules with an highest RFU (363±28) for the
Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-3Galβ1-
4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4(Fucα1-6)GlcNAc- ligand.

The RbmC 2\textsuperscript{nd} lectin, on the other hand, interacted with a similar number of
candidates. It is worth mentioning that the RbmC 2\textsuperscript{nd} domain bound predominantly
doubly-branched N-linked glycans, with the exception of a handful of smaller
oligosaccharides. Interestingly, both the 1\textsuperscript{st} and 2\textsuperscript{nd} RbmC lectins exhibited a
comparable number (eight) of sialated glycans as preferred binding partner in glycan
screening. Another observation is the indifference of the lectins to fucosylation events
on the GlcNAc tail or on the branched arm of the glycan.
Figure 4.3 Representative binding curves for RbmC 2nd lectin domain obtained using ITC and fluorescence assays. Panel A, B, and C show binding curves for methyl α-mannose, trimannose and asialofetuin, respectively, using ITC. The plot in panel D shows the fluorescence binding curve for the pentasaccharide molecule. The K_D values are reported in Table 4.1 and the thermodynamic parameters are shown in Table 4.2.
To determine which part of the N-linked glycan the β-prism lectin binds to, we selected smaller fragments of the N-linked glycan and tested them against β-prism lectins (both RbmC β-prisms and the VCC β-prism lectin) using an ITC based binding assay. We first tested the methyl-α-mannose (MMA) with the β-prism lectins, followed by other monosaccharides such as galactose (Gal), n-acetyl glucosamine (GlucNAc) etc. Next, we measured binding affinities of the β-prism lectins for larger fragments like trimannose, LacNAc, and ManNAc molecules. In addition, we also used the core NGA2 glycan, NA2 glycan, and A2 glycan to measure binding data for β-prism lectins using a fluorescence-based assay.

Amongst the monosaccharides, MMA showed the highest affinity with Gal and GlucNAc failing to show substantial binding under our reaction conditions. The binding reactions for the three lectins with MMA are exothermic in nature, with ΔG values of -6.37, -5.36 and -3.65 kcal/mol for the RbmC 1st lectin, RbmC 2nd lectin, and the VCC β-prism lectin, respectively. Although the enthalpic contribution for sugar molecules binding to the VCC β-prism lectin binding is 20-40 fold higher than the RbmC lectins, these lectins display affinities many fold higher than for the VCC β-prism lectin. Interestingly, the TΔS for the RbmC 1st lectin is slightly favorable (-0.11 kcal/mol), whereas the TΔS values are strikingly unfavorable for the other two lectins. Interestingly, for the VCC β-prism lectin, the considerably higher enthalpy successfully compensates the excessive entropic contribution to keep the overall reaction exothermic. Surprisingly, bigger pieces like LacNAc and ManNAc did not
display substantially stronger binding, and ManNAc failed to display any binding when tested against the VCC β-prism lectin domain (Table 4.1).

Trimannose, or the trisaccharide core of N-linked glycans, binds to the β-prism lectins strongly with high nanomolar–micromolar affinity. The RbmC 1st lectin domain is the tightest binder amongst the group with a $K_d$ of 1.25 μM, followed by the RbmC 2nd lectin, which binds trimannose with a 5.8 μM affinity (Figure 4.3 and Table 4.1). In contrast to these higher nanomolar or lower micromolar affinities, the VCC β-prism lectin exhibits a $K_d$ of 192 μM, which is roughly ~40-100 fold higher than its peers. Following in the footsteps of the MMA binding thermodynamics, the overall free energy for trimannose binding is also favorable in the range of -6.6 to -8.6 kcal/mol. The entropic contribution of trimannose binding is negative for the RbmC 1st lectin and the VCC β-prism lectin, but it is positive for the RbmC 2nd lectin.

The glycoprotein Asialofetuin (AF), which has been previously shown to be a binding partner for the VCC β-prism lectin domain (Saha and Banerjee, 1997), demonstrated a strong affinity of 27.8 nM, 63 nM and 1,282 nM for the RbmC 1st lectin, RbmC 2nd lectin, and the VCC β-prism lectin, respectively. Compared to MMA and the trimannose, as expected, the free energies are higher for the AF binding, with values around 10 kcal/mol. It is worth noting that the enthalpic and entropic costs for binding to AF are negative for all three lectin domains we tested.

We verified the glycan binding affinities of these lectin domains using the NGA2 pentasaccharide core and the NGA2 heptasaccharide core glycans as binding
Table 4.1 Summary of the binding affinities of the β-prism domains for various ligands. Binding affinities for different ligands are determined using ITC and intrinsic fluorescence assays. The purified β-prism domains were titrated with the corresponding ligands and the data were processed using NITPIC v. 1.15 and fit using a single site binding model in SEDPHAT v. 12.1b. The values in parentheses denote the 95% confidence limits and are mentioned only for the ITC experiments. The fluorescence assays were done by titrating the ligand molecules into the protein solution. Normalized fluorescence data were fit in Origin 6.0 using the RandoA function. The schematic representations for each ligand is denoted along with the names. ‘--‘ represents experiment not carried out.

<table>
<thead>
<tr>
<th>Glycan/Sugar</th>
<th>VCC β-prism binding affinity</th>
<th>RbmC 2nd lectin binding affinity</th>
<th>RbmC 1st lectin binding affinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Me-α-mannose</td>
<td>2718 μM (2625, 2777)</td>
<td>178.7 μM (69.6, 268)</td>
<td>502.6 μM (410.4, 725.6)</td>
</tr>
<tr>
<td>LacNac</td>
<td>Did not bind</td>
<td>Did not bind</td>
<td>--</td>
</tr>
<tr>
<td>ManNac</td>
<td>Did not bind</td>
<td>Did not bind</td>
<td>--</td>
</tr>
<tr>
<td>Trimannose</td>
<td>228.6 μM (224.5, 240.6)</td>
<td>4.1 μM (4.0, 5.6)</td>
<td>0.6 μM (0.2, 3.6)</td>
</tr>
<tr>
<td>Pentasaccharide</td>
<td>--</td>
<td>1.1 ± 0.57 nM</td>
<td>--</td>
</tr>
<tr>
<td>NGA2</td>
<td>167 ± 45.9 nM</td>
<td>1.5 ± 0.46 nM</td>
<td>--</td>
</tr>
<tr>
<td>NA2</td>
<td>961 ± 132 nM</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>A2</td>
<td>1.1 ± 0.4 μM</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Asialofetuin</td>
<td>646.1 nM (179.8, 1755.8)</td>
<td>6.4 nM (0.17, 10)</td>
<td>10 nM (9.9, 14.6)</td>
</tr>
</tbody>
</table>
Table 4.2 Thermodynamic parameters of the three β-prism domains with MMA, trimannose and asialofetuin.

<table>
<thead>
<tr>
<th></th>
<th>Ligand</th>
<th>ΔG in kcal/Mol</th>
<th>ΔH in kcal/Mol</th>
<th>TΔS in kcal/Mol</th>
<th>ΔS in cal/Mol*K</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCC β-</td>
<td>Methyl-α-mannose</td>
<td>-2.9</td>
<td>-5.1</td>
<td>-1.5</td>
<td>-5.1</td>
</tr>
<tr>
<td>prism</td>
<td>Trimannose</td>
<td>-5.0</td>
<td>-11.9</td>
<td>-6.8</td>
<td>-23.0</td>
</tr>
<tr>
<td>lectin</td>
<td>Asialofetuin</td>
<td>-8.1</td>
<td>-9.0</td>
<td>0.9</td>
<td>-2.9</td>
</tr>
<tr>
<td>RbmC</td>
<td>Methyl-α-mannose</td>
<td>-5.4</td>
<td>-13.2</td>
<td>-7.8</td>
<td>-26.3</td>
</tr>
<tr>
<td>2nd</td>
<td>Trimannose</td>
<td>-7.3</td>
<td>-2.4</td>
<td>4.9</td>
<td>16.4</td>
</tr>
<tr>
<td>lectin</td>
<td>Asialofetuin</td>
<td>-10.1</td>
<td>-19.0</td>
<td>-8.9</td>
<td>-29.7</td>
</tr>
<tr>
<td>RbmC</td>
<td>Methyl-α-mannose</td>
<td>-6.4</td>
<td>-6.3</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>1st</td>
<td>Trimannose</td>
<td>-8.6</td>
<td>-10.3</td>
<td>-1.6</td>
<td>-5.4</td>
</tr>
<tr>
<td>lectin</td>
<td>Asialofetuin</td>
<td>-10.7</td>
<td>-60.4</td>
<td>-49.7</td>
<td>-166.7</td>
</tr>
</tbody>
</table>
Figure 4.4 **Crystal structure of the RbmC 2nd lectin domain bound to trimannose.** A, Top-down view of the crystal structure of the RbmC 2nd lectin domain bound to trimannose. The trimannose molecule is shown along with the omit map contoured at 2.5σ. The electrostatics potential (in eV) for the surface is calculated using APBS plugin for Pymol and the range is shown in the colored scale bar. B, Schematic generated using LigPlot showing the interactions between residues in the RbmC 2nd lectin domain and trimannose molecule. Hydrogen bonds are depicted by green dotted lines and the hydrophobic interactions are shown as red spoked arcs. The mannose residues in the trimannose molecule (α-D-Man-(1,3)-[α-D-Man-(1,6)]-D-Man) are labelled as 1, 2 or 3 from the left to right. C and D, two different zoomed-in views of the binding pocket in RbmC bound to trimannose. The point of views are depicted with pink arrowheads in the panel A. The protein is shown as a surface representation and the residues involved in the interactions are denoted as sticks. The hydrogen bonds are depicted by black dashed lines.
partners. Both of the above-mentioned glycans bound to the RbmC 2nd lectin with affinities of 0.8 nM and 1.5 nM, respectively. The VCC β-prism lectin domain, on the other hand, displayed a $K_d$ of 167 nM with the NGA2 heptasaccharide (Table 4.1). Using the VCC β-prism lectin domain, we also tested NA2 and A2 glycans and noticed that the addition of sugar molecules on the NGA2 heptasaccharide only slightly reduced the binding affinities.

**Crystal structure of RbmC 2nd lectin with trimannose bound**

We solved the crystal structures of the VCC β-prism lectin domain, RbmC 2nd β-prism lectin domain (apo) and the RbmC 2nd β-prism lectin domain bound to trimannose to high resolution. All three of the structures are overall quite similar and can be superimposed on each other with the rmsd of 1.18 Å. The RbmC 2nd lectin apo protein was crystallized in a C 2 2 2 1 space group with a cell dimension of 90, 92, 100 Å, whereas the trimannose bound protein yielded crystals with a P2 1 2 1 2 1 space group with 39, 59, 113 Å dimension (Table 4.3). The RbmC 2nd β-prism lectin domain apo crystal contained three copies of the protein in the asymmetric unit whereas the RbmC 2nd β-prism lectin domain bound to trimannose had two copies in the asymmetric unit. Both crystals diffracted ~2 Å resolution and both molecules in the trimannose bound crystals displayed unambiguous density for trimannose.

Previous structures from our lab illustrated features like the Asn617 residue, which forms a key hydrogen bond with C4 of Me-α-mannose, but because MMA is a monosaccharide, only a part of the binding pocket was utilized. Our new structure builds on these previous studies and provides clues about how these domains may
interact with more complex cellular glycan targets. In our crystal structure, the trimannose molecule is stabilized by a total of eight hydrogen bonds and at least five major hydrophobic interactions with the protein. The previously identified Asn 617 forms two key hydrogen bonds with the most protruding mannose (Man 401) molecule of the ligand (Figure 4.4), perhaps explaining why removal of this amino acid residue caused a 99% loss in glycan binding (De et al., 2015). Other residues that formed hydrogen bonds with the ligand are Phe 614 (backbone), Thr 634, Asn 635, Gly 636 (backbone), Gly 637 (backbone) and Tyr 658. Along with the polar interactions, some crucial hydrophobic contacts were also observed between the ligand molecule and the amino acid residues Ser 612, Gly 613, Ala 615, Trp 660 and Trp 712. Among these, the Trp 712 analog in the VCC β-prism lectin structure (Tyr 716) was seen in two different conformations – open and closed with respect to the ligand. However in our apo and ligand-bound structures this residue (Trp 712) was only found in a closed. It is possible that structural constraints like this gives RbmC a higher affinity than in the VCC β-prism lectin, as we will see later. Another difference between these old and new structures is a 5 amino acid loop, which is absent in the VCC β-prism lectin domain and RbmC 1st β-prism lectin domain, but present in the RbmC 2nd β-prism lectin. In the apo structure of the RbmC 2nd β-prism lectin domain, this loop (from residues 630-634) is unstructured and hardly visible. Surprisingly, in the trimannose bound structure, the 3rd mannose residue in the trimannose molecule forms multiple interactions (hydrophilic and hydrophobic) with this loop, making it less flexible in the crystal. Interestingly, out of the 7 hydrophilic
interactions previously mentioned between the protein and the ligand, 3 of the amino acids involved resides in the above-mentioned loop. This loop may be one of the causes for the 10-fold stronger binding of sugars/glycans observed in the RbmC 2nd β-prism lectin domain compared to the VCC β-prism lectin. Interestingly, these five residues are only present among five other β-prism lectin domains from different Vibrio species.

**Mutagenesis of the VCC β-prism lectin binding pocket**

To confirm the importance of binding pocket residues, we employed a site-directed mutagenesis based approach utilizing the VCC β-prism domain as a model. We mutated a total of 10 residues in and around the binding pocket to probe potential interaction sites for the heptasaccharide glycans. Six out of ten residues exhibited a significant difference in activity upon mutation, and only four of the residues showed a negative effect on the function of the protein. The equivalent residues which form physical contacts with trimannose residues in the RbmC 2nd lectin structure, when mutated in VCC all revealed a >50% loss of function, with the exception of Tyr 658 (Phe 652 in the VCC). Mutation of Gly 613 to Lys, caused a ~46% loss in activity of the protein (Figure 4.5). Interestingly, this residue (Gly 613 in VCC and the RbmC 2nd lectin) does not directly interact with trimannose in the RbmC 2nd lectin structure. This may be an indication that one of the branches of the glycan may be in close contact with the surface or residue itself, and when we introduced a bulky Lys residue, it may have caused a steric clash inhibiting glycan binding. Where the other branch of the glycan will reside is still unclear from our data, but the absence of any
effect on Asn 630 does point towards the fact that the other branch of the glycan may not interact with the outside surface containing the Asn residue.

Figure 4.5 **Structure aided mutagenesis in the VCC β-prism lectin domain.** A. Site directed mutations were introduced into the full length VCC in the vicinity of the trimannose binding pocket. The mutants proteins were purified as described previously (De et al., 2015) and were used for hemolysis assays. The HD$_{50}$ is defined as the concentration of the protein needed for 50% lysis of rabbit erythrocytes in 30 minutes at room temperature. Among the residues tested, two mutants exhibited > 100-fold loss and four others displayed 10-100 fold loss in hemolytic activity. The error bars represents S.E.M. of the triplicated HD$_{50}$ values. B, The VCC β-prism lectin domain is shown in surface representation and the mutations are depicted according to their loss of hemolytic activity. Residues are colored by their effect in the hemolytic activity; >100 fold loss is denoted in red, 10-100 fold loss is shown in orange and the residues which failed to show any major effect (<10 fold) are shown in green. Individual residues are labeled in black. The trimannose molecule is depicted by superimposing the VCC β-prism lectin domain on the RbmC 2nd lectin structure.
Sequence alignment of the β-prism domains from Vibrio spp. BLAST searches were conducted using the *V. cholerae* β-prism domain protein sequence and the resulting similar sequences were aligned using ClustalW algorithm implemented in MEGA v. 7.0.14. The image is prepared in Jalview v. 2.9. Most of the proteins containing β-prism domains in *Vibrio* spp. are denoted as bacterial toxins except the RbmC from *V. cholerae* (used in this study). The residues in the RbmC 2nd lectin domain involved in either hydrogen bonding or non-polar interactions are designated using red arrowheads above the alignment. All of these residues are identical between the RbmC 1st and 2nd lectin domains except the two residues at the end of the five amino acid loop highlighted using a light blue colored box, which are entirely absent in RbmC 2nd lectin domain. Domains involved in trimannose-RbmC 2nd lectin domain interaction. The conservation score, conservation quality of each residue, and the consensus sequence for the *Vibrio* β-prism lectin domain are calculated using Jalview v. 2.9 and is shown below the alignments.
Figure 4.7 Evolutionary relationships between the different β-prism domains in *Vibrio* spp. An evolutionary tree was constructed in MEGA v. 7.0.14 by using the maximum likelihood method of the Jones-Taylor-Thorton (JTT) matrix based model. Initial tree(s) were calculated by using Neighbor-Join and BioNJ algorithms on the pairwise distances determined by JTT model. The tree with the highest log likelihood (-2516.3794) is shown. The branch lengths of this tree correspond to the number of substitutions per site (scaled-tree). Evolutionary analyses were done using MEGA7 after all gaps and missing data were removed. The β-prism lectin domains containing the five amino acid loop near the trimannose binding site are denoted with red rectangles and the β-prism lectin domains used in this study are marked with a green highlights.
Table 4.3 X-ray and refinement statistics.

<table>
<thead>
<tr>
<th></th>
<th>RbmC 2nd lectin apo</th>
<th>RbmC 2nd lectin trimannose bound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C 2 2 2_l</td>
<td>P 2_l 2_l 2_l</td>
</tr>
<tr>
<td>Cell dimensions a,b,c (Å)</td>
<td>90.6, 92.7, 100.3</td>
<td>39.0, 59.3, 113.6</td>
</tr>
<tr>
<td>Cell dimensions α,β,γ (°)</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>15.9-2.1</td>
<td>16.8-2.0</td>
</tr>
<tr>
<td>R&lt;sub&gt;meas&lt;/sub&gt;</td>
<td>0.34 (1.16)</td>
<td>0.29 (1.20)</td>
</tr>
<tr>
<td>R&lt;sub&gt;pim&lt;/sub&gt;</td>
<td>0.13 (0.43)</td>
<td>0.11 (0.50)</td>
</tr>
<tr>
<td>I/σI</td>
<td>12.2 (2.4)</td>
<td>7.7 (2.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.6 (99.9)</td>
<td>92.4 (95.9)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.2 (7.2)</td>
<td>6.7 (5.5)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.1</td>
<td>2.0</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>14,721</td>
<td>18,579</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.19/0.24</td>
<td>0.20/0.24</td>
</tr>
<tr>
<td><strong>No. of atoms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>6,083</td>
<td>4,174</td>
</tr>
<tr>
<td>Ligand</td>
<td>0</td>
<td>66</td>
</tr>
<tr>
<td>Water</td>
<td>528</td>
<td>336</td>
</tr>
<tr>
<td><strong>R.M.S. deviation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>Bond angles</td>
<td>0.703</td>
<td>0.984</td>
</tr>
<tr>
<td><strong>Ramachandran statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>97.5 %</td>
<td>97.0 %</td>
</tr>
<tr>
<td>Allowed</td>
<td>2.3 %</td>
<td>3.0 %</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.3 %</td>
<td>0.0 %</td>
</tr>
</tbody>
</table>
Discussion

We have shown using glycan screens, binding data, and crystal structures that three β-prism lectins from *V. cholerae* demonstrate specificity for N-linked glycans. Moreover, we established that the β-prism lectins bind to the pentasaccharide core in order to interact with N-linked glycans. Until now, these three are the only bacterial β-prism lectin domains that have been reported to have specificity for N-linked glycans. These interactions are likely to be advantageous for the bacteria since targeting and binding to the membrane gives the bacteria a higher chance of stabilizing microcolonies, and helps PFTs lyse cells by accelerating the search for other monomers on the membrane.

Because of their predominant occurrence on many eukaryotic cell surfaces, other types of lectin domains also exhibit specificity towards N-lined glycans. The nucleocytoplasmic lectin Orysata from *Oryza sativa* (Atlah et al., 2011), PP2-A1 (Phloem protein 2-A1) from *Arabidopsis thaliana* (Beneteau et al., 2010), Galanthus nivalis agglutinin (GNA), and GNA-related lectin from *Zea mays* (Fouquaert et al., 2009) are some examples of plant lectins that display affinity for N-glycans. The precise carbohydrate binding specificity of these lectins varies based on the function they perform. For instance, the orysata and the GNA-related lectin from *Zea mays* are specific for complex type N-glycans, whereas the PP2-A1 and the GNA lectins prefer high-mannose N-linked glycans. Lectins of non-plant origin have also been shown to identify and bind to N-linked glycans (Beneteau et al., 2010; Fouquaert et al., 2009). KSA-2 from *Kappaphycus striatum* (a red algae), mitogenic lectin from *Rhizoctonia*
bataticola (a fungus), and Codakine lectin from Codakia orbicularis (a bivalvia molluscs) are examples of some non-plant lectins with N-glycan specificity (Hung et al., 2011; Nagre et al., 2010; Gourdine et al., 2008).

Although a large amount of genetic and biochemical information is available, there is limited structural information regarding interactions between lectins and N-glycans. Structures have been solved for C-type mannose-binding protein (MBP) bound to the trimannose core and several other monosaccharides and disaccharides (Kenneth et al., 2002): Dendritic cell-specific intracellular adhesion molecule-3 grabbing nonintegrin (DC-SIGN) bound to the pentasaccharide GlcNAc beta(1-2)Man alpha(1-3)[GlcNAc beta(1-2)Man alpha(1-6)]Man (Fienberg et al., 2001), the Pterocarpus angolensis lectin (PAL) in complex with the core pentasaccharide (Buts et al., 2006), the Platipodium elegans lectin with the core penta and heptasaccharides (Benevides et al., 2012), and codaine lectin bound to the nonasaccharide glycan (Gourdine et al., 2008). However, all of these lectins possess structural characteristics distinct from bacterial β-prism lectins and therefore structural information obtained from these lectins cannot be extended to β-prism domains.

Multiple crystal structures have been determined for different β-prism lectin domains in apo forms or bound to smaller ligands, like mannose or dimannose moieties. However, to the best of our knowledge, no β-prism lectin has been crystallized in trimannose or pentasaccharide bound forms. It is worth mentioning that the VCC β-prism lectin domain was crystallized with a methyl-α-mannose residue bound by Levan et al. Interestingly, all of our attempts to co-crystallize or
soak in trimannose or any bigger piece of N-glycans with the VCC β-prism lectin domain have been unsuccessful. Our structure of the RbmC 2nd lectin domain bound to trimannose core provides clues to why crystallizing a bigger piece of an N-glycan was not successful. Firstly, the GlcNAc tail of the glycans, which connects them to Asn residues on cell-surface proteins, is projecting outward from the protein and thereby introducing flexibility into the system. Secondly, although the addition of the two GlcNAc molecules on the trimannose results in an almost 1000-fold higher K_d, the introduction of each saccharide moiety adds to the flexibility and heterogeneity of the system. Similarly, in our binding assays, AF titration to both of the RbmC lectins resulted in a slightly higher free energy of binding (ΔG) when compared to the trimannose binding kinetics (Table 4.2), but a massive entropic penalty needed to be compensated for to proceed forward (ΔS value-166.71 cal/mol*K for RbmC 1nd lectin and 29.71 cal/mol*K for RbmC 2nd lectin). This is not particularly surprising since the dehydration of each sugar molecule contributes significantly to the entropy, and the final enthalpy-entropy compensation determines the directionality of this reaction. Interestingly, for the VCC β-prism lectin, the entropic component of binding is lower for AF than for MMA or trimannose, but the overall ΔG of the reaction is comparable to that of RbmC lectins due to the higher enthalpic contribution seen in ITC assays (Table 4.2). The pentasaccharide and heptasaccharide cores of N-linked glycans exhibit the highest affinity for β-prism lectins amongst all glycans/glycoconjugates that were tested (Table 4.1). This further strengthens our argument that GlcNAc residues in the NGA2 tail hang outside of the structure and do
not physically interact with the lectin domain. The addition of additional monosaccharides to the NGA2 core heptasaccharide caused only a slight reduction in the binding affinity as tested in the VCC β-prism lectin. Taken together, these data suggest that the β-prism lectin binds primarily to the pentasaccharide core of N-linked glycans, and the GlcNAc tail or additional monosaccharide residues do not tightly interact with the β-prism lectins.

From our binding data, we identified that the affinities of the RbmC lectins for both sugars and glycans are significantly higher than in the VCC β-prism lectin. We also identified a five amino acid loop in the RbmC 2nd lectin, missing in the RbmC 1st lectin and the VCC β-prism lectin, which makes crucial interactions with the trimannose molecule in our structure. Interestingly, this loop, with sequence PVQGT, is absent in most of the *Vibrio* spp (Figure 4.6). β-prism lectin containing proteins and is only present in a handful of species like *V. anguillarum*, *V. ordalii*, and *V. albensis*. *V. mimicus* contains a similar loop, but with a degenerative sequence of PIEGT. This loop may be partially responsible for the higher affinity of the RbmC 2nd lectin to trimannose compared to the VCC β-prism lectin, but this cannot explain the tight binding affinity of the RbmC 1st lectin. Phylogenetics based evolutionary studies further showed that this five amino acid insertion came separately only in the group, which contains the RbmC 2nd lectin domain, after they diverged from the common ancestor (Figure 4.7). The absence of the structure of the RbmC first lectin limits our ability to identify precise structural motifs responsible for binding to the lectin, but homology modeling has provided some insight into this. Another possible
explanation for the difference between the binding affinities of RbmC and VCC lectins could lie in the conformational state of the proteins. Previous structures of the VCC β-prism lectin in apo and ligand-bound forms identified a key tryptophan residue (Trp 706), which adopts either an open or closed conformation depending on whether the ligand is present or not, respectively. The open position of Trp 706 may enable the lectin to accommodate additional portions of the glycan molecule for better binding. Interestingly, in the apo RbmC 2nd lectin structure and in the RbmC 1st lectin homology model based on the RbmC 2nd lectin apo structure, the corresponding residue (Trp 712) is in the open form. Lastly, the sequence comparison shows that the RbmC lectins contain more aromatic residues and potentially more accessible hydrogen bonding residues around the binding site, which could be responsible for the differences in the binding affinities.

*V. cholerae* possesses multiple β-prism lectins in its proteome (Figure 4.1). Two of these domains are part of the biofilm matrix protein RbmC, and one is attached to the C-terminal domain of VCC. In both cases, the precise roles and mechanisms of action of the β-prism lectin domains were previously unknown. However, the VCC β-prism lectin was known to bind monosaccharides and show specificity towards N-linked heptasaccharide core. The specificity of the RbmC lectin domains was unknown prior to this study, although there were some proposals regarding RbmC interactions with VPS (*Vibrio* polysaccharide) carbohydrates (Berk et al., 2012). Our glycan screening data identified that the RbmC lectins bind to cell surface glycans with a specificity similar to the VCC β-prism lectin. This is not
surprising as both proteins are interacting with similar types of host cell surfaces. Both of the lectin domains from RbmC, in the glycan screen, interacted mainly with N-linked core glycans or higher glycosylated versions of the same heptasaccharide. Both lectin domains from RbmC bound fucosylated and sialylated versions of N-glycans. The 2nd lectin of the RbmC also showed high specificity for trimannose cores with GlcNAc tails attached and to repeating units of Gal-GlcNAc moieties. As N-linked glycosylation is prevalent in almost all forms of animals and plants, it is safe to assume that the RbmC lectins interact with cell surface glycans in different hosts involved in the life cycle of V. cholerae. By doing so, the lectins aid the RbmC protein to perform its role as an important biofilm matrix protein. Previously, another carbohydrate binding protein called GbpA (GlcNAc binding protein A) was identified, which performs an important role in colony formation in both phytoplanktons as well as in humans (Kirn et al., 2005). Additionally, as Vibrio cholerae is known to attach to other marine invertebrate life forms like cyanobacteria, diatoms, green algae, oysters, blue crabs and water hyacinths, it is possible that while forming a biofilm inside a host, RbmC performs important roles in stability or structural rigidity as a biofilm matrix protein utilizing lectin domains.
CHAPTER 5
CONCLUSION
Summary

Cholera is a human acute gastrointestinal disease caused by the Gram-negative pathogen *Vibrio cholerae*. *V. cholerae* infects the human body through contaminated food or water. An unusually short incubation period of 2-3 days and an abundance of contaminated water make the pathogen highly infective in nature. After reaching the small intestine, the bacteria attaches itself to the epithelial walls and rapidly starts multiplying, generally resulting in the formation of a biofilm on the epithelial lining. Upon attaining a threshold cell density, the bacteria escapes the human body through watery diarrhea caused by a secreted toxin called cholera toxin (CT) (Zhu et al., 2002). Although cholera can be easily treated by simple remedies, more than 100,000 cases resulted in death around the globe in the year 2012 alone. Previously, thought to be an African and Asian disease, currently almost half of the global cholera cases are reported from the American continent. Recent shifts in climate and geo-political situations have created a perfect breeding ground for the disease. For example, in the aftermath of the Haiti earthquake, in 2010, almost 700,000 patients were diagnosed with cholera, showing how vulnerable the modern world is to this ancient pathogen.

The *V. cholerae* life cycle mainly consists of two stages in very diverse environments. Firstly, the bacteria thrive in the marine environment where they attach to aquatic organisms like cyanobacteria, phytoplankton, diatoms, algae, oysters or crabs. Following accidental contamination of drinking water supplies, the bacteria enters the human body and the human intestine. As *V. cholerae* is extremely acid
sensitive, most of the cells die in the stomach and only a handful of them reach the ileum. There, the bacterium swims through the thick mucin layer to reach the epithelial cell surface, where the bacteria adhere to the cell membrane using non-commitment attachment factors (Almagro-Moreno et al., 2015). If the attached surface is suitable for reproduction and growth, the genes responsible for the colonization process are expressed and the pathogen starts dividing rapidly. Additionally, beginning in this stage the bacteria produces specific attachment factors that stabilize interactions between bacteria and cell membranes. The virulence genes in V. cholerae are scattered throughout the genome and are under precise environmental and quorum-sensing control. One of the genes involved in virulence is classical cholera toxin, which is the main toxin responsible for watery diarrhea: the main characteristic of cholera. Through this watery secretion, the bacteria leave the human body and start their aquatic life cycle again. These bacteria can endure harsh marine environments for much longer than other fecal bacteria, like E. coli and Streptococci (Davies et al., 1995). Along with this survivability, V. cholerae’s ability of rapid genetic transformation and adaptability makes this pathogen highly unpredictable and it will certainly cause more outbreaks in the coming future.

As mentioned earlier, the classical cholera toxin (CT) is the main toxic factor that V. cholerae secretes while infecting the human body. In addition to CT, V. cholerae produces a number of accessory toxins to aid the bacteria in the colonization and infection process. Vibrio cholerae cytolysin (VCC), Zot toxin, Ace toxin, new cholera toxin, Shiga-like toxin and thermostable hemolysin are major examples of
these accessory toxins. These toxins perform various important functions during the cholera infection cycle, including killing host immune cells, damaging host cell-cell interactions, and inhibiting the host translation process.

*Vibrio cholerae* cytolysin (VCC) is a pore-forming toxin, which exhibits severe cytotoxic and enterotoxic capabilities. The absence of the VCC gene (*hlyA*) ensues around a 30-fold loss in the infectivity of the pathogen in mouse models (Williams et al., 1993). On the other hand, VCC is able to instigate acute gastroenteritis in human volunteers (Honda et al., 1988; Morris et al., 1984; Zitzer et al., 1997) irrespective of the presence or absence of the CT gene. Moreover, purified VCC stimulates water and mucus build up in rabbit and mice models (Ichinose et al., 1987; Debellis et al., 2009), which confirms the involvement of this pore-forming toxin in cholera pathogenicity. Later, these fluid accretion events were attributed to an efflux of Ca^{2+} ions and cellular non-immune responses to pore formation in the membrane. VCC has been tested on various human immune and non-immune cell lines like neutrophils, erythrocytes, CHO, Vero, MDCK etc. and in all cases the toxin demonstrates cell lysing abilities (Zitzer et al., 1993; Jonas et al., 1994; Debellis et al., 2009;). Additionally, apoptosis, autophagosome formation, and pro-inflammatory effects have also been reported as a result of VCC pore formation on the membrane (Saka et al., 2008; Ou et al., 2009). It has been suggested that VCC plays a key function in cholera pathogenesis by defending the bacteria against human immune cells such as neutrophils (Olivier et al., 2007; Queen and Stachell, 2012).
VCC is a pore-forming toxin that like similar PFTs, is secreted by the bacteria as a water soluble monomer, diffuses to target cell membranes and forms a membrane-spanning pore after binding to cell surfaces using multiple specific interactions. Specifically, VCC is secreted as a ~80 kDa monomer that undergoes proteolytic cleavage for activation. From the secretion site, the protein diffuses freely and attaches to the target cell membrane using cell surface glycans as receptors. This binding is followed by direct membrane interactions through the membrane-proximal rim domain, which enables the toxin to stick to the lipid bilayer. Once on the bilayer, the individual monomers find each other and form a pre-pore complex that ensures a coherent insertion event for the membrane-spanning sequences to construct the heptameric membrane-spanning pore.

VCC belongs to an archetypical β-PFT family, which includes the *Staphylococcus aureus* α-hemolysin. Our VCC heptamer structure disclosed several traits, which were formerly unknown to us. The heptameric pore has dimensions of 140 Å x 135 Å with the seven-fold symmetry axis running through the middle of the pore. Very interestingly, the inner diameter of the channel at the narrowest point is only 8-Å, which is smaller than α-hemolysin’s 10-Å constriction. Moreover, the constriction is formed by a heptad of aromatic tryptophan residues (W318). Since other ion channels generally do not contain aromatic residues due to thermodynamic penalties, it is possible that this residue may perform some unknown function in VCC’s activity. It is also possible that the tryptophan clamp doesn’t have any additional function and exists only as an evolutionary artifact. In another β-PFT, the
anthrax PA channel, a similar aromatic constriction is formed by phenylalanine residues and these residues are instrumental in translocating two other toxin molecules into the cytoplasm. Additionally, we identified comparable salt bridges next to the W318 constriction. The corresponding bridges in the PA channel are imperative for peptide translocation through the pore. Our efforts have so far been unsuccessful in detecting any translocating partner for VCC, but it is still possible that VCC acts as a translocating channel for unknown toxic molecules.

A comparison of the VCC heptamer and monomer structures illustrates four major structural rearrangements that occur during pore formation. These include the removal of the prodomain, an 180° movement of the β-prism lectin domain, a 35° rotation of the β-trefoil lectin domain and the insertion of the prestem region into the membrane. The prestem region contains a large number of hydrophobic residues, which are hidden in the water-soluble form, but become exposed when the domain rearrangements take place. This drives a spontaneous coherent membrane insertion of seven prestem sequences, which are stabilized by more than double the number of hydrogen bonds than in the monomer structure. Furthermore, an enormous number of hydrophobic interactions, along with a gain in hydrogen bonding, makes the channel extremely stable and drives the pore formation process to be irreversible in nature.

To find the proper membrane surface for pore formation, VCC utilizes cell surface glycans as receptors. VCC contains a β-prism lectin domain that has been proposed to interact with terminal galactosyl oligosaccharides on the membrane (Saha and Banerjee, 1997). We investigated VCC-glycan interactions in detail and
discovered that the VCC β-prism lectin domain exhibits specificity towards N-linked glycans and binds to the core heptasaccharide NGA2 glycan with lower nanomolar affinity. We have performed a thorough investigation of the binding affinities of glycans and fragments of glycans with three β-prism lectin domains. As seen in other lectin systems, the binding reactions are very much enthalpy-driven in nature. A single mannose molecule shows an affinity of lower millimolar, whereas the addition of two more mannose molecules improves the affinities >10-fold for all of the β-prism lectin domains. Surprisingly, the addition of two GlcNAc molecules on the growing arms of the trimannose branch causes a large 1000-fold increase in affinity. However, glycans with more residues on the pentasaccharide core complex result in a loss of activity, showing that the β-prism lectin domains exhibit specific binding against the core pentasaccharide molecule of N-linked glycans.

In addition, we identified the amino acid residues that are important in binding and showed the binding mechanism of oligosaccharides by solving the structure of a paralog of the VCC β-prism lectin domain from another V. cholerae protein. Interestingly, the two paralogs we studied are from the biofilm matrix protein RbmC, and both displayed a similar glycan-binding propensity, but with substantially higher affinity for the glycan molecules. Comparing the structures and sequences of the three β-prism lectin paralogs revealed that there is a five amino acid insertion loop present in the RbmC 2nd β-prism lectin, but not in VCC β-prism lectin or RbmC 1st β-prism lectin. This loop makes crucial interactions with the oligosaccharide. Interestingly, structural predictions suggest that the shortened loop in RbmC 1st lectin is almost
identical in position compared to the VCC β-prism lectin. It is possible that this loop, through its flexibility and interactions with the glycan, creates the disparity of binding affinities between the RbmC and the VCC β-prism lectins.

This nanomolar affinity binding permits the toxin molecule to bind to epithelial cell surfaces tightly. Removal of this β-prism lectin domain from VCC, or removing a key glycan-protein interaction, causes 200- to 1,000-fold loss of activity on the cell surface, but the protein is still able to form a fully mature pore on the bilayer with less potency. Although some other functions of the VCC β-prism lectin domain have been reported, like direct involvement in the membrane pore formation, there is no convincing evidence available in support of these claims.

Other than the glycan binding, the toxin molecules interact with lipid components of the membrane. This interaction occurs through membrane-proximal rim domain loops. Other investigators have demonstrated that VCC prefers cholesterol, sphingolipid and ceramide-rich membranes and is able to form heptameric pores on pure cholesterol microcrystals, but does not exhibit specificity for epicholesterol, an enantiomer of cholesterol (Zitzer et al., 1999; Zitzer et al., 2000; Zitzer et al., 2001; Harris et al., 2002; Ikagi et al., 2006; Krasilnikov et al., 2007). The rim domain loops are located well below the membrane-water interface in the VCC heptamer structure, which is equivalent to the inside of the membrane in an in vivo scenario (unless these loops splay out on the membrane surface). The presence of polar and charged residues will perhaps impede such complete insertion events, but in either case, the three loops are likely to be in close proximity to the membrane. In our
study, we systematically looked at the effect of each residue in these three loops on the hemolytic activity of the toxin and identified ten amino acids that when mutated to alanine cause a >90% loss of hemolytic activity. To ascertain if the loss of activity is due to any misfolding or global aggregation events, we performed several experiments on the mutant proteins. We nullified the possibility of these residues interacting with protein receptors by examining them against a liposomal system lacking cell surface glycans and proteins. Furthermore, we uncovered that these three loops contain several putative cholesterol binding motifs and demonstrated that these motifs follow the same patterns as motifs found in perfringolysin O (PFO) and benzodiazepine receptors (Farrand et al., 2010; Li and Papadopoulos, 1998; Epand, 2006; Epand, 2010). The first type of motif is the TL-type motif discovered in another group of pore-forming toxins, the cholesterol-dependent cytolysins. Interestingly, the second motifs (CRAC or inverse CARC motifs) are widespread in various eukaryotic cholesterol-interacting proteins. We speculate that these cholesterol-interacting motifs act as individual modules that work together to bind to the membrane. Mutating these motifs obstructs the primary membrane binding step, as confirmed by SPR and competition lysis assays. We argue that the rim domain mutations cause a lower monomer concentration on the membrane and thereby impede the pore formation process. Upon increasing amounts of toxin in the system, we can compensate for the loss of membrane binding and form similar heptameric pores. In our hands, we did not encounter any evidence for an abortive pre-pore stage suggested by one study (Rai and Chattopadhyay, 2015) and additionally we found clear proof of the additive
and independent nature of glycan-dependent and direct membrane binding events, which also refute the same groups claims. It is worth mentioning here that there is an elementary difference between the cholesterol preference of CDCs and VCC. The CDCs show a strict requirement for a high concentration of cholesterol in membranes before they show any activity. But, in VCC the relationship is linear between the cholesterol content of the membrane and the pore-formation activity of VCC. This dissimilarity signifies that the CDCs interact on the so-called free or flyaway cholesterol molecules, which can only occur after satisfying all potential phospholipid-cholesterol interaction surfaces. Contrary to this, VCC interacts with cholesterol or cholesterol-rich lipid phases directly and does not depend on free cholesterol to bind to the membrane. Generally, after satisfying all of the phospholipid-cholesterol binding potential, cholesterol molecules are left with the possibility of interacting with each other, forming heterogeneous membrane components where the local concentration of cholesterol stays abnormally high. Evidence suggests that this type of cholesterol-rich rafts are important for various cellular and membrane processes.

This study builds on previous structural and biochemical knowledge regarding VCC. Previously, the monomer structure and heptameric end points of assembly of VCC were known. Lower resolution EM studies demonstrated the overall structure of the oligomer, but no precise information was available. Our 2.9Å heptamer structure not only displays a clearer image of the end point of assembly, but also provides us with a precise account of the domain rearrangements during the process. The
investigation of membrane interactions through the β-prism lectin domain and the rim-domain uncover the mechanism by which these sequences bind to the membrane or membrane associated components. The β-prism lectin domain targets N-linked heptasaccharide core sequences and the rim-domain interacts with the cholesterol-rich membrane. We additionally discovered the residues involved in these binding processes and established that these two interactions are independent and additive in nature. In short, our investigation provides a well-defined insight into the construction of the oligomeric pore and helps us to portray an almost complete functional map of the cell surface binding by VCC through glycans and lipid components of the membrane.

Future directions

This report outlines three steps of the vastly complicated multistep process of VCC’s activity. Our investigations have uncovered molecular details about endpoints of the process. However, further studies will be needed to understand the intermediate steps in assembly. The first aspect for future examination could be the pre-pore stage or an intermediate in the pore formation process. We successfully isolated and purified pre-pore complexes, but were unable to solve the structure of these proteins. Future studies using additional mutants arresting the assembly at this stage may prove useful in deciphering the structure.
The next area for future investigation concerns VCC interactions with the lipid membrane. In our study, we demonstrated that VCC uses its rim domain to interact with cholesterol-rich membranes. However, many aspects of the VCC-membrane interaction are still unknown. It will be interesting to learn if VCC directly interacts with cholesterol or if the cholesterol dependence is through an indirect mechanism. Another interesting question is the exact effect of membrane asymmetry on VCC’s activity. Some studies have reported that the presence of cholesterol on both leaflets of the bilayer is necessary for pore formation to occur. Precise effects of the membrane asymmetry in terms of the cholesterol content and lipid composition are yet to be uncovered. These studies will involve planar asymmetric bilayers with variable composition and careful measurements using either electrophysiology or single molecule fluorescence.

The exact mechanism of membrane insertion and the kinetics of each step are also not well-understood. In order to address these questions, properly designed, separate mutational and single molecule studies will be needed. These studies, together with the already published results, will enable us not only to understand VCC’s mechanism of action, but also provide general knowledge about the membrane insertion and interaction processes.


cytolysin by disulfide trapping synchronizes the insertion of the transmembrane beta-sheet from a prepore intermediate. *J Biol Chem*, 276(11), 8261-8268. doi: 10.1074/jbc.M009865200


Li, J. D., Carroll, J., & Ellar, D. J. (1991). Crystal structure of insecticidal delta-endotoxin from Bacillus thuringiensis at 2.5 A resolution. *Nature, 353*(6347), 815-821. doi: 10.1038/353815a0


Queen, J., & Satchell, K. J. (2012). Neutrophils are essential for containment of Vibrio cholerae to the intestine during the proinflammatory phase of infection. *Infect Immun*, 80(8), 2905-2913. doi: 10.1128/IAI.00356-12


using the time derivative of the concentration profile. Anal Biochem, 203(2), 295-301.


gamma-hemolysin reveals the beta-barrel pore formation mechanism by two components. *Proc Natl Acad Sci U S A*, 108(42), 17314-17319. doi: 10.1073/pnas.1110402108


