2013-01-01

Expression And Characterization Of The Major Capsid Protein (MCP) Of A Giant Marine Virus: Cafeteria Roenbergensis Virus

Sayan Chakraborty
University of Texas at El Paso, sayanimmuno@gmail.com

EXPRESSION AND CHARACTERIZATION OF THE MAJOR CAPSID PROTEIN (MCP) OF A GIANT MARINE VIRUS: *Cafeteria roenbergensis* VIRUS (CroV)

SAYAN CHAKRABORTY

Department of Chemistry

APPROVED:

Chuan Xiao, Ph.D., Chair

Jorge Gardea-Torresday, Ph.D.

Mahesh Narayan, Ph.D.

German Rosas-Acosta, Ph.D.

Benjamin C. Flores, Ph.D.
Dean of the Graduate School

Dedication

My parents, Mr. Somnath Chakraborty and Mrs. Mitali Chakraborty

EXPRESSION AND CHARACTERIZATION OF MAJOR CAPSID PROTEIN (MCP) OF A GIANT MARINE VIRUS: *Cafeteria roenbergensis* VIRUS (CroV)

by

SAYAN CHAKROBORTY, B. Tech

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Chemistry

THE UNIVERSITY OF TEXAS AT EL PASO

December 2013

ACKNOWLEDGEMENTS

This research was entirely carried out at the Department of Chemistry, University of El Paso, Texas, USA. I would like to thank my mentor Dr. Chuan Xiao, aka, River (Assistant Professor, Chemistry Department at UTEP) for his help. He has always encouraged me to think critically and advised me to find a new approach for doing experiments. He taught me to keep an open-minded attitude for novel ideas. He was always there to help me with his enormous expertise. I want to thank him once again for leading me to the fascinating world of structural biology. River has also spent tremendous effort to improve the quality of this thesis.

I would like to thank the head of the Department of Chemistry, Professor Jorge Gardea-Torresday for providing a creative and enthusiastic environment for research. I want to express my sincerest thank to him for letting us use his laboratory instruments. I want to thank Dr. Mahesh Narayan (Assistant Chair, Chemistry Department) for helping me with his useful insights and providing all inclusive access to his laboratory. I want to thank Dr. German Rosas-Acosta for reviewing my thesis and fruitful comments on the preparation of this thesis. I want to thank my thesis committee members Dr. Jorge Gardea Torresday, Dr. Mahesh Narayan and Dr. German Rosas Acosta for giving valuable ideas and comments on my studies.

I would like to thank Dr. Louis Echegoyen and Danisha Rivera Nazario, graduate student in his laboratory for giving me an opportunity to carry out some important experiments in their laboratory. I would also like to thank Dr. Jianjun Sun from Department of Biological Sciences for providing us the expression vectors and offering suggestions. I also want to thank Dr. Curtis Suttle at University of British Columbia, Vancouver, Canada and Dr. Matthias Fischer at Max Planck Institute for Medical Research, Heidelberg, Germany for helping us to set up the marine

virus cultures. I also thank the Border Biomedical Research Center (BBRC) providing services for centrifugations and DNA sequencing.

I would like to thank all of the former and present lab members of my group including Gustavo Avila, Duer Bolotaulo, Scott Rush, Jorge Lopez, Emmanuel Silva, Ricardo Parra, Martin Chacon, Lu Xiao, Jasmine Peralta, Christina Alvara, Jennie Choi, David Acosta and Mason Arbogast. Our group is known as 'Eagles', as our mentor River once stated that we should try to be eagles not turkeys. 'Eagles' I enjoyed your company in the laboratory and outside the lab. I want to thank Gustavo for his benevolent nature and always ready to help attitude. Scott, Duer, Emmanuel, Ricardo, Jasmine, Martin, Mason, Christina and Jennie, thank you for your friendship and support. Friendship is a precious gift and I consider myself lucky to have you all as my friends. Finally, I want to thank all the people at the Department of Chemistry for providing a great working atmosphere.

My friends, relatives, family, and students who have given balance to my life outside the laboratory. Special thanks to friends from the childhood, my undergraduate days and UTEP. I want to thank my parents, Mr. Somnath Chakraborty and Mrs. Mitali Chakraborty, for supporting me and believing in me in everything I have decided to do.

ABSTRACT

The oceans play significant roles in maintaining the climate on Earth and providing nutrients to the whole biosphere. Oceanic microbes produce approximately half of the Earth's oxygen. It has been shown that the populations of microbial communities in oceans are largely regulated by viral infection and thus affect the marine ecosystem. *Cafeteria roenbergensis* Virus (CroV) is a giant marine virus with a linear, double stranded, and AT-rich DNA genome. The genome of CroV has a size of 744kb and encodes for 544 predicted genes. CroV derived its name from its host, a unicellular marine zooplankton *Cafeteria roenbergensis* (Cro) that is the major microbial grazer in the ocean. By controlling Cro's population, CroV has a great impact on the marine ecological system. In addition, CroV is a member of the nucleo-cytoplasmic large DNA virus (NCLDV) clade. The presence of genes in NCLDVs encoding enzymes normally only required by cellular organisms for DNA replication, DNA repair, transcription, and translation have stimulated debate over the role played by giant viruses in the evolution of cells. However, despite their ecological and evolutionary importance, limited studies have been reported on the structure of giant marine viruses and their life cycles.

CroV is covered by a protein shell (capsid) comprised of major capsid proteins (MCP) and minor capsid proteins (mCP). The capsid not only protects the genetic material of the virus, but also helps in host-virus recognition, which is the critical initial step for viral infection. This research focuses on obtaining large amount homogenous CroV MCP for future structural studies that will facilitate the understanding of the roles of MCP in the viral life cycle. CroV MCP gene was codon optimized, synthesized, and amplified by PCR followed by cloning into a bacterial expression system in order to obtain purified protein of high homogeneity. Initial trials in the

expression vector without protein chaperone failed to obtain soluble MCP. By fusing trigger factor (TF), a bacterial cold shock chaperone, to the N-terminal of the CroV MCP, we obtained soluble TF-MCP. TF facilitates the proper folding of proteins at low temperature in bacteria. After large-scale expression in bacteria, ultra-high purity TF-MCP fusion protein was obtained by affinity chromatography and size exclusion chromatography. Mass spectrum (MS) and dynamic light scattering (DLS) were performed to analyze the molecular weight and hydrodynamic diameter of the fusion protein. Homologous sequence comparison and structural modeling indicate CroV MCP should have a trimeric double-jelly-roll structure as seen in PBCV-1, another giant NCLDV. The MS and DLS results suggested that pure TF-MCP fusion protein is in trimeric form, which indicates the proper folding of the MCP. This has built a strong foundation for the next step; to crystallize MCP and determining its atomic structure. The crystal structure of MCP will provide insights about the virus-host interaction as well as viral infection strategy and genome delivery mechanism.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION TO THE WORLD OF GIANT VIRUS

1.1 Short history of viruses

Viruses are infectious agents that can only multiply inside a living host cell. They can infect all types of living cellular organisms, ranging from multicellular animals and plants to single cellular bacteria and archaea (Fields, Knipe et al. 2007). In 1890s, Dimitrie Ivanowski and Martinus Beijerinck reported independently a pathogenic agent,(Ivanofsky 1892, Beijerinck 1898) not visible under light microscope, that infects tobacco and can pass through the Chamberland filters designed to trap all known bacteria. Using the same methods, Friedrich Loeffler and Paul Frosch in 1898 discovered that mammalian foot and mouth disease causing agents can also be filtered (Loeffler and Frosch 1989). In the year 1908, Landsteiner and Popper successfully transferred Poliomyelitis from a human sample to monkey in the form of a filterable agent (Landsteiner and E. 1909). The experiments mentioned above has led to the first definition of viruses as filterable pathogens that can pass 0.2μm filters (Landsteiner and E. 1909).

The investigation about virus structure and morphology was accelerated with several major scientific discoveries. In 1931, German scientists Ernst Ruska and Max Knoll developed electron microscopy and later was used to captured the first images of viruses (Kausche, Pfankuch et al. 1939). In 1935, American virologist, Wendell Stanley found that TMV was made up of protein (Stanley 1935). Later, it was demonstrated that viruses are mainly composed of protein and nucleic acid while some have lipids. A mature virus particle known as a virion consists of genetic material in the form of DNA or RNA that is protected by a protein shell (capsid) or an envelope (Fields, Knipe et al. 2007).

It has been found that viruses depict a wide variety of morphologies, based on which they can be classified into three groups: helical, icosahedral, and enveloped (Fields, Knipe et al. 2007, Harrison 2007). Viruses can also be classified according to 1) the nature of genetic material they possess 2) the presence or absence of envelope and 3) the presence or absence of nucleocapsid. For instance, Nobel Prize laureate David Baltimore introduced the classification of viruses (Baltimore 1971) based on their respective genetic material [Table1.1].

1.2 Discovery of giant viruses

As aforementioned, viruses are defined historically as small pathogenic agents that can pass through a 0.2µm filter. Recently, this definition has been challenged after the giant Mimivirus was discovered in 1992 (La Scola, Audic et al. 2003). Large viruses were previously undetected because of the traditional virus isolation process. The classical virus isolation process involves filtration by 0.2µm filters, but these giant viruses cannot pass through the pores and stay with cellular organisms. That is why the giant Mimivirus was first thought to be a bacterium (La Scola, Audic et al. 2003, Van Etten, Lane et al. 2010, Van Etten 2011). Apart from Mimivirus, there are few more giant viruses that have been reported in the literature, e.g., *Cafetreia roenbergensis* virus (CroV) (Fischer, Allen et al. 2010), Mamavirus (Colson, Yutin et al. 2011), Megavirus (Legendre, Arslan et al. 2012), and most recently Pandoravirus (Philippe, Legendre et al. 2013).

All these giant viruses have physical dimensions ranging from 0.2µm to 1µm with a vast genome encoding for hundreds of proteins (Van Etten, Lane et al. 2010). Many of these giant viruses inhabit aquatic environments and infect plankton (Van Etten, Lane et al. 2010). For instance, Mimivirus, which infects amoebae *Acanthamoeba polyphaga*, has a genome size of approximately 1.2 Mb and codes for an estimated 979 proteins (Fischer, Allen et al. 2010,

Colson, Gimenez et al. 2011). CroV, which we are interested in, is considered to be one of the largest marine viruses discovered to date (Fischer, Allen et al. 2010). It has a size of approximately 300nm in diameter and has a genome size of 730kb that encodes for about 544 proteins (Fischer, Allen et al. 2010, Van Etten, Lane et al. 2010, Colson, Gimenez et al. 2011).

1.3 Background and significance

CroV was first discovered in the coastal areas of Texas in the 1990's (Garza and Suttle 1995, Fischer, Allen et al. 2010). It was named after its host, *Cafeteria roenbergensis* (Cro) [Figure1.1 A], a single cellular marine flagellate that feeds on other marine microorganisms (Fischer, Allen et al. 2010). Cro is a kidney shaped biflagellate marine protist with a size of 2 to 6 µm [Figure1.1 A] (Fischer, Allen et al. 2010). Due to its widespread distribution in all oceans, Cro plays a critical role in marine ecosystems by controlling biodiversity in the microorganism community. By infecting its host [Figure 1.1 B], CroV influences ocean biodiversity and accelerates nutrient recycling [Section 1.3.1] (Fischer, Allen et al. 2010). Additionally, the discovery of these giant viruses like CroV has challenged the definition of viruses and raised questions about their evolutionary relationship with other cellular organisms [Section 1.3.2]. The ecological and evolutionary significance of CroV are the major driving forces towards pursuing this research.

1.3.1 Marine virus –impact on ecosystem and biogeochemistry

Viruses are the most abundant biological entities in the oceans that infect a variety of hosts ranging from microorganisms to marine mammals [Section 1.3.1.2] (Suttle 2007). Additionally, it is suggested that viruses play a crucial role in nutrient, food cycles, and fixation of important elements (Fuhrman 1999) [Section 1.3.1.1]. Furthermore, marine viruses control the

population and species distribution [Section 1.3.1.2]. Hence, marine viruses not only control the flow of energy and nutrients but also maintain ocean biodiversity.

1.3.1.1 Virus abundance and nutrition recycling in ocean

Oceans cover 70% of the earth and play an important role in maintaining climate and ecological balance (Suttle 2007). Microorganisms are the major driving force for ocean nutrient cycle, producing 90% of the biomass in the ocean (Suttle 2007). Studies have illustrated that viruses are the most abundant biological entities in ocean (Bergh, Borsheim et al. 1989, Hara, Terauchi et al. 1991). Viral abundance is 10^10 per liter of ocean surface water which is almost 5-25 times the bacterial abundance (Cochlan, Wikner et al. 1993, Hara, Koike et al. 1996) . It is estimated that there are approximately 10^30 viruses present in oceans and around 20% of biomass produced by microorganisms is killed by viruses every day with an infection rate of 10^23 per second (Suttle 2007). Viral lytic infections convert cells into cellular debris, which is made up of simple elements, colloids, and cell fragments that are technically defined as dissolved organic matter (Middelboe and Lyck 2002). In addition, marine viruses convert materials from living entities to particulate and dissolved organic substances to CO_2 (Fuhrman 1999). It has been reported that viruses can convert microbial biomass into dissolved organic matter by cell lysis within the *Photic Zone* (the area of the ocean where light penetrates) and can export more CO_2 out of the photic zone (Lawrence and Suttle 2004). Cell lysis also releases important cellular building blocks like amino acids, nucleic acids, which can be easily incorporated by living organisms (Middelboe and Jorgensen 2006). The release of organic materials from cells via viral lysis influences the amount of carbon exported to the deep sea by biological pump (Fuhrman 1999). Viral lysis of prokaryotes regenerate sufficient amount of biologically available iron that supports the need of phytoplankton (Poorvin, Rinta-Kanto et al. 2004). Most of these lysis

products eventually become available to other microbes. Therefore, viruses play a crucial role in marine ecosystems by accelerating nutrient recirculation in the ocean.

1.3.1.2 Biodiversity regulated by marine viruses

Diverse prokaryotic communities are found in ocean (Thompson, Pacocha et al. 2005). One study shows that host specific and strain-specific viral infections control the composition of the microbial communities (Weinbauer and Rassoulzadegan 2004). For example, studies have revealed bacterial diversity is strictly controlled by bacteriophages (viruses that infect bacteria) (Rodriguez-Valera, Martin-Cuadrado et al. 2009). It is known that viruses prevent the dominance of a single species on earth by a mechanism known as "killing the winner" (Murray and Jackson 1992). Viral infections depend upon the population of the host species. The host with larger population (winner) will have higher chances of viral infection than those with smaller populations. Therefore, the "winning" species will be significantly reduced or killed by epidemic viral infection.

Protozoans residing at a higher level of the marine food chain control the abundance and species composition of prokaryotes (Pedros-Alio, Calderon-Paz et al. 2000). Therefore, viruses that infect protozoans influence the biodiversity in the ocean. *Cafeteria roenbergensis* (Cro) is considered to be the major grazer in the ocean as it feeds upon smaller microbes and is found in the coastal areas of all the oceans (Fischer, Allen et al. 2010). The "killing the winner" model can be best described by the studies on protistan grazers. For example, when the population of Cro increases, the high level of virus mediated infection of the host can be inferred by observing high proportions of visibly infected cells (Brussaard 2004, Martinez, Schroeder et al. 2007, Massana, del Campo et al. 2007). Therefore, viruses are significantly responsible for both the mortality of the marine microbes and the maintenance of the ecological balance in the aquatic

communities. Thus, it is of great importance to understand the underlying mechanisms of virus-host interactions in marine ecosystems. However, the detailed mechanism of marine viruses regulating biodiversity is still elusive due to the lack of knowledge about the marine viruses and their interaction with their hosts.

1.3.2 Giant viruses –impact on evolution

After the discovery of Mimivirus, more giant viruses were discovered in a continuous fashion. Scientists classified these giant viruses (giruses) into a clade known as, Nucleo-Cytoplasmic Large DNA Virus (NCLDV) (Van Etten, Lane et al. 2010). As the name of the group suggests, these viruses are large in physical size, some of which reach to that of small bacteria. This has defied the historical definition of virus as small pathogen that can pass the 0.22 μm filter. More interestingly, NCLDVs contain complex genomes that encode for a variety of "cellular" proteins homologous to those of eukaryotic, prokaryotic and archaeal origin (Van Etten, Lane et al. 2010). It is believed that NCLDV evolution produced a set of 45 common genes and it is suggested that these 45 genes are the minimum genome size of NCLDV common ancestor (Iyer, Balaji et al. 2006). NCLDV is not a taxonomical family but a loosely defined clade that includes viruses from viral families of *Mimivirirdae, Phycodnaviridae, Asfarviridae, Poxviridae, Iridoviridae* and *Ascoviridae* (Koonin and Dolja 2006) [Table 1.2]. For instance, CroV belongs to the *Mimiviridae* family (Yutin and Koonin 2012) along with other giant viruses such as PpV-01, CeV-01, APMV, and ACMV (Colson, Gimenez et al. 2011). Like other NCLDV members, the genome of CroV also contains genes with significant similarities to those found in eukaryotes and prokaryotes (Fischer, Allen et al. 2010).

1.4. Role of NCLDV in evolution

In the literature, it is widely accepted that all NCLDVs are originated from a common ancestor (Iyer, Balaji et al. 2006). However, controversies still exist regarding the morphology and size of the ancestor (Van Etten 2011, Yutin and Koonin 2012). The theory of "gene degeneration" suggests that the common ancestor of NCLDV was a giant virus or cellular organism that evolved via losing unnecessary genes (Forterre 2010, Van Etten 2011). On the other hand, the theory of "gene robbery", suggests that the first NCLDV was a small DNA virus that kept acquiring genes from its host cellular organisms (Yutin, Wolf et al. 2009). For instance, EhV (*Emiliania huxleyi* Virus) possesses the genes encoding a pathway of lipid synthesis (Van Etten 2011). If "gene robbery" theory is correct, we can predict that Horizontal Gene Transfer (HGT) might have played an important role in the evolution of NCLDV (Vogan and Higgs 2011; Moreira and Brochier-Armanet 2008). Nevertheless, the origin of NCLDV is still a mystery but might be very important for understanding the evolution of life as a whole, which stimulate the debate about the "the tree of life" (Zakaib 2011). Some scientists suggested that NCLDV should be classified as the fourth kingdom of life based on different scientific findings, e.g., analyzing the gene sequences, protein folding, and protein motifs (Raoult, Audic et al. 2004, Raoult and Forterre 2008). Ancient protein folds found in giruses suggest that they arrived early in evolution, which further supports the idea of "Fourth kingdom". Ludmir and Enquist (2009) showed that 66% of protein coding genes (CDs) of Mimivirus genome have no similarities to any known proteins. To explain this fact, Balaji et al. (2006) proposed a hypothesis, which states that some NCLDV genes arose from the original gene pool that led to the evolution of prokaryotes and eukaryotes. This hypothesis supports the idea that giruses, such as, Mimivirus may have evolved at an early stage of evolution (Raoult, Audic et al. 2004, Moreira and

Brochier-Armanet 2008, Raoult and Forterre 2008). However, other scientists still consider the three kingdoms of life as more appropriate with bacteria, archaea, and eukaryotes in three different parts without considering viruses as an integral part of life (Moreira and Brochier-Armanet 2008). Therefore, understanding the evolutionary relationship between cellular organisms and giant viruses will provide insight into the origin of life.

Majority of the NCLDVs replicate their viral genomes in the cytoplasm with the exception of Phycodnavirus that replicate their genome in the nucleus (Monier, Pagarete et al. 2009). While replicating in the cytoplasm, these giruses form a special compartment known as "virion factory" (Van Etten 2011) (Figure 1.1). The virion factory is where the viral genomes get replicated independent of the nucleus. Thus the giruses must have their own DNA transcribing machinery (Van Etten 2011). For instance, CroV contains DNA dependent DNA polymerase, machinery required for transcription (initiation, elongation and termination), several helicases, and mRNA capping enzymes (Fischer, Allen et al. 2010). The functional similarities between the virion factory and nucleus led to the hypothesis related to the evolutionary origin of the eukaryotic nucleus. A possible scheme is that an early prokaryotic cell was infected by a girus and the respective virion factory evolved into a nucleus (Forterre 2011).

NCLDVs' genomes encode many genes previously only found in cellular organisms, such as, genes for tRNA, enzymes for synthesizing sugars and lipids, and DNA repair enzymes. The CroV genome contains a 38kb genomic fragment encoding for biosynthesis of 3-deoxy-D-manno-octulosonate. Other giruses encode enzymes essential for sugar synthesis like fucose and rhamnose (Chothi, Duncan et al. 2010). Phycodnavirus EhV encodes enzymes for the synthesis of sphingolipids (Monier, Pagarete et al. 2009). Some giruses contain genes for encoding hyaluronan synthase that is involved in the formation of extracellular matrix (Graves, Burbank et

al. 1999). Hyaluronan was previously only found in vertebrates and bacteria. Some chlorella viruses encode for chitin and polyamine biosynthesis (Kawasaki, Tanaka et al. 2002). CroV contains genes for photolyases (Fischer, Allen et al. 2010), the enzyme necessary to repair DNA damage by UV radiation while some chlorella viruses contain DNA restriction endonucleases (Van Etten 2003). Enzymes responsible for ubiquination are also found in some of the giant viruses. The Chlorella virus encodes for small proteins, such as, histone methyl transferase, type II DNA topoisomerase, and potassium ion channel protein, that might be regarded as an evolutionary pre-cursor of more complex contemporary proteins (Lavrukhin, Fortune et al. 2000).

The viral factory resembles the eukaryotic nucleus and the presence of cellular genes lead to the conclusion that NCLDVs might have a long evolutionary relationship with cellular organisms. More research on NCLDVs might resolve the mysteries of the origin and their contribution to the evolution of cellular organisms.

1.4.1 CroV and its evolutionary relations with other NCLDV viruses and their hosts

CroV has a linear, double stranded AT-rich DNA genome about 740kb long with a 618kb coding region that encodes for 544 predicted genes (Fischer, Allen et al. 2010). Around one-half of the CroV genes display significant similarities to proteins found in eukaryotes, bacteria, archaea, and other giant viruses (Fischer, Allen et al. 2010) . CroV encodes an isoleucyl tRNA synthetase and proteins that resemble eukaryotic translational initiation factors (eIF) as eIF-1, eIF-2α, eIF-2β/eIF-5, eIF-2γ, eIF-4AIII, eIF-4E, and eIF-5B (Fischer, Allen et al. 2010). CroV alos encodes a couple of DNA repair proteins, such as, photolyase and DNA mismatch proteins because of the probability of forming DNA lesions in the AT rich genome as it is exposed to solar radiation on the surface of the water (Fischer, Allen et al. 2010). Moreover, CroV possess

DNA dependent DNA polymerase, transcriptions initiation system machinery, translation elongation and termination machinery, several helicases, and mRNA capping enzymes (Fischer, Allen et al. 2010). Three different topoisomerase genes of types IA, IB, and IIA, topoIA and topoIIA are also found in CroV (Fischer, Allen et al. 2010). All of these discoveries raised the question whether viral gene transcription and translation are independent of the host cell machinery.

In addition to gene transcription and protein translation, post translational modification enzymes were also found in the CroV genome. A set of proteins found in CroV are predicted to be useful in the ubiquitin (Ub) pathway (Fischer, Allen et al. 2010). In a cellular environment, proteins modified by ubiquitin are targeted for protein degradation. The CroV genome contains two of the three enzymes for ubiquitination, E1 Ub-activating enzyme, six E2 Ub conjugating enzymes. In addition, CroV also has two deubiquinating enzymes and one Ub gene (Fischer, Allen et al. 2010). This suggests that CroV may use this ubiquitin pathway to degrade the host proteins for replication.

As a member of NCLDV family, CroV shares genetic similarities with other members. Apart from that, some of the genes and proteins are exclusively found only in CroV. Among the 47 NCLDVs core genes, 13 of which are absent in Marseillevirus, are present in CroV. In comparison to Mimivirus, a RNA ligase and a dUTPase are only present in CroV. Among the 83 selected proteins that are common between Mimivirus and CroV, 47 of them are located in Mimivirus and CroV at different genomic positions. It is also important to mention that some proteins such as a ribosome associated chaperone zuotin, two histone acetyltransferases and a deoxyribo dipyrimidine photolyase were specifically found in CroV but not in any of the NCLDV virus families (Fischer, Allen et al. 2010). Among the 83 proteins mentioned above that

are present in CroV, it is found that 16, 1, 33, and 12 CroV proteins have bacterial, archaeal, eukaryotic, and viral homologs respectively (Raoult and Forterre 2008, Van Etten, Lane et al. 2010). Some part of the amino acid sequence of CroV MCP was found homologous to the genome of *Acanthamoeba castellanii* (Raoult and Forterre 2008), the host of Mimivirus, Megavirus and Pandoravirus. Given that the genome of CroV host is unavailable, comparative genomic analyses between Cro and CroV cannot be performed to detect the gene transfer between the host and the virus (Colson, Gimenez et al. 2011). However, the sequence similarity between CroV and *A. castellanii* suggests that a phagocytic protist may have served as an ancestral host of CroV, which in the course of evolution has been specialized to infect marine flagellates (Colson, Gimenez et al. 2011). All these discoveries concerning the differences in aforementioned genome and proteins among NCLDVs raised debates about the origin of these genes: are they from degeneration of a common viral ancestor with a large gene pool or are they coming from horizontal gene transfer from their hosts (Vogan and Higgs 2011)?

1.5 Summary

In this chapter, we first discussed the discovery of viruses and their early definition. After the recent discovery of giant viruses, the scientific definition about viruses got challenged. The introduction of the NCLDV clade has generated intensive debates about the origin and evolution of these giruses. NCLDVs contain a vast array of genes similar to that of eukaryotes, prokaryotes and archaea. This also leads to the suggestion to add viruses as a part of the phylogenetic tree of life. CroV, a member of NCLDV group, was selected as a research target not only because of its complex genome but also because of its impact on the oceans ecosystem. As a marine virus, CroV plays important role in the marine food web and oceanic biodiversity by killing a marine flagellated grazer in the ocean that controls the bacterial population. Structural and functional

studies will enhance our knowledge of CroV's life cycle such as viral infection and multiplication inside the host. The current research is in its initial stage of understanding the structure of CroV. It will build the foundation for future structural comparison among NCLDVs that will provide insights on the origin and evolution of giruses and their evolutionary relationships with cellular hosts. This research on a giant marine virus CroV will also pave the way to study the structures of other marine viruses. Future study of more marine viruses will provide insights on their roles in marine ecosystem and ocean biodiversity.

Tables and Figures:

Table 1.1 Classification of virus according to their genetic material

Genetic Material	Example of viruses
Ds DNA	Adenovirus, NCLDV
ss DNA	Parvovirus
ds RNA	Reovirus
+ss RNA	Picronavirus
-ss RNA	Rhabdovirus
ss RNA RT	Retrovirus
ds DNA RT	Hepadnavirus

Table 1.2 Classification of NCLDV and their characteristics (Colson, Gimenez et al. 2011)

Virus family	Member	Genome		Host Kingdom
		Size(kb)	CDs	
Mimiviridae	Mimivirus	1181	911	Protozoa
	Mamavirus	~1200	-	Protozoa
	CroV	730	544	Chromista
Myoviride	Phage 201Φ 2-1	317	461	Proteobacteria
	Phage G	670	-	Fermecutes
Nimaviridae	WSSV1	305	531	Animalia
Hycodnaviridae	PBCV-1	331	366	Plantae
	NY2A	369	404	Plantae
	AR158	345	360	Plantae
	FR483	321	335	Plantae
	MT325	321	331	Plantae
	EsV-1	336	240	Plantae
	EhV-86	407	472	Chormista
	CeV-01	~510	-	Chromista
	PoV-01	560	-	Plantae
	PgV-group1	466	-	Chromista
	PpV-01	485	-	Chromista
	HcDNAV	356	-	Protozoa
	EfasV	340	-	Chromista
	MclaV	340	-	Chromista
Poxviridae	CNPV	360	328	Animalia
Marseilleviridae	Marseillevirus	368	457	Protozoa

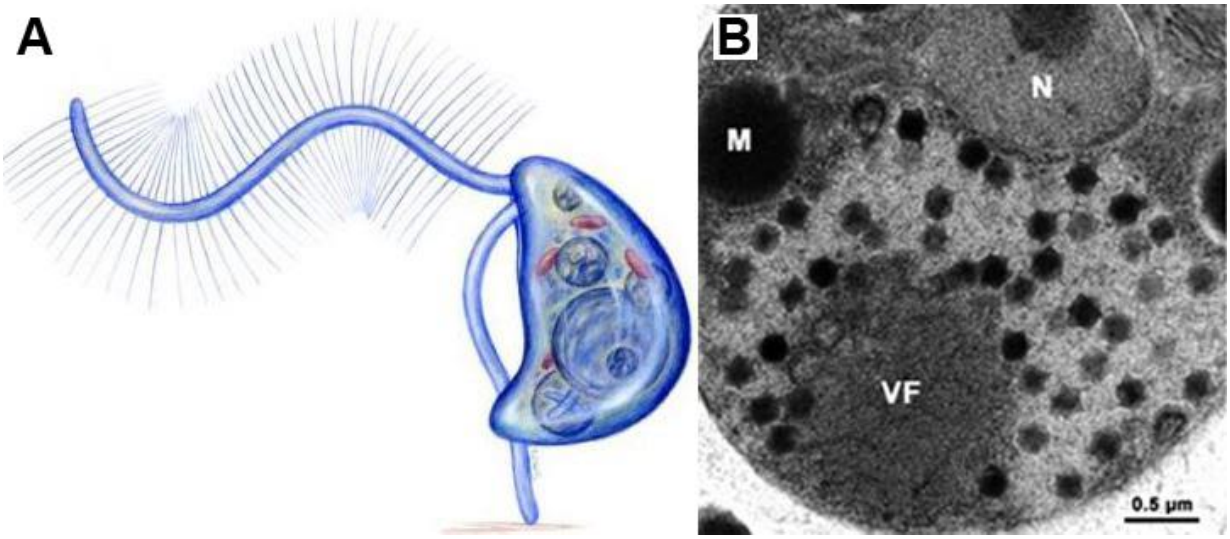

Figure 1.1 *Cafetreia roenbergensis* (Cro) and *Cafeteria roenbergensis* virus (CroV) **A.** a schematic diagram of Cro adapted from encyclopedia of life (http://www.eol.org). **B.** TEM picture of Cro infected by CroV, small black structures are viruses. Virion factory, mitochondria and nucleus are labeled by VF, M and N, respectively. This picture is adapted from (Fischer 2012).

CHAPTER 2

EXPERIMENTAL DESIGN AND METHODS

2.1 Research objectives

The **long-term goal** of this research is to determine the atomic structure of the Major Capsid Protein (MCP) of CroV. Structural elucidation of the MCP will pave the way to generate a pseudo atomic structure of CroV by combining the results from Cryo-Electron Microscopy (Cryo-EM) reconstruction (not included in this thesis). The pseudo atomic structure will provide insights about the host virus interaction, viral assembly pathway, and virus replication cycle inside the host. Due to the time constrains, we only focused on obtaining pure MCP and characterizing the recombinant protein. Two specific objectives have been proposed below:

Objective 1: Expression and purification of MCP

Objective 2: Characterization of MCP by biochemical and biophysical techniques

Objective 1: Expression and purification of MCP

In order to conduct the structural study of CroV, a significant part of this thesis focuses on obtaining soluble Major Capsid Protein (MCP) of CroV. As aforementioned in Chapter 1, a protein shell known as a capsid covers the outside of the virus. The capsid not only protects the genetic material, but also helps in host-virus recognition, the critical initial step for viral infection. After delivering its genome into the host cell, the virus also needs to produce larger quantities of capsid proteins to reassemble the progeny virions. Therefore, determination of the

structure of MCP will provide useful information about the two most important steps in the viral life cycle: infection and viral assembly.

Two major protein motifs/folds are found in most viral capsids, the jelly-roll fold and the HK97 fold (Harrison 2007). The jelly-roll motif was first discovered in small plant RNA viruses and later in human rhinovirus and poliovirus (Rossmann MG, 1989). Since then, jelly-roll fold has been found in many icosahedral viruses. Recent discoveries have suggested that jelly-roll subunit is present in both icosahedral and non-icosahedral viruses (Johnson R. K., 2011). In small icosahedral viruses, single wedge shaped jelly-roll assemble either pentamerically at the fivefold or hexamerically at the pseudo three fold. The jelly-roll fold is a β barrel, wedge shaped structure composed of two beta sheets, each with four anti-parallel β strands. In NCLDV, two jelly-rolls are tangentially connected as a double-jelly-roll motif. Three of these double-jelly-rolls will combine together to form a capsomer, the minimum visible unit for a viral capsid [Figure 2.1]. These capsomeres then come close to assemble the entire virus capsid [Figure 2.2]. These double-jelly-rolls have also been found in smaller dsDNA viruses such as bacteriophages and human adenovirus (Bamford, Grimes, & Stuart, 2005).

The only available structure of MCP from NCLDV is the structure of gene product Vp54 of PBCV-1, another giant algae virus. The capsomer of PBCV-1 is 75Å thick and a diameter ranging between 74 Å to 85 Å (Xiao and Rossmann 2011). Based on the sequence homology, it is predicted that CroV should also have a similar double-jelly-roll structure found in other NCLDV viruses. Sequence homology has shown significant similarities in β strands among NCLDV MCP, but different numbers of amino acid linking strands forming a large insertion loops in strands D and E as well as strands between F and G. In adenovirus these large loop regions are called “Towers” (Athappilly, Murali et al. 1994). Similarly, NCLDV also have large

insertions that appear as towers on the viral surface. These towers differ from one NCLDV virus to another (Xiao and Rossmann 2011). It is presumed that these towers play a crucial role in host-virus recognition. Solving the structure of MCP will also improve the understanding of the function of the towers and how the MCP helps in the viral genome delivery and infection [Figure 2.1; unpublished data from Dr. Chuan Xiao].

Simple icosahedral arrangements of capsomeres in viruses were first described by Watson and Crick in 1956 (Crick and Watson 1956). In 1962, Casper and Klug introduced a mathematical concept known as the "Triangulation (T) number" to quantitatively describe the arrangement of capsomer in icosahedral viruses (Caspar and Klug 1962). The T number measures the number of monomers (triangular area e.g. jelly-roll) in one icosahedral asymmetric unit indicating the size of the virus. According to Casper and Klug (1962), the T number is calculated using the following equation:

$$T= h^2 + hk + k^2$$

Walking from one five fold axis to the closest neighboring five fold, h and k are the two indexes along two axes on a hexagonal array established by a pseudo hexagonal capsomer [Figure 2.2 A, B, C, D]. By using this method, the T number of CroV has been determined to be 499 (h=7, k=18), almost 2 times higher than PpV01 (T= 219) [Figure 2.2 A & B; unpublished data from Dr. Chuan Xiao].

Apart from the Major Capsid Protein (MCP), proteomic studies have also confirmed the presence of at least four minor capsid proteins in CroV [Table 2.1]. It is believed that these minor capsid proteins play a crucial role in stabilizing the virus capsid (Yan, Yu et al. 2009).

In last two and half years, we successfully expressed the CroV MCP protein and established a purification procedure to obtain highly purified CroV MCP. After multiple trials,

the MCP gene of CroV has been successfully amplified and cloned into an expression vector and transformed into bacterial cells. Multiple tests were performed to optimize the optimal conditions for bacterial growth and protein production. Affinity chromatography followed by size exclusion chromatography (SEC) has been utilized to obtain highly purified CroV MCP [Illustration 1].

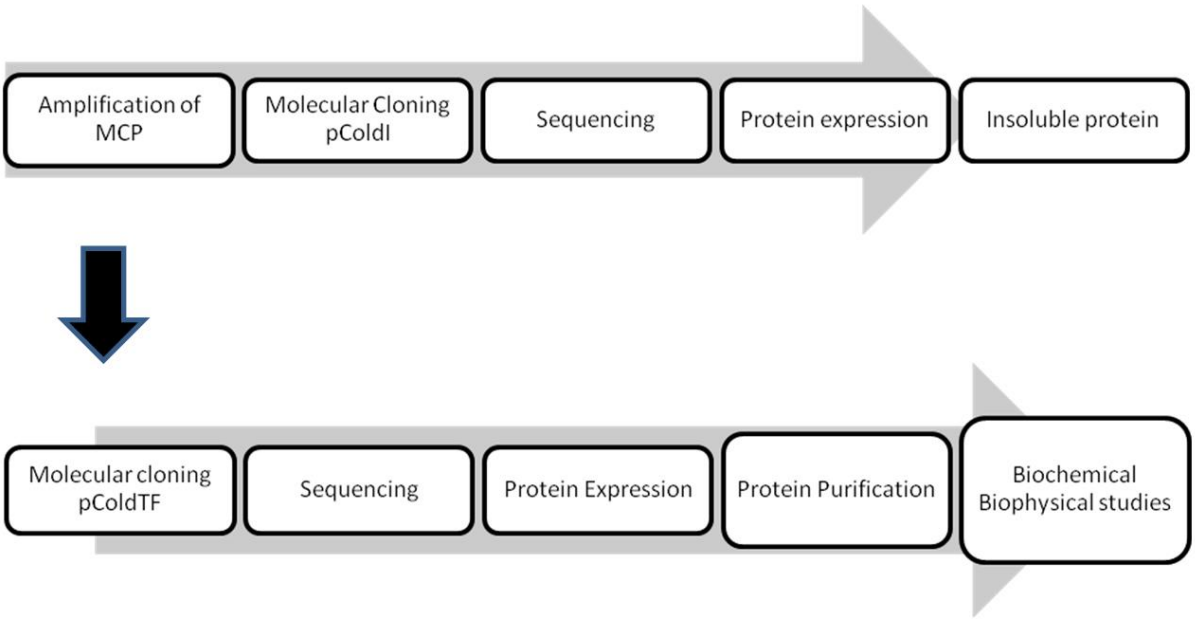

Illustration1. Schematic diagram of experimental design

Objective 2: Characterization of MCP by biochemical and biophysical techniques

Molecular, biochemical, and biophysical tools have been employed to characterize the purified recombinant MCP. These methods include traditional biochemical techniques such as gel electrophoresis and size exclusion chromatography. More advanced methods such as Dynamic Light Scattering (DLS), Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-MS) have been used to determine the size and the molecular weight of

CroV MCP. All the above techniques have been used to confirm that the recombinant CroV MCP is trimeric, which indicates the proper folding of the expressed protein.

2.2 Materials & methods

2.2.1 Molecular cloning and protein expression with pColdI vector

2.2.1.1 Selection of host organism and expression vector

Previously, the MCP of Sulfolobus Turreted Icosahedral Virus (STIV), an archaeal virus, has been successfully expressed in *E. coli* for structural studies (Khayat, Tang et al. 2005). Furthermore, we choose a bacterial expression system to produce large quantities of MCP due to its low cost, fast growth, and easy manipulation. BL21(DE3) cells are modified bacterial cells having T7 promoter region, which can produce more proteins than any other bacterial strain. The reason behind effective protein expression in BL21 is the absence of two genes, *lon* and *ompT*, that code for two key proteases. The protease coded by *lon* degrades misfolded proteins that accumulate in *E. coli* cells. The other protease coded by *ompT* removes extracellular proteins. Therefore, BL21 (DE3) cells are more suitable for the recombinant protein production because of less protease degradation. *E.coli* BL21 (DE3) strains were purchased from the company of Invitrogen.

The pCold series of expression vector are used in cold shock expression systems that express proteins at lower temperatures around 15-16°C. These vectors are advantageous in expressing high amounts of desired protein. The cold temperature increases the stability and solubility of the protein. The lower temperature slows down bacterial metabolism and suppresses the host protein production. pColdI (©Takara) vector was obtained from our collaborator Dr. Jianjun Sun from the Department of Biological Sciences at UTEP. pColdI improves the

expression levels and solubility of the target proteins by the help of a *csp* promoter that enhances transcriptional activity at lower temperatures by stabilizing the 5' UTR region. Therefore, pColdI was a selected initially as a suitable vector for expressing CroV MCP.

2.2.1.2 Polymerase chain reaction and cloning

Polymerase Chain Reaction (PCR) is a widely used tool in molecular biology utilized to amplify specific DNA fragments using two primers that are 5' and 3' of the target DNA sequence. PCR primers [Table 2.2] with additional restriction enzyme sequences at the end were designed to amplify the codon-optimized gene of CroV MCP in the pUC57 entry vector. Phusion (New England Biolabs Inc) high fidelity DNA polymerase was used for the PCR reaction to prevent mutations. Multiple optimization trials were performed to amplify the MCP gene by PCR [Figure 3.1]. The final PCR settings are listed in table 2.3.

Both the pColdI vector and the PCR product of CroV MCP were then digested with restriction enzymes NdeI and HindIII and ligated by T4 DNA ligase enzyme [© Invitrogen] to produce a construct. In this construct the N terminal has a 6X His tag that is followed by the CroV MCP gene. BL21(DE3) *E. coli* cells were treated with cold $RuCl_2$ solutions to produce competent cells (W. Zwerschke's PhD thesis , DKFZ, Heidelberg, Germany, 1997).The ligated product was then transformed into BL21(DE3) competent cells following the traditional 42°C heat shock method (Green, Sambrook et al. 2012). Bacterial clones containing the recombinant DNA were selected on ampicillin LB plates. Bacterial colonies were picked and glycerol stocks were made following traditional protocols (Green, Sambrook et al. 2012). Purified plasmids from the glycerol stocks were verified by both restriction enzyme digestion with HindIII and NdeI [Figure 3.2] and further confirmed by DNA sequencing.

2.2.2 Protein expression with pColdI vector

The transformed colonies were grown in an ampicillin LB liquid media at 37°C until the Optical Density (OD) at 600nm is 0.8. Then the culture was refrigerated down to 16°C for 30 minutes. After 30 minutes, 1mM IPTG was added to the cultures to induce the recombinant protein expression and incubated at 16°C with 220 rpm shaking for 16 hours. The cells were harvested by centrifugation at a speed of 5550xg and the pellets were resuspended in lysis buffer containing 50mM Tris-Cl, 300mM NaCl, 0.02% NaN_3 and 1mM PMSF and 10mM Imidazole.

The resuspended pellets were then subjected to freeze-thaw three times. Lysozyme was added to the cell lyses up to a final concentration of 0.7mg/ml to lyse the cells. Lysozyme is an enzyme that destroys the linkage between N acetyl muramic acid and N acetyl D glucosamine in peptidoglycan and chitodextrins. After the lysozyme reaction, bacterial cell walls are ruptured and the expressed proteins can leak out. After adding lysozyme and five cycles of freeze-thaw, the cell lyses were centrifuged at a high speed of 33,000xg at 4°C for 50 minutes to remove the cell debris, large cellular organisms and unlysed cells.

To estimate the yield of the expressed protein, SDS polyacrylamide gel electrophoresis method was used. The pellet and supernatant both were run on the gel electrophoresis. It was found that the expressed recombinant CroV MCP resides in the cell pellets and it is insoluble. No soluble MCP was detected [Figure 3.3]

2.2.3 Molecular cloning with pColdTF vector

Because the CroV MCP was insoluble in pColdI vector, we chose another pCold expression vector, which contains Trigger Factor (TF). TF is a prokaryotic ribosome associated with a cold shock chaperone of 48kDa that helps in polypeptide folding at a low temperature to reduce the chances of forming misfolded and insoluble protein. As the recombinant MCP with

pColdI was insoluble, we decided to select pColdTF as our new expression vector. As a native product from prokaryote, TF is highly expressed in *E.coli* using the strong *cspA* promoter. Additionally, the pColdTF vector contains a translation enhancing element, an N terminal His-Tag before TF, and a Multiple Cloning Site (MCS).

The pColdTF vector (© Takara) was obtained from our collaborator Dr. Jianjun Sun from Department of Biological Sciences at UTEP. Both the pColdTF vector and the PCR product of CroV MCP were digested with restriction enzymes NdeI and HindIII and ligated by T4 DNA ligase [© Life technologies] enzyme to produce a construct [Figure 2.3]. The clones were screened and confirmed by restriction enzyme digestions [Figure 3.5] and DNA sequencing [Figure 3.4] as mentioned in Section 2.2.1.2.

2.2.4 Protein expression

This constructed plasmid directs the synthesis of the MCP under IPTG inducible conditions. Multiple optimization trials were performed to search for the suitable conditions for protein expression. DNA sequenced and verified clones were cultured for protein expression following a standard protocol explained in section 2.2.2. Both the pellet and supernatant were run on a 10% SDS PAGE to evaluate the expression level and solubility of the recombinant protein [Figure 3.6].

2.2.5 Protein purification trials

Since the constructed clone contains a 6X Histidine tag at the N-terminal before the TF, Nickel-NTA affinity chromatography was used initially to purify the recombinant protein. Test purification trials were performed using Qiagen 1.5ml Ni-NTA superflow columns to identify the binding of the recombinant protein with the Ni-NTA column.

The column was equilibrated with 10 mL of binding buffer containing 50mM Tris, 300mM sodium chloride (NaCl), 0.02% sodium azide (NaN_3). The cell lysate was filtered with a 0.45μm filter and loaded onto the column and run through the column by gravity, allowing the binding of the protein to the column. The flow through was collected in a separate sample tube. The column was then washed with 10ml of binding buffer. A step gradient method was set up to elute the fusion protein MCP bound to the column. The column was washed in steps with 10ml of 40mM, 60mM, and 100mM imidazole solutions at a pH of 7.5. After the wash, the column was eluted with 120mM, 150mM and 200mM imidazole solutions at a pH of 7.5 [Figure 3.7].

His-HiTrap HP column was purchased from the company of GE healthcare Life Science, (catalog number 17-5247-01). His-HiTrap columns are cross-linked sepharose beads packed with a chelating group attached to it. Histidine has an affinity towards Ni^{2+} and the 6X His will have increased affinity for Ni^{2+}. It has been found that Histidine tagged proteins bind very strongly with Ni^{2+} compared to other proteins. Two different buffers, binding buffer containing 50mM Tris, 300mM NaCl, 0.02% NaN_3, 1mM PMSF and 10mM imidazole and an elution buffer containing 50mM Tris, 100mM NaCl, 0.02% NaN_3 and 1M imidazole, were prepared for the purification process using Amersham Pharmacia Biotech (now belongs to GE Healthcare) AKTA FPLC.

The 5mL His chromatography column was first equilibrated by 5 column volumes (25mL) of binding buffer with a flow rate of 1ml/minute. 2L of the cell lyses was loaded onto the column by a peristaltic pump. Elution of the recombinant protein started with the elution buffer by applying a linear gradient from 10mM imidazole to 800mM imidazole. After having a preliminary idea about the concentration of imidazole from the gradient experiment for elution, a

step gradient method was applied. The bound proteins were first washed at 100mM imidazole concentration and eluted at 120mM imidazole to yield relatively pure protein [Figure 3.8].

The eluted TF-MCP was subjected to size-exclusion chromatography for further purification. Size exclusion chromatography (SEC) uses porous polymer beads and separates macromolecules based on its size and hydrodynamic diameter. Therefore, SEC can be used to separate molecules with different molecular weight. SEC can also be used to determine the quaternary structure of the purified protein. When a mobile phase passes through the column, small molecules with a lesser hydrodynamic diameter will travel a longer path through the inside of the beads, while larger molecules will travel a shorter path outside of the beads. Therefore, molecules with a larger hydrodynamic diameter will be eluted faster in SEC, separating them from the smaller molecules.

SEC was employed to remove smaller protein contaminations present in TF-MCP with size exclusion buffer containing 25mM Tris-HCl, 150mM NaCl, 0.02% NaN_3. The chromatogram showed three different peaks at three different elutions [Figure 3.10]. The first peak was our protein of interest, TF-MCP. The other two peaks were contaminants. Samples were prepared with 4X SDS PAGE dye then heated at 100°C for 5 minutes and run on a 10% SDS PAGE stained by coomasie stain. Silver staining technique was also utilized to determine the purity of the MCP [Figure 3.11]. Protein concentration was determined by a BCA Protein Assay kit (©Thermo scientific) using known concentration bovine serum albumin (BSA) as standard.

2.2.6 MALDIMS-TOF in determining the molecular weight of MCP

Mass spectrometry is an analytical tool to determine the mass of the molecules to the accuracy of one Dalton. In order to measure the mass, molecules must be ionized so that their

binding ratio in an electromagnetic field is proportional to their mass and charge ratio. Due to the large size of the proteins, a sophisticated ionization mass spec technique, MALDI is used. Laser ionization radiates the analyte mixed with the matrix material and vaporizes the analyte. Therefore, the matrix plays an important role in MALDI. Organic acids are common matrix materials. The matrix acts both as a proton donor and acceptor that enhances the analyte to ionize in both positive and negative charges. The ionization solely depends upon the matrix analyte combination. Additionally, MALDI is commonly connected with TOF (time of flight) mass analyzer. The analysis is based on the acceleration of a set of ions to reach the detector while all the ions are ionized with same amount of energy. As the ions have same energies but different masses, they will reach the detector at different times. Thus the analyzer is called TOF because the mass is determined by the time of flight of the ions.

MALDI is a useful technique to accurately determine the molecular weight of a protein. Accurate molecular weight is important for the identification and characterization of proteins. Accuracy of the MALDI data depends upon the suitable choice of matrix materials and solvents. Contamination of the sample with excessive salt (>10mM) might affect the sensitivity and accuracy of the data. MALDI TOF can easily examine a complex mixture of peptides. Post translational modifications of proteins (like glycosylation, phosphorylation) can also be recognized by MALDI.

For the analysis of proteins, standard matrices like cinnamic acid (CCA), sinapinic acid (SA), 2, 5 dihydroxy benzoic acid (DHB) are available. The quality of the mass spectra for proteins depends upon the preparation of sample matrix. The most common method used for the sample matrix preparation is called the dried droplet method. This method is relatively simple; a saturated matrix is mixed with the sample to give matrix to sample ratio of 5000:1.

0.5µl of the purified TF-MCP and 0.5µL of the matrix was placed onto the sample plate. The matrix is made up of a combination of a saturated solution of CCA in 50:50 water and acetonitrile with 0.1% TFA. The sample and matrix were well mixed by pipetting and allowed to air dry for 10-15 minutes until the sample-analyte mixture was completely dried. The sample plate was then subjected to MALDIMS TOF (Bruker microFlex LRF model) to measure the molecular weights of different species in the samples [Figure 3.12]

2.2.7 DLS in MCP characterization

DLS is a relatively new technique to determine the hydrodynamic size of particles. DLS utilizes a monochromatic laser beam whose wavelength changes when it passes through a solution with particles in random motion (Brownian motion). This change of wavelength is correlated to the largest dimensional size of the particle. When a particle is moving very fast in a solution, the shape of the particle can be assumed to be spherical with diameter as the longest dimension of the molecule. This change of wavelength is correlated to the largest dimensional size of the particle. Therefore DLS can be used to estimate the size of proteins. For instance, DLS can easily distinguish between monomeric and multimeric proteins. Sample stability, formation of aggregates, and thermal stability can be easily detected by DLS. Therefore, DLS can also be used to monitor the particle size distribution and species, such as aggregation formation, in real time. In addition, the newer models of DLS instruments can also measure the surface charge potential of the protein. This surface charge measurement can offer useful hints regarding whether cation or anion exchange chromatography should be used to purify the protein. DLS can also be utilized to calculate the rate of diffusion of proteins. DLS measurement is fast, in real time, and requires very little amount of sample.

Size of the TF-MCP protein and its native charge was determined by DLS using a Zetasizer nano ZS (©Malvern) instrument. The machine captures the change in the intensity of the scattered light that occurs because of the random motion of the molecules. The analysis of the time dependent change of intensity will determine the particles diffusion coefficient. This is later converted into size distribution.

1mg/ml concentration of the MCP was used for DLS analysis.10µl of sample was diluted with 990µl of the buffer composed of 25mM Tris, 100mM NaCl and 0.02% NaN_3. Time dependent data was acquired every 10s with 12 acquisitions per run. TF-MCP was illuminated with a laser with wavelength of 633nm and the intensity of light scattered at an angle of 173° [Figure 3.13].

The same instrument was used to measure the native charge of TF-MCP. Voltage was applied at both ends of the electrode connected to the sample in a cuvette. Charged particles will be attracted to the electrodes with opposite charges. The velocity of the charged particles is defined by electrophoretic mobility μ_e and based on this, the zeta potential is calculated.

2.3 Alternative approach

In an alternative approach, we tried to purify the MCP of CroV directly from the virus. Previously, it was not feasible due to very low yield of the CroV. Recently, our collaborator has significantly improved the propagation protocol and improved the yield of the virus.

2.3.1 Propagation of virus

The virus was propagated according to the protocol designed by our collaborator, Dr. Matthias Fischer. The initial protocol was developed when he was a Ph.D. student at Dr. Curtis Suttle laboratory at the University of British Columbia, Vancouver, Canada. After graduation,

Dr. Fischer is now at the Max Planck Institute for Medical Research, Heidelberg, Germany. CroV and its host Cro, both were obtained from Dr. Fischer.

Artificial sea water (33 parts per thousand, ppt salinity) was first prepared by adding artificial sea salt to de-ionized water. 35.64g of Instant Ocean sea salt (©Instant Ocean) was very slowly added to 1L of de-ionized water and stirred vigorously overnight.

75g/L $NaNO_3$ and 5g/L NaH_2PO_4 stock solutions were prepared. The solutions were sterilized by filtering through 0.2µm membrane. 4.36 g/L $Na_2EDTA \cdot 2\ H_2O$, 3.15 g/L $FeCl_3$, $6H_2O$ 0.01 g/L, $CuSO_2 \cdot 5H_2O$, 0.18g/L $MnCl_2 \cdot 4H_2O$, 0.006g/L $Na_2MoO_4 \cdot 2H_2O$, 0.022g/L $ZnSO_4 \cdot 7H_2O$ and 0.01g/L $CoCl_2 \cdot 6H_2O$ were added to water to prepare a 1000X trace metal solution. 50,000X vitamin stock was prepared by adding 5 g/L thiamine·HCl (Vitamin B1), 25 mg/L biotin (Vitamin H), 25 mg/L cyanocobalamine (Vitamin B12). Later it was diluted to 2000X stock solution and aliquoted. Later 1 mL of 75 g/L $NaNO_3$, 1 mL of 5 g/LNaH_2PO_4, 1 mL trace metal solution, 0.5 mL Vitamin stock were added to 1L of seawater to make the f/2 media. The f/2 media was sterilized by filtering through 0.2 µm filter (Corning 500mL bottle top filter; catalog number 430049).

25 mL of f/2 media was transferred to a 250mL plastic Erlenmeyer flask; to every 25mL culture, 2.5mg of yeast extract and 500µl-1mL of *Cafeteria roenbergensis* (Cro) inoculant was added to media. The flask was kept in a dark environment with a loosened cap and shaken once every day. Every day, cell samples were stained by Lugols iodine and counted using hemocytometer. Whenever the cell count reached 1x10^6 cells per mL, we split the culture into a new Erlenmeyer flask to 5x10^5 cells per mL, when it is ready to be infected by virus (CroV).

2.3.1.1 Infection of Cro by CroV

1X10^6 cells per mL culture were diluted to 5.0X10^5 cells per mL by 33ppt f/2 media with 0.05% (w/v) yeast extract, 50 µg/mL ampicillin, 50 µg/mL kanamycin, and 25 µg/mL chloramphenicol. The infection process began with 50-100 µl of crude CroV lysate (~1 x 10^11 particle per mL). 50-100 µl of lysate was diluted into 10 ml f/2 without any yeast extract or antibiotic and mixed well. The diluted inoculum was slowly pipetted to the culture while swirling the flask. The infected cells were incubated at room temperature in the dark. The infected cells were swirled once a day. The cells were harvested when the count dropped below 1X 10^4 per mL.

The infected cells were harvested by centrifuging at a speed of 5550xg, for 50 minutes at 4°C. This removed most of the bacteria present in the culture. After centrifugation, the supernatant was transferred to a separate tube and the pellet was discarded. The supernatant was centrifuged again at 10000xg for 10 minutes at 4°C. Then the supernatant was transferred to a separate tube and the pellet was discarded. Once again the supernatant was centrifuged at 10000xg for 2 hours at 4°C. This time the pellet was kept with minimum amount of supernatant. The rest of the supernatant was discarded. The pellet was dissolved in the remaining supernatant.

Since MCP is a major component in the protein shell of CroV, it was expected that the concentrated virus sample will exhibit a strong protein band of the MCP when run into 10% SDS PAGE [Section 3, Figure 3.14].

Tables and figures

Table 2.1 Major and Minor Capsid Proteins

Protein	Gene	Length(bp)
Major capsid protein (MCP)	CroV342	1518
Minor Capsid protein	CroV187	424
Minor Capsid protein	CroV240	280
Minor Capsid protein	CroV318	280
Minor Capsid protein	CroV320	604

Table 2.2 Primer designed for MCP

Sense Primer	5'-gcggcatatggcaggcggtctgatgcagctggttgcttatg-3' NdeI
Antisense Primer	5'- cggcaagcttctagttggaatacgccagaccgccca-3' HindIII

Table 2.3 PCR settings for MCP amplification

Process	Temperature	Time
Initial Denaturation	98°C	2 minutes
Denaturation	98°C	10seconds
Annealing	64°C	30seconds
Extension	72°C	25seconds
Final Extension	72°C	10minutes

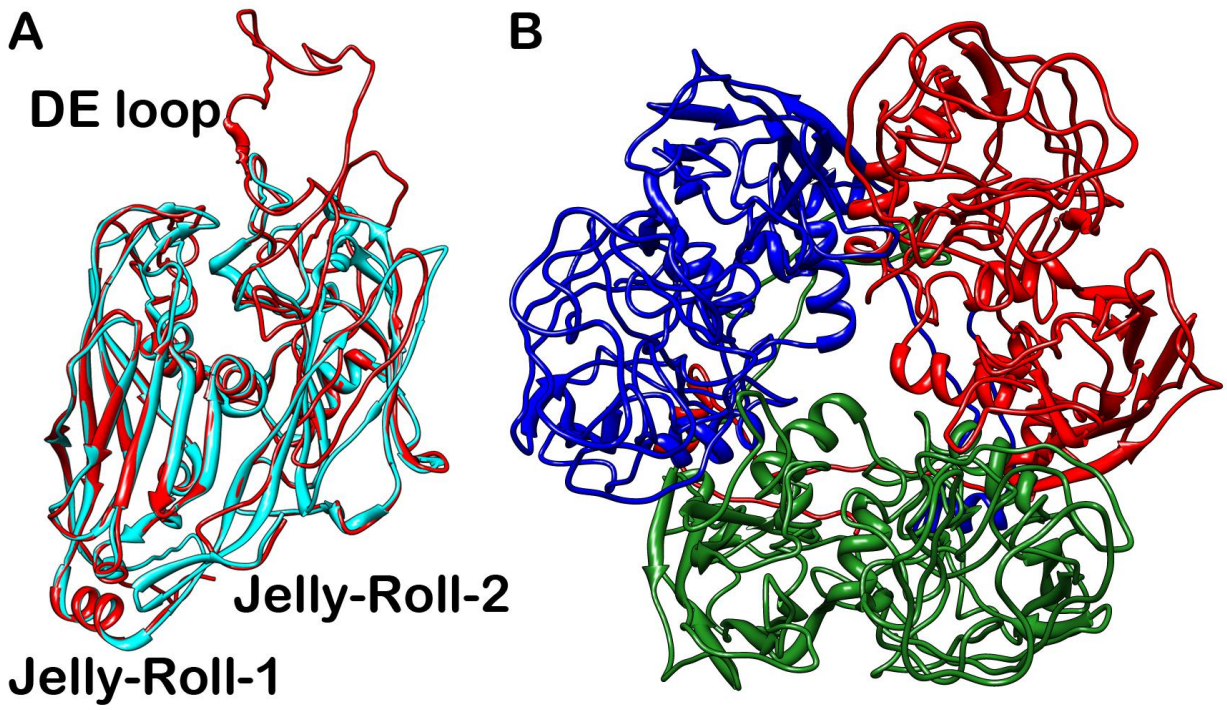

Figure2.1 Double-jelly-roll MCP **(A)** Homologous modeled CroV MCP (red) superimposed with PBCV-1 MCP (cyan). The jelly-rolls and the extended DE loop in CroV are labeled. **(B)** Modeled CroV capsomer. Individual monomeric CroV MCP is colored red, blue and green respectively. [Unpublished data from Dr. Xiao]

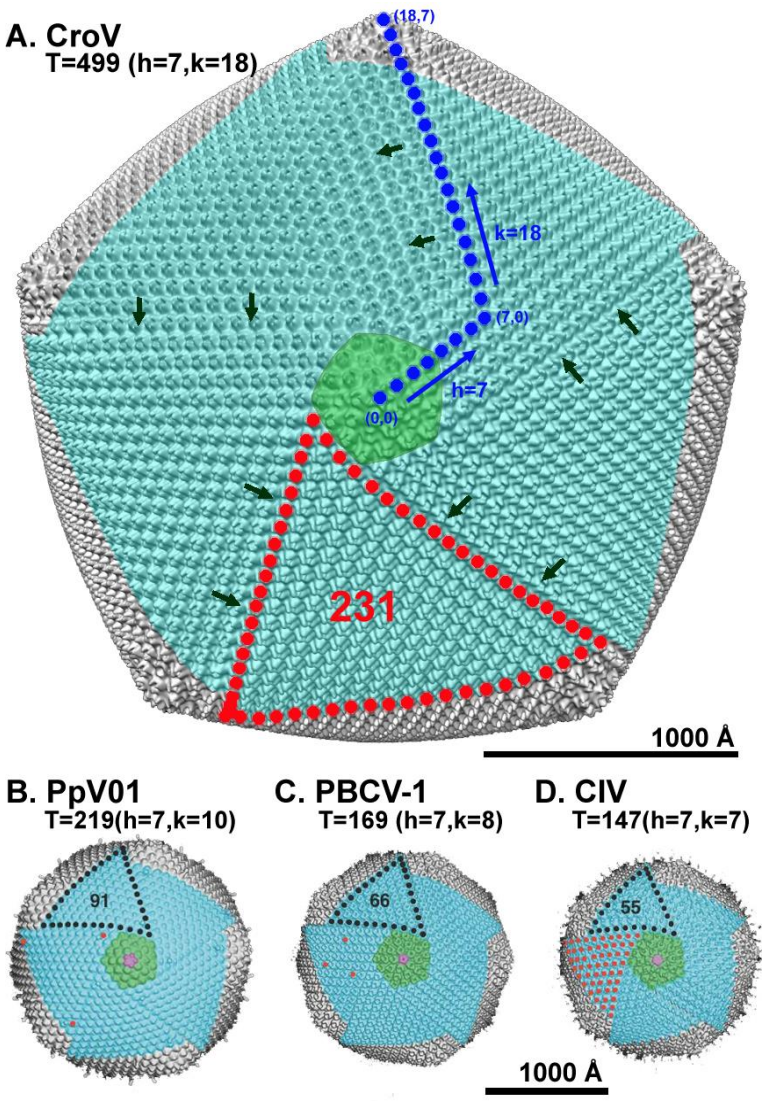

Figure 2.2 CroVcapsomer arrangements **(A)** CroV: Grid points of hexagonal array labeled in blue indicating h=7 and k=18 from one 5-fold to the neighboring one. Pentasymmetron is colored in green and one trisymmetron is outlined by red dots on top of the capsomer. The discontinuity lines between neighboring trisymmetrons are indicated by dark green arrows. **(B)** PpV01 **(C)** PBCV-1 and **(D)** CIV, showing the pentasymmetron in green and one trisymmetron outlined in black dots. **(A)** unpublished data from Dr. Xiao; **(B) (C) (D)** are adapted from (Yan, Chipman et al. 2005).

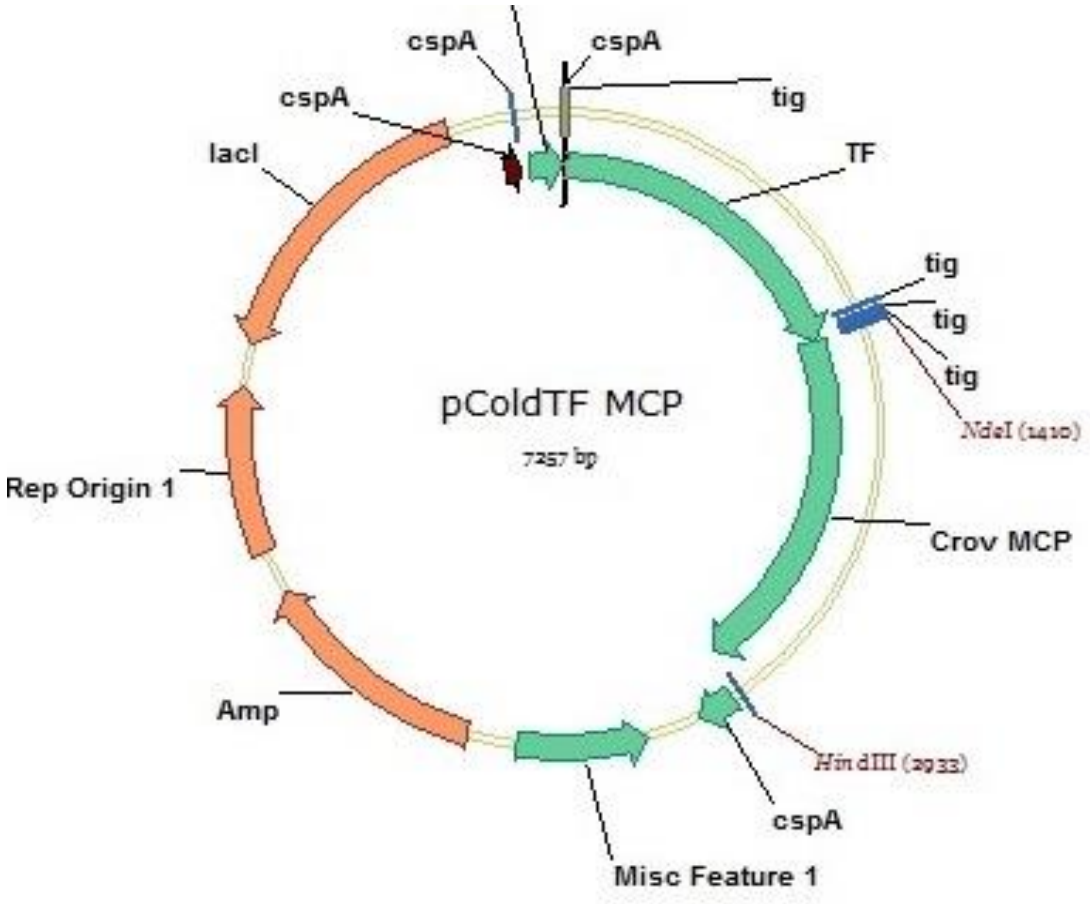

Figure 2.3 Expression construct of CroV TF-MCP with vector pColdTF. Genes are presented by arrows. Cloning restriction sites are labeled.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Cloning of major capsid protein (MCP) gene

Nancy Ulloa-Rondeau, a former graduate student, worked on the amplification of the MCP gene from the genomic DNA obtained from our collaborator Dr. Matthias Fischer of UBC. Long Amp PCR kit (© New England Biolabs) was used for the amplification, and it was cloned using the Invitrogen Gateway (©Life technologies) system. The Gateway cloning system uses two recombination reactions known as LR and BP, both of which are sequence-specific recombination catalyzed by Vaccinia virus topoisomerase (Shuman 1994). LR recombination reaction takes place in between an entry clone and a destination vector while BP recombination reaction occurs in between an expression clone and a donor vector. During the verification of the construct sequence, a mutation from glycine to arginine was found due to the low fidelity of the Long Amp polymerase. Later MCP was codon optimized and synthesized into an entry vector pUC57.

MCP gene was then amplified [Figure 3.1] from pUC57 by a standard PCR technique [PCR settings table 2.3, Section 2.2.1] using Phusion (New England Biolabs Inc) polymerase. Phusion polymerase is a high fidelity polymerase that prevents the introduction of mutation during PCR. The expression vector pColdI and the CroV MCP were successfully treated with restriction endonuclease enzymes, NdeI and HIndIII. After purified from agarose gel electrophoresis, the insertion and vector were ligated by T4 DNA ligase enzyme (©Invitrogen) After successfully transforming the ligated product into the *E. coli* BL21 (DE3) bacterial cells, the plasmids were purified for construct verification. Restriction enzyme digestion of NdeI and HindIII verified the cloned construct [Figure 3.2] showing 2 bands of 4.4kb (vector) and 1.5kb

(insert) in an agarose gel electrophoresis. DNA sequencing also showed complete sequence identity between the optimized MCP and the cloned construct. Later, small scale protein expression trials were performed, which showed that the MCP protein was insoluble [Figure 3.3].

There was a large fraction of insoluble protein observed for the pColdI expression trials. Insolubility of the foreign protein expressed in *E.coli* is quite common (Wilkinson and Harrison 1991). There are various factors that lead to the insolubility of the recombinant protein (Wilkinson and Harrison 1991). It is hypothesized that the MCP may be rich in hydrophobic patches at the N-terminal region, which may be anchored into the membrane. Elimination of that region can greatly increase the stability and solubility of MCP. Another possibility is that during the induction by IPTG, a large quantity of the protein is produced in a very short time. The recombinant proteins do not get enough time to fold properly. Hopefully, co-expressing with a molecular chaperone will solve this issue.

Due to the insoluble nature of the MCP in the expression trials with pColdI vector, an alternative expression vector pCold Trigger Factor (TF) was selected to express the CroV MCP. TF itself is a prokaryotic ribosome associated cold shock chaperone with a molecular weight of 48kDa. TF facilitates correct protein folding under low temperature, thus enabling the efficient soluble protein production.

CroV MCP was cloned into the pColdTF using the same method as described above for the pColdI vector. The resulting plasmid purified from *E. coli* BL21 (DE3) cells was verified by restriction enzyme digestion. It indicated two DNA fragments at the expected sizes of 5.7kb and 1.5kb [Figure 3.5] in an agarose gel electrophoresis. Furthermore, the recombinant DNA was

verified by DNA sequencing [Fig. 3.4]. The sequence of TF-MCP construct completely matched the codon optimized sequence.

3.2 Protein expression and purification

Furthermore, multiple confirmed colonies were selected and induced with IPTG to express the recombinant TF-MCP under the stimulus of a *cspA* promoter. As shown in Figure 3.6, a recombinant protein band corresponding to approximately108kDa (TF+MCP) protein was present in both the insoluble fractions and supernatant of the lysed cells by implementing both the freeze-thaw and microfluidized method. The non-induced pellet and supernatant did not show a protein band corresponding to the size of TF-MCP. The SDS PAGE showed that the protein TF-MCP was present in the soluble fraction. It was assumed that TF facilitated the proper folding of MCP and increased the solubility of the protein. TF might also bind to the N-terminal hydrophobic region, preventing MCP aggregation.

After multiple successful small-scale protein expression trials, large scale protein expression trials were performed. The lysate was tested on the gravity flow affinity chromatography using a step elution with different imidazole concentrations to determine the optimum imidazole concentration to elute TF-MCP. After washing with 40mM, 60mM and 100mM imidazole concentrations and eluting with 120mM, 150mM and 200mM imidazole relatively pure TF-MCP [Figure 3.7] was obtained.

Moreover, the presence of a 6X-His tag at the N terminal of the recombinant protein facilitates the possibility of obtaining relatively pure protein via Nickel^{2+} NTA column chromatography. Ni^{2+} is the most commonly used bivalent metal for purifying the His tagged proteins. Ni^{2+} exhibits higher affinity towards His tagged proteins but it also binds nonspecifically to the host cell proteins with histidine clusters. To remove the nonspecific

binding of the host cell proteins, 10mM imidazole was added to the binding buffer. Affinity chromatography of supernatants from the cell extracts enhances the purification of TF-MCP with yields of 1mg/L with the presence of detectable contamination of *E. coli* proteins [Figure 3.8].

A larger step gradient was also applied with 190mM imidazole for washing and 400mM imidazole for eluting the TF-MCP. It was expected that 190mM imidazole would remove most of the host proteins and 400mM elution will provide pure TF-MCP. It was observed that 190mM imidazole concentration washed away sufficient amount of unnecessary proteins while eluting with 400mM imidazole concentration yielded relatively pure TF-MCP [Figure 3.9] with very little trace of other contaminating proteins from *E.coli*.

In order to obtain highly purified protein, 400mM eluted TF-MCP was subjected to size-exclusion chromatography (SEC). In native form CroV MCP is a trimer, during multiple trials of SEC, a large peak was observed during the early elution fraction at around 120ml. There were also 2 small peaks in later elutions [Figure 3.10]. Based on the calibration using other proteins that have been eluted from the same column, the elution volume of 120mL is correspondent to a protein at about 300kDa, which indicates a trimeric form of TF-MCP (monomeric molecular weight of 108kDa). It was concluded from the SEC that TF-MCP folded properly and was present in its native trimeric form. To determine the purity of TF-MCP after SEC, the samples were run in a SDS PAGE stained by silver [Figure 3.11].

The silver staining process is very sensitive; it can detect the presence of trace amount of proteins in acrylamide gel. Traditional coomasie blue staining can detect 50ng of protein while silver stain is 50 times more sensitive than coomasie technique. The SDS PAGE stained by silver showed that TF-MCP is very pure showing no trace of other contaminant proteins [Figure 3.11].

3.3 Characterization of MCP by MALDIMS-TOF & DLS

After pure TF-MCP was obtained, MALDIMS-TOF and DLS studies were performed to determine the molecular weight and the hydrodynamic size of the recombinant protein. After MALDIMS-TOF, using universal matrix with TFA, two major peaks were observed showing molecular weights of MCP at 242kDa and 307kDa respectively [Figure 3.12]. It was suggested that pure MCP obtained after affinity chromatography and SEC, had a molecular weight of 307kDa. The 242kDa fragment might be the dimeric form generated during the ionization, while the monomeric fragment was not determined due to the measurement range.

To determine whether the MCP has formed a trimer or not, we also performed a DLS study. Proteins are macromolecules attached together by non-covalent interactions such as van der Waals interactions, hydrogen bonds, salt bridges, and hydrophobic interactions. Disruption of any of these forces can lead to the formation of aggregates. DLS and SEC are both analytical tools that can be utilized to determine multimeric formation of the proteins. The DLS measurement showed a typical z average size (hydro dynamic diameter) of 9.3 nm (PDI=1.0). The size distribution analysis in figure 3.13 showed a single peak near the expected size for the trimeric TF-MCP. As mentioned above in [Section 2.1], the trimeric MCP of NCLDV has the diameter ranging between 74 Ǻ to 85 Ǻ. After fused with TF, the hydrodynamic size might be larger. The size of 93 Ǻ indicates that the purified TF-MCP is in its trimeric form because monomeric form will be much smaller. The sharp peak also indicates that there is no presence of other large aggregates.

3.4 Alternative method

Alternatively we tried to purify the MCP directly from virus. We obtained two different strains E and R of the host Cro. We infected both E and R host strains but it was observed that

the R strain is resistant to infection. After infecting the E strain of Cro cells by CroV, it was observed that the cell count decreased after 3 days. After 7 days post infection the Cro population was crashed down to lower than 1x10^4 cells per mL. After extracting the virus following the procedures mentioned in Section 2.3.1.1, the concentrated virus sample was loaded in a 10% SDS PAGE to see if the MCP band was observed. As MCP covers most of the virus particle, it was expected that the MCP band would be visible in SDS PAGE. It is known that the MCP has a molecular weight of 58kDa. The concentrated 20µL virus sample was loaded on SDS PAGE and showed two bands just below 63kDa protein marker [Figure 3.14].

Tables & Figures

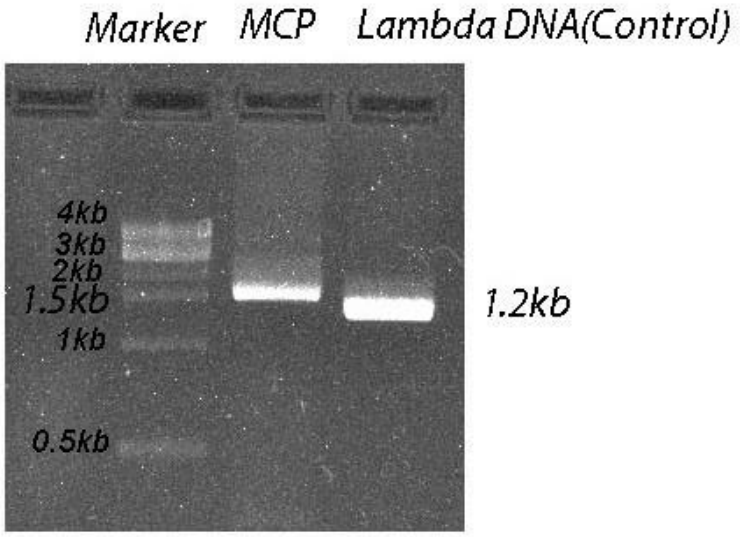

Figure 3.1 Amplification of CroV MCP gene by PCR. Agarose gel electrophoresis following the amplification of CroV MCP gene by PCR. MCP gene is in the second lane with correct size of 1.5kb. The third lane is the positive control PCR of λ DNA.

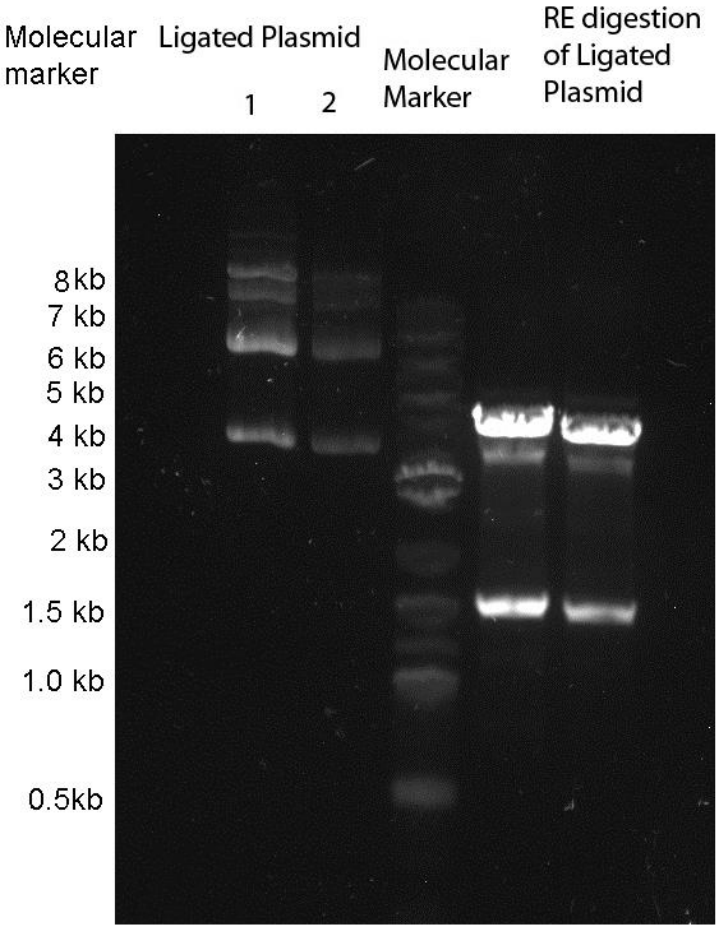

Figure 3.2 Verification of pColdI-MCP constructs by restriction enzyme (RE) digestion. Agarose gel electrophoresis of pColdI MCP plasmid after a duel RE digestion using NdeI and HindIII. Undigested plasmids were load on the left two lanes of the molecular marker. The RE digestion produces fragments (right two lanes of the marker) of expected size of 4.4 kb (vector) and 1.5 kb (insertion of the MCP).

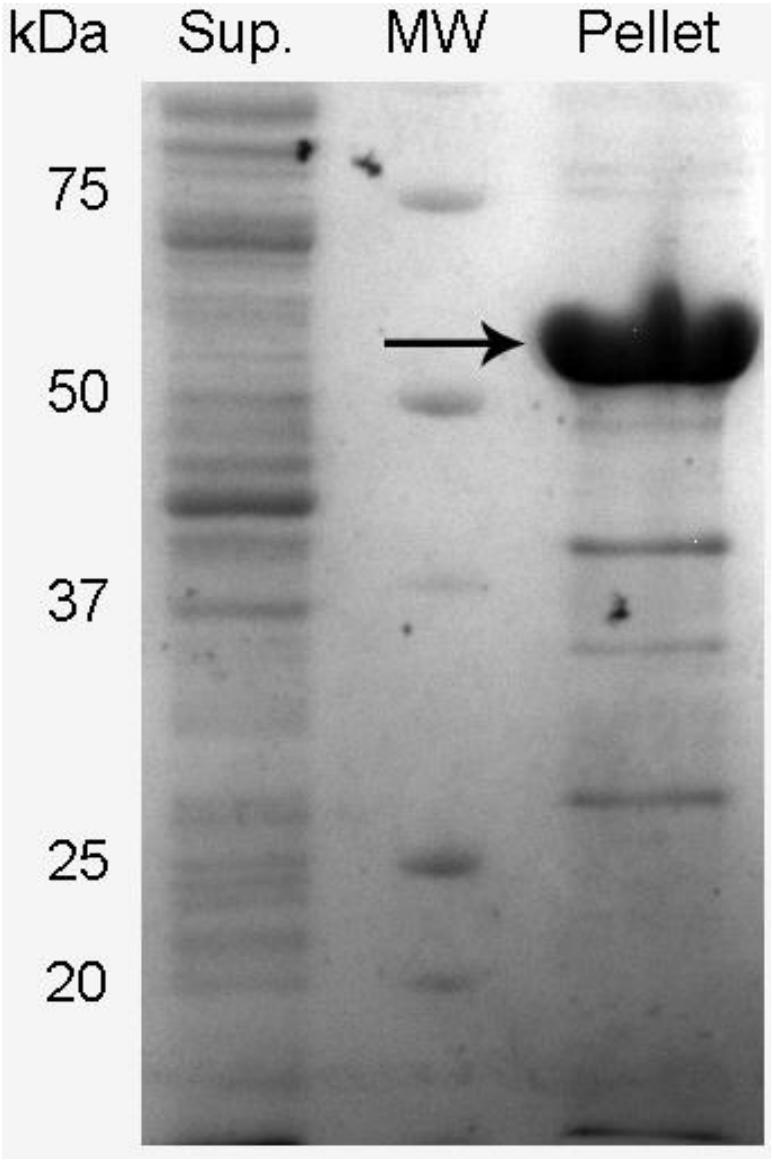

Figure 3.3 SDS-PAGE of protein expression trials of CroV MCP using pColdI vector. Supernatant and cell pellet were loaded on the left and right of the molecular marker (MW) respectively. Arrow depicts the expected size CroV MCP, which is only present in the pellet fraction.

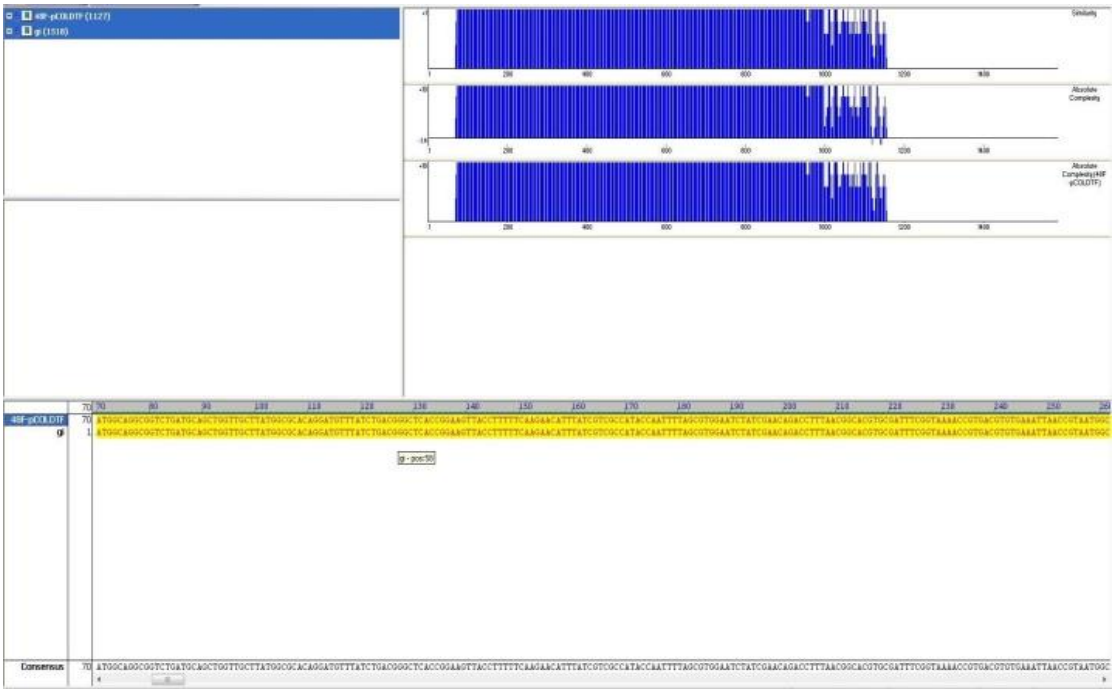

Figure 3.4 DNA sequencing verification of TF-MCP expression constructs. The height of blue-colored histogram on the upper-right panel indicates the sequence homologous among three MCP sequencing results. Yellow sequences show alignment between the two sequences with 100% identity.

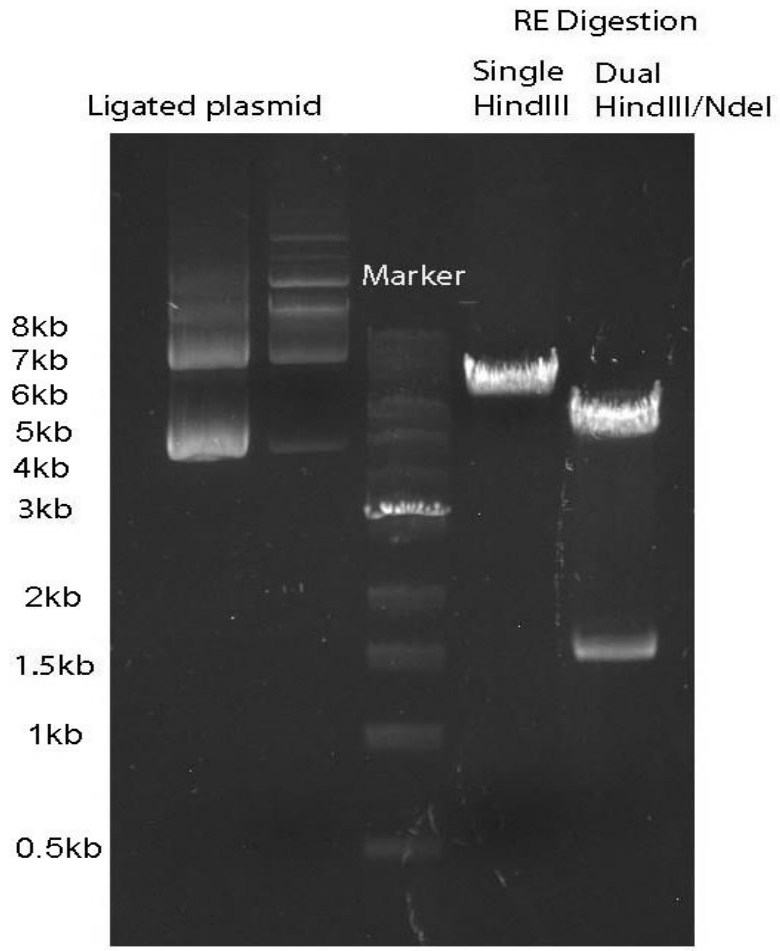

Figure 3.5 Verification of pColdTF-MCP cloned construct by restriction enzyme digestion. Agarose gel electrophoresis of undigested plasmid (left two lanes), molecular weight marker (3rd lane), single (4th lane) and dual (5th lane) restriction enzyme digestion of TF-MCP constructs showing the expected size of the vector (5.7kb) and insert MCP (1.5kb).

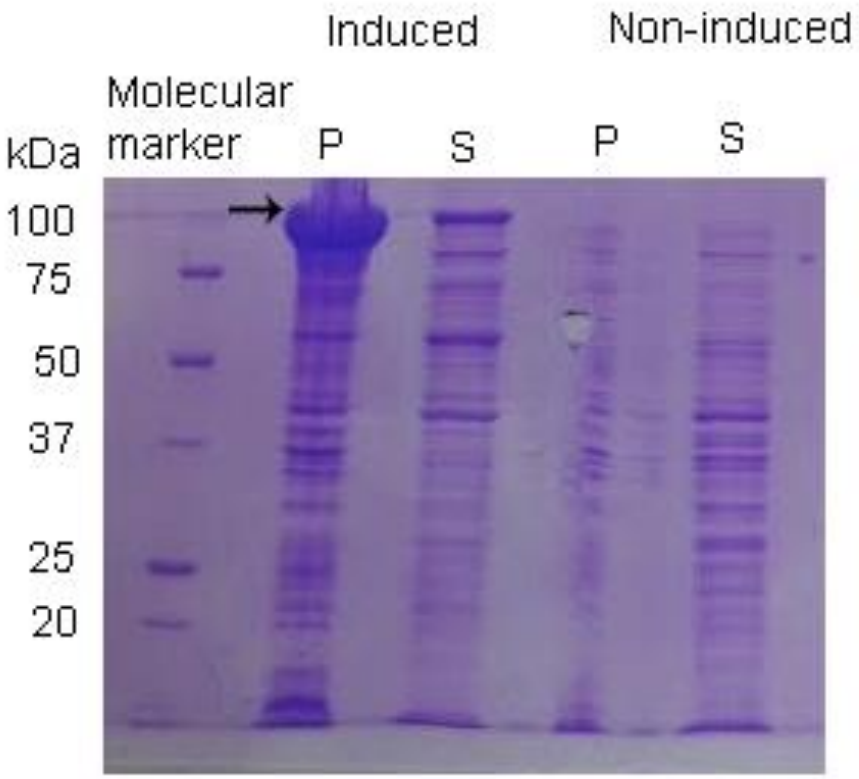

Figure 3.6 SDS PAGE of CroV TF-MCP expression trials. Lane 2 to lane 5 show IPTG induced and non-induced cell lyses supernatant (S) and pellet (P). The induced pair displays a strong protein band of molecular weight around 100kDa where the arrow points to; correspond to the expected size of the TF-MCP.

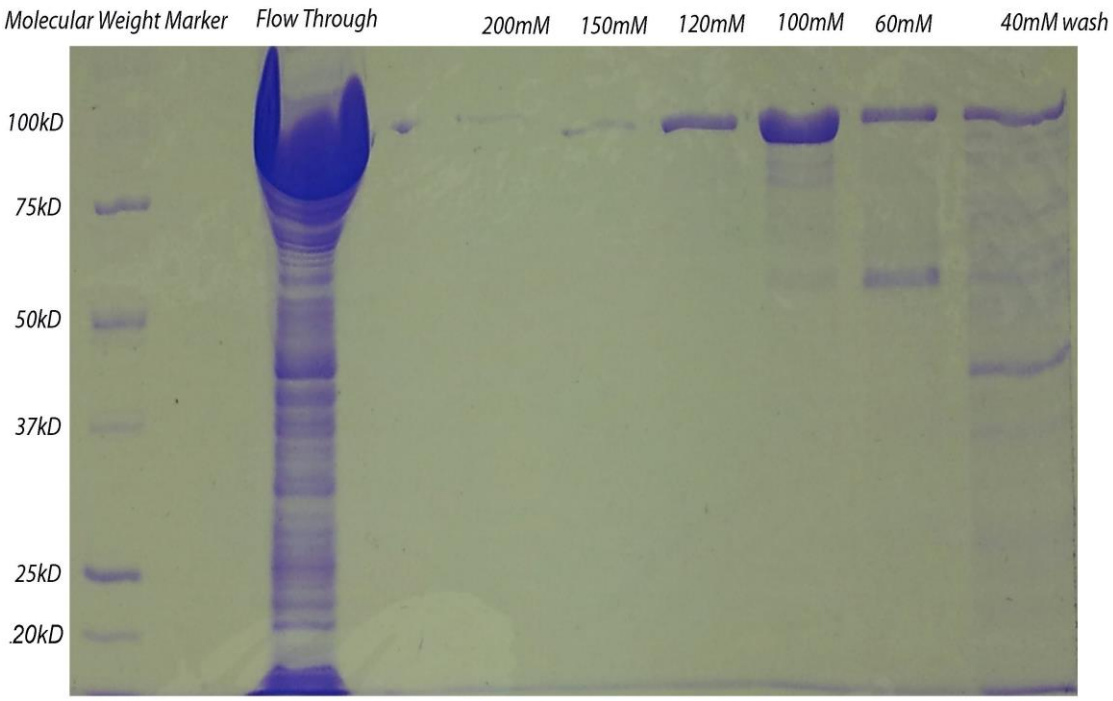

Figure 3.7 SDS PAGE of CroV MCP purification using gravity flow affinity chromatography. Lane 1 is molecular marker. Lane 2 is empty. Lane 3 shows the flow through. Lane 4 is empty. Lane 5 to 10 depicts 200mM, 150mM, 120mM, 100mM, 60mM and 40mM imidazole elution respectively.

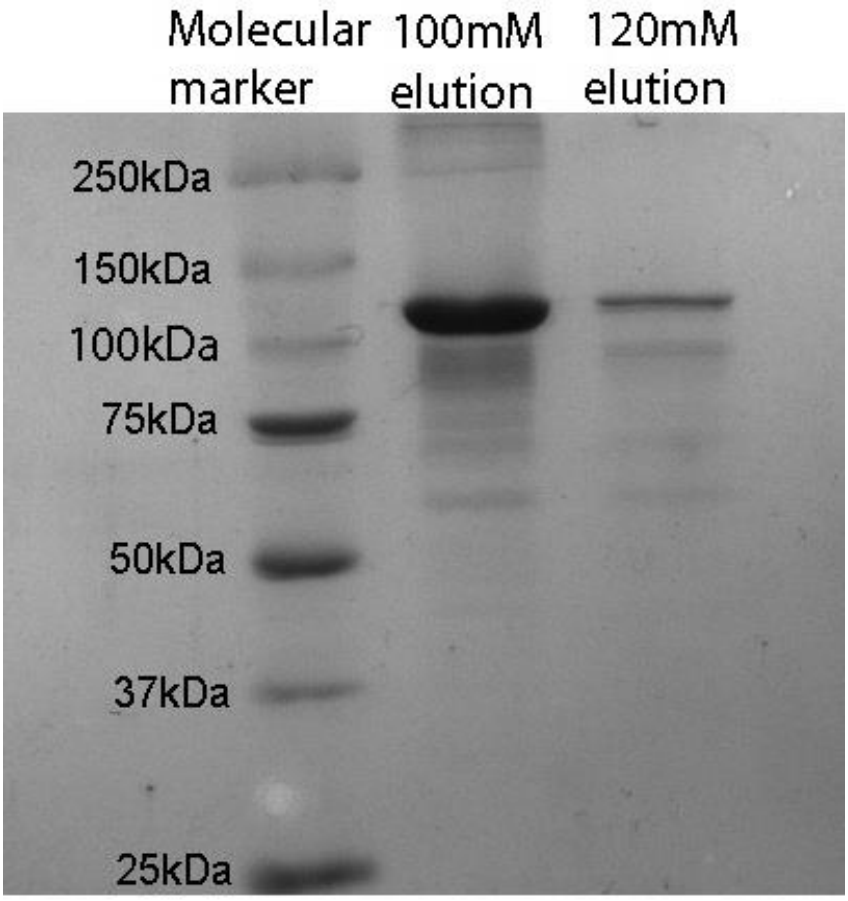

Figure 3.8 SDS PAGE of CroV TF-MCP purification using Ni-NTA affinity chromatography. Lane 1 shows the molecular marker. Lanes 2 and 3 shows 100mM and 120mM imidazole elutions after concentrating it down by centrifugation.

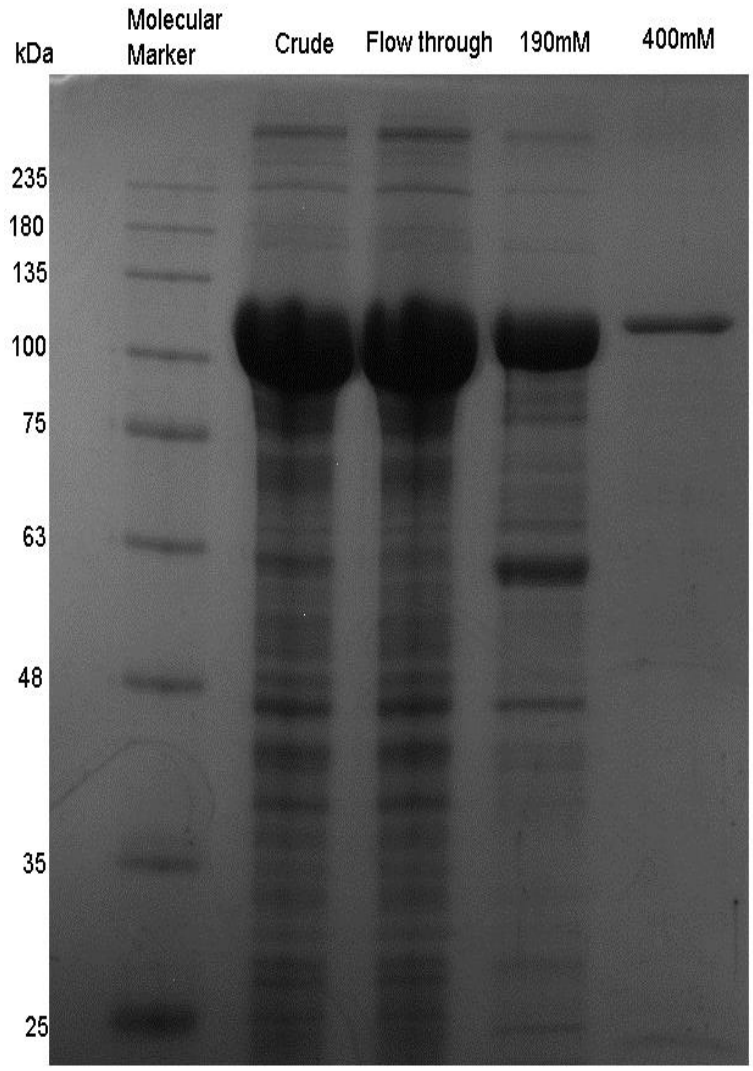

Figure 3.9 SDS PAGE of CroV TF-MCP purification using higher imidazole concentrations to elute. Lane 1 is molecular marker. Lane 2 and 3 are crude and flow through, respectively. The loaded affinity column was washing with 190mM imidazole concentration (Lane 4) and then elute with 400mM imidazole (Lane 5).

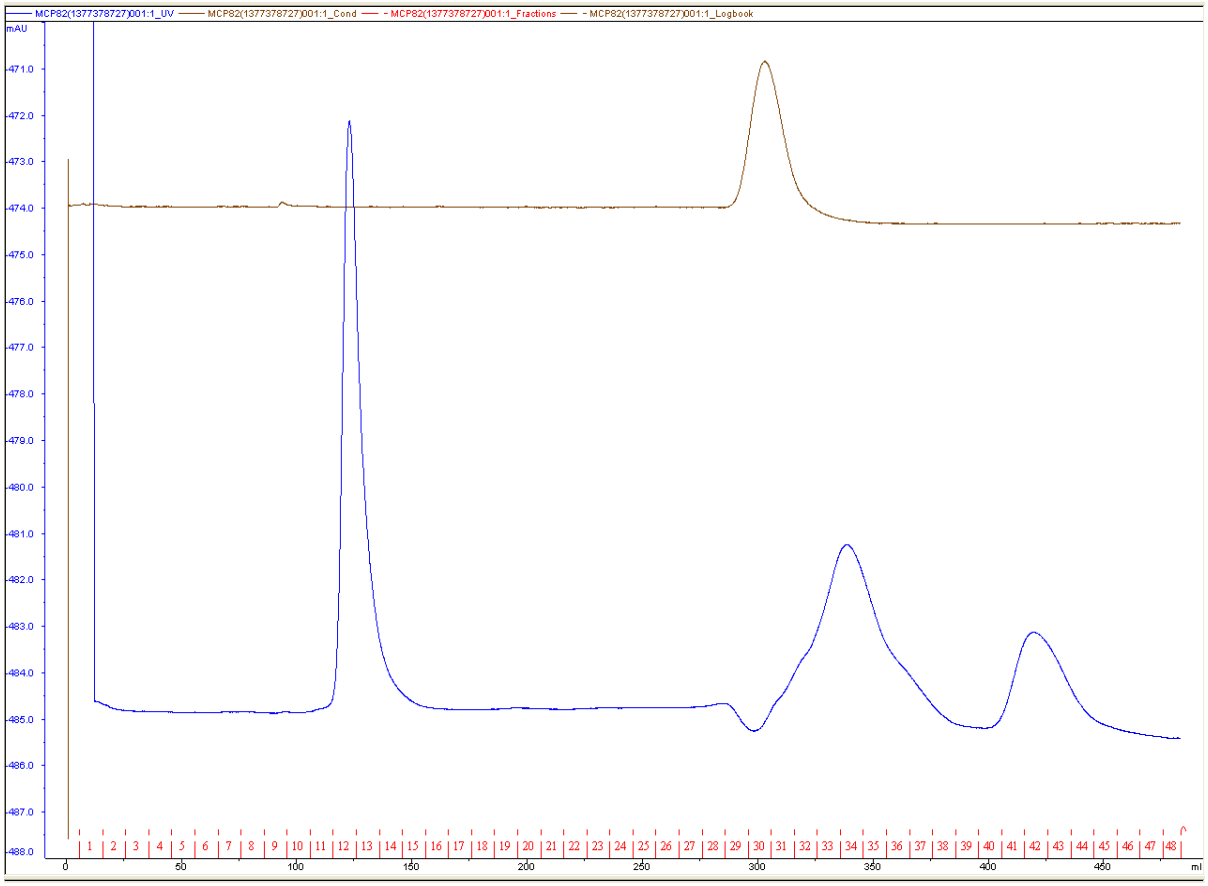

Figure 3.10 Chromatogram of TF-MCP purification using size exclusion chromatography. The blue curve is the UV absorption at 280nm wavelength. The brown curve is the conductivity of the elution. The major peak on the left is the TF-MCP (verified by SDS-PAGE) while the last two peaks are contamination protein (also verified by SDS-PAGE). The fraction connection (indicated by the bottom red numbers) started at elution volume of 110mL and the TF-MCP peak appeared at elution volume around 120mL.

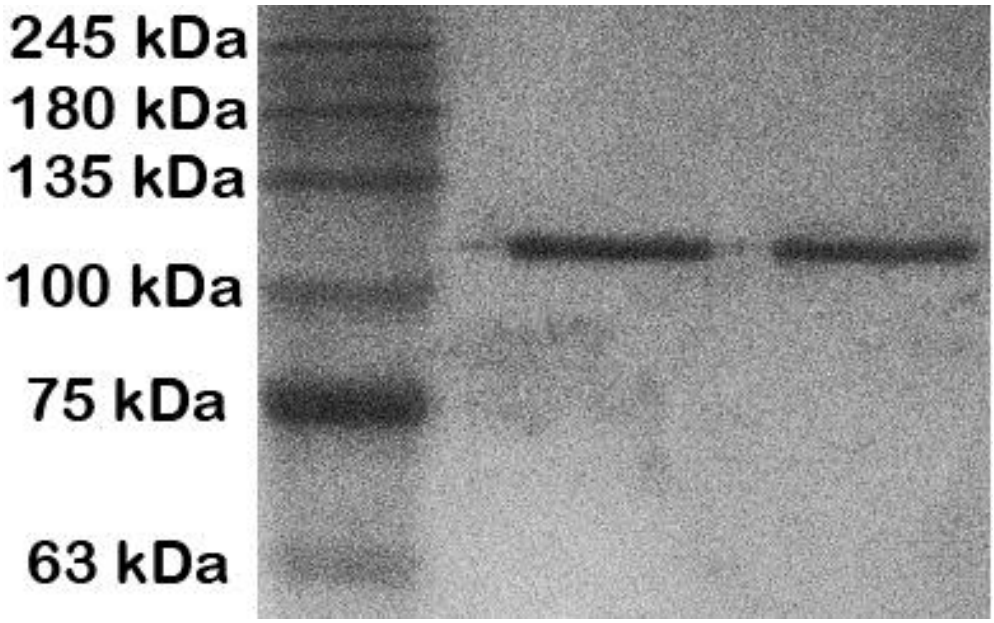

Figure 3.11 Silver stained SDS PAGE of CroV TF-MCP after affinity and size exclusion chromatography purification. Lane 1 is molecular marker. Two batches of purified TF-MCP were loaded on Lane 2 and Lane3.

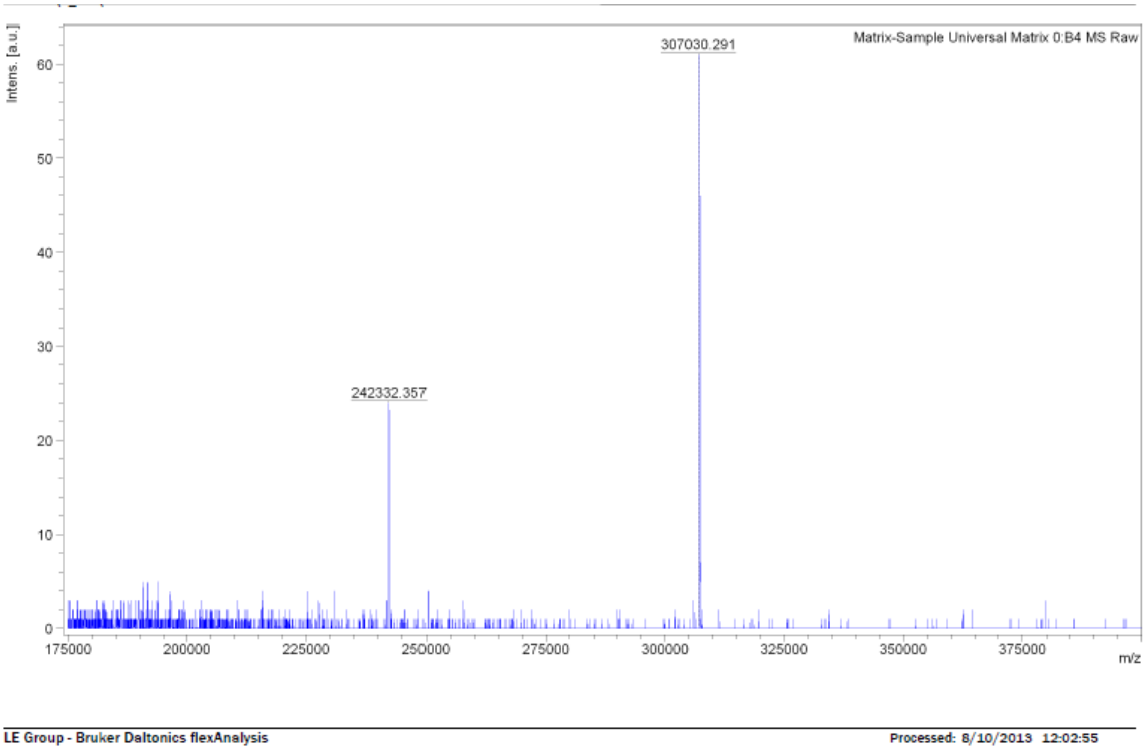

Figure 3.12 MALDIMS-TOF spectrum of recombinant TF-MCP. Two peaks at 242.33kDa and 307.03kDa were observed.

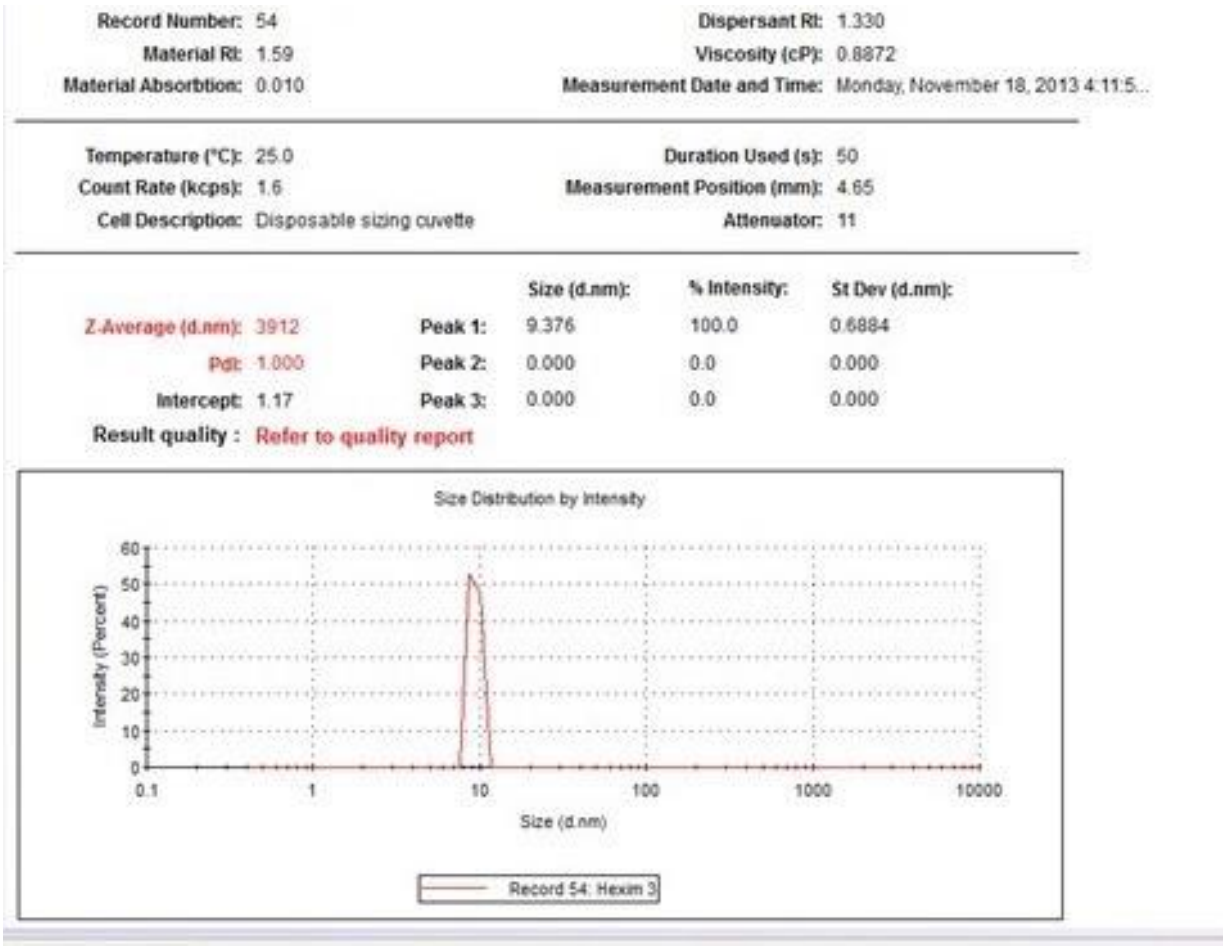

Figure 3.13. Average and standard deviation of DLS data of TF-MCP. Z average and standard deviation of Z average were showed above the size distribution profile.

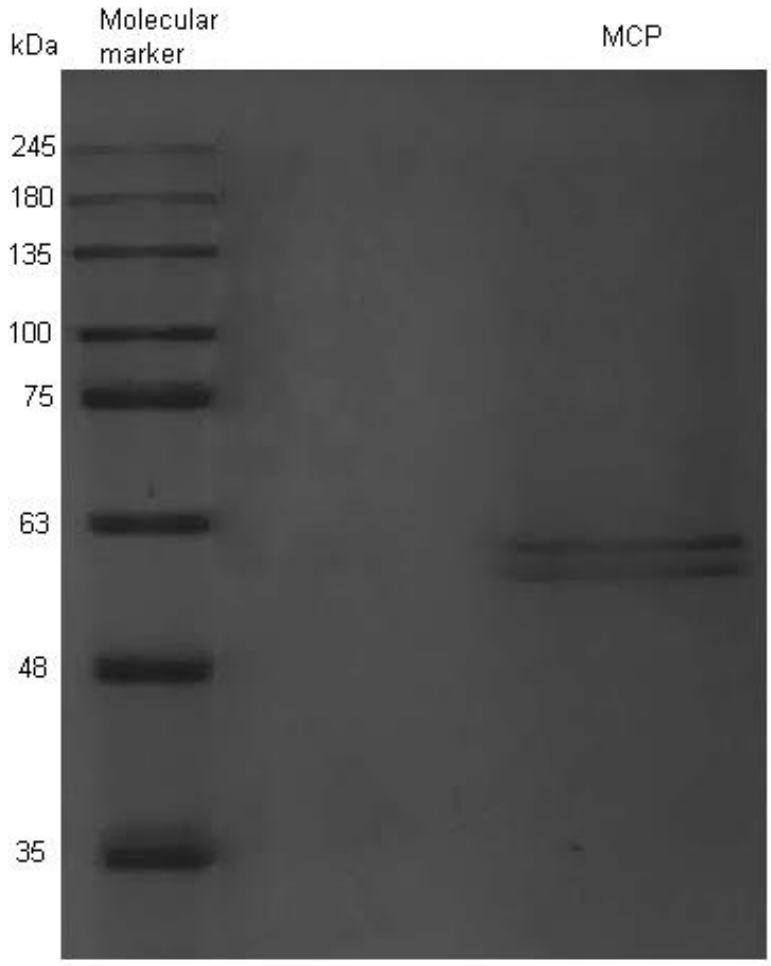

Figure 3.14 SDS PAGE of CroV MCP directly from purified virus sample. Lane 1 is molecular marker, 20µl of purified virus sample were loaded on Lane 3. Two major bands were observed.

CHAPTER 4

CONCLUSION AND FUTURE WORK

The aim of this project was to express, purify and characterize the CroV MCP protein. In this research, the major capsid protein (MCP) of a giant marine virus, CroV, was cloned and expressed in *E. coli* Bl21 (DE3) cells. It was observed that the recombinant MCP was insoluble when it was cloned and expressed using pColdI vector. Later MCP gene was cloned in pCold-TF expression vector in which the MCP was fused at its N-terminal with a cold shock protein chaperone TF that facilitated the proper folding of the protein. Sufficient amount of TF-MCP was detected in the soluble fraction of the cell lyses. TF-MCP was then purified by affinity chromatography and size exclusion chromatography, resulting in recombinant TF-MCP with high purity. The fusion protein was characterized by different biochemical and biophysical techniques, e.g., MALDIMS-TOF and DLS, which indicates that TF-MCP is in its native trimeric form. The results obtained from this research build a very solid foundation for structural studies of CroV MCP using X-ray crystallography and Cryo-EM.

X-ray crystallography is a powerful tool to determine the structure of a molecule to atomic resolution. Protein with high homogeneity will be crystallized and X-ray diffraction data will be collected by using the protein crystals. The diffraction pattern will be used to decipher the structure based on Bragg's law. Double-jelly-roll major capsid proteins of PBCV-1 have been crystallized and the structure has been determined by X-ray crystallography. PBCV-1 is another giant virus in NCLDV (Table 2.1). As CroV MCP is homologous to PBCV-1 MCP, it is very feasible to solve the structure by X-ray crystallography. Highly pure MCP has been obtained and crystallization screening trials will start shortly in future.

Cryo-electron microscopy (Cryo-EM) is a relatively new technique, in which biological samples are examined at a cryogenic temperature. Unlike X-ray crystallography, crystallization of the sample is not required in Cryo-EM. Cryo-EM is a useful tool for determining the structure of a large dynamic macromolecule that cannot be crystallized. Recently, the resolution of PBCV-1 and CIV Cryo-EM map has been significantly improved (Yan, Yu et al. 2009). Higher resolution of these structures has revealed the presence of minor capsid proteins(Yan, Yu et al. 2009). Higher resolution of Cryo-EM reconstruction of CroV will be used to compare with PBCV-1, CIV, and other NCLDVs.

X-ray crystallography and Cryo-EM will be combined together to study the intact virion structure of CroV. Since the dynamic giant virus particle cannot be crystallized, atomic structure of the individual viral component determined from X-ray crystallography can be fitted into Cryo-EM to generate a pseudo atomic model of the intact virus. The pseudo atomic model of CroV will build the foundation for future understandings about the capsid assembly, host virus recognition and genome delivery. Structural comparison between CroV and other NCLDVs will shed light on the evolutionary relationship among these giant viruses as well as between these giruses and cellular organisms.

REFERENCES:

1.Athappilly, F. K., R. Murali, J. J. Rux, Z. Cai and R. M. Burnett (1994). "The refined crystal structure of hexon, the major coat protein of adenovirus type 2, at 2.9 A resolution." J Mol Biol **242**(4): 430-455.

2. Baltimore, D. (1971). "Expression of animal virus genomes." Bacteriol Rev **35**(3): 235- 241.

3. Beijerinck, M. W. (1898). "Concerning a contagium vivum fluidum as a cause of the spot-disease of tobacco leaves." Verh Akad Wetensch, Amsterdam, **II**(6): 3-21.

4. Bergh, O., K. Y. Borsheim, G. Bratbak and M. Heldal (1989). "High abundance of viruses found in aquatic environments." Nature **340**(6233): 467-468.

5. Brussaard, C. P. D. (2004). "Viral control of phytoplankton populations - a review." Journal of Eukaryotic Microbiology **51**(2): 125-138.

6. Caspar, D. L. and A. Klug (1962). "Physical principles in the construction of regular viruses." Cold Spring Harb Symp Quant Biol **27**: 1-24.

7. Chothi, M. P., G. A. Duncan, A. Armirotti, C. Abergel, J. R. Gurnon, J. L. Van Etten, C. Bernardi, G. Damonte and M. Tonetti (2010). "Identification of an L-Rhamnose Synthetic Pathway in Two Nucleocytoplasmic Large DNA Viruses." Journal of Virology **84**(17): 8829-8838.

8. Cochlan, W. P., J. Wikner, G. F. Steward, D. C. Smith and F. Azam (1993). "Spatial-Distribution of Viruses, Bacteria and Chlorophyll-a in Neritic, Oceanic and Estuarine Environments." Marine Ecology Progress Series **92**(1-2): 77-87.

9. Colson, P., G. Gimenez, M. Boyer, G. Fournous and D. Raoult (2011). "The giant Cafeteria roenbergensis virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life." PLoS One **6**(4): e18935.

10. Colson, P., N. Yutin, S. A. Shabalina, C. Robert, G. Fournous, B. La Scola, D. Raoult and E. V. Koonin (2011). "Viruses with more than 1,000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes." Genome Biol Evol **3**: 737-742.

11. Crick, F. H. and J. D. Watson (1956). "Structure of small viruses." Nature **177**(4506): 473-475.

12. Fields, B. N., D. M. Knipe and P. M. Howley (2007). Fields virology. Philadelphia, Wolters Kluwer Health/Lippincott Williams & Wilkins.

13. Fischer, M. G. (2012). "Sputnik and Mavirus: more than just satellite viruses." Nature Reviews Microbiology **10**(1): 78-U92.

14. Fischer, M. G., M. J. Allen, W. H. Wilson and C. A. Suttle (2010). "Giant virus with a remarkable complement of genes infects marine zooplankton." Proceedings of the National Academy of Sciences of the United States of America **107**(45): 19508-19513.

15. Forterre, P. (2010). "Giant viruses: conflicts in revisiting the virus concept." Intervirology **53**(5): 362-378.

16. Forterre, P. (2011). "Manipulation of cellular syntheses and the nature of viruses: The virocell concept." Comptes Rendus Chimie **14**(4): 392-399.

17. Fuhrman, J. A. (1999). "Marine viruses and their biogeochemical and ecological effects." Nature **399**(6736): 541-548.

18. Garza, D. R. and C. A. Suttle (1995). "Large double-stranded DNA viruses which cause the lysis of a marine heterotrophic nanoflagellate (Bodo sp) occur in natural marine viral communities." Aquat Microb Ecol **9**: 203-210.

19. Graves, M. V., D. E. Burbank, R. Roth, J. Heuser, P. L. DeAngelis and J. L. Van Etten (1999). "Hyaluronan synthesis in virus PBCV-1-infected chlorella-like green algae." Virology **257**(1): 15-23.

20. Green, M. R., J. Sambrook and J. Sambrook (2012). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

21. Hara, S., I. Koike, K. Terauchi, H. Kamiya and E. Tanoue (1996). "Abundance of viruses in deep oceanic waters." Marine Ecology Progress Series **145**(1-3): 269-277.

22. Hara, S., K. Terauchi and I. Koike (1991). "Abundance of viruses in marine waters: assessment by epifluorescence and transmission electron microscopy." Appl Environ Microbiol **57**(9): 2731-2734.

23. Harrison, S. C. (2007). Chapter 2, Principles of Virus Structure. Fields' Virology. D. M. Knipe and P. M. Howley. Philadelphia, Wolters Kluwer Health/Lippincott Williams & Wilkins. **1:** 59-98.

24. Ivanofsky, D. (1892). "Concerning the mosaic disease of the tobacco plant." St. Peterburg Acad. Imp. Sci. Bull.(35): 67-70.

25. Iyer, L. M., S. Balaji, E. V. Koonin and L. Aravind (2006). "Evolutionary genomics of nucleo-cytoplasmic large DNA viruses." Virus Res **117**(1): 156-184.

26. Kausche, G. A., E. Pfankuch and H. Ruska (1939). "The visualisation of herbal viruses in surface microscopes." Naturwissenschaften **27**: 292-299.

27. Kawasaki, T., M. Tanaka, M. Fujie, S. Usami, K. Sakai and T. Yamada (2002). "Chitin synthesis in chlorovirus CVK2-infected chlorella cells." Virology **302**(1): 123-131.

28. Khayat, R., L. Tang, E. T. Larson, C. M. Lawrence, M. Young and J. E. Johnson (2005). "Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses." Proc Natl Acad Sci U S A **102**(52): 18944-18949.

29. Koonin, E. V. and V. V. Dolja (2006). "Evolution of complexity in the viral world: the dawn of a new vision." Virus Res **117**(1): 1-4.

30. La Scola, B., S. Audic, C. Robert, L. Jungang, X. de Lamballerie, M. Drancourt, R. Birtles, J. M. Claverie and D. Raoult (2003). "A giant virus in amoebae." Science **299**(5615): 2033.

31. Landsteiner, K. and P. E. (1909). "Ubertragung der Poliomyelitis acuta auf Affen." Z Immunitatsforsch Orig **2**: 377-390.

32. Lavrukhin, O. V., J. M. Fortune, T. G. Wood, D. E. Burbank, J. L. Van Etten, N. Osheroff and R. S. Lloyd (2000). "Topoisomerase II from Chlorella virus PBCV-1. Characterization of the smallest known type II topoisomerase." J Biol Chem **275**(10): 6915-6921.

33. Legendre, M., D. Arslan, C. Abergel and J. M. Claverie (2012). "Genomics of Megavirus and the elusive fourth domain of Life." Commun Integr Biol **5**(1): 102-106.

34. Loeffler, F. and P. Frosch (1989). "Report of the commission for research on foot-and-mouth disease." Zentrabl. Bacteriol. Parastenkunde Infektionkrankh **23**: 371-391.

35. Martinez, J. M., D. C. Schroeder, A. Larsen, G. Bratbak and W. H. Wilson (2007). "Molecular dynamics of Emiliania huxleyi and cooccurring viruses during two separate mesocosm studies." Applied and Environmental Microbiology **73**(2): 554-562.

36. Massana, R., J. del Campo, C. Dinter and R. Sommaruga (2007). "Crash of a population of the marine heterotrophic flagellate Cafeteria roenbergensis by viral infection." Environmental Microbiology **9**(11): 2660-2669.

37. Middelboe, M. and N. O. G. Jorgensen (2006). "Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds." Journal of the Marine Biological Association of the United Kingdom **86**(3): 605-612.

38. Middelboe, M. and P. G. Lyck (2002). "Regeneration of dissolved organic matter by viral lysis in marine microbial communities." Aquatic Microbial Ecology **27**(2): 187-194.

39. Monier, A., A. Pagarete, C. de Vargas, M. J. Allen, B. Read, J. M. Claverie and H. Ogata (2009). "Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus." Genome Res **19**(8): 1441-1449.

40. Moreira, D. and C. Brochier-Armanet (2008). "Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes." BMC Evol Biol **8**: 12.

41. Murray, A. G. and G. A. Jackson (1992). "Viral Dynamics - a Model of the Effects of Size, Shape, Motion and Abundance of Single-Celled Planktonic Organisms and Other Particles." Marine Ecology Progress Series **89**(2-3): 103-116.

42. Pedros-Alio, C., J. I. Calderon-Paz and J. M. Gasol (2000). "Comparative analysis shows that bacterivory, not viral lysis, controls the abundance of heterotrophic prokaryotic plankton." Fems Microbiology Ecology **32**(2): 157-165.

43. Philippe, N., M. Legendre, G. Doutre, Y. Coute, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J. M. Claverie and C. Abergel (2013). "Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes." Science **341**(6143): 281-286.

44. Poorvin, L., J. M. Rinta-Kanto, D. A. Hutchins and S. W. Wilhelm (2004). "Viral release of iron and its bioavailability to marine plankton." Limnology and Oceanography **49**(5): 1734-1741.
Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan and J. M. Claverie (2004). "The 1.2-megabase genome sequence of Mimivirus." Science **306**(5700): 1344-1350.

45. Raoult, D. and P. Forterre (2008). "Redefining viruses: lessons from Mimivirus." Nat Rev Microbiol **6**(4): 315-319.

46. Rodriguez-Valera, F., A. B. Martin-Cuadrado, B. Rodriguez-Brito, L. Pasic, T. F. Thingstad, F. Rohwer and A. Mira (2009). "OPINION Explaining microbial population genomics through phage predation." Nature Reviews Microbiology **7**(11): 828-836.

47. Shuman, S. (1994). "Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase." J Biol Chem **269**(51): 32678-32684.

48. Stanley, W. M. (1935). "Isolation of a Crystalline Protein Possessing the Properties of Tobacco-Mosaic Virus." Science **81**(2113): 644-645.

49. Suttle, C. A. (2007). "Marine viruses--major players in the global ecosystem." Nat Rev Microbiol **5**(10): 801-812.

50. Thompson, J. R., S. Pacocha, C. Pharino, V. Klepac-Ceraj, D. E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D. L. Distel and M. F. Polz (2005). "Genotypic diversity within a natural coastal bacterioplankton population." Science **307**(5713): 1311-1313.

51. Van Etten, J. L. (2003). "Unusual life style of giant chlorella viruses." Annu Rev Genet **37**: 153-195.

52. Van Etten, J. L. (2011). "Another Really, Really Big Virus." Viruses-Basel **3**(1): 32-46.

53. Van Etten, J. L., L. C. Lane and D. D. Dunigan (2010). "DNA Viruses: The Really Big Ones (Giruses)." Annual Review of Microbiology, Vol 64, 2010 **64**: 83-99.

54. Vogan, A. A. and P. G. Higgs (2011). "The advantages and disadvantages of horizontal gene transfer and the emergence of the first species." Biol Direct **6**: 1.

55. Weinbauer, M. G. and F. Rassoulzadegan (2004). "Are viruses driving microbial diversification and diversity?" Environ Microbiol **6**(1): 1-11.

56. Wilkinson, D. L. and R. G. Harrison (1991). "Predicting the solubility of recombinant proteins in Escherichia coli." Biotechnology (N Y) **9**(5): 443-448.

57. Xiao, C. and M. G. Rossmann (2011). "Structures of giant icosahedral eukaryotic dsDNA viruses." Current Opinion in Virology **1**(2): 101-109.

58. Xiao, C. and M. G. Rossmann (2011). "Structures of giant icosahedral eukaryotic dsDNA viruses." Curr Opin Virol **1**(2): 101-109.

59. Yan, X., Z. Yu, P. Zhang, A. J. Battisti, H. A. Holdaway, P. R. Chipman, C. Bajaj, M. Bergoin, M. G. Rossmann and T. S. Baker (2009). "The capsid proteins of a large, icosahedral dsDNA virus." J Mol Biol **385**(4): 1287-1299.

60. Yan, X. D., P. R. Chipman, T. Castberg, G. Bratbak and T. S. Baker (2005). "The marine algal virus PpV01 has an icosahedral capsid with T=219 quasisymmetry." Journal of Virology **79**(14): 9236-9243.

61. Yutin, N. and E. V. Koonin (2012). "Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes." Virol J **9**: 161.

62. Yutin, N., Y. I. Wolf, D. Raoult and E. V. Koonin (2009). "Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution." Virol J **6**: 223.

63. Zakaib, G. D. (2011). "The challenge of microbial diversity: Out on a limb." Nature **476**(7358): 20-21.

VITAE

Sayan Chakraborty received his undergraduate degree, Bachelor in Technology from West Bengal University of Technology, India, in 2010 with a GPA of 7.68/10. After the completion of undergraduate studies, Sayan joined a research laboratory in Calcutta School of Tropical Medicine as a research assistant. One of his research works was selected for poster presentation in 14th International Congress of Immunology 2010 in Kobe, Japan. Another research work was also selected for poster presentation in British Society of Immunology Conference 2011 held in Liverpool, UK. Later he decided to pursue higher studies and in 2011 he joined graduate program in Chemistry Department in University of Texas at El Paso under the guidance of Dr. Chuan Xiao (River). During his graduate studies, he served as a teaching assistant for general chemistry laboratory (Chemistry 1105). In his Master's thesis, entitled "Expression and characterization of the major capsid protein (MCP) of a giant marine virus: *Cafeteria roenbergensis* virus (CroV)", Sayan has mainly focused on obtaining pure MCP expressed in bacterial cell line.

Since his high school days, Sayan has participated in different extra-curricular activities. He has won several prizes in interprovincial debate and quiz competitions. Sayan was an active member of rural education program organized by UNESCO. Currently, Sayan is an active member of Indian student association here in UTEP and serving as treasurer of the organization.

Achievements:

- Was selected for POSTER presentation in 14th International Congress of Immunology 2010 (ICI 2010) was held in Kansai, Japan on 22nd August – 27th August

- Was selected for POSTER presentation in British Society of Immunology Conference 2011 held in Liverpool, UK on December 5th -8th.
- Participated in a seminar on "Innovative Alleviations and Potential Explorations in Cancer"
- Was a part of a seminar on – "ONCON-cancer research"
- Presented a Seminar on SCID(severe combined immunodeficiency) at Haldia Institute of Technology

Permanent address: 134/2/2 Upen Banerjee Road

Kolkata, West Bengal, India, zip-700060

This thesis/dissertation was typed by Sayan Chakraborty.