Purdue University
Purdue e-Pubs

Open Access Dissertations | Theses and Dissertations

Fall 2013

Structural Studies On The Rubella Virus Capsid Protein And Its Organization In The Virion

Vidya Mangala Prasad
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

 Part of the Biophysics Commons, Molecular Biology Commons, and the Virology Commons

Recommended Citation
Mangala Prasad, Vidya, "Structural Studies On The Rubella Virus Capsid Protein And Its Organization In The Virion" (2013). *Open Access Dissertations*. 89.
https://docs.lib.purdue.edu/open_access_dissertations/89

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Vidya Mangala Prasad

Entitled
STRUCTURAL STUDIES ON THE RUBELLA VIRUS CAPSID PROTEIN AND ITS ORGANIZATION IN THE VIRION

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Wen Jiang
Chair

Michael G. Rossmann

Richard J. Kuhn

Marc Loudon

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Michael G. Rossmann

Approved by: Peter Hollenbeck 11/22/2013
Head of the Graduate Program Date

STRUCTURAL STUDIES ON THE RUBELLA VIRUS CAPSID PROTEIN AND ITS ORGANIZATION IN THE VIRION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Vidya Mangala Prasad

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2013

Purdue University

West Lafayette, Indiana

To my parents and my sister

ACKNOWLEDGEMENTS

I would like to start by thanking Dr. Michael G. Rossmann for giving me the opportunity to do my graduate studies in his laboratory. He has been a major source of inspiration for me. His love for his work, his enthusiasm for science and his curiosity towards everything around him has been a motivational factor for my work. I thank him for his guidance and for instilling my self-confidence in my abilities. I am grateful for all the interesting discussions and challenges that I have been exposed to during my years in his laboratory. These have been crucial for the success of my projects and also for my personal development.

I would also like to thank my committee members, Dr. G. Marc Loudon, Dr. Wen Jiang and Dr. Richard J. Kuhn; for their advice, suggestions and comments that have helped me through my PhD training. I would also like to thank our collaborator, Dr. Tom Hobman, and his graduate student, Steven Willows, for their help and involvement in my projects.

I consider myself lucky to have had the chance to spend my PhD life in a big lab. All the lab members have not only been very nice to me, but I have also had the opportunity to observe, experience and learn many things from them, scientifically and otherwise. I thank Andrei Fokine for helping me learn crystallography and also for answering all my queries during this process. I thank Siyang Sun for helping me getting started with my project in the lab and for teaching me to work with crystals and collect x-ray data at the Advance Photon Source. I thank Anthony Battisti for teaching me cryo-electron tomography and for the interesting discussions regarding the rubella virus. I thank Pavel Plevka for helping me with crystallography and electron microscopy, and for the various stimulating scientific discussions that I have had with him. I thank Thomas Klose for all his help and suggestions for my project work, and also for his insights on various topics

that have helped me shape my mind. I would also like to thank my other former and current lab members, Ye Xiang, Long Li, Anastasia Aksyuk, Moh Lan Yap and Jimson Wong for their help, suggestions and advice; which have all in different ways positively influenced my graduate life.

I would like to extend special thanks to Sheryl Kelly for helping with my manuscripts and other administrative works. I thank Steve Wilson for his excellent IT support. I also thank Valorie Bowman and Agustin Avila-Sakar for their exceptional help and support with using the electron microscopes.

Finally, I would like to thank my family and friends for their support. I thank my parents and my sister for their trust and confidence in me. Their high expectations and complete faith in my abilities to achieve them, have constantly motivated me to push myself to do my best. I thank my fiancé for being my anchor in life, for sharing my dreams and aspirations. I am eternally grateful to all of them for their unwavering support to my decisions in life.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Appendix Figure

LIST OF ABBREVIATIONS

µcal	micro calories
µm	micro meter
µM	micro molar
3D	three dimensional
2D	two dimensional
Å	angstrom
ADE	Antibody dependent enhancement
BHK	baby hamster kidney
BSA	bovine serum albumin
CCD	charge-coupled detector
cDNA	complementary DNA
CRS	congenital rubella syndrome
C-terminal	carboxy terminal
CTF	contrast transfer function
DENV-2	dengue virus type 2
DMEM	Dulbecco modified Eagle medium
DMSO	dimethyl sulfoxide
E.coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
ER	endoplasmic recticulum
Fab	Fragment antigen binding
FBS	fetal bovine serum
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase

GST	Glutathione S-transferase
g/l	grams/liter
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropyl thiogalactoside
ITC	Iso-thermal-calorimetry
kcal	kilo calories
kDa	kilodaltons
kV	kilo volts
LB	*Luria Bertani*
LDAO	lauryl-dimethylamine oxide
MAD	multiple wavelength anomalous dispersion
mAu	milli absorption units
MEM	Minimum Essential Medium
mg	milli grams
MIR	multiple isomorphous replacement
ml	milli liters
mM	milli molar
MOG	myelin oligodendrocyte glycoprotein
MR	molecular replacement
mRNA	messenger RNA
mV	milli volts
NCBI	National Centre for Biotechnology Information
NCS	non-crystallographic symmetry
NEAA	non-essential amino acids
NHS	N-Hydroxysuccinimidyl
nm	nano meter
N-terminal	amino terminal
O.D	optical density

Par-4	prostate apoptosis response - 4
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pI	iso-electric point
RMSD	root mean square difference
rpm	revolutions per minute
RV	rubella virus
RVC	rubella virus capsid protein
S	sedimentation coefficient
SAD	single wavelength anomalous dispersion
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sE	soluble E ectodomain
Se-Met	selenomethionine
SIRAS	single isomorphous replacement with anomalous scattering
TRAIL	TNF (tumor necrosis factor) related apoptosis inducing ligand
UV	ultra violet
VLP	virus-like particles
Wt	wild type

ABSTRACT

Mangala Prasad, Vidya. Ph.D., Purdue University, December 2013. Structural studies on the rubella virus capsid protein and its organization in the virion. Major Professor: Michael G. Rossmann.

Rubella virus is a leading cause of birth defects due to infectious agents. When contracted during pregnancy, rubella infection leads to severe damage in fetuses. Despite its medical importance, very little is known about the structure of the pleomorphic rubella virus as compared to its alphavirus relatives. The rubella capsid protein is a critical structural component of virions as well as a key factor in virus-host interactions. Three crystal structures of the structural domain of the rubella capsid protein have been described here. The polypeptide fold of the capsid protomer has not been observed previously. The capsid protein structure, along with cryo-electron tomograms of rubella virus particles and mutational studies on the capsid protein, provides a low resolution structure of the virus particle. Nucleocapsid assembly studies on the rubella capsid protein has given additional information on the factors affecting formation of the virion cores. Together, these studies enumerate critical differences between rubella virus and alphaviruses that might affect their specific pathogenicities.

CHAPTER 1. INTRODUCTION TO RUBELLA VIRUS

1.1 Overview

Rubella virus is the only member of the Rubivirus genus, which, together with alphaviruses including Chikungunya virus, Sindbis virus and Ross River virus, make up the virus family *Togaviridae. 'Toga'* literally means cloak in Latin. Rubella virus is a human pathogen that causes "German measles", a relatively mild disease characterized by rashes and low-grade fever. Its symptoms include fever, rashes, joint pains and swollen glands. The virus spread occurs through nasal or throat secretions from infected individuals. It is generally mild in case of children and adults. Rubella infection has serious implications during the first trimester of pregnancy which results in birth defects. This condition is called congenital rubella syndrome (CRS). Rubella virus has only one serotype and a one-time infection provides life-long immunity against it (Hobman and Chantler, 2007). The prevention of the disease has been highly successful in the United States and other parts of the world with the introduction of live attenuated virus vaccines since 1969 (Centers for Disease Control and Prevention, 2010; Hobman and Chantler, 2007). Despite these measures, the virus still exists as an important human pathogen (Frey, 1994). There is no specific treatment available for the disease.

Rubella virus is an enveloped, single-stranded, RNA virus. The particle diameters of the virions range from 600-800Å, with most of the roughly spherical virions having a diameter of approximately 700Å (Battisti *et al*., 2012a) (**Figure 1.1**). The virus consists of a 40S single-stranded positive sense RNA genome enclosed within a nucleocapsid core, which is surrounded by a membrane-envelope with glycoproteins embedded in it. The RV genome is non-segmented and has a GC content (69.5%) higher than any other known RNA virus (Frey, 1994). The structural proteins of the virus are formed as a polyprotein precursor from a 24S sub-genomic RNA found in infected cells. This polyprotein is later

cleaved to give 3 proteins; the capsid protein (~32kDa), the E1 (58kDa) and the E2 (42-54kDa) glycoproteins (Oker-Blom *et al*., 1984a). The RV genome also encodes two non-structural proteins, p150 and p90, which are translated from the genomic RNA (Frey, 1994).

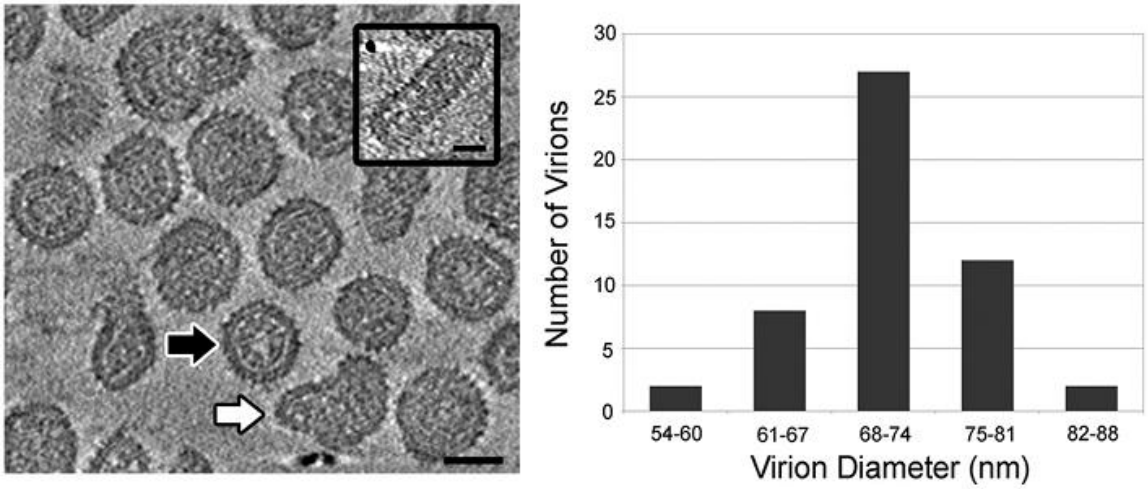


Figure 1.1 Rubella virion morphology. Left: Cryo-electron microscopy image of rubella virions. The black arrow indicates a roughly spherical virion and the white arrow shows an irregular shaped virion. The inserted panel on the right-top of the image shows a cylinder-shaped virion. Scale bar indicates 500 Å. Right: Bar graph showing distribution of rubella virion diameters (Battisti.*et al*., 2012a).

1.2 Rubella virus proteins

1.2.1 Structural proteins

Rubella virus capsid protein

The rubella virus capsid protein forms the nucleocapsid which contains and interacts with the viral genome. The capsid protein is present as a disulphide linked dimer in the virion. The capsid monomer consists of 277 amino acids and has a molecular weight of about 31 kDa. It has a high pI (iso-electric point) value of 10.6 and its amino acid sequence is rich in arginine and proline residues. The amino terminal region of approximately 100 amino acids is highly basic and interacts with the viral genomic RNA to form the nucleocapsid (Baron and Forsell, 1991; Frey, 1994). The amino acid region of residues 28 to 56 of the capsid protein binds specifically to nucleotides 347-375 of the viral RNA (Liu *et al*., 1996). Phosphorylation of capsid protein, especially Serine-46, restricts binding of RNA to the capsid protein. Thus, phosphorylation and de-phosphorylation of the capsid protein is regulated in infected cells to coincide with budding of the virus (Law *et al*., 2003).

The capsid protein is synthesized along with the glycoproteins as a polyprotein precursor. The carboxy terminus of the capsid protein is connected to the amino terminus of the E2 glycoprotein by a signal peptide. This signal peptide targets both these proteins to the endoplasmic recticulum (ER). The E2 glycoprotein gets transported to the ER lumen and gets cleaved off by a host cell signal peptidase. The signal peptide remains attached to the capsid protein anchoring it to the ER membrane (Suomalainen *et al*., 1990) (**Figure 1.2**). The above mentioned capsid protein length excludes the E2 signal peptide.

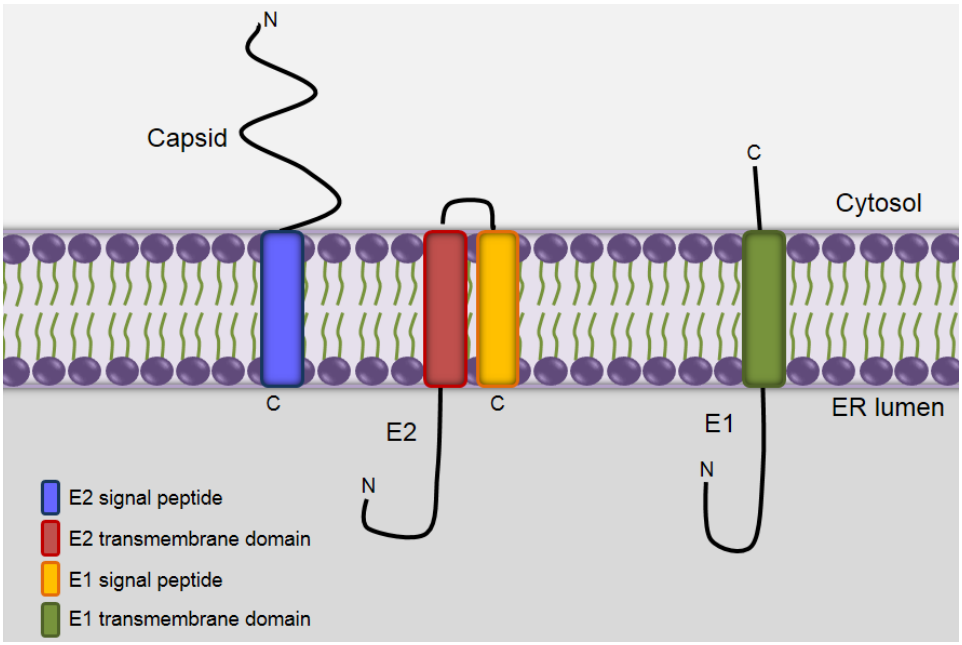


Figure 1.2 Position of the rubella virus structural proteins inside a cell. The different proteins are shown after post-translational cleavage, along with the signal peptides and transmembrane domains. The letters N and C denote the amino and carboxy terminus of the proteins respectively. Figure reproduced from Mangala Prasad, *et al*., 2013.

Apart from forming the nucleocapsid, interaction of the capsid protein with the E1 and E2 glycoproteins is required for assembly and budding of the virion (Qiu *et al*., 1994). The capsid protein also has other non-structural functions in the infected host cell. The capsid protein has been reported to regulate the viral RNA transcription and replication (Chen *et al*., 2004). It inhibits protein translation of host-cell mRNAs by binding to the initiation factor PABP (Ilkow *et al*., 2008). The capsid protein causes dramatic changes in the morphology and distribution of mitochondria in infected cells (Beatch *et al*., 2005). It also inhibits mitochondrial import in infected cells (Ilkow *et al*., 2010). It has also been shown to bind to proteins such as mitochondrial matrix protein (p32), prostate apoptosis response (Par-4) protein and Bax, which are involved in apoptotic pathways in host cells, thereby influencing apoptosis mechanisms in infected cells (Beatch and Hobman, 2000; Beatch *et al*., 2005; Ilkow *et al*., 2011) . This function may form the basis for the observed

delay in lytic release of virus during rubella infection as compared to other RNA viruses (Ilkow *et al*., 2011). The multiple functions of the capsid protein are depicted in **Figure 1.3**.

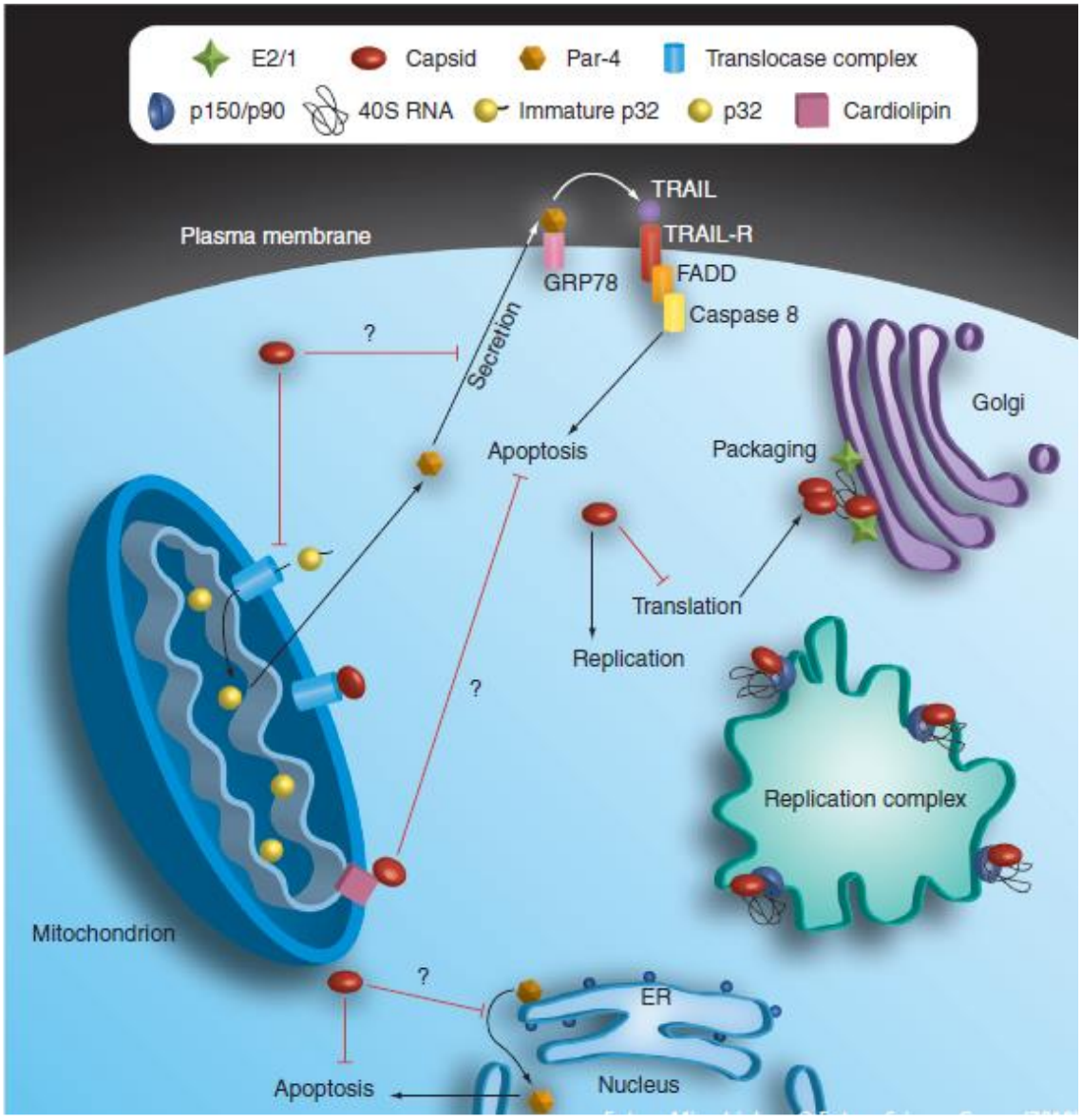


Figure 1.3 Different functions of rubella capsid protein in an infected cell. Figure reproduced from Ilkow *et al*., 2010. ER - Endoplasmic reticulum, TRAIL- TNF related apoptosis inducing ligand.

Rubella virus envelope glycoproteins

Rubella virus contains two glycoproteins, E1 and E2, that are arranged on the exterior of the viral membrane. Both proteins are N-glycosylated with the E2 protein having O-glycosylation sites also. The rubella virus glycoproteins are class I membrane proteins that form heterodimers.

The E1 glycoprotein is 481 amino acids long and is more exposed on the virion surface than E2. Accordingly, the E1 protein is the major antigenic determinant (Katow *et al*., 1985) on the virus. It is responsible for receptor recognition (Cong *et al*., 2011) and membrane fusion of the virus with the endosomal membranes at low-pH (Yang *et al*., 1998; Katow *et al*., 1988). The E1 glycoprotein binds to the recently identified receptor, myelin oligodendrocyte glycoprotein (MOG) (Cong *et al*., 2011).

The E2 glycoprotein is 282 amino acid residues in length. It is essential for proper folding and maturation of the E1 glycoprotein (Garbutt *et al*., 1999; DuBois *et al*, 2013). E2 is not easily accessible to glycosidases (Ho-Terry and Cohen, 1980) and monoclonal antibodies (Pettersson *et al*., 1985), which implies that E2 is probably buried in the membrane and/or concealed by the E1 on the surface.

1.2.2 Non-structural proteins

The non-structural proteins of rubella virus are translated as a polyprotein precursor from the viral genomic RNA. The polyprotein is later cleaved into two products, p90 and p150, through the proteolytic activity of p150's carboxy terminal region. These two proteins together form the viral replicase (Hobman and Chandler, 2007).

1.3. Virus replication and assembly

On binding to a cellular receptor, rubella virions are internalized through receptor-mediated endocytosis (Petruzziello, 1996). Uncoating of the virus takes place in the acidic environment within the endosome leading to release of viral genomic RNA into the cytoplasm. Replication of the viral genome and subsequent production of viral proteins in the infected cells causes big changes in the intracellular organelle distribution and organization (Lee *et al*. 1996). Formation of cytopathic vacuoles occurs within infected cells. This is a characteristic of togaviruses (Hobman and Chantler, 2007). These cytopathic vacuoles are derived from the lysosomes in cells and are sites for the viral replication complexes (Lee *et al*., 1994; Magliano *et al*., 1998). The cytopathic vacuoles are frequently found in association with mitochondria, the ER and Golgi stacks (Risco *et al*., 2003).

On translation, the structural polyprotein of rubella virus is transported to the ER. The glycoproteins E1 and E2 are transported into the ER lumen while the capsid protein remains attached to the cytoplasmic side of the ER membrane (Suomalainen *et al*., 1990). The structural polyprotein is cleaved into its corresponding proteins by the host cell signalase (Clarke *et al*., 1987; Oker-Blom *et al*., 1990). Glycosylation and other post-translational modifications of the glycoproteins occur within the ER. The structural proteins subsequently accumulate in the Golgi complex, where the assembly of virion particles takes place. The transmembrane regions of the E2 glycoprotein contain a Golgi retention signal and is predicted to determine the site of virion assembly (Garbutt *et al*., 1999). Interactions between the cytoplasmic domains of the glycoproteins and the capsid protein possibly form the trigger for virus budding (Garbutt *et al*., 1999; Qiu *et al*., 1999). The physical association observed between the Golgi stacks and the viral replication complexes form the basis for passage of the newly synthesized viral genome RNA to the sites of virus assembly (Fontana *et al*., 2010). Rubella virus-like particles can be formed by expressing only the structural proteins in cultured cell lines. This indicates that the presence of viral genomic RNA is not essential for the assembly of virions (Hobman *et al*., 1994).

The budding of rubella virus particles occurs on the Golgi membranes. The nucleocapsid assembly of rubella virions occurs in synchronization with budding of the

virus particles (Frey, 1994). Freeze-substitution electron microscopy studies of cells infected with rubella virions show that the rubella virions first assemble and bud as an immature form of the virus. These ‘immature’ virions are uniformly dense particles. These later undergo morphogenesis to form the mature virus particles during its passage through the Golgi. The mature virion particles are then secreted out of the cells through the trans-Golgi network (Risco *et al*., 2003).

1.4 Comparison between the Togavirus genera

Togaviruses are enveloped, positive-sense, single-stranded RNA viruses. The two genera under this family are the alphaviruses and the rubivirus. The rubivirus genus has only one member, the rubella virus. The alphavirus genus, on the other hand, includes a diverse class of viruses that include 26 different species. These infect both birds and mammals. Some of the well known alphaviruses are the Chikungunya virus, Sindbis virus, Semiliki forest virus, Ross River virus and the equine encephalitis viruses (Griffin, 2007).

Alphaviruses and rubella virus have a similar sized genome. The gene order is also similar (Oker-Blom *et al*.,1984b). Both these sets of viruses have membrane associated replication complexes and virus assembly sites (Strauss and Strauss, 1994; Fontana *et al*., 2010). There is also significant structural similarity between the E1 glycoproteins from alphaviruses and rubella virus (DuBois *et al*., 2013).

Beyond these similarities, the alphaviruses and rubella virus differ on many counts. Though they share similar genome organization, alphaviruses and rubella virus have very little homology in their nucleotide or amino acid sequences (Oker-Blom *et al*., 1984b). Alphaviruses are uniformly sized, icosahedral particles whereas the rubella virions occur in different shapes and sizes. The alphavirus capsid protein is an auto-protease that assembles into nucleocapsids in the cytoplasm. The virus budding occurs at the plasma membrane (Strauss and Strauss, 1994). In contrast, the rubella virus capsid protein does not possess any protease activity and is associated with the intracellular membranes. Nucleocapsid assembly of the rubella virus takes place simultaneously with virus budding on the Golgi membranes (Frey, 1994).

CHAPTER 2. STRUCTURE DETERMINATION OF THE RUBELLA VIRUS CAPSID PROTEIN

The work described in this chapter has been accepted for publication (Mangala Prasad *et al*., PNAS, 2013 (in press)). Parts of the chapter that were taken verbatim from the paper are marked by vertical lines in the right margin.

2.1 Introduction

High resolution structural information on the pleomorphic rubella virus has long been elusive. This is mainly because the currently available high resolution techniques for virus studies, x-ray crystallography and cryo-electron microscopy, depend on sample homogeneity. Studying the individual structural components of the virus is an indirect method to gather information about the structure of the whole virus.

The rubella capsid protein is a key structural protein in the rubella virion. It forms the inner nucleocapsid core of the virus and contains the viral RNA genome. The capsid protein is also known to have several other cellular functions that are required for the successful replication and assembly of the virus (Ilkow *et al*., 2010). Secondary structure prediction for the rubella capsid protein indicates that about 120 residues at the amino terminus of the protein are disordered. This disorder is similar to that in viruses such as the picornaviruses (Rossmann and Johnson, 1989) and alphaviruses (Cheng *et al*., 1995; Mukhopadhyay *et al*., 2006), in which the amino termini of the capsid protein are disordered and interact with the viral RNA. Comparison of rubella capsid protein secondary structure prediction with the known secondary structure of the capsid protein of Sindbis virus (an alphavirus) suggested that approximately 100 amino terminal residues are highly basic and disordered in both proteins. In alphaviruses, approximately 100 amino terminal residues of its capsid protein are associated with viral RNA (Geigenmuller-Gnirke *et al*., 1993, Cheng *et al*., 1995). Thus, it is likely that in the rubella virus nucleocapsid

also, the structural unit is formed by the carboxy terminal fragment of its capsid protein

The atomic structure of the capsid protein may help to gain some information on the molecule and its arrangement in the nucleocapsid core. It might also help to explain the diverse functions attributed to the protein.

To obtain the high resolution structural information on the rubella capsid protein, x-ray crystallography has been used. In this technique, an x-ray diffraction pattern obtained from the protein crystal is utilized to determine the arrangement of atoms in the protein structure. To obtain structural information, both intensities and phases of the diffracted x-rays are required. X-ray diffraction data of a protein crystal gives only the intensity measurements of the diffracted x-ray spots. To obtain phase information, different methods such as Molecular Replacement, Multiple Isomorphous Replacement and Multiple wavelength Anomalous Dispersion are used.

2.1.1 The Molecular Replacement method (MR)

The MR method can be used to obtain trial phases for an unknown protein structure if a suitable atomic model is available (Rossmann and Blow, 1962). This may be any other protein or part of it with a similar structure to the protein of interest. This method consists of two major steps, the rotation function and the translation function. The rotation function is used to determine the orientation of the structure in the unit cell by super-imposition of the Patterson function of the known structure model with the Patterson function of the unknown structure.

A Patterson function is a mathematical function with a form similar to the electron density function, but with phases as zero. The peaks in this map denote the inter-atomic vectors between atoms in the unit cell (see McPherson, 2003). Thus, when the relative orientation of the structure model is similar to the unknown structure in the crystal unit cell, the overlap between the two Patterson functions will be highest.

Mathematically, the rotation function (R) can be expressed as (Rossmann and Blow, 1962):

$$R = \Sigma\, [|F_h|^2\ \{\Sigma|F_p|^2\, G\}]$$

where F_h and F_p are the structure factor amplitudes of the unknown structure and the model respectively. G is an interference function which has large values only when the position of a reciprocal lattice point 'p' corresponding to the model structure is brought close by rotation to the position of reciprocal lattice point 'h' corresponding to the unknown structure.

Once the orientation of the molecule is determined, the translation function is used to place the model at a position in the unit cell that best fits the observed data. Different methods can be used for this calculation. Typically, many of these methods are Fourier methods. The Phaser program uses the Maximum Likelihood Translation function (McCoy, 2005).

2.1.2 The Multiple Isomorphous Replacement method (MIR)

The MIR technique involves using native crystal diffraction data along with diffraction data from different isomorphous heavy atom derivatives of the native protein to determine protein structures. This technique was used to determine the first protein structures, myoglobin (Kendrew *et al*., 1960) and hemoglobin (Perutz *et al*., 1960), which marked the start of the protein crystallographic era.

Heavy atom derivatives of protein crystals can be obtained either by soaking the crystal or co-crystallization of the protein with heavy atom compounds. A difference Patterson map between the structure factor amplitudes of native and derivative data can then be used to obtain the position of the heavy atoms. The heavy atom positions can be used to calculate their corresponding phases.

Assuming that the presence of a heavy atom does not change the structure of the protein significantly, the structure factors of the derivative crystal ($\mathbf{F}_{PH}$) can be expressed

as a vector sum of the structure factors of the native crystal ($\mathbf{F}_P$) and the heavy atom ($\mathbf{F}_H$) (see Ladd and Palmer, 2003):

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

where bold characters indicate vectors.

The above equation can be represented graphically as in **Figure 2.1**.

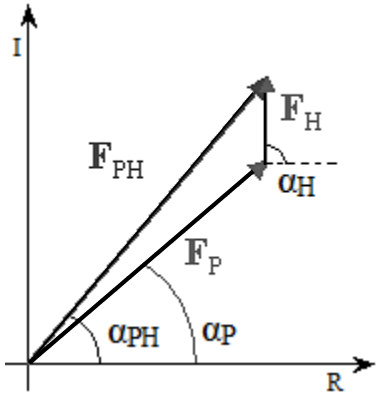


Figure 2.1 Graphical representation of the isomorphous replacement equation. The phase angles are denoted by α. The R and I represent the real and imaginary axes respectively.

Using the structure factors of the heavy atom, the phases for the protein structure can be calculated. This can be represented geometrically using the Harker diagram (Harker, 1956). An example of a Harker diagram is shown in **Figure 2.2**.

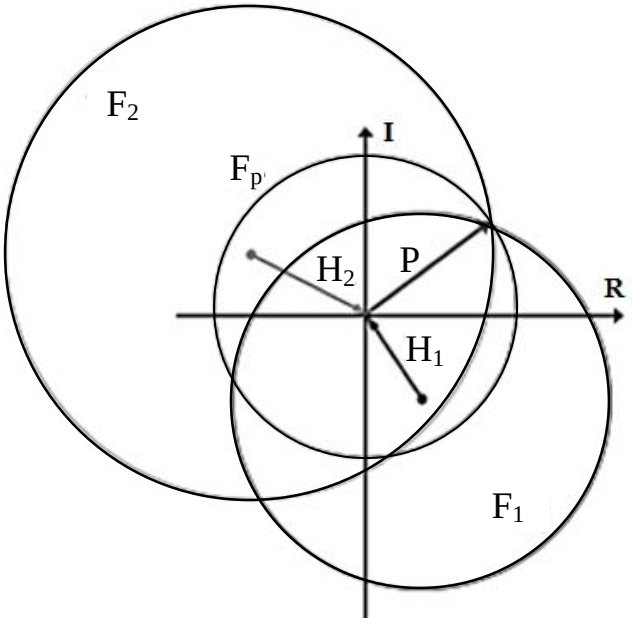


Figure 2.2 Harker diagram using two heavy atom derivatives. The circle denoted as F_P is drawn using the origin as centre with a radius equal to the amplitude of the native protein structure factor. The circles denoted F_1 and F_2 are drawn with radius equal to the amplitude of the respective heavy atom derivative protein structure factors. However, the centres for these circles are at the end of the vector for their respective heavy atom structure factors (H_1 and H_2). The point at which the three circles co-incide gives the phase angle for the protein structure factor.

As can be deduced from the above descriptions, isomorphous replacement can also be done with only one heavy atom derivative for the protein. In this case, an ambiguous result for the phase angle is obtained. This can be seen from the Harker diagram (**Figure 2.2**). If there is only one derivative, there will only be two circles in the diagram, which will result in two positions of intersection between them. The correct phase angle then needs to be determined by testing the different solutions. Another solution for finding the correct phase angle would be to use the anomalous scattering from the heavy atoms to resolve the phase ambiguity (Blow and Rossmann, 1961). This method is known as SIRAS

(Single isomorphous replacement with anomalous scattering). The concept of anomalous scattering is explained in detail in the next section.

2.1.3 The Multiple Anomalous Disperson method (MAD)

The MAD method is most useful when the native protein crystal and the heavy atom derivative protein crystal are non-isomorphous. MAD datasets are also collected on heavy atom derivative crystals, but at different wavelengths. The MAD method exploits the anomalous scattering property of heavy atoms.

All elements absorb x-rays. The absorption increases with increasing atomic number of the element. For every element, there is an energy value when the energy of the x-ray radiation is sufficient to eject an electron from a specific electron shell in its atom. This point is called the element's absorption edge. When x-rays at wavelengths close to the absorption edge of an atom are used, the scattering by the atom becomes anomalous. Anomalous scattering of an atom introduces changes in the phase of its atomic scattering factor. The atomic scattering factor (f) can now be expressed as a complex number (Rossmann, 1961):

$$f = f_0 + \Delta f' + i\,\Delta f''$$

where f_0 is the wavelength independent scattering factor and f' and f'' are the two wavelength-dependent scattering factors. The $\Delta f'$ and $\Delta f''$ values can be calculated for a specific atom based on the energy of the x-ray wavelength used.

Thus, at the absorption edge, an atom's diffraction spectra no longer follows Friedel's law. Friedel's law can be simply stated as (see Ladd and Palmer, 2003):

$$|F(h, k, l)| = |F(-h, -k, -l)|$$

where F is the amplitude of the structure factor of the diffraction spot and (h, k, l) denotes the position of the spot in reciprocal space.

A Pattersson synthesis with coefficients as the square of the difference between the amplitudes of the Friedel pairs can then be used to identify the position of the heavy atoms

(Rossmann, 1961). Phases for the heavy atoms can then be calculated and subsequent trial phases for the native protein data obtained.

Once the phases are obtained using any of the techniques described above, an electron density map can be calculated using all reflections and an atomic model built to fit the electron density. Iterative refinement procedures can then be used to improve the model and reduce differences between the observed and calculated structure factors.

2.2 Methods

2.2.1 Cloning

The nucleotide sequence encoding the full length rubella capsid protein derived from the rubella virus strain M_{33} was provided as a pET23 plasmid construct (Ilkow *et al*., 2008) by Dr. Tom Hobman, University of Alberta, Canada. This sequence was used for subsequent cloning of the capsid protein into various other constructs.

Full length and truncated forms of the rubella capsid protein were cloned into pET28a (Novagen) and pGEX-KG (ATCC-77103) vectors. The pET28a vector encodes for amino and carboxy terminal His tag. The pGEX-KG vector encodes for a Glutathione S-transferase (GST) tag at the amino terminus of the protein of interest. Preparation of the plasmid constructs was carried out using standard cloning protocols provided in the pET system manual (Novagen). Briefly, the DNA sequence of the desired protein length was amplified through polymerase chain reaction (PCR) with the addition of suitable restriction endonuclease cleavage sites at the 5' and 3' end of the sequence. The amplified DNA fragment was then digested with two restriction endonucleases corresponding to the cleavage sites at the 5' and 3' end. This DNA fragment or insert was then ligated into the similarly digested plasmid vector. The ligation mixture was transformed into *Escherichia coli* (*E.coli*) Novablue competent cells (Novagen). Colonies were selected using the antibiotic resistance conferred by the vector. Colonies positive for the desired clone were confirmed through restriction endonuclease digestion and DNA sequencing. A list of the different capsid protein constructs made, along with their primer sequences are given in **Table 2.1**.

Table 2.1 List of capsid protein constructs

Capsid protein length	Fusion tag	Vector	Primer sequences (5'-3') (**F**-forward, **R**-reverse)
RVC-N1-277[1]	N-terminus 6X His	pET28a	**(F)**TAACATATGGCTTCCACTACCCCC **(R)**TAAAAGCTTTCAGCGGATGCGCCAAGG
RVC-N1-277[1]	C-terminus 6X His	pET22b	**(F)**TAACATATGGCTTCTACTACCCCCATC **(R)**TAAAAGCTTTCAGCGGATGCGCCAAGG
RVC-N1-277[1]	N-terminus GST	pGEX-KG	**(F)**TAAGGATCCATGGCTTCCACTACC **(R)**TAAAAGCTTTCAGCGGATGCGC
RVC-N127-277[1]	N-terminus 6X His	pET28a	**(F)**TAACATATGAACCCGTTCCAGGCA **(R)**TAAAAGCTTTCAGCGGATGCGC
RVC-N127-277[1]	C-terminus 6X His	pET 28a	**(F)**TAACCATGGAACCCGTTCCAGG **(R)**TAACTCGAGGCGGATGCGC
RVC-N127-277[1]	N-terminus GST	pGEX-KG	**(F)**TAAGGATCCATGAACCCGTTCCAG **(R)**TAAAAGCTTTCAGCGGATGCGC
RVC-N148-277[1]	N-terminus GST	pGEX-KG	**(F)**TAAGGATCCATGGCACCCACTGAG **(R)**TAAAAGCTTTCAGCGGATGCGC

[1]RVC denotes rubella capsid protein and N refers to the amino terminus of the protein followed by the amino acid residue numbers.

2.2.2 Expression of capsid protein constructs

The different constructs listed in **Table 2.1** were tested for soluble protein expression in a bacterial expression system. The plasmid constructs were transformed into *E.coli* BL21 (DE3) cells (Novagen). 0.5 µl of the plasmid clone was incubated on ice with 30-50 µl of *E.coli* BL21(DE3) competent cells for 10 minutes. The mixture was then heat shocked at 42 ºC for 1 minute and placed on ice again for further 10 minutes. 450 µl of *Luria Bertani* (LB) medium was added to the mixture and cultured by shaking at 250 rpm for 1 hour at 37 ºC. The cultured cells were then plated onto LB plates with the corresponding antibiotic for the plasmid vector used.

Colonies were picked from the cultured LB plates and grown in 5 ml of LB medium at 37 ºC till an O.D of 0.6 was reached. The cultures were then induced with 1mM Isopropyl thiogalactoside (IPTG) at temperatures of 37 ºC, 30 ºC, 25 ºC and 16 ºC for time durations of 2, 4, 6 and 16 hours respectively. The cells were pelleted and lysed using Bugbuster master mix solution (Novagen) for 20 minutes at room temperature. The cell lysate was centrifuged at 15000 rpm for 10 minutes to obtain a clarified supernatant and pellet. The supernatant and pellet fractions were analyzed using SDS-polyacrylamide gel electrophoresis (SDS-PAGE) to determine the expression and solubility of the capsid protein constructs.

2.2.3 Large-scale protein expression and purification

Soluble expression of the capsid protein construct with amino acid residues 127-277 fused with an amino terminal GST tag was observed in the protein expression tests. Henceforth, this clone will be denoted as RVC-N127-277-GST. For large scale expression, the cells containing the plasmid for RVC-N127-277-GST were grown in 1 liter LB medium at 37 °C. Expression of the fusion protein was induced when the culture O.D reached 0.6 with 1 mM IPTG. The induced cultures were grown at 25 °C for 6 hours before harvesting the cells. The cells were pelleted by centrifugation at 6000 rpm for 15 minutes at 4 °C. The cell pellets were re-suspended in phosphate buffer saline (PBS, pH 7.4) and lysed in ice using a sonicator. The lysed cells were centrifuged at 15,000 rpm for 25 minutes using a Beckman Coulter centrifuge. The supernatant was incubated with GST beads (GE

Lifesciences) under constant stirring for 1 hour at 4 °C. The supernatant with the beads was then packed into a 5 ml chromatography column. The column was washed with 5 column volumes of PBS. The GST beads with the fusion protein were then incubated with 50 units of thrombin (GE Lifesciences) overnight at room temperature. Cleaved rubella capsid protein was collected as flow-through from the beads. The GST tag bound to the beads was eluted using 5 column volumes of 20 mM reduced Glutathione in 100 mM Tris, pH 8.0. The previously collected flow-through from the GST column was further purified using a Superdex 75 gel filtration column (GE Lifesciences). The elution fraction containing the capsid protein was collected and concentrated to 5 mg/ml. The yield of the purified protein was approximately 0.25 mg/ml

2.2.4 Large scale purification of Selenomethionine (Se-Met) derivatized capsid protein

The RVC-N127-277-GST protein was produced as a Se-Met derivatized protein similar to the protocol outlined in VanDuyne *et al*., 1993. The *E.coli* BL21(DE3) cells were transformed with the plasmid clone and grown overnight in 50 ml of LB medium at 37 ºC. The cells were pelleted and added to M9 minimal medium (6 g/l Sodium phosphate dibasic, 3 g/l Potassium phosphate monobasic, 1 g/l ammonium chloride and 0.5 g/l sodium chloride in distilled water). The culture was grown at 37 ºC till an O.D of 0.3 was reached. Se-Met and other amino acids were then added to the culture. Fifteen minutes later, over-expression of protein was induced with 1 mM IPTG at 25 ºC. The cells were grown for 6 hours and harvested. The target protein was purified with the same procedure as the native protein (Section 2.2.3).

2.2.5 Crystallization and data collection

Crystallization trials were set-up using Hampton crystallization screen-kits. Initial crystals of the capsid protein were obtained as clusters of thin plates in 50 mM HEPES at pH 7.5, containing 10% 2-propanol and 25% PEG-4000. This crystal condition was then optimized by varying the buffer, 2-propanol and PEG-4000 concentrations. The Se-Met protein crystallized in similar conditions as the native protein. These crystals formed clusters similar to the native protein but, in some cases, formed longer cuboid-shaped

crystals after two months of setting up the crystallization condition. After extensive screening, thin-plate crystals of rubella capsid protein were also obtained in the presence of 50 mM Bis-Tris at pH 6.5, 200 mM NaCl, 25% PEG-3350 and 1% LDAO (lauryl-dimethylamine oxide).

For phase determination, crystals were tested with different heavy atom compounds through co-crystallization or soaking. The different heavy atoms that were tested include mercury, platinum, iodine, gold, lead, elements of the lanthanide series and uranium. Co-crystallization trials were set-up at heavy atom compound concentrations of 2-5 mM in the crystallization drop. For soaking experiments, heavy atom compounds were added to the crystals at different concentration ranges (1 mM-10 mM) for different lengths of time (20 minutes to 1 day).

All crystals were screened at the Advanced Proton Source, GM/CA, beamline 23 ID-D and ID-B, Chicago, USA.

2.2.6 Structure determination and refinement

The diffraction data obtained from the different crystal forms of the rubella capsid protein were indexed, integrated and scaled using HKL2000 (Otwinowski and Minor, 1997). The output intensity files were converted to structure factor files using the CCP4i program package (1994).

Based on sequence comparisons, the rubella capsid protein does not have any homologues in the NCBI protein database. However, by functional similarities and grouping of rubella virus with alphaviruses, the Sindbis virus capsid protein structure was used as a molecular replacement model for the rubella capsid protein structure determination. The Phaser program (McCoy, 2007) from the CCP4i package was used for these trials.

Crystals soaked with heavy atom compounds were used to collect diffraction data at different wavelengths for MAD and MIR trials. Datasets from various heavy atom soaked crystals were analyzed for the presence of anomalous signal. Unmerged datasets of the heavy atom derivative crystals were analyzed by comparing the absolute difference between the structure factor amplitudes of the friedel pairs from the centric reflections and

the non-centric reflections. Higher differences in amplitudes between friedel pairs of the non-centric reflections when compared to the centric reflections indicated the presence of anomalous scatterers (heavy atoms) in the crystal.

SIRAS was used to determine initial phases from the uranium-soaked protein crystal datasets. This was carried out using the Autosolve program (Terwilliger *et al*., 2009) in the Phenix software package (Adams *et al*., 2010). The heavy atom parameters were determined at 3.7 Å resolution and the initial phases were calculated at 3 Å resolution. An electron density map was calculated using Phenix.solve (Terwilliger, 2003). The map was then solvent flattened and averaged using non-crystallographic symmetry (NCS) detected by Phenix.resolve (Terwilliger, 2003). The initial protein chain was built into the electron density map manually using COOT (Emsley and Cowtan, 2004) and further refined using Phenix.refine (Afonine *et al*., 2012). The refined map was then used to further build and correct the initial protein chain. This procedure was used iteratively till all the electron density in the map could be interpreted.

The above structure was used as a molecular replacement model to determine the structures in the other crystal forms obtained for the capsid protein. The procedure was carried out using the Phaser program from the Phenix software package.

2.3 Results and Discussion

2.3.1 Expression and purification

The different constructs of the rubella capsid protein were tested for soluble protein expression. Of all the conditions tested, soluble capsid protein expression was obtained only for the clones RVC-N127-277-GST and RVC-N148-277-GST at 25 ºC for 6 hours on induction with 1 mM IPTG. Thus, the longest soluble construct only accounted for approximately 2/3rd of the whole capsid protein length.

Based on the expression tests, RVC-N127-277-GST was used for further purification and crystallization trials as it was the longest capsid construct. The RVC-N127-277-GST clone was expressed in *E.coli* BL21(DE3) cells. The fusion protein was purified using GST affinity chromatography and cleaved with thrombin to give the rubella capsid protein. The cleaved protein was further purified using gel filtration chromatography. The rubella capsid protein eluted as a single peak at an elution volume which indicated that the protein was a dimer (**Figure 2.3**).

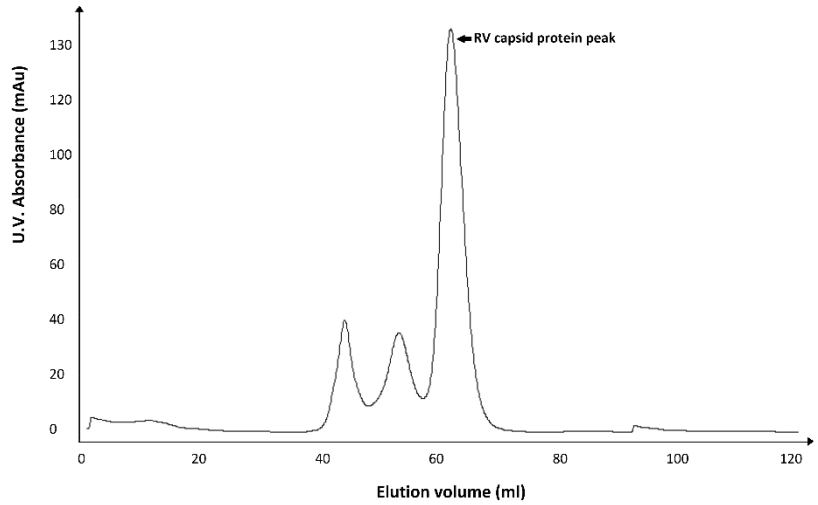


Figure 2.3 Gel filtration chromatography profile of the recombinant rubella capsid protein. The y-axis is the UV absorption measurement at 280 nm (mAU) and the x-axis is the elution volume (ml). Based on the standardized elution profile for Superdex 75 16/600 prepgrade (GE Lifesciences), the capsid protein peak has an apparent molecular weight of ~40 kDa which indicates that the protein forms a dimer in solution. The molecular weight of the capsid monomer is 17 kDa. The two smaller peaks in the elution profile are un-identified protein impurities in the sample.

2.3.2 Crystallization, data collection and processing

The first crystals of the truncated rubella capsid protein were obtained as clusters of thin plates in 50 mM HEPES pH 7.5, 10% 2-propanol and 20% PEG-4000. Improved crystals were obtained in 50 mM Tris pH 7.5, 24% 2-propanol and 25% PEG-4000 (**Figure 2.4**). Thin plates were broken from the clusters and used for data collection. The crystals diffracted to 2.3 Å resolution and were of space group $P22_12_1$ with two molecules in the asymmetric unit.

Figure 2.4 Thin cluster crystals of rubella capsid protein. These crystals were obtained in 50 mM Tris pH 7.5, 24% 2-propanol and 25% PEG-4000.

Cuboidal shaped crystals of Se-Met derivatized capsid protein (**Figure 2.5**) diffracted to 3.25 Å and had space group $C222_1$with five molecules in the asymmetric unit. These crystals were non-isomorphous with the native crystals. A Patterson function using the Se-Met crystal dataset had large non-origin peaks at (0,1/5,0) that were about 50% of the height of the origin peak. Thus, this crystal form had pseudo-translational symmetry (**Figure 2.6**). Analysis of the structure factor amplitudes showed that all reflections with a Miller index, k, divisible by 5 were much larger than other reflections, indicating that the actual unit cell was approximately divided into five smaller unit cells that repeat the structure in the **b** (axis) direction. Each of these smaller unit cells had a $C222_1$ space group, but with only one molecule in the asymmetric unit. Although the Se-Met RV capsid protein crystal showed anomalous signal, it was very weak.

Automated phasing program packages, Phenix and CCP4, were unable to locate heavy atom sites in the Se-Met dataset due to the weak anomalous signal of the Se atoms and the strong pseudo-translational symmetry. Manual interpretation of peaks in the Harker sections of the anomalous difference Patterson maps were used to identify 1 heavy atom site per protein molecule.

Though there were two methionine residues per protein molecule, only one Se site was identifiable with confidence. The other site may have been disordered in the crystal. This single site was used in a (F^+-F^-) difference map, where F^+ and F^- denote the structure

factors of the friedel pairs for a given reflection, to determine initial phases. However, the phases obtained were not accurate enough to produce an interpretable electron density map. Thus, more crystallization conditions were screened to obtain new crystal forms.

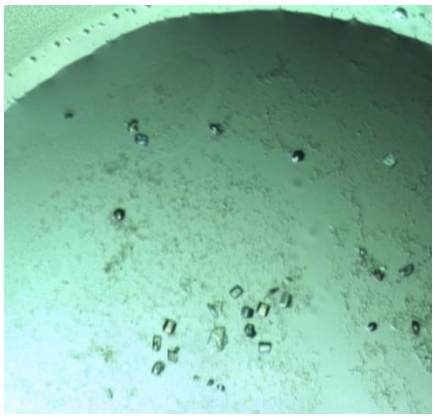


Figure 2.5 Crystals of the Se-Met derivatized rubella capsid protein.

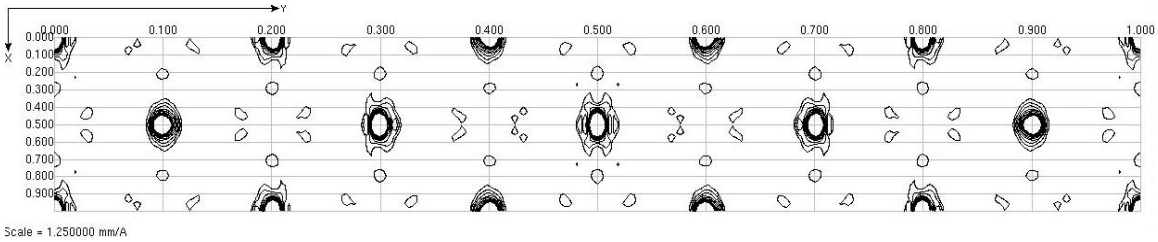


Figure 2.6 Native Patterson map generated map section from the Se-Met capsid protein derivative data. The section shown is of the x-y plane at z = 0, where x,y,z are the Cartesian axes. The section divisions are in fractional co-ordinates.

Thin-plate crystals were obtained in PEG-3350 (**Figure 2.7**) and were found to diffract inconsistently with best diffraction to 2.66 Å resolution. The space group of these crystals was monoclinic C2 with six molecules in the asymmetric unit.

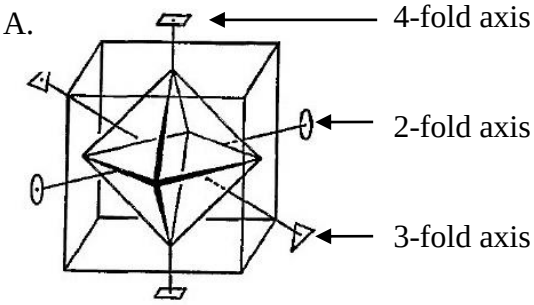
Figure 2.7 Thin plate crystals of rubella capsid protein in PEG-3350.

Rotation function analysis was used to determine the non-crystallographic symmetry between the six molecules in the asymmetric unit (Tong and Rossmann, 1997). A rotation function depends on multiplying the crystal's Patterson function with its rotated version. The rotation is represented by a set of polar (φ, ψ, κ) angles. The angles φ and ψ define the orientation of a rotation axis relative to a Cartesian coordinate system and κ defines the angle of rotation around this axis.

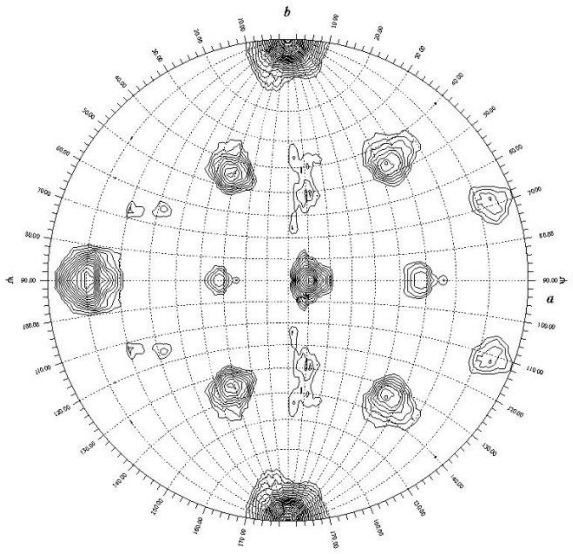
In the self-rotation functions shown below (**Figure 2.8**), the radius of integration was 25 Å, and the resolution range was between 4-12 Å. The radius of integration specifies the integration volume for the rotation function. An optimum value for the radius of integration will exclude vectors between atoms from molecules outside the asymmetric unit of the crystal. All peaks shown have values three standard deviations above the mean intensity values (sigma). On analysis, the prominent peaks in the self-rotation function for this dataset corresponded to peak patterns observed in objects with octahedral symmetry. However, when the structure had been determined, it was discovered that the molecular arrangement in the crystal did not exhibit any octahedral symmetry. Hence, the presence of self-rotation peaks exhibiting octahedral symmetry was evidently an artifact of crystal packing (Akervall *et al*., 1971).

Figure 2.8 GLRF analysis. A. Octahedron showing the symmetry axes. B-D. Self-rotation function plots at κ values of 180°, 120° and 90° respectively. The radius of integration is 25 Å. The resolution range is 12-4 Å. The contour levels range from one to ten sigma with a step size of 0.4 sigma. All peaks shown have values three sigma and above. The Ψ angles are denoted by the latitudinal lines and marked 0° to 180° along the perimeter of the sphere. The Φ angles are denoted by the longitudinal lines on the surface of the circle. Each longitude marks 10° rotation along the Φ axis from 0° to 180°.

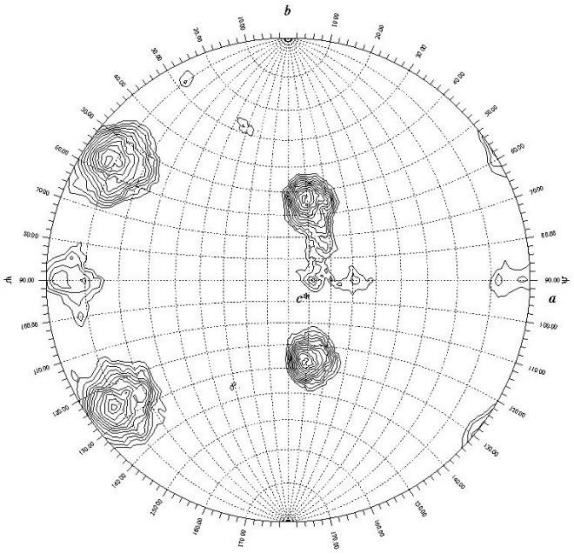
A.



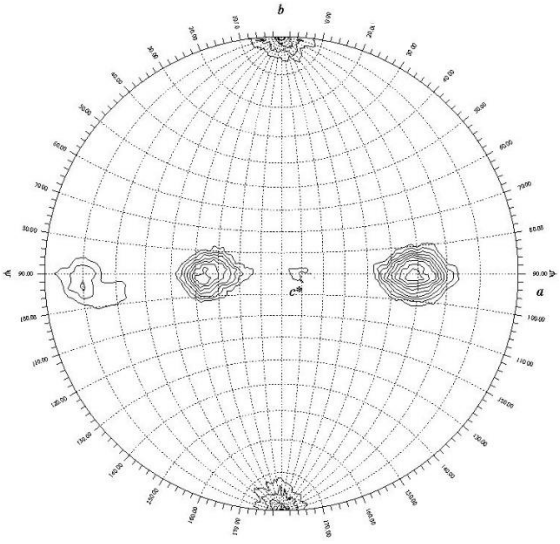
B. Two-fold rotation axis



C. Three-fold rotation axis



D. Four-fold rotation axis



The C2 space group crystals were soaked in 2 mM uranium nitrate hexahydrate ($UO_2(NO_3)_2.6H_2O$) overnight at room temperature. The crystals soaked in the uranyl compound diffracted to 3.7 Å resolution. A single wavelength anomalous dispersion (SAD) dataset was collected from these crystals. The uranyl compound-soaked crystal data was isomorphous to the native crystal and was used for determination of heavy atom sites.

2.3.3 Structure determination and refinement

The determination of the heavy atom parameters and subsequent phasing was performed with the phenix.autosol software. Phases were improved by density modification using non-crystallographic symmetry with the same program. The initial protein chain was built into the electron density map manually using COOT (**Figure 2.9**) and refined using Phenix.refine. The refined map was then used to further build and correct the initial protein chain. This procedure was used iteratively till all the electron density in the map was interpreted. The uranyl moiety was found to be interacting with Glu-229, Arg-230, Asp-231 and His-240 in each of the six RV capsid proteins in the asymmetric unit. Parts of the protein that seem to be disordered include the 30 amino acids of the carboxy terminal end and approximately 20 amino acids of the amino terminal end. The protein chain that could be built extended from residue 151 to 247 (**Figure 2.10**).

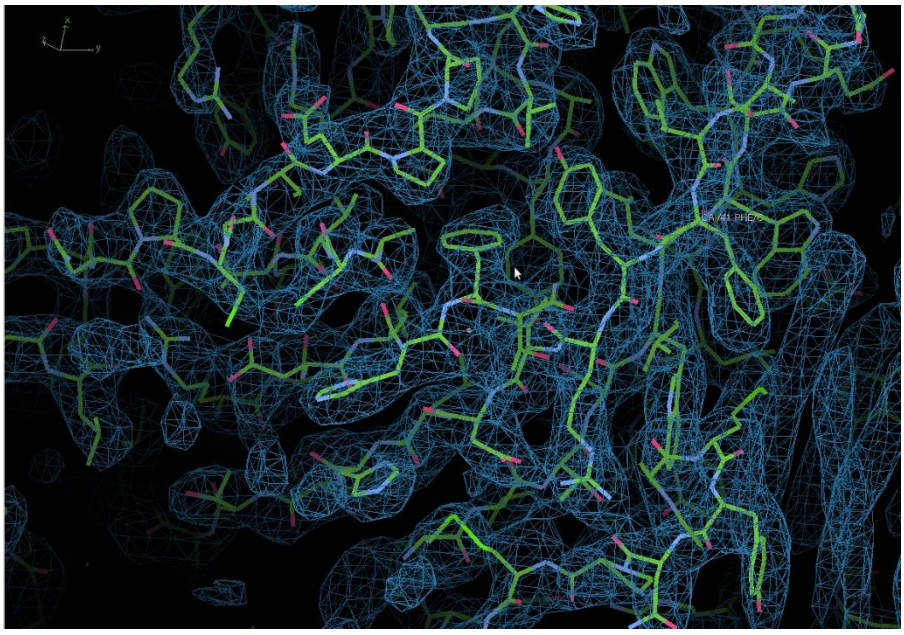


Figure 2.9 Snapshot of initial protein chain built into the electron density map.

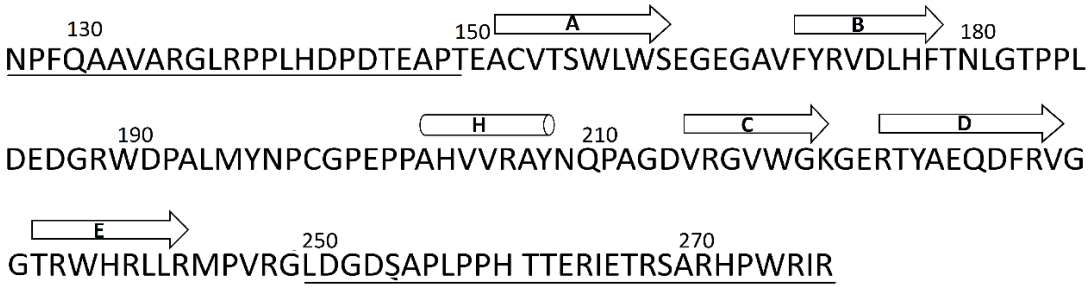


Figure 2.10 RV capsid protein secondary structure. Amino acid sequence (127-277) with β-strands shown as arrows and the helix as a cylinder. The disordered regions are underlined.

The structures of RV capsid protein in the $P22_12_1$ and $C222_1$ space groups were determined by MR using the Phaser software. Refinement of the structure in the $C222_1$ space group had some difficulty due to the strong pseudo-translation symmetry in the crystal. The problem was to differentiate the small differences in structure between the five slightly different molecular positions in the five very similar smaller cells. Whereas the

overall structure was given by the strong reflections when the Miller index k was divisible by 5, the other weak reflections determined the differences between the different sub-unit cells. At the completion of refinement, the R factor for the strong k/5 reflections was 26% and for the other weak reflections it was 39%.

The crystallographic data and refinement statistics are given in **Table 2.2**.

Table 2.2. X-ray data and crystallographic information

	Crystal RVC[1] – Native-1	Crystal RVC[1] – Native-2	Crystal RVC[1] – Se-Met derivative
Data collection			
X-ray wavelength (Å)	1.006	1.072	0.988
Space group	$P22_12_1$	C2	$C222_1$
Cell dimensions			
a,b,c (in Å)	37.73, 71.48, 79.04	147.76, 74.78, 96.07	42.79, 279.7, 76.73
α,β,γ (in degrees)	90, 90, 90	90, 97.529, 90	90, 90, 90
Resolution (Å)[2]	2.28 (2.32– 2.28)	2.66 (2.71 - 2.66)	3.25 (3.31 - 3.25)
Rmerge[2]	0.082 (0.441)	0.08 (0.428)	0.061(0.522)
$I/\Delta I$[2]	21.09 (2.26)	26.71 (5.58)	14.27 (3.74)
Completeness (%)[2]	95.6 (99.4)	100 (86.9)	99.1 (92.7)
Overall Redundancy	8.1	6.7	12.7
Refinement			
No. of reflections	9961	29202	13688
Rwork/Rfree	0.199/0.258	0.185/0.23	0.339/0.367
Ordered residues	150 – 247	150 – 250	150 – 246

Table 2.1 Continued

R.m.s deviations			
Bond length (Å)	0.008	0.008	0.007
Bond angles (deg.)	1.23	1.34	1.197
Ramachandran plot (%)			
Favored	93.8	97.2	80.7
Allowed	6.2	2.5	13.5
Outliers	0	0.3	5.8

[1] Rubella virus capsid protein (RVC)
[2] Numbers in parentheses represent values in the highest resolution shell

2.3.4 Structure of the rubella virus capsid protein

The structure of each monomer of the rubella capsid protein structural domain consists of five anti-parallel β-strands A, B, C, D, E as well as a two turn α-helix H between strands B and C (**Figure 2.11**). The strands and helix are connected by loops AB, BH, HC, CD and DE. The BH loop is by far the largest loop consisting of residues 176 to 197. Although the core structure (missing the BH loop) of each of the thirteen independent RV capsid monomers from the three different crystal forms are essentially the same (**Table 2.3**), the BH loop structure is different in each of the crystal forms (**Figure 2.11**) and is disordered in the $C222_1$ crystal form.

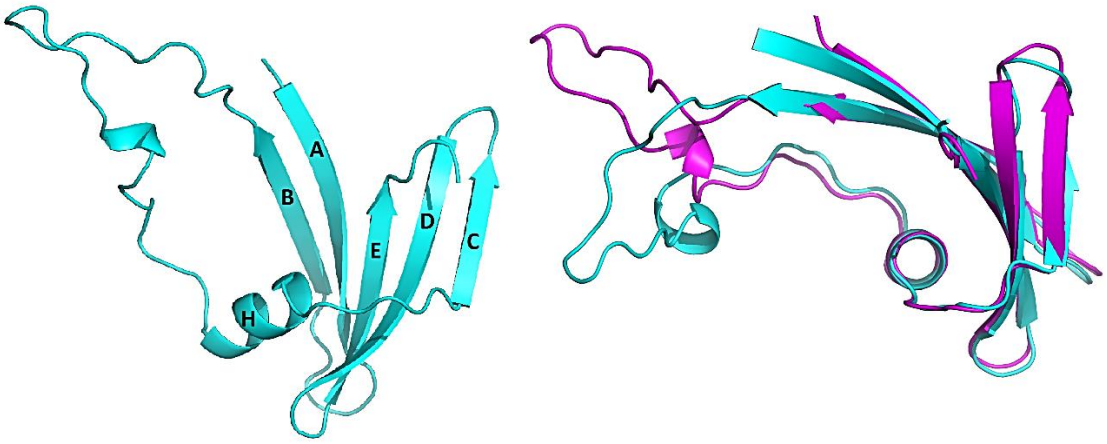


Figure 2.11 Rubella capsid monomer structure. Left: Ribbon diagram of the capsid monomer. The β-strands are marked A to E, starting from the N-terminus. The α-helix is marked as H. Right: Superimposed ribbon diagrams of the monomer as seen in the C2 (cyan) and $P22_12_1$ (magenta) crystals.

Table 2.3 Root mean square difference (RMSD) (in Å) between Cα atoms in the core structures of the capsid monomers in the three crystal forms[1]

	mC-A	mC-B	mC-C	mC-D	mC-E	mC-F	oP-A	oP-B	oC-A	oC-B	oC-C	oC-D	oC-E
mC-A	0	0.07	0.07	0.07	0.06	0.13	0.53	0.40	0.49	0.99	0.83	0.76	0.64
mC-B		0	0.08	0.06	0.06	0.13	0.51	0.40	0.48	0.99	0.82	0.77	0.64
mC-C			0	0.07	0.05	0.15	0.49	0.38	0.46	0.64	0.84	0.78	0.67

Table 2.3 Continued

mC-D				0	0.07	0.11	0.51	0.40	0.47	0.98	0.81	0.76	0.63
mC-E					0	0.14	0.50	0.39	0.47	0.63	0.83	0.77	0.66
mC-F						0	0.51	0.39	0.47	1.0	0.77	0.77	0.64
oP-A							0	0.53	0.37	0.34	1.10	1.08	0.42
oP-B								0	0.43	1.07	1.03	0.98	0.96
oC-A									0	0.85	1.48	0.95	1.16
oC-B										0	1.13	0.52	1.07
oC-C											0	1.43	2.27
oC-D												0	0.42
oC-E													0

[1]The different crystal structures are identified by the space group of the crystals in which they occurred (mC for monoclinic C centered, oP for orthorhombic primitive, oC for orthorhombic centered) and chain names (A, B etc.). The RMSDs were calculated using the Homology software (Rossmann and Argos, 1976).

All three crystal forms indicate that the RV capsid protein forms a 2-fold symmetric dimer. The monomers are associated by disulfide bonds between Cys-153 and Cys-198. Within a dimer the β-sheet is continuous with the residues 170-176 of β-strand B from one

monomer being hydrogen bonded with the same but anti-parallel β-strand B in the other monomer. The resultant ten-stranded, left-handed twisted sheet forms a partially open β-barrel with the two helices lying in the center of the barrel. The β-strands A and B from one monomer are inserted into the BH loop of the other monomer, forming a tightly bound dimeric structure (**Figure 2.12**). A DALI search (Holm and Rosenstrom, 2010) using the RV capsid structure did not give any significant similarity with known protein structures.

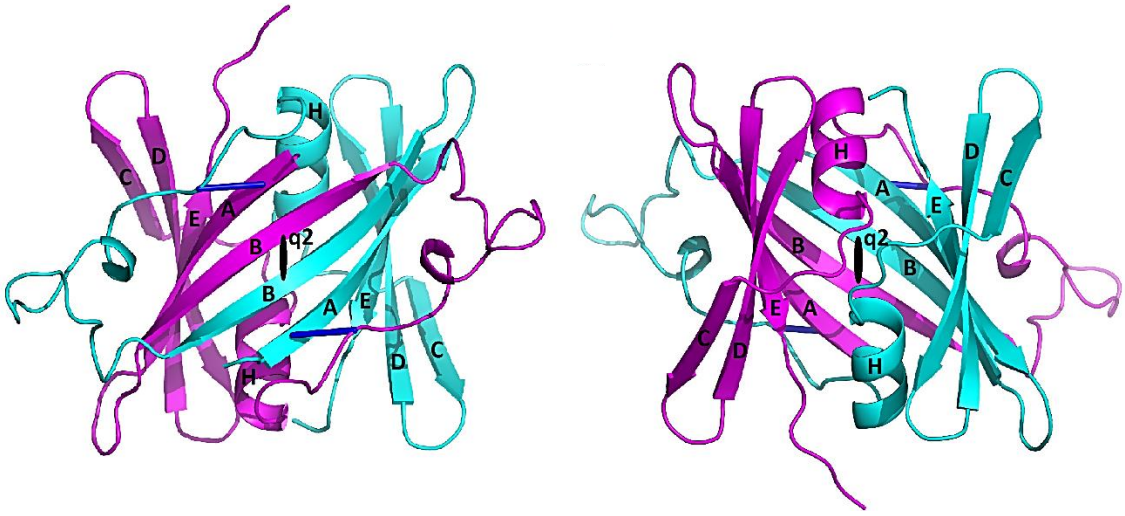


Figure 2.12 Ribbon diagram of the RV capsid protein dimer. The two figures are 180° related to each other. The black oval shape (q2) indicates a quasi-two-fold axis. The monomer chains are represented in cyan and magenta.

In the $P22_12_1$ crystal structure, there is a one residue translation in β-strand C, at the edge of the β-sheet, with respect to strand D in one of the monomers (**Figure 2.13**). This translation requires a 180º flip of the peptide bonds in the C strand. A variety of tests were conducted to confirm these unexpected amino acid assignments. The presence of Trp, Arg and Lys residues in the C β-strand helped to confirm the correctness of the structure. The impact of this change was not as large as might have been expected because the two basic residues are oriented in the plane of the sheet rather than being perpendicular to the sheet.

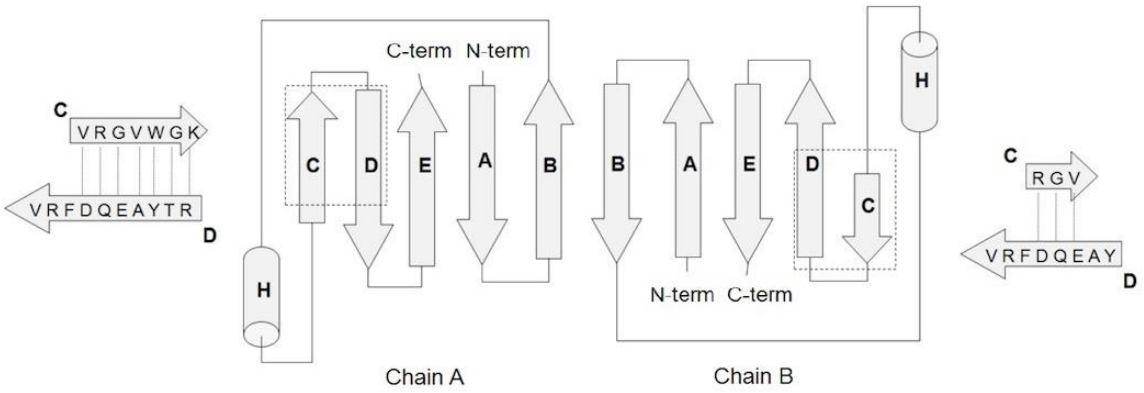


Figure 2.13 Diagram of the capsid dimer structure as seen in the $P22_12_1$ crystal form. Bold letters are names assigned to the secondary structural elements, the letters in the magnified C and D β-strands refer to amino acid abbreviations.

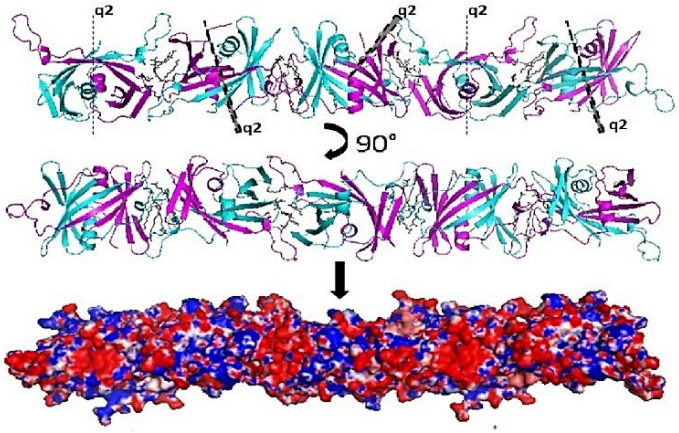
All three crystal forms of the RV capsid protein contain rows of closely packed dimers (**Figure 2.14**) in which neighboring dimers are held together by hydrogen bonds and hydrophobic contacts. The interactions between rows are fewer and formed mostly by hydrophilic contacts. In the C2 crystal form, adjacent dimers in a row make hydrophobic contacts between Trp-221 in one dimer with Trp-189 andPro-191 in the neighboring dimer's BH loop. In the $C222_1$ crystal form, the dimers interact through hydrogen bonds between the terminal β-strands C (amino acids 218-222) in neighboring molecules. In the $P22_12_1$ crystal form, the dimers are organized into a staggered row-like arrangement similar to that in the $C222_1$ crystal form and also involve hydrogen bond interactions between Val-220 of one monomer with Gly-219 of the neighboring monomer. The similar rows of dimers in the $P22_12_1$ and $C222_1$crystal forms have a surface that is mostly positively charged (**Figure 2.14B,C**).

2.3.5 Orientation of the Capsid Protein in the Virion

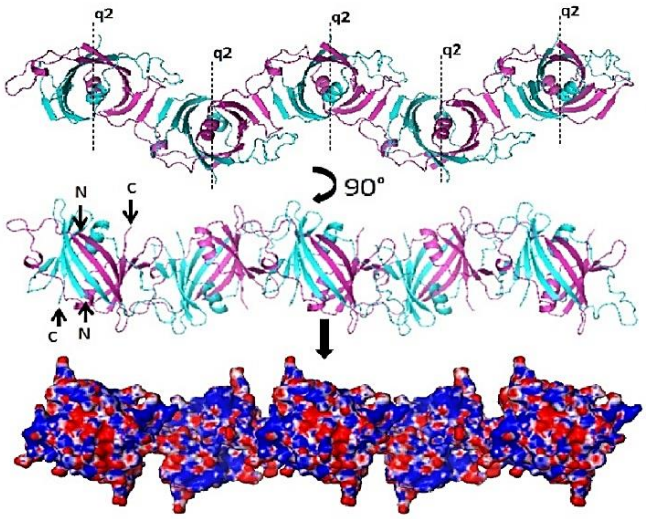
In infected cells, the carboxy terminus of the RV capsid protein remains attached to the E2 signal peptide which anchors this viral protein to intra-cellular membranes (Suomalainen *et al*., 1990). This implies that, the carboxy terminus of the RV capsid structure should be oriented towards the membrane surface in the virions. Hence, the optimum way to orient the dimer structure would be for the dimer axes to be perpendicular to the membrane, thus placing the carboxy termini of the monomers at approximately equal distances from the membrane (**Figure 2.14C**). The distance to the end of the E2 signal peptide in the membrane can then be easily spanned by the disordered 27 amino acids at the carboxy termini of the capsid protein dimer. As the amino terminal region of the RV capsid protein interacts with the viral RNA (Hobman and Chantler, 2007), placing the dimer axis perpendicular to the membrane would also position the exposed amino terminus of both monomers within the dimer at a similar distance from the interior of the virus (**Figure 2.14C**).

Figure 2.14 Rows of RV capsid protein dimer. Rows in the crystal forms (A) C2, (B) $P22_12_1$and (C) $C222_1$ with dimer axis parallel and perpendicular to the plane of the paper, respectively. For each crystal form, the molecular surface that presumably faces the viral membrane is shown, colored according to its electrostatic potential. The potentials (Baker *et al*., 2001) at 300K are in the range +125 mV (+5 kT/e) (blue) to -125 mV (-5 kT/e) (red), where k = Boltzmann's constant, T= temperature, e = electronic charge and mV = millivolts. The dotted lines (q2) correspond to quasi-two-fold axes. Figure panels were prepared using PyMOL (Schrodinger, 2010).

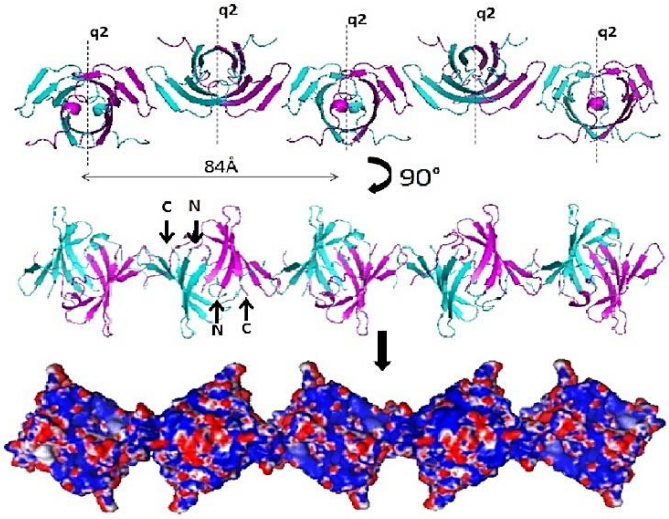
A.



B.



C.



2.3.6 Location of the Glycoprotein Binding Region on the Capsid Protein

In alphaviruses the glycoproteins have a cytoplasmic domain that is anchored in the nucleocapsid (Mukhopadhyay *et al*., 2006). Similarly, in RV the short predicted transmembrane and/or cytoplasmic domains of the RV E2 glycoprotein may serve as an anchor on the capsid protein (Garbutt *et al*., 1999). The cytoplasmic domain of E2 has been shown to be essential for budding of RV into the Golgi apparatus (Garbutt *et al*., 1999). In the C2 crystal structure, three detergent molecules derived from the crystallization solution, are associated with each capsid monomer. It is possible that this binding site in the RV capsid functions similarly to the pocket in alphavirus capsid proteins that interact with the carboxy terminal region of E2 (Lee *et al*., 1996). The binding site for the detergent molecules in the RV capsid protein consists of mostly hydrophobic residues and a few conserved acidic residues. Appropriately, the carboxy terminal cytoplasmic region of the RV E2 protein contains mostly positively, highly conserved, charged residues that complement the binding site. However, this site is only occupied in the C2 crystal form, as this is the only form in which there is a suitable ligand in the crystallization solution that might bind into the pocket.

2.4. Evolution of the Viral Structural Proteins

Although the core structure of the RV capsid protein is the same in all the crystal forms reported here, there is variability in the relative positions of the secondary structure elements in the different crystal structures (**Table 2.3**). This variability indicates flexibility in the capsid protein structure which may be required for its numerous cellular functions (Ilkow *et al*., 2010b) by facilitating its interaction with a large cohort of host cell proteins such as Bax, mitochondrial matrix protein p32, prostate apoptosis response protein and poly(A) binding protein. . However the RV capsid structure has not yet yielded any detailed insight into the mechanisms of its cellular functions, but its function as a viral capsid protein provides fascinating insight into the evolution of virus structure.

The E1 glycoprotein of alphaviruses and RV (DuBois *et al*., 2013; Lescar *et al*., 2001) as well as the E glycoprotein of flaviviruses (Rey *et al*., 1995) have similar structures. This common surface glycoprotein motif suggests that these structures evolved from an ancestral structure (DuBois *et al*., 2013) that can attach to potential host cells and initiate infection by fusion with the host cell. In contrast, the capsid protein structures of alpha-, flavi- and rubella viruses are completely different. The alphavirus capsid protein is a β-sheet structure with a chymotrypsin-like fold (Choi *et al*., 1991) which forms an icosahedral shell around the viral genome, whereas the flavivirus capsid protein is an α-helical structure (Ma *et al*., 2004) which is closely associated with the genome but does not form a protective shell. The RV capsid protein structure as reported here has a polypeptide fold that is different to both alpha- and flaviviruses. The common function of the capsid proteins in these viruses is to neutralize the charge on the genome, but the task of confining the genome is most obvious for alphaviruses, less clear for RV, and absent for flaviviruses. In addition, the capsid proteins of these viruses have gained other cellular functions that facilitate virus replication and assembly inside host cells. These observations show that the ability to attach and fuse with potential host cells has been conserved in the glycoproteins of these viruses while the capsid proteins have evolved independently, resulting in acquisition of different non-structural functions.

CHAPTER 3. CRYO-ELECTRON TOMOGRAPHY OF RUBELLA VIRUS PARTICLES

The work described in this chapter has been accepted for publication (Mangala Prasad *et al*.(in press)). Parts of the chapter that were taken verbatim from the paper are marked by vertical lines in the right margin.

3.1 Introduction

Cryo-electron tomography is an electron microscopic technique in which a three dimensional structure of an object can be determined by obtaining sequentially tilted two dimensional images. This technique is useful for studying samples which are heterogeneous as it does not require averaging between particles. However, cryo-electron tomography is still a relatively low resolution tool when compared to single-particle cryo-electron microscopy and x-ray crystallography for studying biological macromolecular complexes.

To obtain a cryo-electron tomogram, tilt-series images are obtained of biological samples frozen in vitreous ice (**Figure 3.1**). Images are generally taken around a single axis of the specimen stage, though dual-axis tilts are also possible (Iancu *et al*., 2005). The tilt-series images are then aligned with the help of fiducial gold markers that are added during sample preparation. A three-dimensional reconstruction of the sample can then be calculated using weighted back-projection of the aligned tilt-series.

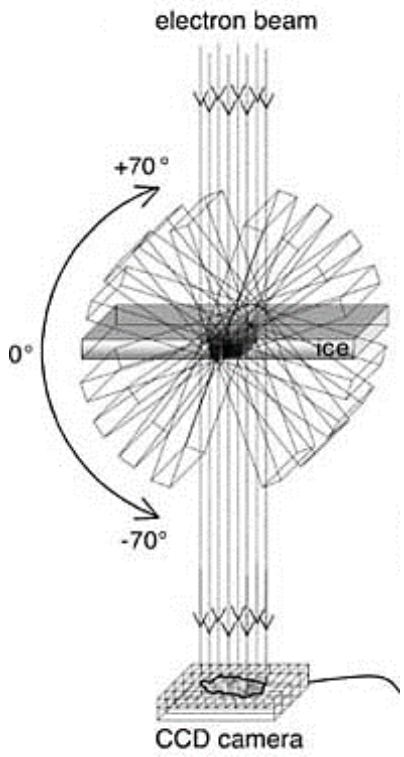


Figure 3.1 Tilt-series image collection for cryo-electron tomography. Figure modified from Grunewald *et al*., 2003.

The possibility of generating a three-dimensional (3D) reconstruction from two-dimensional (2D) projections of the object is given by the projection theorem. According to this theorem, "*the 2D Fourier transform of a projection of an object is identical to a central section of the object's 3D Fourier transform*". Thus, by obtaining images of an object at different tilt angles, the entire Fourier transform of the object can be measured. The weighted back projection method is one of the most commonly used methods for cryo-electron tomographic reconstructions. It works on Fourier transform data and real space data of the object or its projections (Frank 2006). As a first step in this method, the projections of an object are smeared out to form back-projection bodies.

A back projection body b^i of a projection p^i along the direction *z* in the *xyz* plane can be given as,

$$b^i (x,y,z) = p^i (x,y)q(z)$$

where q(*z*) is 1 within a given boundary radius and 0 otherwise. In general, the boundary radius of the back-projection volume is at least twice the object radius.

Summing up multiple back-projection bodies from different angles will result in a three-dimensional volume that resembles the object of study. This can be expressed as,

$$b(x,y,z) = \Sigma\, b^i (x^i,y^i,z^i)$$

where i is the i^{th} projection in a set of projections.

For more accurate rendering of the object, the back projected volume needs to be corrected using a weighting function. For a tilt series about a single axis taken at regular angular intervals, the correction can be applied to the Fourier amplitudes of the back-projections by multiplication with a weight factor W. This factor can be calculated as the radius of the back-projection in Fourier space perpendicular to the tilt axis.

Though, cryo-electron tomography has its merits for studying macromolecular assemblies that are not homogeneous, it also has some serious limitations. The highest resolution achievable through cryo-electron tomography is about 40Å without any sub-tomogram averaging. Biological samples are sensitive to electron dose. As the same specimen is exposed to electrons at different tilt angles in tomography, the electron dosage per image has to be kept minimal to avoid radiation damage of the sample. Typically, the total electron dose per tilt-series is kept below 200 electrons/ $Å^2$ (Tocheva *et al*., 2010).This results in low signal to noise ratio in the images. Higher defocus values are then needed to improve contrast in the images, which in turn affects the final achievable high resolution in the three-dimensional reconstruction.

Another issue is the physical limitations of sample holders in currently available microscopes that allow tilt angle ranges of only ± 70º. This results in a ‘missing wedge’ of data that causes smearing of the reconstructed volume in the direction perpendicular to the un-tilted sample grid (Frank, 2006). Sample thickness at higher tilt angles also affects

quality of the images obtained. As the distance for the electron path through the sample increases with tilt, the likelihood of an inelastic scattering of electrons increases. Inelastic scattering of electrons refers to a loss of energy of the electrons on interacting with the specimen while passing through the microscope lens system. Thus, inelastically scattered electrons have a different trajectory as compared to the elastically scattered electrons and can confuse the acquired image (Coslett, 1978). However, the invention and use of energy filters has greatly helped to correct aberrations in images taken at high tilt angles. An energy filter is a magnetic prism that can alter the path of electrons based on their energy. Thus, the electrons with energy in the desired range can be selected and used to form an image (Schroder *et al*, 1990).

For pleomorphic viruses such as rubella virus, cryo-electron tomography is the most viable choice to obtain a three-dimensional structure of the whole virion, albeit at low resolution. Tomographic studies on the rubella virus by Battisti *et al*.,2012a have shown that the rubella virion structure deviates quite markedly from that of alphaviruses. As opposed to the icosahedral arrangement of glycoprotein spikes on the surface of alphaviruses, the rubella viral glycoproteins appear as long rows of density on the virion surface. These rows of density are separated by 90 Å. Also, the rows of density appear to be paired. Based on these data, a model for the arrangement of glycoproteins on the rubella virion surface has been proposed (**Figure 3.2**).

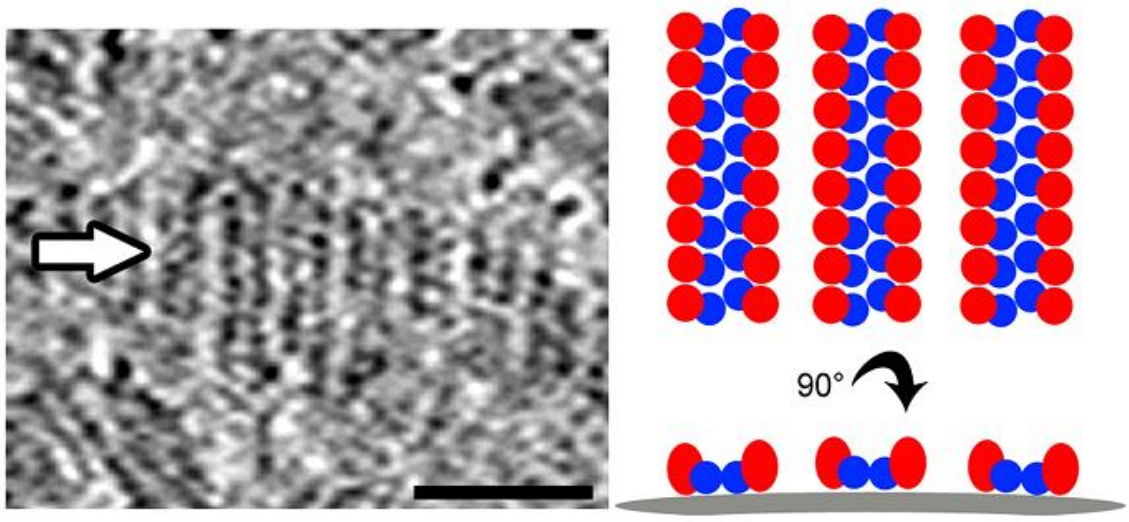


Figure 3.2 Interpretation of rubella surface glycoproteins. Left: Tomographic section on surface of virion. Right: Depiction of a model of E1 (red) and E2 (blue) glycoproteins on virion surface (Battisti *et al*.,2012a).

To delineate the positions of glycoproteins E1 and E2 individually on the viral surface, the use of Fabs from monoclonal anitbodies against E1 and E2 separately was initially proposed to act as position markers for the two glycoproteins on the viral surface. Monoclonal antibodies are divided into two regions, the Fc (Fragment crystallizable) (50 kDa) part and the Fab (Fragment antigen binding) (50 kDa) part. The Fab part corresponds to the region on antibodies that bind to their antigens. The extra density contribution from the Fab fragments was expected to help with the positioning of the individual glycoproteins. However, it was later observed that the length of the glycoprotein spikes is variable on the viral surface. This was also observed by Battisti *et al,* 2012a, where it was reported that the length of spikes from the membrane surface can vary between 50 Å to 80 Å. In light of this information, the above proposal was no longer feasible.

Although, the location of Fabs on the rubella virion surface could not be deduced using the cryo-electron tomograms, the tomographic reconstructions were analyzed more carefully on the basis of new information obtained from the atomic structures of the rubella capsid protein. This has helped to establish a more detailed, low resolution, three-dimensional structure of the rubella virion.

3.2 Materials and methods

3.2.1 Virus preparation and purification

Rubella virus was propagated and purified in a similar way as outlined by Battisti et, al, 2012. Vero cells were grown in Dulbecco's modified Eagle medium with 5% fetal bovine serum (FBS) and 1% L-glutamine at 37 ºC in a 5% CO_2 incubator. Passage number of the cells was kept below thirty. The cells were infected with the F-therein strain of rubella virus when the cells were approximately 85% confluent. Infection with the virus was carried out by overlaying the cells with Dulbecco modified Eagle medium (DMEM) supplemented with 1 mM L-glutamine and the virus stock for 2 hours at 35 ºC. The multiplicity of infection was maintained between 0.1 to 0.5. The medium containing the virus was then removed and the cell medium was changed to DMEM supplemented with 1 mM L-glutamine and 2% FBS. The cells were then incubated for 2 days at 35 °C. Deterioration of the cells was observed, with cell death occurring at a rapid rate beyond 48 hours of cell growth after infection.

The supernatant from the cell culture was collected and centrifuged for 10 min at 6000 rpm in a Beckman Coulter centrifuge (rotor ID: JA-10). PEG precipitation was carried out by the adding PEG-8000 to a final concentration of 8% and stirring overnight at 4 °C. The precipitates from this step were collected by centrifuging for 1 hour at 9500 rpm (rotor ID: JA-10). The pellets were re-suspended in NTE buffer (20 mM Tris, pH 8.0, 120 mM NaCl and 1 mM EDTA) and then concentrated to less than 2 ml volume. This mixture was layered onto a 0 to 54% Opti-prep-NTE gradient and centrifuged at 32000 rpm for 2 hours in an ultracentrifuge (rotor ID: SW-41). Illumination of the gradient centrifuge tubes after this step showed the presence of purified virus as a single white band. The band region was collected and buffer exchanged into NTE buffer using a 100 kDa-molecular-mass cutoff centrifuge filter (Amicon).

3.2.2 Preparation of Fab fragments

Monoclonal antibody against the E1 glycoprotein named 1B9, was kindly provided by Dr. Tom Hobman, University of Alberta, Canada. The Fab fragments were generated from the antibody using the manufacturer protocol from the Pierce Fab purification kit.

The antibody solution at a concentration of 4 mg/ml was digested using immobilized papain protease for 4 hours at room temperature with gentle stirring. The digested sample was purified over a protein-A affinity column that binds to the Fc fragment alone, thus, purified Fab fragments can be collected in the flow-through from the column. The purified Fab fragments were concentrated to a final concentration of 10 mg/ml.

3.2.3 Preparation of samples for cryo-electron tomography

Purified rubella virus was mixed with 1B9 Fab in a 1:4 molar ratio and incubated for 30 minutes at room temperature. Bovine serum albumin-conjugated 10 nm gold particles (Aurion, Wageningen, Netherlands) were added to aliquots of purified rubella virus particles and the virus-Fab mixture. 3 µl of these mixtures were applied to holey EM grids (R 2/2 Quantifoil; Micro Tools GmbH, Jena, Germany), and frozen in liquid ethane.

3.2.4 Cryo-electron tomography of rubella virions

The sample grids were imaged under low-dose conditions at 300 kV with an FEI Titan Krios microscope equipped with a Gatan Tridiem energy filter (GIF 863). Tilt series images were collected over a range of ±66° at 2.0° intervals at a magnification of 19500X. The total electron dose for a particular tilt series was kept in the range of 120 electrons/$Å^2$. Images were acquired at a defocus of around 7 µm underfocus. Pixels of the tilt series images were bin averaged to yield a final pixel size near 1.5 nm. Tilt-series images were aligned using the colloidal gold particles as fiducial markers and reconstructed using the weighted back-projection method as implemented in the IMOD software package (Kremer *et al*., 1996). The tomographic reconstructions were low-pass filtered to a resolution of 40Å.

3.3 Results and Discussion

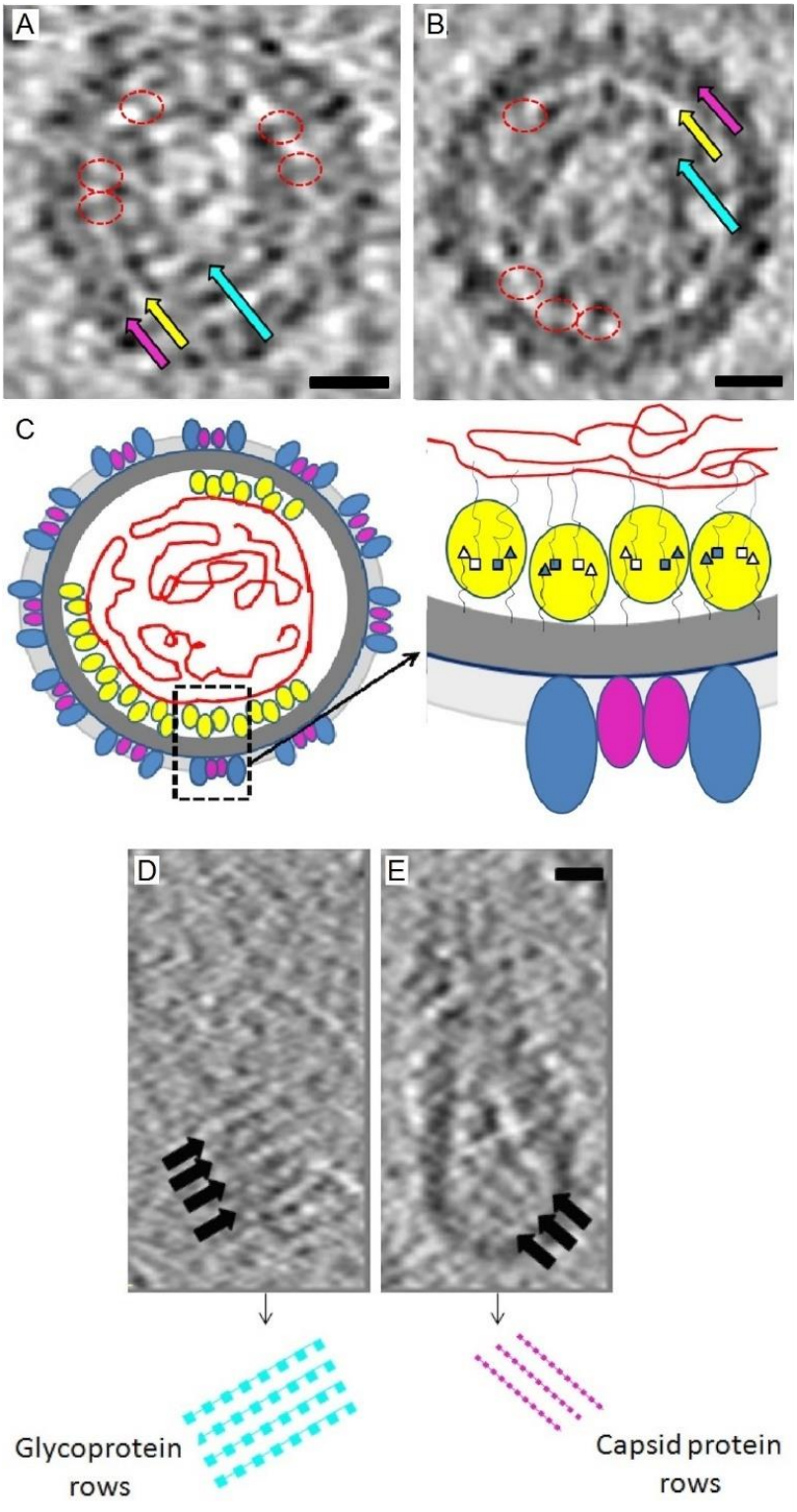
Cryo-electron microscopic (Cheng *et al*., 1995; Mukhopadhyay *et al*., 2006) and tomographic (Battisti *et al*., 2012b) reconstructions of alphaviruses show that there is a 40 Å-wide gap at a radius of approximately 190 Å between the lipid membrane and the nucleocapsid RNA. The alphavirus capsid protein pentamers and hexamers are situated in this gap and produce regions of higher density crossing the gap at intervals of around 80 Å. A similar gap with thickness ranging from 40 Å to 70 Å, crossed by regions of higher density separated by about 90 Å, exists in rubella virus tomograms (Battisti *et al*., 2012b) (**Figure 3.3A,B**), except that the mean radius of the gap is at approximately 240 Å instead of 190 Å. The similarity of these features in alphaviruses and rubella virus suggests that the positive surface of the capsid protein rows of dimers is close to the lipid membrane and the alternate capsid protein dimers span the gap in tomograms. This observation overrides our previous conclusions based on volume calculations to determine the site of the capsid protein in the virion (Battisti *et al*., 2012b).

In many enveloped viruses, the lipid membrane faces a positively charged protein surface. Of particular relevance to rubella virus is the nucleocapsid of Sindbis virus (Mukhopadhyay *et al*., 2006) and Ross River virus (Cheng *et al*., 1995), alphaviruses whose nucleocapsid surface is positively charged and surrounded by lipid membranes. A second example is the double-stranded DNA containing Paramecium bursaria Chlorella virus-1 whose capsid protein hugs the exterior of the internal lipid membrane (Nandhagopal *et al*., 2002). A third example is the matrix protein of the pleomorphic Newcastle disease virus which is closely associated with the interior of the viral lipid envelope (Battisti *et al*., 2012b). Thus, it seems likely that the rubella capsid protein also assembles with a predominantly positive surface close to and facing the viral membrane.

If we assume that the positive surface of the capsid protein rows in $P22_12_1$and $C222_1$ crystal forms (Section 2.3.4 and **Figure 2.14**) is positioned parallel to the rubella virion membrane, then the capsid protein dimer axes would be roughly perpendicular to the viral membrane. Furthermore, alternate dimer axes are in opposite directions, implying that every second dimer makes close contacts with the membrane. These contacts are separated by approximately 85 Å (**Figure 2.14C**).

Some rubella virus tomograms show short, parallel ridges separated by approximately 90 Å below the viral membrane. These ridges likely correspond to the rows of capsid protein imers that occur in the crystal structures (**Figure 3.3D,E**). Although the rows of dimers found in the crystals are linear, it is possible that in virions, the rows are slightly bent as would be required for the formation of spherical or elongated RV particles. Tomograms of rubella virus particles show ridges, separated by about 90 Å, on the viral surface formed by the glycoproteins (Battisti *et al*., 2012a). These ridges are approximately orthogonal to the rows of capsid protein dimers (**Figure 3.3D,E**). Thus the arrangement of the glycoproteins E2 and E1 in rubella virions are correlated with the organization of the capsid proteins (**Figure 3.3C**).

Figure 3.3 Rubella virus tomograms. (A, B) Central section of two rubella virus particle tomograms. Magenta arrows: glycoprotein plus membrane layer (average thickness = 100 Å).Yellow arrows: gap between membrane and internal RNA core. The C-terminal domain of the capsid protein is seen as regions of density (dotted arrows) crossing the gap. Cyan arrows: region occupied by RNA plus N-terminal domain of the capsid protein. (C) Diagram of a RV virion cross-section showing the ectodomains of E1 (blue) and E2 (magenta) glycoproteins, the capsid protein dimer (yellow) and genomic RNA (red). The grey region denotes the membrane layer (dark grey) plus the glycoproteins. In the magnified section, shown are the polypeptide connections (thin lines) of the N-termini (□) and C-termini (Δ) of the capsid protein to the RNA and membrane, respectively. The filled and open shapes denote the termini pointing towards and away from the plane of the paper. (D) A section near the surface of a RV virion, at a distance of 345 Å from the particle center. (E) A section at a distance of 210 Å from the particle center and below the external surface of the virion shown in D. The separation of pixels in the tomogram is 15 Å. The scale bar represents 170 Å. Dark regions indicate high density.


A
B
C
D
E
Glycoprotein rows
Capsid protein rows

Based on the above structure of the rubella virion, it was hypothesized that the oligomerization of capsid protein dimers, as seen in the $C222_1$ crystal form (Section 2.3.4), should be essential for virion assembly and infection. Experiments were conducted by Steven Willows in Dr.Tom Hobman's group at the Univeristy of Alberta, Canada, to test the above hypothesis. Results and conclusions from those experiments are briefly described below.

If the organization of capsid protein into rows of dimers is important for the assembly of the rubella virion (**Figure 2.14C**), disrupting the hydrogen bonding between adjacent dimers should have a detrimental effect on virus assembly and/or infectivity. The inter-dimer molecular interface involves amino acid residues 218-221, Gly-Val-Trp-Gly (GVWG). To determine whether the inter-dimer interface of the capsid is important for infectivity, the rubella capsid gene was modified to encode two proline substitutions (PVWP) that should disrupt the hydrogen bonding interactions. Plasmid-based expression of the PVWP capsid protein in transfected cells verified that this mutant was identical in apparent molecular weight to wild-type (Wt) capsid protein (**Figure 3.4A**). The mutant protein also exhibited an identical subcellular localization to the Wt capsid protein (**Figure 3.4B**).

Cells were transfected with in vitro synthesized Wt and PVWP genomic RNAs. The resulting viral titers were determined at 1-3 days post-transfection. Whereas the specific infectivity of Wt RV genomic RNA approached 10^6 plaque-forming units per microgram, cells transfected with PVWP genomic RNA did not produce detectable amounts of infectious virus (**Figure 3.4C**). The production of p150, a viral replicase protein that is translated directly from the viral genomic RNA, was assayed to monitor the efficacy of RNA transfection. Expression of p150 was similar for the Wt and mutant after 1 day. The p150 levels remained relatively constant over time in cells transfected with the mutant RV genome. In contrast, in cells transfected with Wt RV RNA, levels of p150 steadily increased along with capsid levels in a time-dependent manner, a scenario that is consistent with assembly of infectious virus and their infection of neighboring cells (**Figure 3.4D**). Compared to capsid in the Wt RV RNA transfected cells, levels of PVWP capsid were low

because no infectious virus was being made and therefore, mutant capsid is only made in the cells that were originally transfected with the "infectious" RNA (**Figure 3.4D**). To determine if the PVWP capsid protein could support assembly and secretion of virus particles, Vero cells were transfected with plasmids encoding the Wt or PVWP mutant capsid protein plus glycoproteins E2 and E1. It has previously been shown that co-expression of the rubella virus structural proteins results in the assembly of virus-like particles in the Golgi, followed by secretion into the medium (Hobman *et al*., 1994). The virus-like particles, which are antigenically and immunologically indistinguishable from infectious virions (Qiu *et al*., 1994), can easily be recovered from the cell media by ultracentrifugation (Garbutt *et al*., 1999; Law *et al*., 2001). Using this assay, it was observed that virus-like particles were only secreted from cells expressing Wt capsid, E2 and E1 (**Figure 3.4E**). Together, these data suggest that the interface between capsid dimers, as seen in the orthorhombic crystal forms, is critical for assembly of capsid protein into infectious virions.

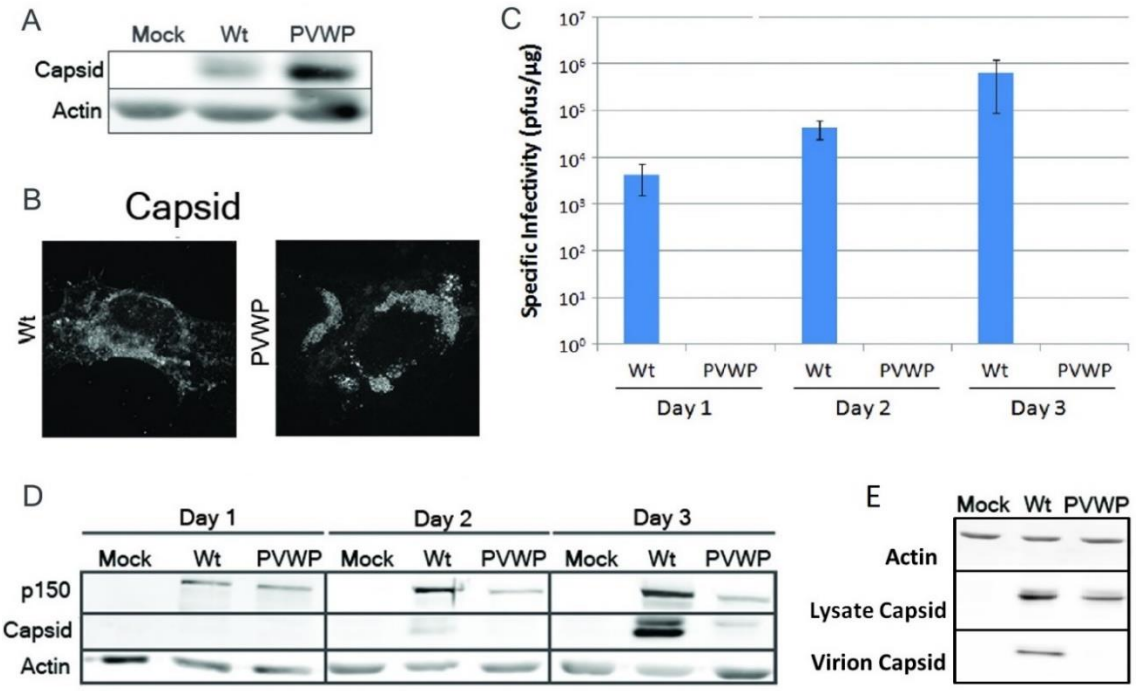


Figure 3.4 Interaction between neighboring capsid dimers is essential for infectious virus production. (A) Lysates of Vero cells transfected with plasmids encoding wild type (Wt) or the PVWP mutant capsid protein were subjected to immunoblotting with anti-RV serum. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a loading control. (B) Indirect immunofluorescence was used to detect localization of Wt and the mutant capsid protein in transfected cells. (C) Vero cells were transfected with in vitro synthesized infectious RNA encoding Wt or PVWP mutant RV genome. Viral titers were determined 1-3 days post-transfection and the specific infectivities (plaque-forming units/µg DNA) of the infectious clones were calculated. (D) Immunoblot analyses of cells transfected with viral RNAs showing expression of capsid and the viral replicase protein p150. Actin was used as a loading control. (E) Vero cells were transfected with plasmids encoding Wt or PVWP capsid and glycoproteins E2 and E1 or vector alone (V). Forty-eight hours post-transfection, virus-like particles were recovered from media by ultracentrifugation. Cell lysates and virus particle (virion) fractions were subjected to immunoblot analyses.

CHAPTER 4. IN VITRO ASSEMBLY OF RUBELLA NUCLEOCAPSID CORES

4.1 Introduction

Rubella virus assembly and budding occurs on the Golgi membrane. This is different from the budding process of the related alphaviruses in which virus budding occurs from the plasma membrane. The alphavirus nucleocapsids are assembled in the cytoplasm whereas the rubella virus nucleocapsid assembly occurs on the Golgi membrane in synchronization with virus budding (Hobman and Chantler, 2007).

Even though, *in vivo* the rubella capsid protein is found in association with the membranes during virus assembly (Frey, 1994), there is no study reporting the requirement of membrane attachment for nucleocapsid assembly. Studies reporting production of rubella VLPs (virus-like particles) using cDNA encoding only the structural proteins of RV have shown that the viral genomic RNA is not required for particle assembly (Hobman *et al*., 1994). The amino acid region 28 to 56 on the capsid protein has been reported to specifically bind to the rubella viral RNA, while a few other regions on the capsid protein have been shown to bind non-specifically to nucleic acids (Liu *et al*., 1996).

In vitro assembly systems for Sindbis virus (type member of alphavirus genus) nucleocapsid using recombinantly produced capsid protein have been established (Tellinghuisen *et al*., 2000). The aim of the work reported in this chapter is to establish conditions to form rubella nucleocapsid assemblies *in vitro*. These nucleocapsids would then be used for electron microscopy studies to gain further information on the structure of the capsid cores and could be correlated with the structural organization of the whole virion.

4.2 Materials and Methods

4.2.1 Cloning

Different lengths of the rubella capsid protein were cloned into pET28a (Novagen) and pET44a (Novagen) vectors. The pET28a vector constructs contain an N-terminal His tag. The pET44a vector constructs contain a His tag plus NusA tag at the N-terminus of the protein of interest. NusA is a transcription termination/antitermination protein from *E.coli* which has high solubility when expressed in the bacterial system. Preparation of the plasmid constructs was carried out similarly as described in section 2.2.1. A list of the constructs along with their primer sequences are given in **Table 4.1**.

Table 4.1 List of rubella capsid protein constructs

Rubella capsid protein residues	Fusion Tag	Primer sequence (5'-3')** (**F**- forward, **R**- reverse)
RVC-N9-277[1]	N-terminus 6X His	**(F)**TAACATATGATGGAGGACCTCCAGAAG **(R)**TTAAAGCTTTCAGCGGATGCGCCA
RVC-N9-277[1]	N-terminus 6X His+ NusA	**(F)**TAAGGATCCATGGAGGACCTCCAGAAG **(R)**TTAAAGCTTTCAGCGGATGCGCCA
RVC-N30-277[1]	N-terminus 6X His	**(F)**TAACATATGGGCGCCTCGCAGTCG **(R)**TTAAAGCTTTCAGCGGATGCGCCA
RVC-N30-277[1]	N-terminus 6X His + NusA	**(F)**TAAGGATCCATGGGCGCCTCGCAGTCG **(R)**TTAAAGCTTTCAGCGGATGCGCCA
RVC-N44-277[1]	N-terminus 6X His	**(F)**TAACATATGGACTCCAGCACCTCCG **(R)**TTAAAGCTTTCAGCGGATGCGCCA

Table 4.1 Continued

RVC-N44-277[1]	N-terminus 6X His + NusA	**(F)**TAAGGATCCATGGACTCCAGCACCT CCG **(R)**TTAAAGCTTTCAGCGGATGCGCCA
RVC-N65-277[1]	N-terminus 6X His	**(F)**TAACATATGAACCGGGGCCGT **(R)**TTAAAGCTTTCAGCGGATGCGCCA
RVC-N87-277[1]	N-terminus 6X His	**(F)**TAACATATGACTCGCTCCCAGACT **(R)**TTAAAGCTTTCAGCGGATGCGCCA
RVC-N109-277[1]	N-terminus 6X His	**(F)**TAACATATGATGCAAACCGGGCG **(R)**TTAAAGCTTTCAGCGGATGCGCCA

[1]RVC denotes rubella capsid protein and N refers to the amino terminus of the protein followed by the amino acids residue numbers

4.2.2 Expression of plasmid constructs

Expression tests were carried out with the positive plasmid clones similar to the procedure described in Section 2.2.2.

4.2.3 Large-scale expression and purification of pET44a constructs

The pET-44a capsid protein constructs RVC-N9-277, RVC-N30-277 and RVC-N44-277, that showed soluble expression of the fusion protein, were used for larger scale expression of the protein. The clones were grown in 1 liter LB medium at 37 °C to an O.D reading of 0.6-0.8. The expression of the fusion protein was then induced with 1 mM IPTG. The induced cultures were grown at 25 °C for 6 hours before harvesting the cells. The cells were pelleted by centrifugation (Rotor ID: JLA-9.1) at 6000 rpm for 15 minutes at 4 °C. The cell pellets were re-suspended in 50 mM Tris, 500 mM NaCl, 20 mM imidazole (buffer A) and lysed in ice using a sonicator. The lysed cells were clarified by centrifugation at 15,000 rpm for 25 minutes using a Beckman Coulter centrifuge (Rotor ID: JA-20). The supernatants were incubated with Ni-NTA beads (Qiagen) under constant stirring for 1 hour at 4 °C. The supernatant with the beads were then packed into a 5 ml chromatography

column. The column was washed with 10 column volumes of buffer A followed by 2 column volumes of cleavage buffer (20 mM Tris at pH 8.0, 120 mM NaCl, 10 mM $CaCl_2$). The beads with the fusion protein were then incubated with 0.02 µg/ml of enterokinase (New England Biolabs) for different time intervals (2, 5 and 16 hours) at room temperature to test the efficiency of protease cleavage. The protease cleavage products were then collected and analyzed using 12% SDS-PAGE.

4.2.4 Large-scale expression and purification of pET28a constructs

The soluble pET28a constructs, RVC-N65-277 and RVC-N85-277, were scaled up for larger volume expression and purification. The procedure for protein expression and purification was similar to the procedure outlined in the previous section. However, the expressed proteins that were bound to the Ni-NTA column were not treated with enterokinase. The bound proteins were eluted from the column using 5 column volumes of buffer containing 50 mM Tris, 500 mM imidazole and 500 mM NaCl at pH 8.0. The eluted proteins were applied to a Superdex 200 gel filtration column (GE Lifesciences) and analyzed using negative-stain cryo-electron microscopy.

4.2.5 Negative-stain cryo-electron microscopy

About 3µl of the sample was added to home-made holey carbon grids (EMS, G400-Cu) and incubated for 1 minute. 3 µl of 1% uranyl acetate solution was added to this and incubated for 30 seconds. The grid was washed thrice with 15 µl of distilled water and blotted with Whatmann filter paper. The grids were then imaged at 200 kV in the FEI CM200 microscope at a magnification of 35,000X and 17,000X with defocus between 5-7 µm underfocus.

4.2.6 Purification of RVC-N65-277-His protein dimers

The RVC-N65-277 construct in pET28a was expressed as outlined in Section 4.2.4. Twenty units of benzonase (Sigma) was added along with 5 mM $MgCl_2$ (final concentration) and protease-inhibitor cocktail mini-tablet (Roche) to the cell pellets during re-suspension in the binding buffer (Section 4.2.4). The cells were lysed and purified as outlined in

Section 4.2.4. The only difference was that after cell lysis, the lysate was incubated under constant stirring at room temperature for 1 hour before clarifying the lysate by centrifugation. The purified protein sample was analyzed using Superdex 200 (GE Lifesciences) size-exclusion column. The buffer condition for the size-exclusion experiment was 50 mM Tris at pH 8.0 with 500 mM NaCl.

4.2.7 Analysis of capsid cores

Size-exclusion chromatography

The capsid oligomers were applied to Hi-prep Sephacryl S-400 column (GE Lifesciences). The column was run at 0.4 ml/min at 4 ºC in 50 mM Tris, 100 mM NaCl at pH 8.0.

Agarose gel

The capsid oligomers were run on a 1% agarose gel in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA at pH 8.0) at 90 volts for 2 hours at room temperature. The gel was then imaged for presence of nucleic acids using UV transillumination and then subsequently stained with Coomassie blue to detect presence of protein.

Cryo-electron electron microscopy

The RVC-N65-277-His protein complex was frozen as a vitrified sample in liquid ethane on 200 mesh EM grids (R 2/2 Quantifoil; Micro Tools GmbH, Jena, Germany). The grids were imaged at 200 kV under the FEI CM200 microscope at a magnification of 38000X (effective magnification 51500X) and defocus of 5 μm underfocus. The grids were also imaged at 300 kV with an FEI Titan Krios microscope equipped with a Gatan Tridiem energy filter (GIF 863). CCD (charge-coupled detector) data was collected at a magnification of 47000X (effective magnification 79100X) at defocus of about 3.5 μm underfocus. The collected images were boxed using the EMAN software package (Ludtke *et al*., 1999).

4.3 Results

4.3.1 Cloning and expression

The positive clones for the different constructs of the rubella capsid protein were tested for soluble expression at 25 ºC for 6 hours on induction with 1mM IPTG. The results for the protein expression tests are given in **Table 4.2**.

Table 4.2 Expression tests on the rubella capsid protein constructs.

	N-terminal His Tag	N-terminal (His + NusA) Tag
RVC-N9-277[1]	Insoluble	Soluble
RVC-N30-277[1]	Insoluble	Soluble
RVC-N44-277[1]	Insoluble	Soluble
RVC-N65-277[1]	Soluble	N/A
RVC-N87-277[1]	Soluble	N/A
RVC-N109-277[1]	Insoluble	N/A

[1]RVC denotes rubella capsid protein and N refers to the amino terminus of the protein followed by the amino acids residue numbers

4.3.2 Large-scale expression and purification of pET44a constructs

The pET44a constructs of the rubella capsid protein were all soluble on expression. But protease treatment (using enterokinase) to remove the fusion tag always resulted in non-specific cleavage of the proteins in all constructs. Different buffer conditions, incubation temperatures and length of incubation were tried but were unsuccessful. The most obvious protein bands after enterokinase treatment for the fusion protein RVC-N9-277 with NusA tag was sent for mass-spectrometry analysis (Electrospray ionization, LC-MS/MS) at the proteomics facility, Bindley Biosciences, Purdue University. The mass-spectrometry results indicated that band#3 is the rubella capsid protein (**Figure 4.1**). The size of the rubella capsid band on the gel was much lower than the actual construct, but similar to the stable capsid protein fragment which had been crystallized. Thus, these

results showed that the amino terminal region of the rubella capsid protein was disordered and hence enterokinase treatment resulted in non-specific peptide cleavage.

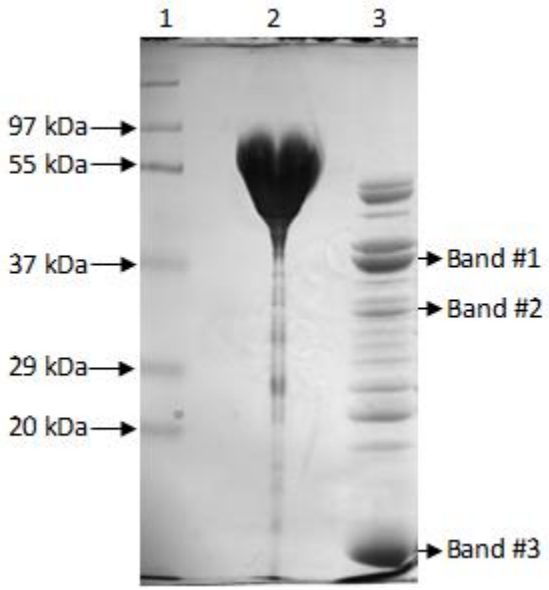


Figure 4.1 12% SDS-PAGE. Lane 1: Standard marker, Lane 2: Fusion protein (RVC-N9-277-NusA), Lane 3: Protein products after enterokinase treatment (0.5% (v/v) enzyme after incubation at 25 °C for 5 hours). Band 1,2,3 indicate bands sent for mass-spectrometry analysis.

4.3.3 Large-scale expression and purification of the pET28a constructs

Soluble expressions of the pET28a constructs were obtained for capsid protein regions 65-277 and 87-277 amino acids. These constructs were purified with Ni-NTA affinity column using a linear gradient for elution (**Figure 4.2**), followed by gel filtration chromatography. Gel-filtration chromatography showed that the purified proteins were forming large aggregates (**Figure 4.3**).

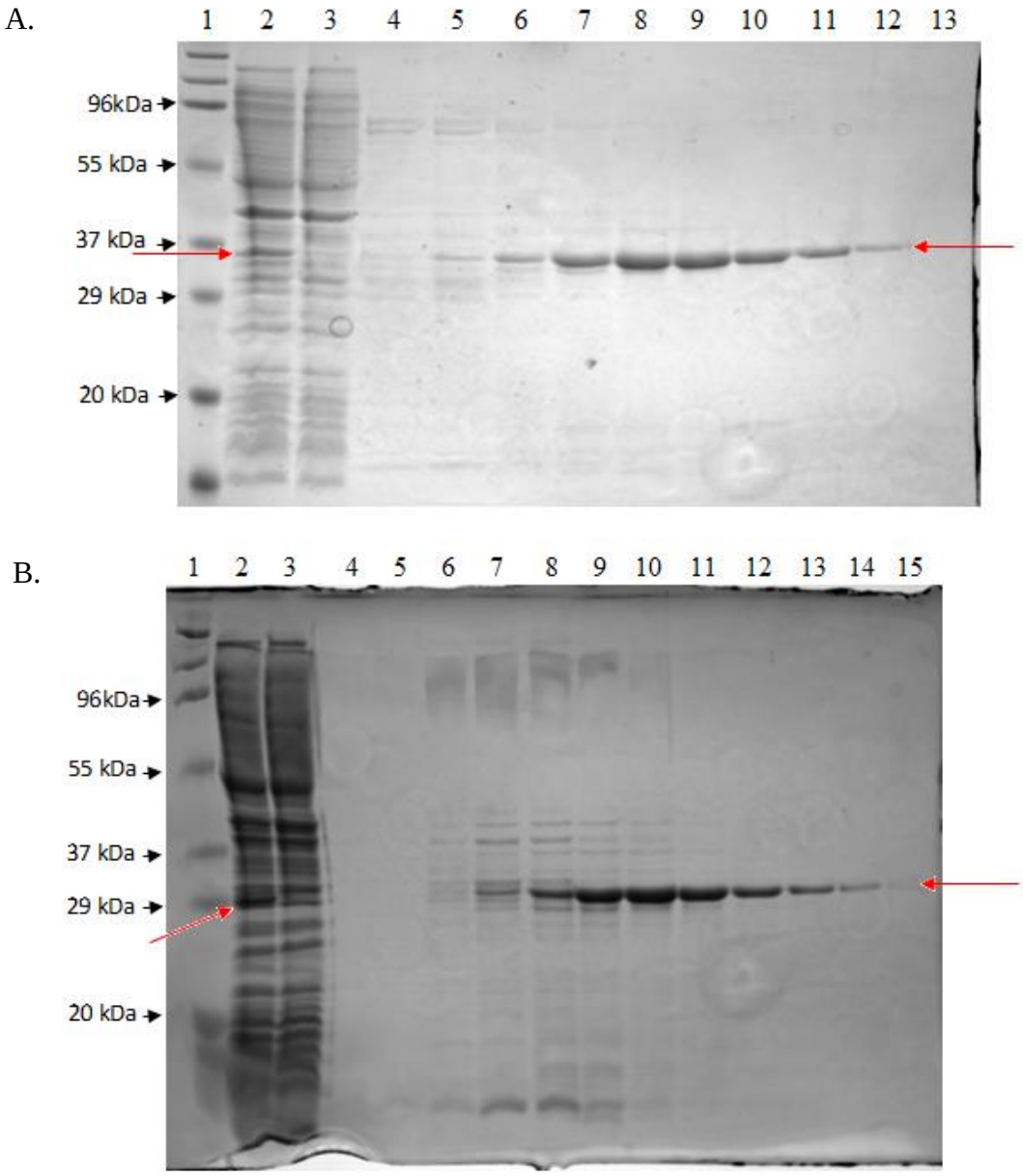


Figure 4.2 12% SDS-PAGE of Ni-NTA purified proteins. A. Purified RV-N65-277-His protein. Lane 1: Protein Ladder. Lane 2: clarified lysate. Lane 3: flowthrough from the Ni-NTA column. Lane 4: wash from the column. Lane 5-13 are the elution fractions. B. Lane 1: Protein Ladder marker, Lane 2: clarified lysate, Lane 3: flowthrough from the Ni-NTA column, Lane 4: wash from the column, Lanes 5-13: elution fractions. B. Purified RV-N87-277-His protein. Lane 1: Protein ladder, Lane 2: clarified lysate, Lane 3: flowthrough from the Ni-NTA column, Lanes 4 and 5: wash from the column, Lanes 6-15 are the elution fractions. Red arrows indicate the position of the protein of interest.

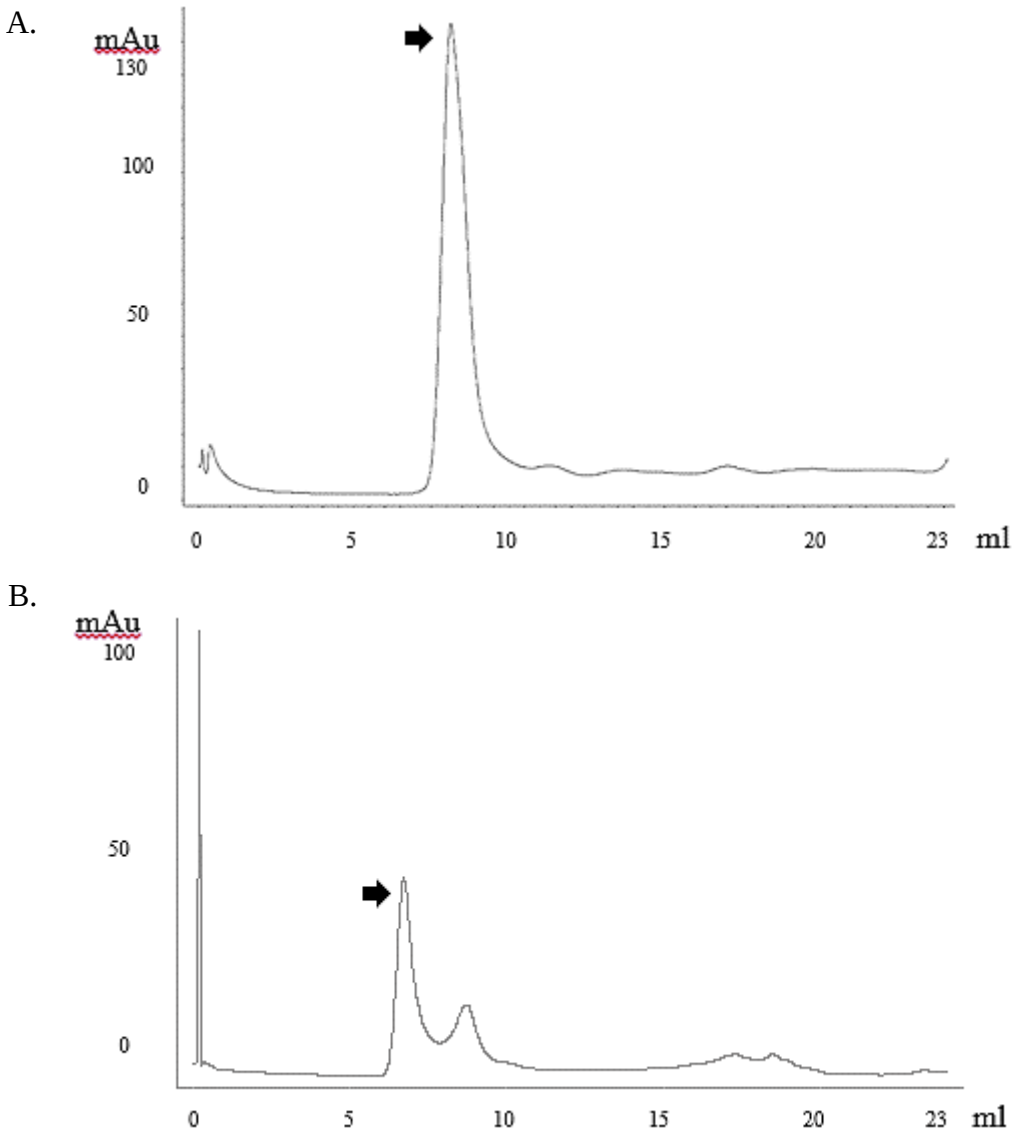


Figure 4.3 Gel filtration absorption curves at 280nm. The y-axis is in mAu (milli absorption units) and the x-axis is the elution volume (ml). A. Absorption curve for RVC-N65-277-His on superdex-200 column. B. Absorption curve for RVC-N87-277-His on superdex-75 column. The black arrow indicates the elution position of the protein of interest in each. Both proteins elute in the void volume.

As the capsid protein is expected to form large assemblies, these protein aggregates were analyzed by negative staining under the CM200 transmission electron microscope

(**Figure 4.4**). The images showed some spherical shaped particles that might resemble capsid cores. Based on these initial experiments, further work was mainly carried out with the RV capsid 65-277 amino acids construct (RVC-N65-277-His) as it was the longer construct.

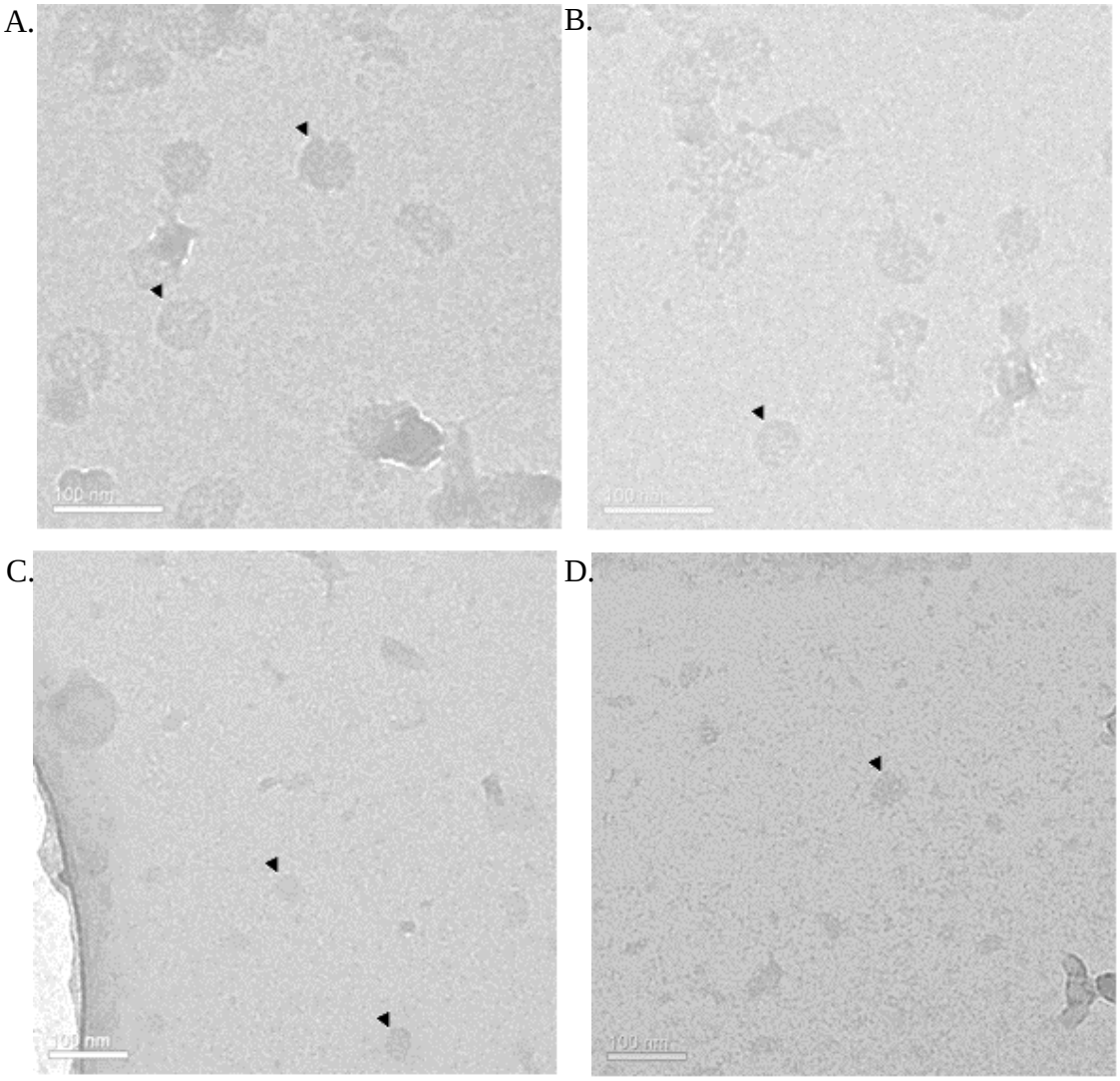


Figure 4.4 Negative stain images of capsid cores. A and B. Negative stain images of RVC-N65-277-His aggregates (Magnification 35000X). C and D. Negative stain images of RVC-N87-277-His aggregates (Magnification 17000X). Black arrowheads point at some of the core-like particles. Scale bar is 100nm.

4.3.4 Purification of RVC-N65-277-His capsid protein dimer

RVC-N65-277-His protein can be purified in its dimeric form by treating the bacterial cells with benzonase in the presence of 5 mM $MgCl_2$ during cell lysis. The cell lysate was then incubated at room temperature for 1 hour before centrifugation to remove cell debris. The RVC-N65-277-His protein was purified using Ni-NTA column. The dimer form could be separated from the larger complexes using a size-exclusion column (**Figure 4.5**). The dimeric RVC-N65-277-His was stable only at salt (NaCl) concentrations of 500 mM or above. Lowering of salt concentration resulted in precipitation of the protein. The yield of dimeric RV-N65-277-His with respect to the oligomeric form could be increased by incubating the cell lysate for longer periods with benzonase.

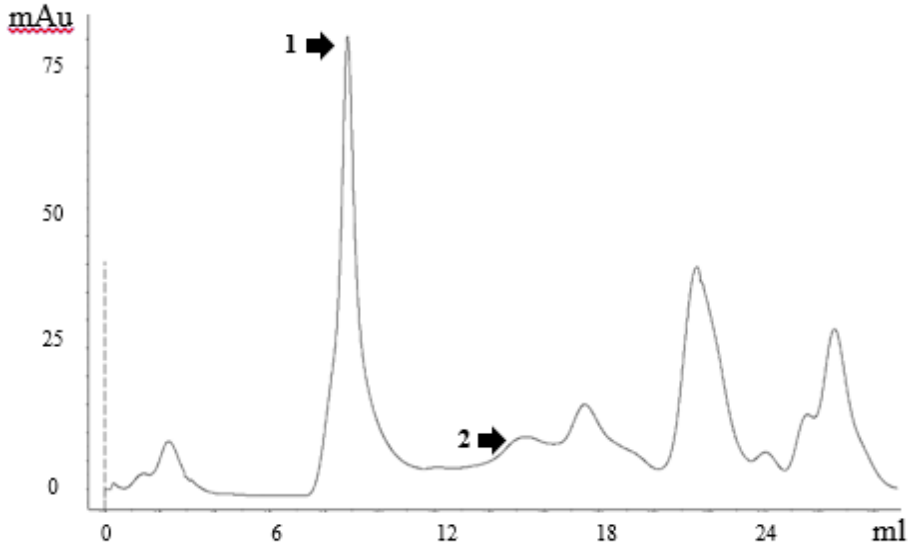


Figure 4.5 Elution profile at UV wavelength of 280nm from superdex-200 column. Peak 1 is the RVC-N65-277-His nucleocapsid complex and peak 2 is the dimeric form of the protein. The dotted line indicates the sample injection point.

4.3.5 Analysis of capsid cores

Size-exclusion chromatography

The RVC 65-277 protein-nucleic acid complex was run on a Sephacryl S-400 column which is used for separating particles with very large molecular weights (in the kilo and mega Dalton ranges). The elution profile from this column (**Figure 4.6**), when compared with the standard elution curves for this column, indicated that the RVC-N65-277-His protein-nucleic acid complex had a molecular weight in the range of 3 mega daltons.

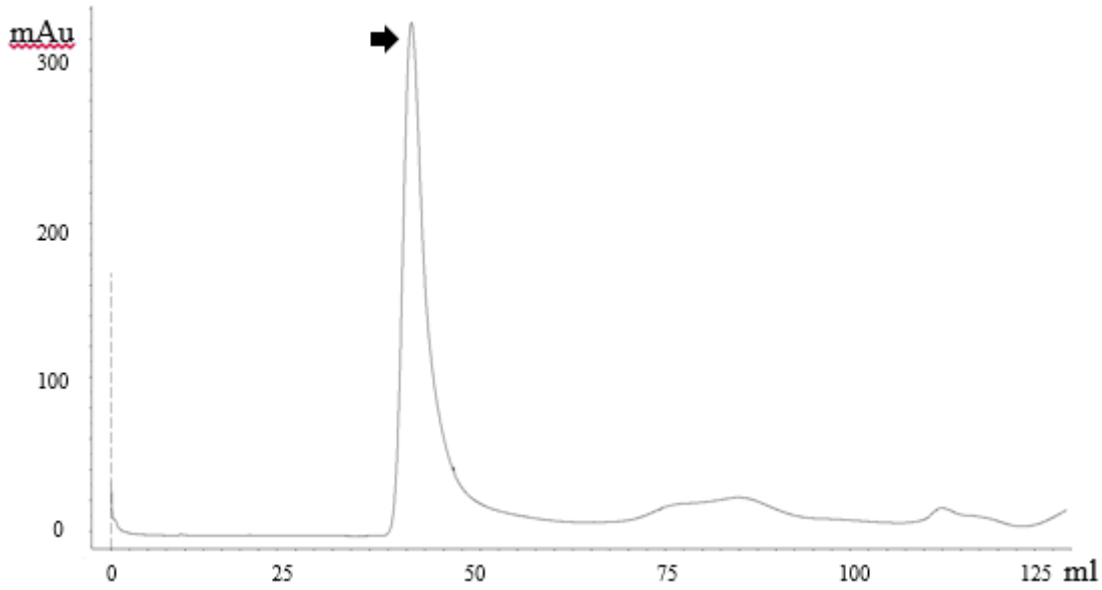


Figure 4.6 Gel filtration absorption curve for RVC-N65-277 at wavelength of 280nm. The y-axis is in mAu (milli absorption units) and the x-axis is the elution volume (ml). The black arrow indicates the elution position of the protein of interest. The dotted line indicated the position of sample injection.

Agarose gel assay

The purified RVC 65-277 aggregate complexes were analyzed using agarose gel. The gel was imaged for presence of nucleic acids using UV transillumination and then

subsequently stained with Coomassie blue to detect presence of protein. The results are shown in **Figure 4.7**. The RVC-N65-277-His complexes did not move much into the gel, but the position of the nucleic acid band overlapped well with the position of the protein band, indicating that the complexes were between protein and nucleic acids. These complexes were not homogenous as could be seen from the smearing of the bands. As the complexes were formed during lysis and purification of the recombinant protein, these nucleic acid fragments must be random DNA/RNA from the bacterial cell. Presence of nucleic acids in the nucleocapsid complexes could also be detected using an UV spectrophotometer at 260 nm wavelength.

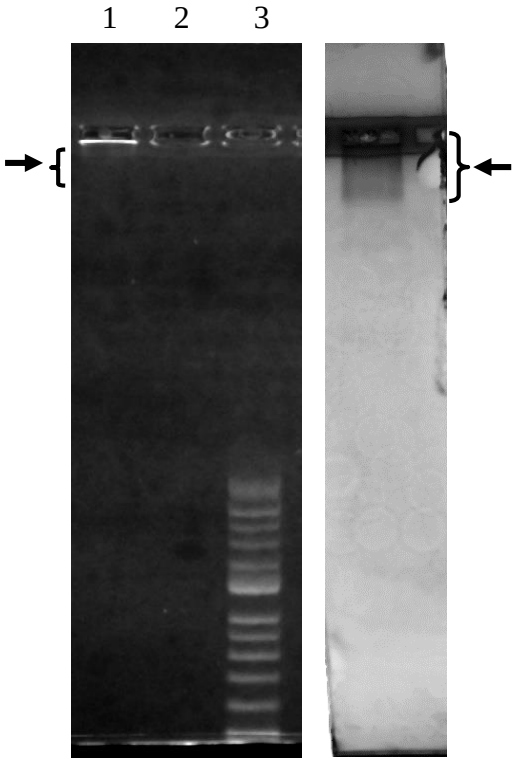


Figure 4.7 Agarose gel assay. Left: UV transillumination of agarose gel. The bright bands indicate nucleic acids. Lane 1: RVC-N65-277-His, Lane 3: Standard ladder with highest band size of 2 kb. Right: Lane 1 of agarose gel after staining with Coomassie blue. Dark bands indicate protein content. The black arrows indicate the position of bands on both the gel.

Cryo-electron microscopy

The capsid core complexes were next prepared for analysis using cryo-electron microscopy. Freezing of the grids had some issues with forming good ice. This was because the protein complexes were not uniform as seen from negative stain images. The presence of some very large aggregates caused very thick ice formation on freezing. After testing different blotting times during freezing, grids were frozen with workable ice thickness. These grids were analyzed under the CM200 transmission electron microscope. The contrast of sample on the grid was very low, hence high defocus was needed to identify regularly shaped particles. Some of these particles are shown in **Figure 4.8**. These particles have varied sizes and appear hollow. Regularly shaped particles made up approximately 10% of the total sample on the grids.

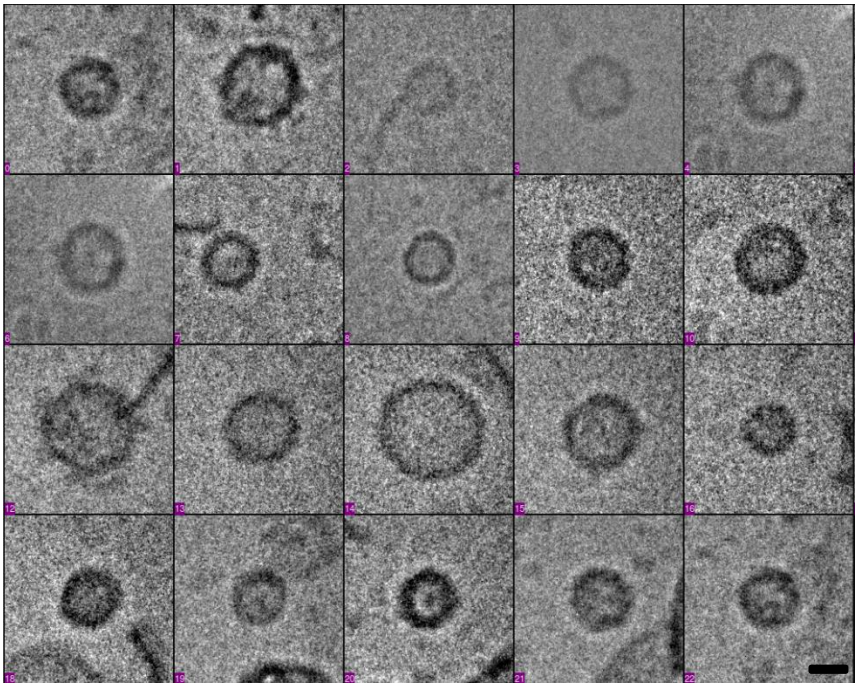


Figure 4.8 Boxed particles of RVC-N65-277 cores. Particles from CCD images collected at 38000X magnification on the CM200 microscope. Black is high density. Scale bar indicates 200 Å.

Frozen grids of the nucleocapsid particles were then analyzed under the Titan-Krios microscope and CCD data was collected. Using the Titan microscope, particles with different shapes and sizes were identified. About 100 CCD images were collected with 2-3 particles per image. These were then boxed using the EMAN software package and analyzed for similarities between particles. Some of the boxed particles are shown in **Figure 4.9**. The rubella nucleocapsid cores appeared to have a smooth surface. The diameter of the particles ranged from 200-900 Å. The core particles also showed two concentric layers of high density. The inner part of the core appeared to be less dense than the outer layers.

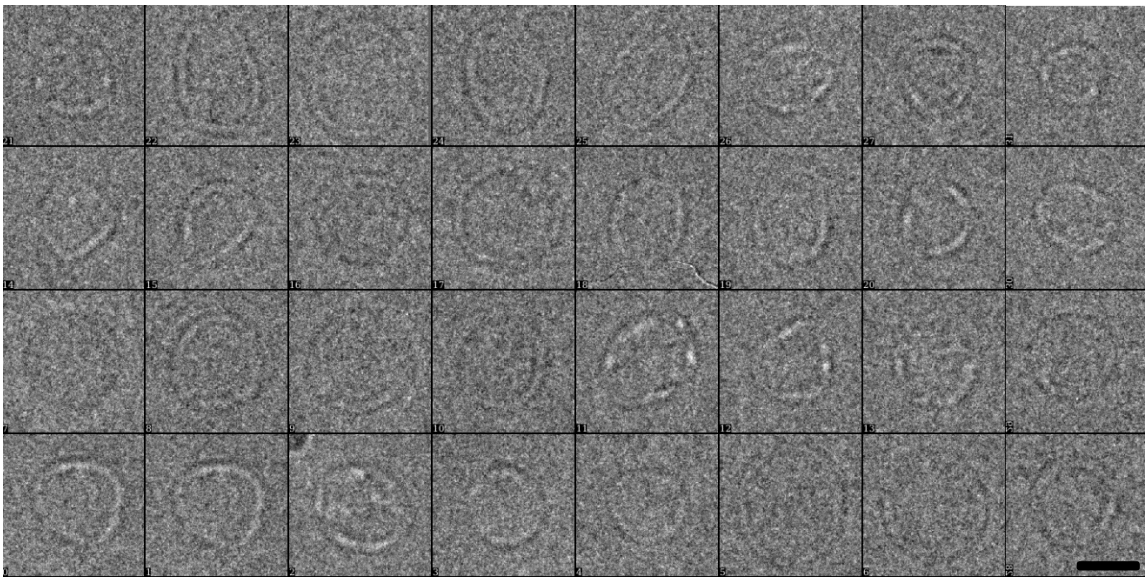


Figure 4.9 Boxed particles of RVC-N65-277 cores. Particles from CCD images collected on the Titan-Krios microscope. Black is high density. Scale bar indicates 200 Å.

4.4 Discussion

Nucleocapsid core formation has been well studied for alphaviruses. For the rubella virus, there are no previous reports of *in vitro* assembly of nucleocapsid cores. From the above study, it is clear that though the formation of rubella virus nucleocapsid cores occurs on the Golgi membrane in infected cells, presence of a membrane is not required for formation of the nucleocapsids. In alphaviruses, capsid protein constructs with truncations of up to 81 amino-terminal residues can still interact with nucleic acids, but core formation occurs spontaneously only with constructs having truncations up to 19 amino-terminal residues (Tellinghuisen *et al*., 2000). In contrast, with rubella virus, capsid constructs with truncations up to 87 amino-terminal residues have the ability to form nucleocapsid cores. Moreover, the recombinantly produced rubella capsid protein constructs seem to spontaneously form nucleocapsid cores during cell growth and lysis. This is different from the alphavirus capsid protein, which can be purified in its protomeric form from cells without using RNase/DNAse treatment. However, in rubella virus and alphaviruses, capsid cores can be formed in the presence of non-specific nucleic acids just as in alphaviruses.

The rubella nucleocapsid cores have different shapes and sizes as seen from electron microscopic images. This again stresses the inherent pleomorphic nature of the virus. The EM images show particles with less density in the interior of the particles, which suggests that the interior of the nucleocapsids are nearly empty with probably small amounts of nucleic acids within it.

Imaging of the nucleocapsid particles with the Titan microscope show that the cores have a smooth outer surface and a double layer architecture. Better quality and detail in the images collected from the Titan microscope can be attributed to the fact it has a 4K CCD detector as compared to a 1K CCD detector in the CM200. The Titan also has better beam coherence, stage stability and higher voltage when compared to the CM200 microscope. These properties further contribute to obtaining higher resolution images from the Titan microscope.

Alphaviruses have a single capsid protein layer in their nucleocapsids. For rubella virus, there is no direct structural information available on its nucleocapsid features. Thus, the double layer feature of the rubella nucleocapsid, observed in the EM images, is an

interesting and new observation. This suggests that the rubella nucleocapsid cores may have a very different organization when compared to alphaviruses.

CHAPTER 5. PURIFICATION OF AN IMMATURE FORM OF RUBELLA VIRUS

5.1 Introduction

Freeze-substitution cryo-electron microscopy studies of cells infected with rubella virus (Risco.*et al*., 2003) have identified virion particles that are morphologically different from the mature rubella virions. The particles have been named 'immature' particles. These immature particles have been found close to budding sites inside infected cells and appear as dense, homogenous precursors to the mature rubella virus particles. Only the mature form of the particles are released from the cells. This maturation process takes place in the trans-Golgi network. There are, however, no reports on factors that may be responsible for maturation of rubella virions. As the Golgi network is at a different pH compared to the cytoplasm, it is possible that the change in pH has an effect in maturation of the precursor RV virions. In flaviviruses, the acidic pH in the Golgi network is required for processing the prM protein on the immature virus to pr peptide and M protein which leads to maturation of the virion (Yu *et al*., 2008). Thus, a potential way to isolate the immature RV particles maybe to disrupt the change in pH in the Golgi network and analyze its effect on virus production.

5.2 Materials and Methods

5.2.1 Virus preparation and purification from cell medium

Ammonium chloride was used to neutralize the pH gradient in the secretory pathway of cells. It has been used successfully to purify immature dengue virus, maturation which requires low pH induced cleavage of its pr peptide (Yu *et al*., 2008). Thus, the rubella virus particles were propagated and purified in a similar way as described in Section 3.2.1 with a modification to include 20mM ammonium chloride (final concentration) to the cell-growth medium after virus infection.

5.2.2 Purification of intra-cellular virions

Protocol for purification of the intracellular virions was adapted from Aboud *et al*., 1982. Vero cells infected with rubella virus and grown in cell medium containing 20mM ammonium chloride for 48 hours, were washed 2-3 times with cold PBS (phosphate buffer saline, pH 7.4). The cells were then treated with 1% trypsin for 2-3 minutes at room temperature to disperse the cells. Medium containing DMEM supplemented with 5% FBS was added to the cells and the mixture was pelleted at 1000 x g for 10 minutes. The cell pellets were washed with cold PBS 2-3 times. The cell pellets were then re-suspended in a hypotonic buffer containing 20 mM Tris, 10 mM NaCl and 15 mM $MgCl_2$ at pH 8.0 for 15 minutes at 4 ºC. The suspension was homogenized using 20 strokes in a Dounce glass homogenizer with a tight-fitting pestle (7 ml volume). The homogenate was then vortexed for 5 minutes. The vortexed mixture was centrifuged for 15 minutes at 4000 rpm to remove cell debris. The supernatant was again centrifuged at 15000 rpm (rotor ID: JA-20) for 20 minutes to remove organelles and broken cell debris. Sodium chloride was then added to the supernatant to bring the salt concentrations to 120 mM in the sample. This was mixed with 8% of PEG-8000 (final concentration) for precipitation of the virus particles. Further purification was carried out in a similar way to the virus purification from cell medium (Section 3.2.1).

5.2.3 Cryo-electron microscopy

The virus samples were frozen with BSA-conjugated 10 nm colloidal gold solution as outlined in Section 3.2.3. The grids were imaged using CCDs in the FEI Titan-Krios microscope at a magnification of 19500X and approximately 5 μm underfocus.

5.3 Results

Rubella infected cells treated with ammonium chloride (NH_4Cl) exhibited degradation of cellular structure and cell death similar to untreated rubella infected cells. However, the yield of purified virus from the NH_4Cl treated rubella-infected cell medium was nearly ten times lower than the un-treated rubella preparations. Very few particles could be observed with cryo- EM (**Figure 5.1**). Thus, the addition of NH_4Cl had a detrimental effect on virus yields during purification.

It was hypothesized that this reduction in virus yield may be due to the inability of the 'immature' rubella virus particles to bud out of the cell. Thus, the infected Vero cells (treated with ammonium chloride) were lysed and analyzed for the presence of intracellular virus particles. Preparations from the Vero cell lysates were applied to an Opti-prep gradient. Analysis of the gradient column after ultra-cetrifugation did not show the presence of any bands in the region normally occupied by rubella virus particles.

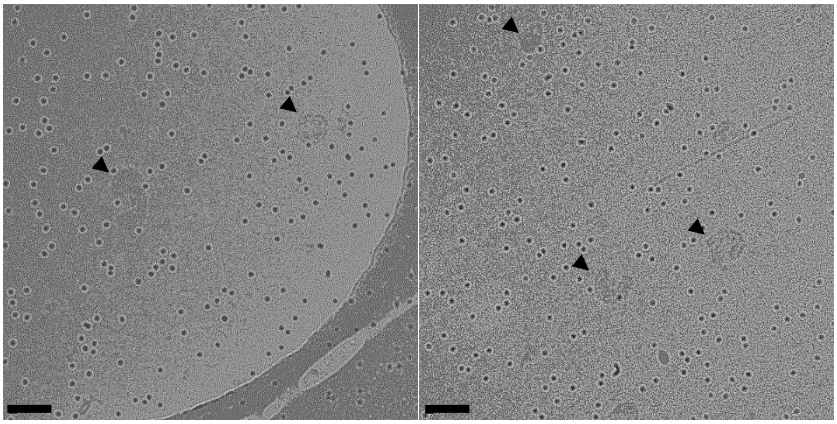


Figure 5.1 Cryo-electron microscopic images of immature rubella particles. The black arrowheads denote the virus particles. The other black dots are the 10 nm BSA-gold particles. The scale bar represents 100 nm.

5.4 Discussion

Occurrence of maturation in the rubella virus life-cycle has been reported for the first time by Risco *et al*., 2003. The maturation process has only been described phenotypically, and factors affecting the process are unknown. In this chapter, the effect of destroying the pH gradient in infected cells was tested as a possible factor for the maturation of rubella virions. The experimental observations show that the addition of ammonium chloride to disrupt the pH change in infected cells had a detrimental effect on the rubella virion production. As the assembly of virions occurs on the Golgi membranes and budding of the virus takes place into the Golgi, it is probable that disrupting the pH of the infected cells has an adverse effect on virus assembly and, in turn, inhibits virus production. The lower pH of the Golgi, thus, may be required for triggering the virion assembly process and not for the maturation process. The maturation of the rubella virions could be due to other factors, such as, some form of proteolytic processing of the virion components in the Golgi leading to re-arrangement of the components in the virion. However, further studies would be required to test this hypothesis. At this stage, it is not yet clear as to what other factors might be involved in triggering the maturation of the rubella virion particles in infected cells.

LIST OF REFERENCES

LIST OF REFERENCES

(1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr Pt 5 *50* **:** 760-763.

Aboud, M., Wolfson, M., Hassan, Y., and Huleihel, M. (1982). Rapid purification of extracellular and intracellular moloney murine leukemia virus. Arch. of Virol. *71*, 185-195.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallography *D66*, 213-221.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352-367.

Akervall, K., Strandberg, B., Rossmann, M.G., Bengtsson, U., Fridborg, K., Johannisen, H., Kannan, K.K., Lovgren, S., Petef, G., and Oberg, B. (1971). X-ray diffraction studies of the structure of satellite tobacco necrosis virus. Cold Spring Harbor Symp. Quant. Biol. *36*, 469-488.

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

Battisti, A.J., Yoder, J.D., Plevka, P., Winkler, D.C., Mangala Prasad, V., Kuhn, R.J., Frey, T.K., Steven, A.C., and Rossmann, M.G. (2012a). Cryo-electron tomography of rubella virus. J. Virol. *86*, 11078-11085.

Battisti, A.J., Meng, G., Winkler, D.C., McGinnes, L.W., Plevka, P., Steven, A.C., Morrison, T.G., and Rossmann, M.G. (2012b). Structure and assembly of a paramyxovirus matrix protein. Proc. Natl. Acad. Sci. U.S.A. *109*, 13996-14000.

Baron, M.D., and Forsell, K. (1991). Oligomerization of the structural proteins of rubella virus. Virology *185*, 811-819.

Beatch, M.D., Everitt, J.C., Law, L.J., and Hobman, T. C. (2005). Interactions between rubella virus capsid and host protein p32 are important for virus replication. J. Virol. *79*, 10807–10820.

Beatch, M., and Hobman, T. (2000). Rubella virus capsid associates with host cell protein p32 and localizes to mitochondria. J. Virol. *74*, 5569-5576.

Blow, D.M., and Rossmann, M.G. (1961). The single isomorphous replacement method. Acta Crystallpgraphy *14*, 1195-1202.

Centers for Disease Control and Prevention (2010). Progress toward control of rubella and prevention of congenital rubella syndrome - worldwide, 2009. Morbidity and Mortality Weekly Report *59*, 1307-1310.

Chen, M., Icenogle, J. (2004). Rubella virus capsid protein modulates viral genome replication and virus infectivity.J. Virol. *78*, 4314–4322.

Cheng, R., Kuhn, R.J., Olson, N.H., Rossmann, M.G., Choi, H., Smith, T.J., and Baker, T.S. (1995). Nucleocapsid and glycoprotein organization in an enveloped virus. Cell *80*, 621-630.

Choi, H.K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M.G., and Wengler, G. (1991). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature *354*, 37-43.

Clarke, D.M., Loo, T.W., Hui, I., Chong, P., and Gillam , S. (1987). Nucleotide sequence and *in vitro* expression of rubella virus 24 subgenomic messenger RNA encoding the structural proteins E_1, E_2 and C. Nucleic Acids Research *15*, 3041-3057.

Cosslett, V.E. (1978). Radiation damage and chromatic aberration produced by inelastic scattering of electrons in the electron microscope: statement of the problem. Annals of the New York Academy of Sciences *306*, 3-13.

Cong, H., Jiang, Y., and Tien, P. (2011). Identification of the myelin oligodendrocyte glycoprotein as a cellular receptor for rubella virus. J. Virol. *85*, 11038-11047.

Crowther, R.A. (1972). The fast rotation function. In The Molecular Replacement Method, M.G. Rossmann, ed. (New York: Gordon & Breach), pp. 173-178.

DuBois, R.M., Vaney, M.C., Tortorici, M.A., Kurdi, R.A., Barba-Spaeth, G., Krey, T., and Rey, F.A. (2013). Functional and evolutionary insight from the crystal structure of rubella virus protein E1. Nature *493*, 552-556.

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

Fontana, J., Lopez-Iglesias, C., Tzeng, W., Frey, T.K. (2010). Three-dimensional structure of rubella virus factories. Virology *45*, 579-591.

Frank, J., Ed. (2006). Electron tomography: Methods for three-dimensional visualization of structures in the cell. 2nd Edition. (New York, Spinger Science and Business Media, LLC).

Frey, T.K. (1994). Molecular biology of rubella virus. Adv. Virus Res. *44*, 69-160.

Garbutt, M., Law, L.M., Chan, H., and Hobman, T.C. (1999). Role of rubella virus glycoprotein domains in assembly of virus-like particles. J. Virol. *73*, 3524-3533.

Griffin, D.E. (2007) Alphaviruses. In Fields Virology, D.M. Knipe, and P.M. Howley, eds. (Philadelphia, Lippincott Williams & Wilkins), pp 1023-1068.

Grunewald, K., Medalia, O., Gross, A., Steven, A.C., and Baumeister, W. (2003). Prospects of electron cryotomography to visualize macromolecular cmplexes inside cellular components: implications of crowding. Biophy. Chem. *100*, 577-591.

Ho-Terry, L., and Cohen, A. (1980). Degradation of rubella virus envelope components. Arch. Virol. *65*, 1-13.

Hobman, T.C., and Chantler, J. (2007). Rubella virus. In Fields Virology, D.M. Knipe, and P.M. Howley, eds. (Philadelphia, Lippincott Williams & Wilkins), pp. 1069-1100.

Hobman, T., Lundstrom, M.L., Mauracher, C.A., Woodward, L., Gillam, S., and Farquhar, M.G. (1994). Assembly of rubella virus structural proteins into virus-like particles in tansfected cells. Virology *202*, 574-585.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucl. Acids. Res. *38*, W545-549.

Iancu, C.V., Wright, E.R., Benjamin, J., Tivol, W.F., Dias, D.P., Murphy, G.E., Morrison, R.C., Heymann, J.B., and Jensen, G.J. (2005). A flip-flop rotation stage for routine dual-axis electron cryotomography. J. Struct. Biol. *151*, 288-297.

Ilkow, C., Weckbecker, D., Cho, W., Meier, S., Beatch, M., Goping, I., Herrmann, J., and Hobman T. (2010). The rubella virus capsid protein inhibits mitochondrial import. J.Virol. *84*, 119–130.

Ilkow, C., Goping, I., and Hobman, T. (2011). The rubella virus capsid is an anti-apoptotic protein that attenuates the pore-forming ability of Bax. PLOS Pathogens *7*.

Ilkow, C.S., Mancinelli, V., Beatch, M.D., and Hobman, T.C. (2008). Rubella virus capsid protein interacts with poly(A)-binding protein and inhibits translation. J. Virol. *82*, 4284-4294.

Ilkow, C.S., Willows, S.D., and Hobman, T.C. (2010). Rubella virus capsid protein: a small protein with big functions. Future Microbiology *5*, 571-584.

Katow, S., and Sugiura, A. (1985). Antibody response to individual rubella virus proteins in congenital and other rubella virus infections. J. Clin. Microbiol. *21*, 449-451.

Katow, S., and Sugiura, A. (1988). Low-pH induced conformational change of rubella virus envelope glycoproteins. J. Gen. Virol. *69*, 2797-2807.

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C. (1960). Structure of myoglobin: A three dimensional Fourier synthesis at 2Å resolution. Nature *185*, 422-427.

Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. *116*, 71-76.

Ladd, M., and Palmer, R. (2003). Structure determination by x-ray crystallography. (New York, Kluwer Academic/Plenum Publishers), p.345-404.

Law, L.M., Duncan, R., Esmaili, A., Nakhasi, H.L., and Hobman, T.C. (2001). Rubella virus E2 signal peptide is required for perinuclear localization of capsid protein and virus assembly. J Virol *75*, 1978-1983.

Law, L.M.J., Everitt, J.C., Beatch, M.D., Holmes, C.F.B., and Hobman, T. (2003). Phosphorylation of rubella virus capsid regulated its RNA binding activity and virus replication. J. Virol. *77*, 1764-1771.

Lee, J.Y., Bowden, D.S., and Marshall, J.A. (1996). Membrane junctions associated with rubella virus infected cells. J. Submicrosc. Cytol. Pathol. *28*, 101-108.

Lee, J.Y., Marshall, J.A., and Bowden, D.S. (1994). Characterization of rubella virus replication complexes using antibodies to double-stranded RNA. Virology *1*, 307-312.

Lee, S., Owen, K.E., Choi, H.K., Lee, H., Lu, G., Wengler, G., Brown, D.T., Rossmann, M.G., and Kuhn, R.J. (1996). Identification of a protein binding site on the surface of the alphavirus nucleocapsid and its implication in virus assembly. Structure *4*, 531-541.

Lescar, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., and Rey, F.A. (2001). The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. Cell *105*, 137-148.

Liu, Z., Yang, D., Qiu, Z., Lim, K., Chong, P., and Gillam, S. (1996). Identification of domains in rubella virus genomic RNA and capsid protein necessary for specific interaction. J. Virol. *70*, 2184-2190.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: Semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82-97.

Ma, L., Jones, C.T., Groesch, T.D., Kuhn, R.J., and Post, C.B. (2004). Solution structure of dengue virus capsid protein reveals another fold. Proc Natl Acad Sci U.S.A *101*, 3414-3419.

Mangala Prasad, V., Willows, S., Fokine, A., Battisti, A.J., Plevka, P., Sun, S., Hobman, T., and Rossmann, M.G. (2013). Rubella virus capsid protein structure and its role in virus assembly and infection. Proc Natl Acad Sci U.S.A (*in press*).

Magliano, D., Marshall, J.A., Bowden, D.S., Vardaxis, N., Meanger, J., and Lee, J.Y. (1998). Rubella virus replication complexes are virus-modified lysosomes. Virology *240*, 57-63.

McPherson, A. (2003). Interpreting Patterson maps. In Introduction to Macromolecular Crystallography. (New Jersey, Wiley-Lis Inc.), pp. 179-201.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood enhanced fast translation functions. Acta Crystallogr. D *61*, 458-464.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of Applied Crystallography *40*, 658-674.

Mukhopadhyay, S., Zhang, W., Gabler, S., Chipman, P.R., Strauss, E.G., Strauss, J.H., Baker, T.S., Kuhn, R.J., and Rossmann, M.G. (2006). Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. Structure *14*, 63-73.

Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V., Van Etten, J.L., and Rossmann, M.G. (2002). The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proc Natl Acad Sci U.S.A. *99*, 14758-14763.

Oker-Blom, C., Ulmanen, D., Kaariainen, L., and Pettersso, R.F. (1984a). Rubella virus 40S genomic RNA specifies a 24S subgenomic mRNA that codes for a precursor to structural proteins. J. Virol.*49*, 403-408.

Oker-Blom, C., Jarvis, D.L., and Summers, M.D. (1990). Translocation and cleavage of rubella virus glycoproteins: identification and role of the E2 signal sequence. J. Gen. Virol. *71*, 3047-3053.

Oker-Blom., C. (1984b). The gene order of rubella virus structural proteins is NH_2-C-E2-E1-COOH. J. Virol. *51*, 354-358.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, Pt A *276***,** 307-326.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C.T. (1960). Structure of haemoglobin: A three dimensional Fourier synthesis at 5.5Å resolution. Nature *185*, 416-422.

Petruzziello, R., Orsi, N., Macchia, S., Rieti, S., Frey, T.K., and Mastromarino, P. (1996). Pathway of rubella virus infectious entry into Vero cells. J. Gen. Virol. *77*, 303-308.

Pettersson, R.F., Oker-Blom, C., Kalkkinen, N., Kallio, A., Ulmanen, I., Kaariainen, L., Partanen, P., and Vaheri, A. (1985). Molecular and antigenic characteristics and synthesis of rubella virus structural proteins. Rev. Infec. Dis. *7*, S140-149.

Qiu, Z., Ou, D., Wu, H., Hobman, T.C., and Gillam, S. (1994). Expression and characterization of virus-like particles containing rubella virus structural proteins. J. Virol. *68*, 4086-4091.

Rossmann, M.G., and Johnson, J.E. (1989). Icosahedral RNA virus structure. Annu. Rev. Biochem. *58*, 533-573.

Rossmann, M.G., and Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallography *15*, 24-31.

Rossmann, M.G. (1961). The position of anomalous scatterers in protein crystals. Acta Crystallography *14*, 383-388.

Rossmann, M.G., and Argos, P. (1976). Exploring structural homolgy of proteins. J. Mol. Biol. *105*, 75-95.

Risco, C., Carrascosa, J.L., and Frey, T.K. (2003). Structural maruation of rubella virus in the Golgi complex. Virology *312*, 261-269.

Schiltz, M., and Bricogne, G. (2010). Broken symmetries in macromolecular crystallography: phasing from unmerged data. Acta Crystallography *66(Pt 4)*, 447-457.

Schroder, R.R., Hofmann, W., and Menetret, J.F. (1990). Zer-loss energy filtering as improved imaging mode in cryo-electron microscopy of frozen hydrated specimens. J. Struct. Biol. *105*, 28-34.

Schrodinger, LLC (2012). The PyMOL Molecular Graphics System, Version 1.5.

Strauss, J.H., and Strauss, E.G. (1994). The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. *58*, 491-562.

Suomalainen, M., Garoff, H., and Baron, M. (1990). The E2 signal sequence of rubella virus remains part of the capsid protein and confers membrane association in vitro. J. Virol. *64*, 5500-5509.

Taylor, G. (2003). The phase problem. Acta Crystallography *D59*, 1881-1890.

Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P.V., Zwart, P.H., and Hung, L.W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr. D Biol. Crystallogr. *65*, 582-601.

Tellinghuisen, T., Hamburger, A.E., Fisher, B.R., Ostendorp, R., and Kuhn, R.J. 2000. *In vitro* assembly of alphavirus cores using nucleocapsid protein expressed in *Escherichia coli*. J.Virol. *73*, 5309-19.

Terwilliger, T.C. (2003). SOLVE and RESOLVE: Automated structure solution and density modification. Methods Enzymol. *374*, 22-36.

Tocheva, E.I., Li, Z., and Jensen, G.J. (2010). Electron cryotomography. Cold Spring Harb. Perspect. Biol. *2*, a003442.

Tong, L., and Rossmann, M.G. (1997). Rotation function calculations with GLRF program. Methods Enzymol. *276*, 594-611.

VanDuyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. *229*, 105-124.

Yang, D., Hwang, D., Qiu, Z., and Gillam, S. (1998). Effects of mutations in the rubella virus E1 glycoprotein on E1-E2 interaction and membrane fusion activity. J. Virol. *72*, 8747-8755.

Yu, I.M., Zhang, W., Holdaway, H.A., Li, L., Kostyuchenko, V.A., Chipman, P.R., Kuhn, R., Rossmann, M.G., and Chen, J. (2008). Structure of immature dengue virus at low pH primes proteolytic maturation. Science *319*, 1834-1837.

APPENDIX

APPENDIX
ANALYSIS OF DRUG CANDIDATES AGAINST DENGUE VIRUS

1. Introduction

Flavivirus is a genus under the family *Flaviviridae* which also includes the genera pestivirus and hepacivirus. There are nearly 70 different species of flaviviruses, many of which cause human diseases. Some of the well-known flaviviruses are the dengue virus, West-Nile virus, yellow fever virus, Japanese-encephalitis virus and tick-borne encephalitis virus (Lindenbach *et al*., 2006).

Dengue virus is a mosquito-borne flavivirus. It is an icosahedral, enveloped virus containing an un-segmented, positive-sense, single stranded RNA genome. The mature virus particle has a diameter of about 500 Å (Kuhn *et al*., 2002). The virus is transmitted by the *Aedes aegypti* mosquito. Dengue fever, which is caused by infection of any of its four serotypes, is characterized by high fever, severe headache and body pain. Infection of an individual with more than one serotype of dengue virus causes dengue hemorrhagic fever, which can be fatal if unrecognized (Lindenbach *et al*., 2006). A possible reason for the increased severity during a secondary infection of dengue virus is antibody dependent enhancement (ADE) of the virus infection. ADE can be described as an enhancement in virus infection in the presence of sub-neutralizing concentrations of the virus antibodies. The presence of significant variations between the different dengue virus strains can result in a scenario where the antibody against the first infection is not suitable to neutralize a secondary infection. However, partial cross-reactivity of the virus strain causing the secondary infection with the antibody to the first infection can result in ADE (Halstead and O'Rourke, 1977). There are about 100 million cases of dengue infection reported worldwide every year (Shu and Huang, 2004).

Dengue virus, including other flaviviruses, has a genome size of 10.7 kb and encodes three structural proteins and seven non-structural proteins. The three structural proteins are the capsid protein, the E glycoprotein and the prM protein. The seven non-structural proteins, NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 are involved with viral replication. The capsid protein interacts with the viral RNA and forms the inner core of the virus. The capsid plus RNA core is surrounded by a lipid envelope. The prM and E glycoproteins are arranged on the outer surface of the virion. The prM protein has three domains: pr, M and a carboxy terminal transmembrane region. The E glycoprotein is the major glycoprotein that is involved in receptor attachment and fusion of the virion with the infected cell (Lindenbach *et al*., 2006; Perera and Kuhn, 2008).

Dengue virus exists predominantly in an immature and mature form. The immature form exists in nature only inside the host cell. It has 180 copies of the prM and E heterodimer on its surface forming a spiky exterior to the viral particle (Zhang *et al*., 2003) (**Figure A 1A**). The dengue virus particles assemble and bud into the endoplasmic reticulum as immature particles. These subsequently travel via a secretory pathway to the Golgi apparatus, during which, they are exposed to low pH. This acidic environment causes a conformational change in the E protein which disassociates it from the prM protein and induces it to form E homodimers. These homodimers lie flat against the viral surface giving the maturing virion a smooth appearance. During maturation, the pr peptide is cleaved from the M peptide by the host protease, furin. The pr peptide stays associated with the E protein until the mature virus particle is released into the extracellular environment (Yu *et al*., 2008). The glycoprotein shell of the mature virion consists of 180 copies each of the E protein and M protein (Kuhn, *et al*., 2002) (**Figure A 1B**).

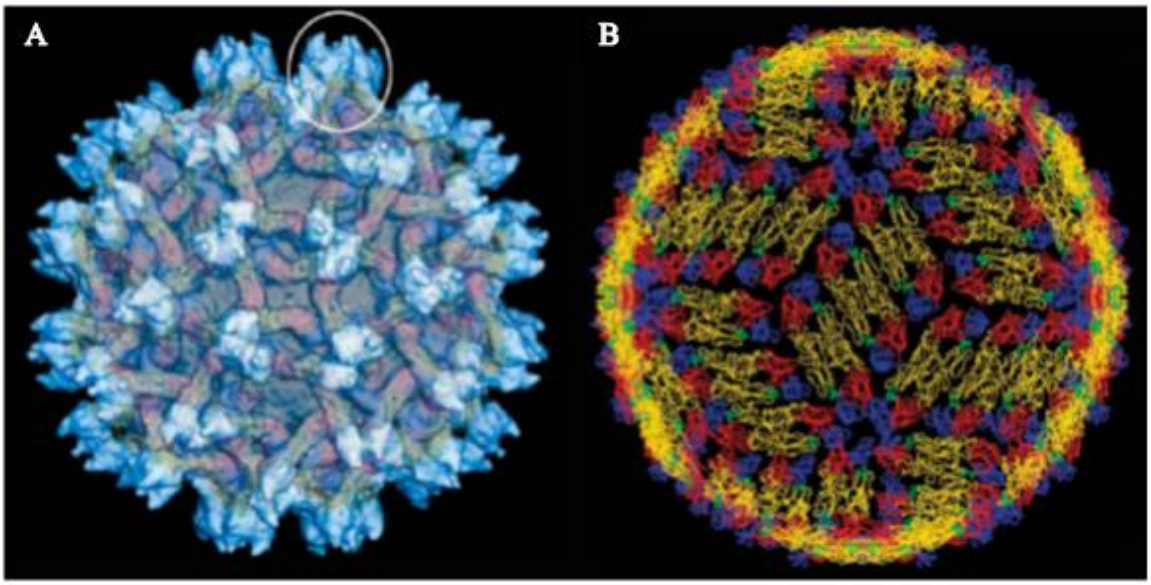


Figure A 1 Dengue virus structures. A. Immature dengue virion. The spike structure on the virion surface is indicated by a circle. B. Mature dengue virion. Figure reproduced from Mackenzie *et al*., 2004.

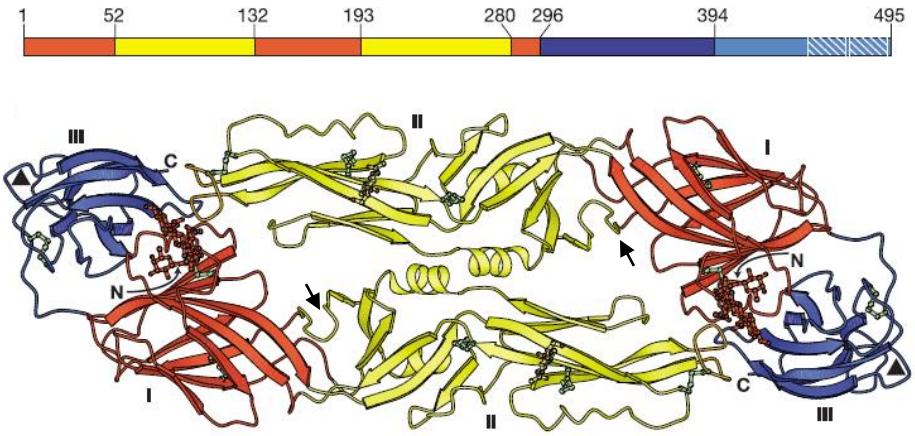
There are currently no vaccines or anti-viral therapies available for dengue that can provide immunity against all the four serotypes. Thus, considerable effort is being focused on arriving at a solution to combat the virus infection and spread. For antiviral drug development against dengue virus, one of the highly targeted regions on the virus structure is the region between domain I and II of the E protein on the virus envelope.

The E protein is a 495 amino acid long glycoprotein. It consists of a transmembrane region, a stem region and an ectodomain region. The ectodomain region consists of three domains. Domain I is the central domain and consists of an eight-stranded beta barrel. Domain II is linked to domain I by four peptide chains. This is the dimerization domain with a fusion loop at its distal end. Domain III is also connected to domain I by a peptide. Domain III is an immunoglobulin-like domain that harbors the receptor binding region (Modis *et al*., 2003; Rey *et al*., 1995). The hinge region between domain I and II has been observed to have different angles during the life cycle of the virus. These movements and corresponding structural changes that occur are crucial for the virus maturation and fusion

to the cell. Thus, the hinge region between domains I and II is a popular site for drug targeting, as effectively hindering the motion between these two domains can have a detrimental effect on virus fusion and virus maturation (Rey, 2003) (**Figure A 2A**).

The project discussed in this chapter was initiated in collaboration with the National Institute of Tropical Diseases (NITD), Novartis Co, Singapore, to study the effects of drug candidates on dengue virus. The NITD had developed drug candidates that had shown efficacy in inhibiting the dengue virus entry into host cells and propagation in tissue culture (Wang *et al*., 2009). The drug candidates were designed to target the hinge region between domain I and II of the E protein. Three of these compounds were selected by the team at NITD, based on plaque reduction capabilities in cell based assays. The three compounds had similar structures and were named NITD-276, NITD-029 and NITD-249. NITD-276 was the first compound to be sent to the Rossmann and Kuhn laboratories to verify its binding position to dengue E protein. It is an organic molecule with a molecular weight of about 500Da (**Figure A 2B**). The other two compounds were sent subsequently for similar studies. All three compounds were stored as a 10mM solution in DMSO (Dimethyl Sulfoxide) at room temperature.

A.



B.

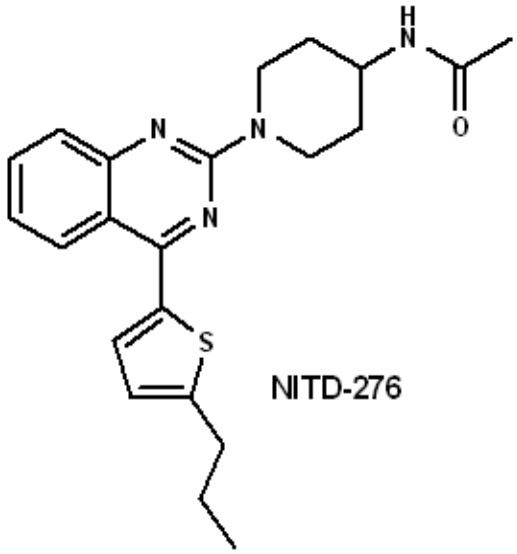


Figure A 2 Anti-viral drug candidate against Dengue E protein. A. Structure of Dengue E protein dimer. The domains I, II and III are shown in red, yellow and blue respectively. Figure modified from Modis *et al*., 2004. The black arrows indicate the target position of the drug candidates. B. Structure of the Novartis compound NITD-276.

2. Material and Methods

2.1 Recombinant expression of dengue E protein

The soluble ectodomain of dengue E protein (sE) was produced recombinantly in the *Drosophilla melanogaster* derived S2 cell line using a protocol established by Long Li, a previous graduate student and post-doctoral associate in the Rossmann laboratory. The S2 cells are semi-adherent cells that can be either grown as adherent cultures or as suspension cultures. Frozen S2 cells transiently expressing dengue sE (to be referred to as S2-sE cells further on) were obtained from Long Li. The frozen S2-sE cells were thawed and starter cultures were established in a T-25 flask containing 5 ml of Schneider's *Drosophila* medium supplemented with 10% fetal bovine serum (FBS) and 300 µg/ml hygromycin-B (Invitrogen). The cells were grown at room temperatures between 22-27 ºC without supplemental carbon dioxide. Once the cells were revived, they were adapted to the Ex-cell 420 medium (Sigma) supplemented with 300 µg/ml hygromycin-B. The cells were sequentially scaled up to 120 ml in four T-150 flasks. Care was taken to maintain the cell density between 2-10×10^6 cells/ml during each passage. After the cells had been grown in the T-150 flasks for 4 days, they were spun down and transferred to two spinner flasks with 250 ml, each, of Ex-cell 420 medium without hygromycin. The spinner flasks were stirred at approximately 130 rpm. After 5-6 days, the medium volume was scaled up to 2 litres with initial cell concentration maintained between 2-4×10^6 cells/ml. The suspension cell culture was split into larger spinner flasks and stirred at about 300 rpm under room temperature. The cultures were induced for protein expression with 500 µM Copper sulfate once the cell concentrations had reached 6-10×10^6 cells/ml. The medium was harvested after 3-5 days post-induction by centrifuging at 9000 rpm for 25 minutes. The supernatant was filtered using a 0.45 µm cut-off membrane.

2.2 Purification of dengue E protein

Dengue sE (soluble E ectodomain) was purified from the cell culture supernatant using an immuno-affinity column. The immuno-affinity column utilized a monoclonal antibody, 4G2, with specificity for binding to the dengue sE protein.

The monoclonal antibody, 4G2, was expressed using hybridoma cells (Henchal *et al*, 1982). The cells were initially grown in a 6 well plate with DMEM plus 20% FBS. Once the cells were 80% confluent, they were seeded into 40% BD cell MAb medium (BD Biosciences) plus 40% DMEM and 20% FBS. The initial cell density was maintained around 3×10^5 cells/ml. The cultures were gradually adapted to grow in BD medium with 20% FBS. The cells were re-suspended in 15 ml BD medium plus 20% FBS to attain a concentration of $1.5\text{-}2\times10^6$ cells and transferred to the cell compartment of a CELLine flask (INTEGRA Biosciences). The medium compartment of the flask was filled with 1 litre BD medium. After seven days, the supernatant was collected for antibody purification. If more production of antibody was required, the cultures from the cell compartment in the CELLine flask was collected, spun down and added again to the cell compartment at a concentration of only 5×10^6 cells/ml. The medium compartment was topped with 1 litre of BD medium. This process could be repeated at least 2-3 more times without significant drop in antibody expression.

The antibody containing supernatant medium was filtered using a 0.45 µm cut-off membrane filter (Amicon). The medium was then loaded on a 1 ml Protein-A column (GE Lifesciences). The column was washed with PBS at pH 7.4. The bound antibody was eluted using 100 mM glycine at pH 2.7. The elutions were neutralized immediately using 1M Tris at pH 9.0 (100 µl of 1M Tris, pH 9.0 for 900 µl of elution volume).

The purified 4G2 antibody was then coupled to a 5 ml NHS-activated sepharose column (GE Lifesciences) using the manufacturer's protocol. Briefly, the column was washed with 1 ml ice-cold hydrochloric acid and immediately injected with 5 ml antibody solution of concentration about 10 mg/ml. The column was sealed and incubated at room temperature for 20 minutes. The column was then washed with 100 mM Tris at pH 8.0, followed by 100 mM sodium acetate plus 500 mM NaCl at pH 4.0. The prepared column was stored in PBS plus 0.05% sodium azide.

Dengue sE protein was purified by applying the sample to the 4G2-bound NHS column. The column was washed with 10 column volumes of PBS at pH 7.4. The sE protein was eluted using 100 mM glycine at pH 2.7 which was immediately neutralized using 1 M Tris at pH 9.0. The protein was further purified using an anion exchange column, Mono-Q

(GE lifesciencs) with a sodium chloride gradient from 0 to 1 M. The protein was then purified using the Superdex 200 gel filtration column (GE) under buffer conditions of 50 mM Tris and 100 mM NaCl at pH 8.0. The protein was concentrated to a final concentration of 3 mg/ml.

2.3 Co-crystallization of dengue E with NITD-276

The dengue sE protein was crystallized in 100 mM HEPES, pH 7.5, 100 mM ammonium sulfate and 30% PEG-3350 in the presence of 1 mM NITD-276 in the crystallization drop.

2.4 Data collection and structure determination

Crystals of the dengue sE protein with NITD-276 were tested at the Advanced Photon Source, GM/CA, beamline 23 ID-D and ID-B, Chicago, USA. The datasets were indexed using HKL2000. The structure was determined using the molecular replacement (MR) method with the program PHASER in the CCP4 software package. The dengue sE structure was used as a search model (Modis *et al*., 2003).

2.5 Cryo-electron microscopy:

Sample grids of dengue virus incubated with NITD-276 at room temperature for 30 minutes had been made by Heather Holdaway, a previous electron microscopist in the biological sciences department. Imaging data from the grids were collected by Ye Xiang, a former post-doctoral associate in the Rossmann laboratory. The micrographs were collected at a magnification of 50,000 X in the 200 kV CM200 microscope using underfocus values between 0.8-2.0 µm. The data was processed, from this point on, by myself.

The micrographs were digitized using a Nikon Scanner. The pixel size of the images were 1.244 Å/pixel. About 300 particles were boxed from the images. The particles were boxed with a box size twice the diameter of the dengue virus particles (about 1000 Å). CTF (Contrast Transfer function) correction was applied to each image. The orientation of the boxed particles was determined using the common lines method without any starting

model. The aligned 2D images were then reconstructed into a 3D volume. This volume was then used as a starting model for the next orientation search. The procedure was repeated till the reconstruction converged and did not change anymore. The 3D reconstruction procedure was carried out using the EMAN software package (Ludtke *et al*., 1999).

2.6 Iso-thermal calorimetry (ITC) studies using NITD-249

ITC is a technique used to measure interactions between ligands, such as proteins and peptides, or proteins and other compounds. It is based on the fine changes of heat absorption or release that accompanies a binding reaction between two molecules. The heat gain or loss is measured as a change in temperature between a reference and sample cell. The ligand is injected into the cell containing the protein in pulses (at specific time intervals) and the data is collected as a series of spikes of heat flow. The ITC measurements with dengue sE and NITD-249 was carried out at the proteomics facility in Birck Nanotechnology Centre at Purdue University by Dr. Lake Paul. In this experiment, soluble dengue E protein was placed in the sample cell and the drug NITD-249 was injected as the ligand.

2.7 Purification of immature dengue virus

The propagation and purification of immature dengue virus (strain 16881) was conducted in a similar way as outlined by Kuhn *et al*., 2002. C636 mosquito cells were grown in Minimum Essential Medium (MEM) supplemented with 10% FBS, 1% non-essential amino acids (NEAA) and 1% L-glutamine at 30 ºC in a 5% CO_2 incubator. 25-50 mM HEPES buffer was added to the medium if the cells appeared unhealthy or when the medium pH was observed to change. Passage number of the cells before infection was maintained well below thirty. The cells were infected with the dengue virus strain when the cells were close to 60-70% confluent. Infection was carried out with overlaying the cells with MEM containing 1% L-glutamine, 1% NEAA and the virus stock for 2 hours at 30 ºC. The medium containing the virus was then removed and the cells were overlaid with MEM plus 5% FBS, 1% L-glutamine and 1% NEAA. After 24 hours post-infection, the

medium was changed to MEM supplemented with 2% FBS, 1% L-glutamine, 1% NEAA and 20 mM ammonium chloride. The infected cells were then grown for 72 to 96 hours at 30 ºC before harvesting the cell medium.

The collected cell medium was centrifuged for 10 min at 6000 rpm in a Beckman Coulter centrifuge (rotor ID: JA-10). PEG precipitation was carried out by the addition of PEG-8000 to a final concentration of 8% and stirring overnight at 4 °C. The precipitates from this step were collected by centrifuging for 50 minutes at 9,500 rpm (rotor ID: JA-10). The pellets were re-suspended in 10 ml of NTE buffer (20mM Tris (pH 8.0), 120 mM NaCl and 1 mM EDTA). The mixture was centrifuged at 3000 rpm for 5 mintues. It was then layered on top of a 10-35% potassium tartarate step gradient. 2 ml of a 22% sucrose solution in NTE buffer was added to the bottom of the gradient tubes using a Pasteur pipette. The tubes were then centrifuged at 32000 rpm for 2 hours in an ultracentrifuge (rotor ID: SW-41). Illumination of the gradient centrifuge tubes after this step showed the presence of purified virus as a single white band. The band region was collected and buffer exchanged into NTE buffer using a 100 kDa cut-off centrifugal filter (Amicon).

2.8 Cryo-electron microscopy studies with NITD-249

The immature virus was incubated with NITD-249 at room temperature for about 1 hour. The incubation conditions were based on information given by the Novartis Co. team. The pH of the virus-compound mixture was then lowered to pH 6.0 using citrate-phosphate buffer. This mixture was applied to C-flat (1.2/1.3-4C-50) grids (EMS) and frozen as vitrified ice samples in liquid ethane. The grids were then analyzed at 200 kV in the Phillips CM200 microscope. CCD images were collected at a magnification of 11000X and defocus of about 3-5 µm.

2.9 Plaque assay studies with NITD-029 and NITD-249

Plaque assays were set up using BHK-15 cells. The BHK-15 cells were grown till confluence on 6-well titre plates to form a single continuous layer of cells. These wells were then infected with 10-fold serial dilutions of mature dengue virus (DENV-2, strain 16881) for 1 hour at 30 ºC. The control plates did not have any drug compound. The test

plates were infected with virus pre-incubated for 1 hour with 50 μM (final concentration) of either NITD-029 or NITD-249. The final DMSO concentration in the wells was 5%. The wells were then overlaid with 2-3 ml of MEM plus 5% FBS and 0.8% agar. After the agar solidified, the wells were overlaid with 1 ml of MEM plus 5% FBS. The well-plates were incubated at 37 ºC in a 5% CO_2 incubator for 6 days. After 6 days, the agar was removed gently from the well-plates. 0.1% crystal violet dye was used to stain the wells. Crystal violet stained and fixed the living cells for visualization.

3. Results and Discussion:

The results obtained from the individual drug candidates have been discussed separately in the following sections.

3.1 Drug candidate: NITD-276

3.1.1 Co-crystallization studies with dengue-E protein

Co-crystallization of the soluble ectodomain of the dengue-E protein with 1 mM of the NITD-276 compound gave regular shaped crystals (**Figure A 3**). The molar ratio of the dengue sE protein to the drug compound was 1:10.

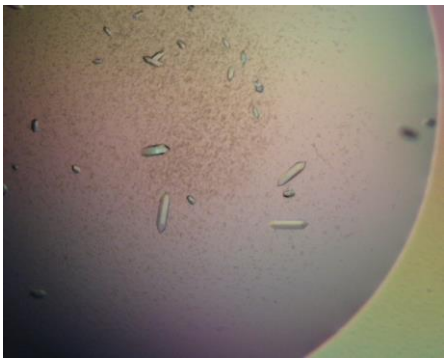


Figure A 3 Crystals of dengue E protein co-crystallized with NITD-276.

X-ray datasets with resolution of about 3.4 Å were collected from these crystals. The space group was $P4_32_12$ (a = 105.680 Å, b = 105.680 Å, c = 167.717 Å) with 2 molecules in the asymmetric unit. The structure was determined using the molecular replacement (MR) technique with the known dengue E protein structure as a search model. However, the electron density map produced after structure determination did not show any extra density which might correspond to the density of the drug compound. Thus, it was concluded that the drug does not bind to the dengue E protein in solution. The crystallographic data statistics are given in **Table A 1**.

Table A 1 X-ray data and refinement statistics for the dengue E crystals with NITD-276

Data collection	
X-ray wavelength (Å)	0.98
Space group	$P4_32_12$
Cell dimensions	
a,b,c (in Å)	105.724, 105.724, 169.208
α,β,γ (in degrees)	90, 90, 90
Resolution (Å)[2]	3.40 (3.46-3.40)
Rmerge[2]	0.169 (0.73)
$I/\Delta I$[2]	26.4 (6.04)
Completeness (%)[2]	96.98 (93.93)
Overall Redundancy	14.3
Refinement	
No. of reflections	13633
Rwork/Rfree	0.34/0.42
R.m.s deviations	
Bond length (Å)	0.0053
Bond angles (deg.)	1.598
Ramachandran plot (%)	
Favored	85.52
Outliers	3.62

3.1.2 Cryo-electron microscopy studies

Simultaneous experiments using cryo-electron microscopy were carried out with NITD-276 in order to locate the binding position of the drug on the mature dengue virus. However, the quality of the micrographs was not very good. The final reconstruction obtained had a resolution worse than 25 Å and lacked detail. This could have been due to the low number of particles, ice contamination that was seen on the micrographs or also due to the large number of broken virus particles seen in the images.

On obtaining the results from the co-crystallization studies, followed by the discovery of solubility issues with NITD-276 in aqueous solutions with less than 10% DMSO concentration, further experiments were not carried out with this compound.

3.2 Drug candidate: NITD-029

The NITD-029 compound was soluble in 5% DMSO solution. Plaque assay studies were conducted using BHK-15 cells infected with dengue-2 virus that had been treated with NITD-029. The results from these assays are shown in Section 3.33.

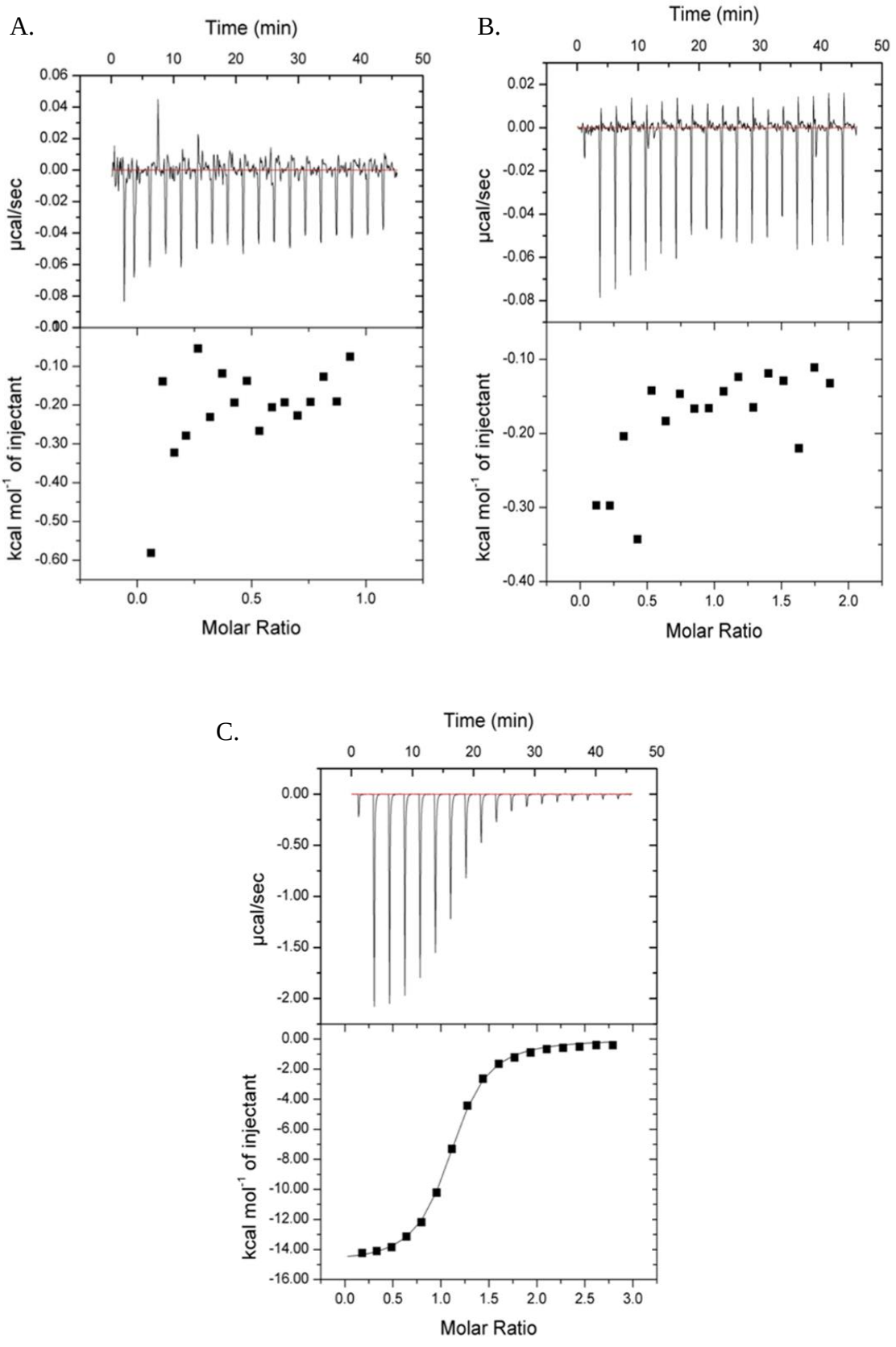
3.3 Drug candidate: NITD-249

3.3.1 Iso-thermal calorimetry studies

The NITD-249 compound was also soluble in water in the presence of 5% DMSO. In order to ascertain binding of the compound to dengue E before conducting structural studies, Isothermal Titration Calorimetry (ITC) was carried out with NITD-249 in 10% DMSO solution. The results from the ITC experiments are given in **Figure A 4**.

From the ITC experiments it was concluded that NITD-249 did not bind to E protein in solution.

Figure A 4 ITC binding experiments. A. Control experiment with 1mM NITD-249 injected into sample cell containing 10% DMSO solution. B. Test experiment with 1mM NITD-249 injected into sample cell containing 100μM dengue-E protein. Both ligand and protein solution contained 10% DMSO. C. An example of an ideal ITC curve for 1:1 ligand to protein binding.


A.
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar Ratio
B.
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar Ratio
C.
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar Ratio

3.3.2 Cryo-electron microscopy studies with NITD-249

From the results of the ITC experiments, it was hypothesized that the NITD-249 compound could bind to epitopes on the E protein that may be present only on the dengue virion and not on the recombinant soluble ectodomain of E protein. As the compound was designed to bind the hinge region between domain I and II in the E protein, it would be expected to block conformational changes in the E protein. Major conformational changes occur in the E protein on the fusion of the virus with the host cell endosomal membrane. In this step, the homodimeric E protein rearranges to form trimeric spike structures that resemble the glycoprotein spike structures of the immature virus. Low pH induces the above mentioned conformational changes. Low pH conformational changes can also be induced with immature virus particles, in which case, the spikes on the immature virus adopt a more flattened conformation as in the mature virus. It was thus hypothesized, that the drug should also be able to bind to immature dengue virus particles. Lowering of pH would then negatively affect the motions of the E protein and inhibit it from forming the smooth, mature-particle like conformation. Immature dengue was chosen as a starting point because it was easier to produce and more stable. Lowering of pH induces aggregation in case of mature virus as this exposes the fusion loop on the E protein. Aggregation at low pH can be avoided with immature virus particles as the pr peptide is present to mask the fusion loop.

Analysis of the grids showed that on lowering pH, the immature virus-drug mixture showed conformational changes of the E glycoprotein spikes towards the flattened conformation of mature virus in about 35% of the virus particles (**Figure A 5**). This was similar to the percentage seen on lowering pH with immature virus alone. Thus, it was concluded that either the drug compound did not block the expected conformational change or it did not bind to immature virus.

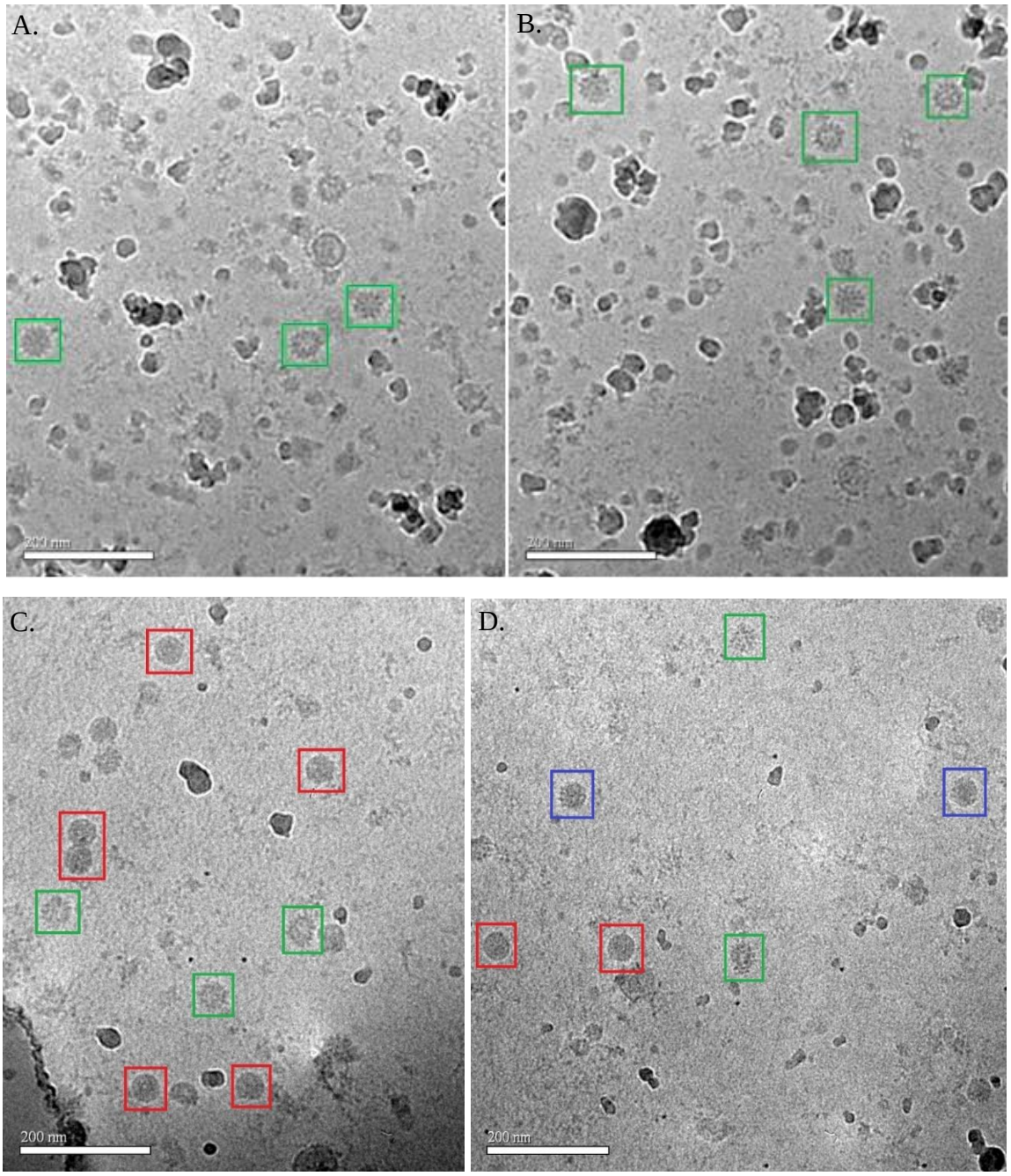


Figure A 5 Dengue-2 treated with NITD-249. A and B. Dengue virus plus NITD-249 at neutral pH. C and D. Dengue virus plus NITD-249 at pH 6.0. The green boxes indicate immature particles, red boxes indicate the mature particles and the blue boxes indicate partially mature or partially immature particles. The scale bar in white is 200nm.

3.3.3 Plaque Assay studies with NITD-029 and NITD-249

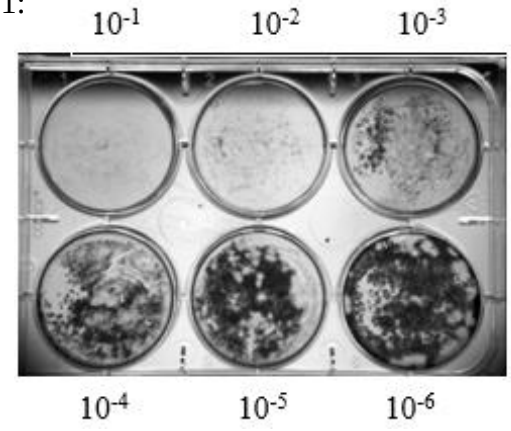
The NITD-029 and NITD-249 compounds were incubated with dengue virus separately and then used to infect BHK-15 cells. In the plaque assay results shown below (**Figure A 6**), the greater the amount of darker color corresponds to higher amount of living cells, which in turn means lower virulence of the virus particle. Thus, if the compound inhibited dengue virus infection, there would be lower amount of plaques or higher amount of living cells when compared to the control experiments.

The plaque assay results clearly showed that the compounds NITD-029 and NITD-249 promoted better cell growth (or inhibit virus) only at very low virus concentrations. Hence, the inhibition effects of the compounds were not very good. At this point, information was also received from Novartis Co. (NITD) that they had stopped working on these drug candidates, and further work on this project was halted both at the Rossmann laboratory and at the NITD, Singapore.

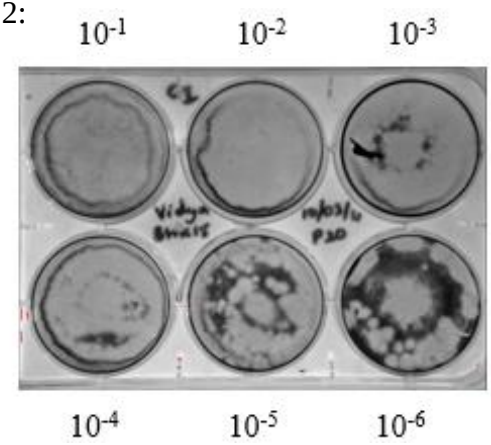
Figure A 6 Plaque assays using BHK-15 cells infected with Dengue virus type-2 (DENV-2) in presence of NITD-249 and NITD-029.

Control 1: BHK-15 + DENV-2

Set 1:

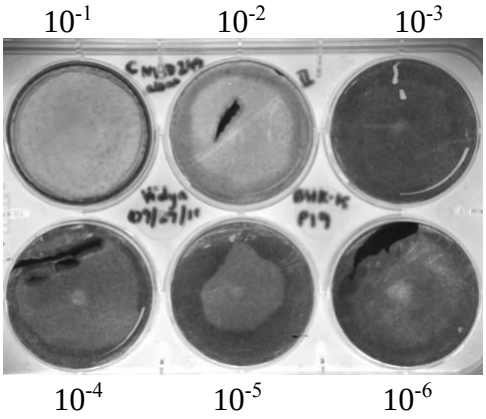


Set 2:



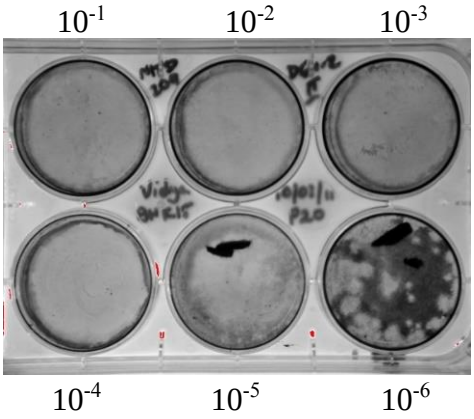
Control 2: NITD-249 drug alone

Set 2:



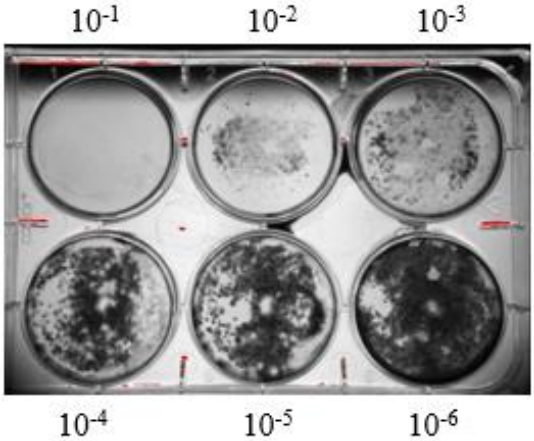
Sample 1: BHK-15 + NITD-029 + DENV-2

Set 2:

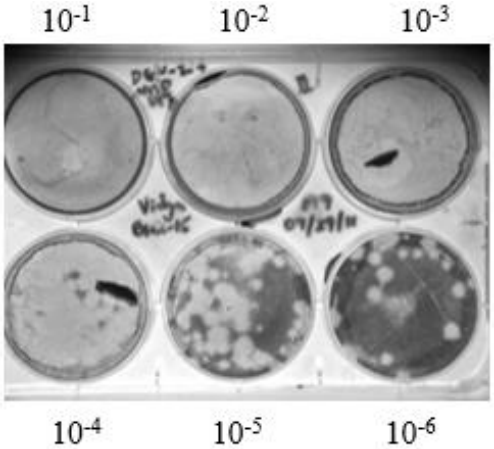


Sample 2: BHK-15 + NITD-249 + DENV-2

Set 1:



Set 2:



4. References

Halstead, S.B., and O'Rourke, E.J. (1977). Dengue viruses and mononuclear phagocytes. I. Infection enhancement by non-neutralizing antibody. J. Exp. Med *146*, 201-217.

Henchal, E.A., Gentry, M.K., McCown, J.M., and Brandt, W.E. (1982). Dengue virus-specific and flavivirus group determinants identified with monoclonal antibodies by indirect immunofluorescence. Am. J. Trop. Med. Hyg. *31*, 830-836.

Kuhn, R.J., Zhang, W., Rossmann, M.G., Pletnev, S.V., Corver, J., Lenches, E., Jones, C.T., Mukhopadhyay, S., Chipma, P.R., Strauss, E.G., Baker, T.S., and Strauss, J.H. (2002). Structure of dengue virus: implications for flavivirus organization, maturation and fusion. Cell *108*, 717-725.

Lindenbach, B.D., Thiel, H., Rice, C.M. (2006). *Flaviviridae*: The viruses and their replication. In Fields Virology. 5th edition. Knipe, D.M., Griffin, D.E., Lamb, R.A., Martin, M.A. eds. (Philadelphia: Lippincott, Williams & Wilkins), pp 1069–1100.

Mackenzie, J.S., Gubler, D.J., and Peterson, L.R. (2004). Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue virus. Nature Medicine *10*, S98-109.

Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. Proc Natl Acad Sci U S A *100*, 6986-6991.

Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue envelope protein after membrane fusion. Nature *427*, 313-319.

Perera, R., and Kuhn, RJ (2008). Structural proteomics of dengue virus. Curr. Opin. Microbiol. *11*, 369-377.

Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2Å resolution. Nature *375*, 291-298.

Rey, F. (2003). Dengue virus envelope glycoprotein structure: New insight into its interactions during viral entry. Proc. Natl. Acad. Sci. *100*, 6899-6901.

Shu, P., and Huang, J. (2004). Current Advances in Dengue Diagnosis. Clin. Diagn. Lab Immunol. *11*,642-650.

Wang, Q., Patel, S.J., Vangrevelinghe, E., Xu, H.Y., Rao, R., Jaber, D., Schul, W., Gu, F., Heudi, O., Ma, N.L., Poh, M.K., Phong, W.Y., Keller, T.H., Jacoby, E., and Vasudevan, S.G. (2009). A small-molecule dengue virus entry inhibitor. Antimicrobial Agents and Chemotherapy *53*, 1823-1831.

Zhang, Y., Corver, J., Chipman, P.R., Zhang, W., Pletnev, S.V., Sedlak, D., Baker, T.S., Strauss, J.H., Kuhn, R.J., and Rossmann, M.G. (2003). Structures of immature flavivirus particles. EMBO *22*, 2604-2613.

VITA

VITA

I was born in Chennai, state of Tamil Nadu, India in 1987. I studied till middle school in Bhaktavatsalam Vidyashram, and did my high school education in S.B.O.A School and Junior College; both in Chennai. I received my degree in Bachelor of Technology (B.Tech) with major in Industrial Biotechnology from Shanmugha Arts, Science, Technology and Research Academy (SASTRA) at Thanjavur, Tamil Nadu in May, 2009. After my bachelors, I applied and got admitted to the graduate program in Biological Sciences department at Purdue University. At Purdue, I got the opportunity to work under Dr. Michael Rossmann for my doctoral studies. My graduate thesis work mainly involves the study of rubella virus and its capsid protein using x-ray crystallography and cryo-electron tomography. I hope to receive my Ph.D. degree by December, 2013.