MODULATION OF WEST NILE VIRUS CAPSID PROTEIN AND VIRAL RNA INTERACTION THROUGH PHOSPHORYLATION

CHEONG YUEN KUEN
(B.Sc. (Hons) University of Toronto, Canada)

A THESIS SUBIMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MICROBIOLOGY

NATIONAL UNIVERSITY OF SINGAPORE

2010

PUBLICATIONS AND PRESENTATIONS GENERATED DURING THE COURSE OF STUDY

Publications:

Bhuvanakantham, R., Cheong, Y. K. and Ng, M. L. (2010). West Nile virus capsid protein interaction with importin and HDM2 protein is regulated by protein kinase C-mediated phosphorylation. *Microbes Infect* **12**, 615-625

Tan, L.C.M., Chua, A.J.S., Goh, L.S.L., Pua, S.M., Cheong Y.K. and Ng M.L. (2010). A membrane chromatography method for the rapid purification of recombinant flavivirus proteins. *Protein Expression and Purification.* [Epub ahead of print]

Manuscript Under Revision:

Cheong, Y.K. and Ng M.L. Dephosphorylation of West Nile virus capsid protein enhances the processes of nucleocapsid assembly. *Microbes Infect.*

Conference Presentations (Oral):

Cheong, Y.K. and Ng M.L. (2009). Phosphorylation is a key modulator of flaviviral capsid protein functions. *Duke-NUS Emerging Infectious Diseases Symposium*, Singapore.

Conference Presentations (Poster):

Cheong, Y.K. and Ng M.L. (2010). Phosphorylation of West Nile virus capsid protein is essential for efficient viral replication. *10th Nagasaki-Singapore Medical Symposium on Infectious Disease,* Singapore.

Goh, L.S.L., Tan, L.C.M., Chua A.J.S., Pua, S.M., Cheong, Y.K. and Ng M.L. (2009) A membrane chromatography method for the rapid purification of recombinant flavivirus proteins. *Duke-NUS Emerging Infectious Diseases Symposium,* Singapore.

Cheong, Y.K. and Ng M.L. (2009). The role of West Nile virus capsid phosphorylation and its consequence in viral replication. *8th Asia Pacific Congress of Medical Virology*, Hong Kong.

Cheong, Y.K. and Ng M.L. (2008). Phosphorylation of West Nile virus capsid protein attenuates its interaction with viral RNA. *14th International Congress of Virology*, Istanbul.

Cheong, Y.K. and Ng M.L. (2008). West Nile Virus capsid protein and RNA interaction. *13th International Congress of Infectious Disease*, Kuala Lumpur.

Cheong, Y.K. and Ng M.L. (2008). Flaviviral capsid protein and RNA interaction. *1st Annual Singapore Dengue Consortium,* Singapore.

Cheong, Y.K. and Ng M.L. (2007). Characterization of West Nile virus capsid protein. *6th National Health Group Annual Congress,* Singapore.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Professor Ng Mah Lee for the immense amount of support and guidance she has provided throughout this study. Professor Ng's insights into this project and patience towards me have been a true blessing. In addition, I would like to thank her for her generosity during festivities and birthday celebrations. It has made this journey much more vibrant and enjoyable.

I would also like to thank Han Yap for preparing the samples and viewing the sections for electron microscopy. Kim Long for helping me to establish a method for measuring the relative amount of positive-sense and negative-sense viral RNA from purified virions. Bhuvana for her assistance and technical advise on kinase activators and inhibitors.

I thank all the members of the Flavivirus Laboratory: Boon, Mun Keat, Xiao Ling, Li Shan, Shu Min, Samuel, Vincent, Melvin, Edwin, Terence, Anthony and Audrey for their friendship, company on lonely weekends, technical advice and constructive criticism.

Confocal microscopy would have been challenging if not for the assistance of Clement Khaw at the Nikon-Singapore Bio-imaging Consortium.

I am grateful to my parents who never ceased loving and supporting me.

Last but not least, I thank U-Shaun for always being there for me and for encouraging me in Christ always whenever I felt I was going nowhere.

Thank you.

TABLE OF CONTENTS

APPENDICES

Summary

West Nile virus (WNV) is a mosquito-borne flavivirus, which can cause fatal meningoencephalitis. Although studies have been done to examine virus life cycle, the mechanisms that regulate its assembly are unknown. This study aims to characterize the interaction between WNV viral RNA and the capsid (C) protein and also elucidate how the processes of nucleocapsid assembly are regulated.

Initial *in vitro* C protein and viral RNA interaction studies showed that this interaction was not specific. This suggested the presence of a regulatory mechanism to regulate C protein and RNA interaction. There are 5 putative phosphorylation sites on C protein hence; functions of C protein could be regulated by phosphorylation. Western blot analysis with anti-phosphoserine antibodies confirmed that C protein was phosphorylated. Experiments using kinase inhibitors and activators identified protein kinase C to be one of the kinases responsible for the phosphorylation of C protein.

Mutations were introduced into C protein to abolish phosphorylation. The effects of hypophosphorylation with regards to the nucleocapsid assembly processes like RNA binding; nuclear localisation and oligomerization were investigated. Phosphorylation of C protein attenuated RNA binding and this study showed that C protein was dephosphorylated later during infection to facilitate C protein and viral RNA interaction. It is also known that WNV C protein localises predominantly in nuclei of infected cells. Hypophosphorylation of C protein was shown to disrupt this process. Tracking of the cellular localisation of C protein during an infection showed that C protein localised in the nucleus during the early phase of infection but was found in the cytoplasm during late phase of infection. This correlated with the gradual dephosphorylation of C protein in

infected cells. Capsid protein also oligomerized to form the nucleocapsid but hypophosphorylation did not affect the formation of oligomers. However, the rate of oligomerization was retarded by phosphorylation. These data showed that hypophosphorylation favoured the processes of nucleocapsid assembly.

The biological significance of these hypophosphorylated mutants was further investigated by introducing the same mutations into a WNV infectious clone. Characterization of the mutant viruses showed that mutant viruses produced a lower viral yield and also experience a lag in viral production compared to wild type virus. Complementing mutant virus infected cells with wild type C protein partially restored viral yield, however, the lag in virus production was still apparent. The lag in mutant virus replication was only abolished through the transfection of infectious viral RNA into cells. This suggested that phosphorylation was critical for the early events of viral replication linked to nuclear localisation. In addition, it was found that, while wild type virus packaged 10 times more positive-stranded viral RNA than negative-stranded viral RNA, mutant virus packaged only twice as much positive-stranded viral RNA to negative-stranded viral RNA.

This study shows that WNV C functions as a phospho-protein and proposed the dynamics of phosphorylation and dephosphorylation of C protein prevents the premature assembly of the nucleocapsid. This allows assembly to occur during the late phase of an infection where large pools of positive-stranded virus RNA are present.

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

<	Less than
µg	Microgram
µl	Microlitre
ºC	Degrees Celsius
%	Percent
x g	Times gravitational force
AP	Alkaline-phosphatase
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
CPE	Cytopathic effect
kDa	Kilodalton
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic acid
g	Grams
His-tag	Histidine tag
hr	Hours
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kbp	Kilo base pairs
kDa	Kilodalton
L	Litre
LB	Luria Bertani
lbs	Pounds
M	Molar
min	minutes
MOI	Multiplicity of infection
ml	Millilitre
mM	Milimolar
mg	Milligram

ng	Nanogram
O.D.	Optical Density
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PFU	Plaque forming unit
P.I	Post infection
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT-PCR	Reverse transcription polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TAE	Tris acetate EDTA
TBST	Tris buffered saline Tween-20
TEMED	N,N,N',N'-Tetramethylethylenediamine
UV	Ultraviolet
WCL	Whole cell lysate
WB	Western blot
WNV	West Nile virus
w/v	Weight/Volume
x	Times

1.0 LITERATURE REVIEW

1.1 Introduction to West Nile virus

West Nile virus (WNV) is a mosquito-borne virus that was first isolated and identified in 1937 from the blood of a febrile adult woman in the West Nile District of Uganda (Smithburn *et al.,* 1940). In an outbreak in 1957 in Israel, it became recognised as a cause of severe human meningoencephalitis in elderly patients. There was an increased interest in the virus when an outbreak, which began in New York City in 1999, spread to the rest of the United States of America, Canada, Mexico the Caribbean and Central America. Since then, it has become an important emerging disease globally.

The virus is classified as a flavivirus within the Japanese encephalitis sero-complex by a cross-neutralisation test (Calisher *et al.*, 1989; Wengler & Rey, 1999). The flavivirus is a genus of the family *flaviviridae*. The flavivirus complex includes the Dengue virus, Tick-borne encephalitis virus, Japanese encephalitis virus and Yellow fever virus and other viruses. Genetic analysis techniques such as *in situ* hybridization and real time polymerase chain reaction (PCR) are needed to unequivocally identify WNV as the causative agent in infections due to antigenic cross-reactivity (Briese *et al.*, 2002; Lanciotti *et al.*, 2002).

1.2 West Nile virus epidemiology

WNV isolates are genetically grouped into Lineages 1 and 2 on the basis of signature amino acid substitutions or deletions in their envelope protein (Berthet *et al.*, 1997). Lineage 1 viruses are associated with human diseases while lineage 2 viruses are restricted to endemic enzoonotic infections (Jia *et al.*, 1999; Lanciotti *et al.*, 2002)

Until recently, WNV is an infectious disease endemic in Africa, Europe, the Middle East, Central Asia and Oceania (Brinton, 2002). There were brief epizoonotic outbreaks (Romania, Russia, Algeria, Madagascar France, Senegal and South Africa) and infrequent disease outbreaks in human. However, WNV was recently introduced to the American continent.

It is unclear how WNV was first established in the American continent but phylogenic analysis of the envelope protein of the New York WNV outbreak isolated in 1999 suggested that it was closely related to a goose isolate from Israel (Jia *et al.*, 1999; Lanciotti *et al.*, 2002). Hence, it seemed likely that ornithophilic mosquitoes that fed on infected migratory birds introduced WNV to the American continent (Rappole *et al.*, 2000).

Wild bird species are the reservoir hosts in endemic regions and high levels of viremia have been detected in a number of wild birds. Viremic levels of WNV are sustained at a minimum level of 10^5 plaque forming units/ml (PFU/ml) of serum for several days to weeks (Bernard & Kramer, 2001). There are a number of mosquito species that have been infected with WNV and they include *Culex*, *Aedes*, *Minomyia*, *Mansonia* and *Anopheles* but the *Culex* species remain the most susceptible to infection (Ilkal *et al.*, 1997). Natural vertical transmission of the virus in *Culex* mosquitoes have been reported in Africa and this is believed to enhance virus maintenance in nature (Miller *et al.*, 2000). Humans, mammals and amphibians are incidental host and do not play a role in the transmission cycle because of the level of viremia is too low to infect mosquitoes (Anderson *et al.*, 1999; Hubalek, 2000; Rappole *et al.*, 2000). Although

alternate modes of human to human transmission is possible through blood transfusion, organ transplantation and ingestion of breast milk (Hayes & O'Leary, 2004).

The WNV strain that caused the outbreak in New York city is characterized by a high avian death rate and an increase in human disease severity (Solomon *et al.*, 2003). It has been hypothesized that it had acquired some changes to its neurovirulent properties (Ceccaldi *et al.*, 2004). In 2002, there were 4156 human cases of WNV infection in the United States of America (O'Leary *et al.*, 2004) and this continued to increase to 9862 cases resulting in 264 deaths in 2003 (CDC, 2004). But the numbers of cases have begun to decline. As of 2009 there were 720 cases with 32 fatalities (CDC, 2009).

Incubation period of a WNV infection in humans is typically 2 – 6 days usually accompanied with a fever. The course of the fever is sometimes biphasic and about 50% of the infected individuals report a maculopapula or pale roseolar rash (Petersen & Roehrig, 2001). Encephalitis, meningoencephalitis, pan-meningoencephalitis, (Omalu *et al.*, 2003) myocarditis, pancreatitis or hepatitis have all been reported in severely infected individuals. Histopathatological studies in birds have revealed that WNV could be detected in all major organs and viral antigens were also found in 88% of the brains examined (Steele *et al.*, 2000). Because WNV is neuroinvasive, it has caused fatalities in immunocompromised individuals (George *et al.*, 1984) and also in older individuals aged 60 and above (Chowers *et al.*, 2001). The neurological manifestations of WNV are very similar to other flaviviruses such as the Japanese encephalitis virus and Murray Valley virus. The damage caused to the meninges, brain parenchyma and spinal cord had caused poliovirus-like flaccid paralysis in some infected patients (Sejvar *et al.*, 2003).

1.3 Virus morphology

Mature virion particles are ~50 nm in diameter spherical, enveloped (Fig. 1-1A) and have a buoyant density of ~1.2g/cm^3. Their cores are about 25 – 30 nm in diameter and the projections on the envelope are about 5 – 10 nm long. The core is composed of multiple capsid (C) proteins encapsidating the viral RNA (Fig. 1-1B). A lipid bilayer, a lipid membrane (M) protein and a large envelope glycoprotein (E) surround the core (Fig. 1-1C). In immature virions, the pre-membrane (prM) protein is cleaved away from the E protein to produce mature virions (Fig. 1-1D). The two viral surface proteins, M and E, are Type I integral membrane proteins with C-terminal membrane anchors (Mukhopadhyay *et al.*, 2003) which form homodimers on the surface (Fig. 1-1D).

The determination of the structure of the entire Dengue type 2 virion by cryo-electron microscopy (Kuhn *et al.*, 2002) and structure determination of the E glycoprotein of Tick-borne encephalitis virus by X-ray crystallography (Rey *et al.*, 1995) have increased our understanding of the structure and function of the flavivirus virion. The E glycoprotein contains a fusion peptide responsible for fusing the viral lipid bilayer with that of the host. It is the principal antigenic site, which stimulates the production of neutralizing antibodies.

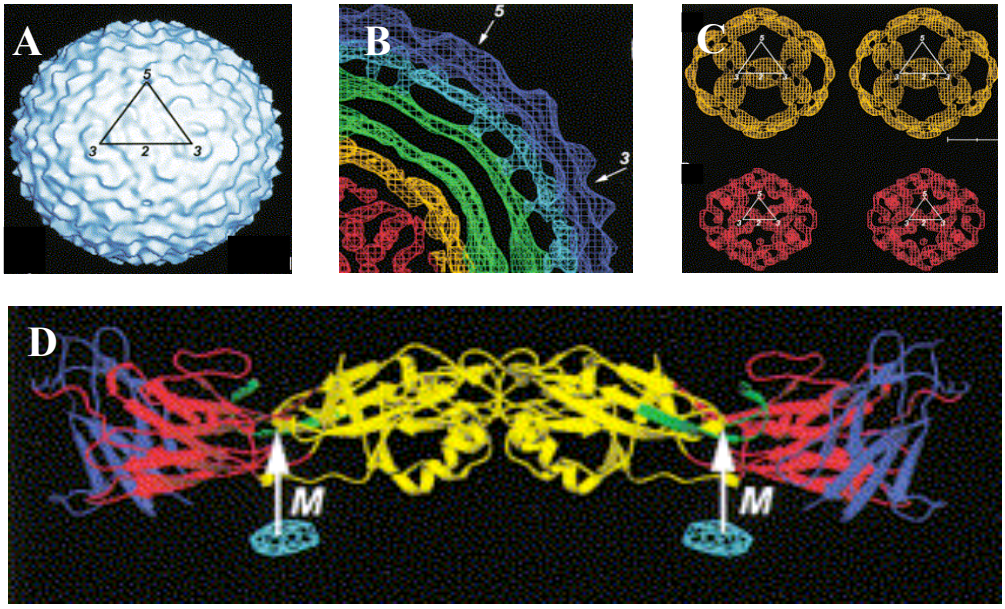

Figure 1-1. Structure of flavivirus. (A) A surface reconstruction of the Dengue type 2 virus at 24 Å resolution by cyro-electron microscopy. The 5-fold and 3-fold icosohedral symmetry axes are labelled. (B) An electron density cross-section reconstruction of the virion. The dark blue layer is the envelope protein, light blue region is the membrane protein, green is the lipid bilayer, yellow is the capsid protein and red is the RNA. (C) Stereo reconstruction of the capsid and RNA, yellow is the capsid while red is the RNA. (D) Homodimer of the envelope protein. The light blue represents the M protein which is cleaved in a matured virus and the arrows indicate the holes between the dimers where the M protein sits (Kuhn *et al.*, 2002).

1.4 West Nile virus RNA genome organisation and viral proteins

Flaviviruses are single positive-stranded RNA virus. The WNV genome is approximately 11, 029 bases in length and it contains a single open reading frame (ORF) of about 10,301 bases. The 5' untranslated region (UTR) and the 3' UTR are 96 and 631 bases in length, respectively. The conserved repeated sequences on the 3' UTR in flaviviruses are believed to form a pseudo-knot structure necessary for RNA cyclization (Hahn *et al.*, 1987; Shi *et al.*, 1996). Since the genomic RNA is positive-stranded, it is infectious (Ada & Anderson, 1959). However unlike mammalian mRNA, the genomic RNA of mosquito-borne flaviviruses lack the 3' terminal polyadenine tract (Brinton *et al.*, 1986; Westaway *et al.*, 1997) instead terminates with the conserved dinucleotide CU_{OH}.

The ORF encodes for 3 structural proteins, the Capsid (C), premembrane/membrane (prM/M) and envelope (E) and 7 non-structural (NS) protein, NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5. The translation of viral proteins is initiated at the 5' end of the genome. Each viral protein is believed to be cleaved from the precursor polyprotein during or after translation (Castle *et al.*, 1985; Wengler *et al.*, 1985).

1.4.1 Structural proteins

The structural proteins are responsible for the assembly of the virus. The first structural protein to be translated is the C protein. Details of the C protein will be dealt with in Section 1.6. The second protein is the prM/M protein. It is a relatively small protein of about 18-19 kDa. This protein is a glycoprotein precursor of the

unglycosylated structural M protein (Mr of 8 kDa). The prM protein is present on the intracellular immature virion and the hydrophilic amino-terminal portion of the protein is cleaved in the trans-Golgi network by cellular furin or related protease to produce mature virions with M protein (Stadler *et al.*, 1997). It has been proposed that the function of prM protein is to chaperone the E protein so that it does not undergo irreversible conformational changes in the acidic compartments of the secretory pathway (Heinz & Allison, 2000). The prM and E interaction are hypothesized to maintain the E protein in a stable, fusion-active conformation during assembly and exocytosis mature virions.

The E protein is the largest structural protein of WNV (50-60 kDa). The E protein is present on the surface of the virus as club-like projections (Westaway, 1987). It is a type I integral membrane protein where the amino-terminal is exposed on the surface and the hydrophobic carboxyl-terminal is buried within the membrane bilayer (Chambers *et al.*, 1990). The cysteine residues in the E protein are highly conserved amongst the flaviviruses since they are crucial for the formation of intracellular disulphide bonds (Nowak & Wengler, 1987).

Crystallography data revealed that the E protein of the Tick-borne encephalitis virus formed homodimers arranged in a head-to-tail orientation (Rey *et al.*, 1995). The E protein can be divided into three domains - a central domain, designated as domain I, an elongated dimerization region designated as domain II and an immunoglobulin-like module designated as domain III (Rey *et al.*, 1995). Based on the structure of Dengue virus, the E protein dimers are closely packed on the surface of the virus making fusion impossible at neutral pH. However, it has been proposed that conformational changes of the E proteins occurred at reduced pH environment (Modis *et al.*, 2004). The reduced pH

in the endocytic vesicles causes the E protein dimers on the virion surface to dissociate so that the fusion peptide in domain II is exposed. This allows domain II to insert its fusion peptide into the target cell membrane. Subsequently, the interactions of the E protein with the endocytic vesicle membrane form hemifusion pores through which the viral nucleocapsid is released.

1.5 Virus cellular life cycle

The WNV can replicate in a wide variety of cell cultures from primary chicken, duck and mouse embryo cell cultures and secondary cell lines derived from human, monkeys, pigs rodents, amphibians and insects, however, it does not cause any apparent cytopathology in many of these cell lines (Brinton, 1986). The receptor utilized by the WNV is likely to be conserved since it is transmitted between an insect and a vertebrate host (Brinton, 2002). This notion is reinforced by evidence that the receptor utilized by the WNV in Vero and murine neuroblastoma 2A cells is a 105-kDa protease sensitive, N-linked glycoprotein (Chu & Ng, 2003). This was later identified as the alpha v beta 3 integrin receptor (Chu & Ng, 2004)

Viral entry into host cell is facilitated by clathrin-mediated endocytosis pathway (Marsh & McMahon, 1999). This event is followed by low-pH fusion of the viral membrane with the lysosomal vesicle membrane (Heinz & Allison, 2000). The low pH causes a conformational change in the E proteins, which allow for the interaction between the E proteins and the lysosomal membrane to form hemifusion pores for the release of viral nucleocapsid into the cytoplasm (Modis *et al.*, 2004).

The nucleocapsid is believed to be uncoated and the viral RNA released into the cytoplasm. Viral RNA is transported to the endoplasmic reticulum (ER), where WNV protein translation and RNA replication occur (Mackenzie *et al.*, 1996)

Host protease and the NS2B-NS3 viral serine proteases cleave the translated polyprotein at multiple sites to generate mature viral proteins (Chambers *et al.*, 1990). The function of most of the non-structural proteins in replication is ill-defined, except for NS3 and NS5, which are the most conserved of the flavivirus protein. The C-terminal portion of the NS5 protein contains a motif characteristic of an RNA-dependent RNA polymerase (RdRps). It has also been suggested that NS1 and NS4A proteins are necessary for viral RNA synthesis too (Lindenbach & Rice, 1999). Flavivirus replication was thought to occur exclusively in the cytoplasm (Grun & Brinton, 1988; Takegami & Hotta, 1990; Westaway *et al.*, 1997) however, recent evidence suggests that as much as the 20 % of the replication complexes are localized in the nucleus (Uchil *et al.*, 2006). Nuclear localisation factors of these complexes are unknown.

In the replication complex, the RdRps, NS5 protein together with other viral non-structural and possibly host proteins begin to use the positive strand RNA to make a complementary strand RNA. In turn these negative sense RNA serve as a template for the positive strand genomic RNA which will be incorporated into the nucleocapsid later (Brinton, 2002).

During WNV infection, it is observed that there are extensive reorganisation and proliferation of the smooth and rough endoplasmic reticula (Lindenbach & Rice, 1999; Westaway & Ng, 1980). Membranous structures are also induced but their function remains unknown (Westaway *et al.*, 2002). One of these structures seen in both vertebrate

and invertebrate cells is the formation of vesicles packets that contain bi-layered membrane vesicles of 50-100nm in size. Within these structures are single- or double-stranded "thread-like" structures (Ng, 1987).

There is at present very little information on the assembly and packaging of the nucleocapsid and this will be discussed in a later section. Nucleocapsid assembly with the E and prM protein occurs in association with the endoplasmic reticulum membrane. It has been observed that the intracellular immature virions accumulated in vesicles, which were then transported through the host secretory pathway (Heinz *et al.*, 1994; Wengler, 1989). Matured virions are generated when the glycosylated and hydrophilic amino-terminal portion of the prM is cleaved in the trans-Golgi network by cellular furin and other host proteases (Stadler *et al.*, 1997) while the carboxyl terminal remains inserted in the envelope of mature virus (Murray *et al.*, 1993). This process is essential for the infectivity of some flaviviruses (Elshuber *et al.*, 2003) but not Dengue (Murray *et al.*, 1993). The mature virions were observed by electron microscopy to reside within the lumen of the endoplasmic reticulum (Hase *et al.*, 1989; Matsumura *et al.*, 1977; Ng, 1987; Sriurairatna & Bhamarapravati, 1977) at the perinuclear area of the cytoplasm (Westaway & Ng, 1980).

However, the Sarafend strain of WNV has a slightly different process of assembly as described above. Instead of associating with the endoplasmic reticulum, the nucleocapsid was observed by cryo-immunoelectron microscopy to associate with the E proteins at the host cell's plasma membrane (Ng *et al.*, 2001). Unlike the rest of the flaviviruses, which have a trans-mode of maturation where the mature virion particles are release by exocytosis, Sarafend strain of the WNV matures cis-mode at the plasma

membrane (Mason, 1989; Nowak *et al.*, 1989). The egress of the WNV (Sarafend) was observed to occur primarily at the apical surface of polarized Vero cells, this suggested that a microtubule-dependent polarized sorting mechanism exists for WNV proteins (Chu & Ng, 2002b). A later study demonstrated that both the E and C proteins associated strongly with microtubules and were transported to the plasma membrane for assembly (Chu & Ng, 2002a). This interaction between the WNV proteins and the microtubule was suggested to be ionic in nature since the interaction was sensitive to high salt extraction but resistant to Triton X-100 and octyl glycoside extraction (Chu & Ng, 2002b).

A phenomenon observed with viruses in the *flaviviridae* family is that virion-like particles (VLPs) were found in infected sera (Kaito *et al.*, 1994; Mizuno *et al.*, 1995; Shimizu *et al.*, 1996). These VLPs are made up of the E and M proteins embedded in the lipid bilayer but it does not have a nucleocapsid core and thus non-infectious. Such observations point out that nucleocapsid assembly of the C protein and the budding of the E and M protein in the lipid bilayer are two processes independent of each other.

1.6 The capsid (C) protein

The WNV C protein is the first structural protein found in the ORF and has a molecular weight of approximately 12-15 kDa. The C protein is the basic building block of the nucleocapsid which encapsidates viral RNA. It is assumed that C protein has RNA binding properties and this was elucidated by Khromykh and Westaway in 1996. A precursor of the C protein is 123 amino acids long and it contains a hydrophobic region at the carboxyl terminal, which anchors the protein to the membrane of ER. This region is cleaved off to generate a mature 105 amino acid C protein by viral protease NS2B-NS3

(Chambers *et al.*, 1990; Wengler & Gross, 1978). The sequences of the C proteins are poorly conserved among flaviviruses however it is biochemically and structurally similar (Fig. 1-2). The flavivirus C protein is rich in basic residues and it has a distinct hydrophobic segment in the middle (Markoff *et al.*, 1997).

Although the C protein is thought to primarily package the viral genome, a functional nuclear localisation signal characterized on the WNV, Dengue virus and Japanese encephalitis virus C protein has called into question its function in the nucleus (Bhuvanakantham *et al.*, 2009; Mori *et al.*, 2005; Wang *et al.*, 2002). Studies on the Hepatitis C virus (a member of the *Flaviviridae* under the genus Hepacivirus) C protein suggested that the C protein is multifunctional. It has been implicated in cell transformation, lipid metabolism, transcription, immune presentation and regulation of apoptosis (McLauchlan, 2000).

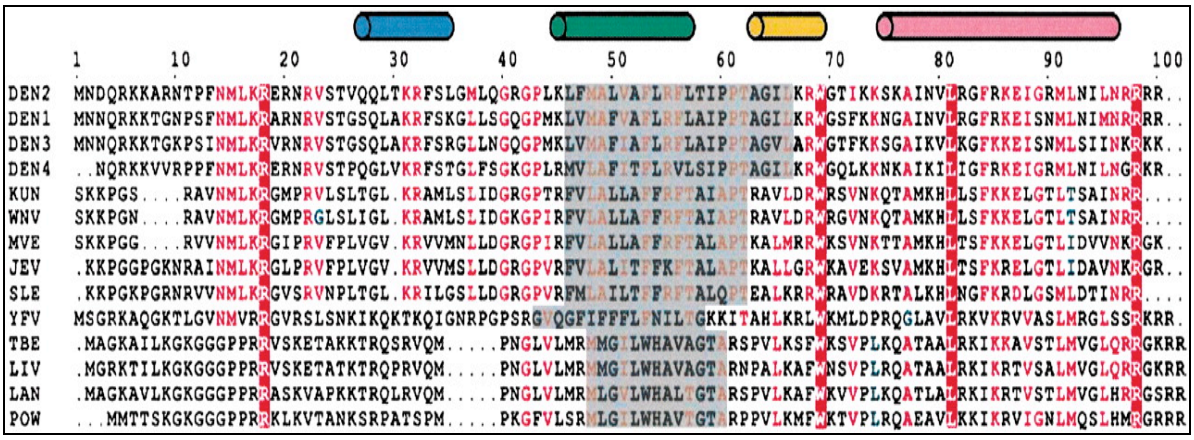

Figure 1-2. Multiple sequence alignment of flavivirus C protein. The 4 alpha-helices are indicated at the top and colour coded. The colours correspond to the 3 dimensional structure of the C protein in Fig 1-3. The conserved regions are shaded grey. Residues with similarity greater than 50 % are in red while conserved residues are highlighted in red. DEN2, Dengue type 2; DEN1, Dengue type 1; DEN3, Dengue type 3, DEN4, Dengue type 4; KUN, Kunjin; WNV, West Nile virus; MVE, Murray Valley encephalitis; JEV, Japanese encephalitis; SLE, St. Louis encephalitis; YFV, yellow fever; TBE; LIV, louping ill; LAN, Langat; POW, Powassan virus (Ma *et al.*, 2004).

1.6.1 Structure of capsid (C) protein

Cryo-electron microscopy analyses of both the WNV and Dengue viruses showed that E protein on the surface is well-ordered but indicated that the C protein had no ordered density (Kuhn *et al.*, 2002; Mukhopadhyay *et al.*, 2003) This suggests that C protein does not share the organization of E proteins on the surface.

Structural studies done on Yellow Fever and Dengue Virus C protein revealed that flavivirus C is a dimeric alpha-helical protein (Jones *et al.*, 2003). It was demonstrated that the flavivirus C protein has a novel fold and the monomer comprises of 3 helices, alpha-1 to 3 and a fourth helix, alpha-4 extending away from the protein and can form tetramers [(Fig. 1-3) (Ma *et al.*, 2004)]. Ma and colleagues (2004) proposed that the positively charged regions would extend into the centre to interact with viral RNA while the hydrophobic region would interact with the membrane of endoplasmic reticulum (Fig. 1-4). The dimeric structure of the Dengue virus C protein was confirmed by the crystal structure of the C protein from the Kunjin strain of WNV (Dokland *et al.*, 2004). The crystal structure revealed the alpha-1 helix, which corresponded to the amino acid terminal, was flexible. It was proposed that this flexibility allowed for a conformation switch so that amino- and carboxyl-terminal of C protein could be brought together. This is significant since the RNA binding region on Kunjin C protein was found on both the amino and carboxyl terminals (Khromykh & Westaway, 1996). In addition, it was revealed that the dimers formed tetramers in the crystal structure (Dokland *et al.*, 2004). The authors proposed that this tetrameric structure actually shielded the hydrophobic region, creating a positively charged surface for RNA binding.

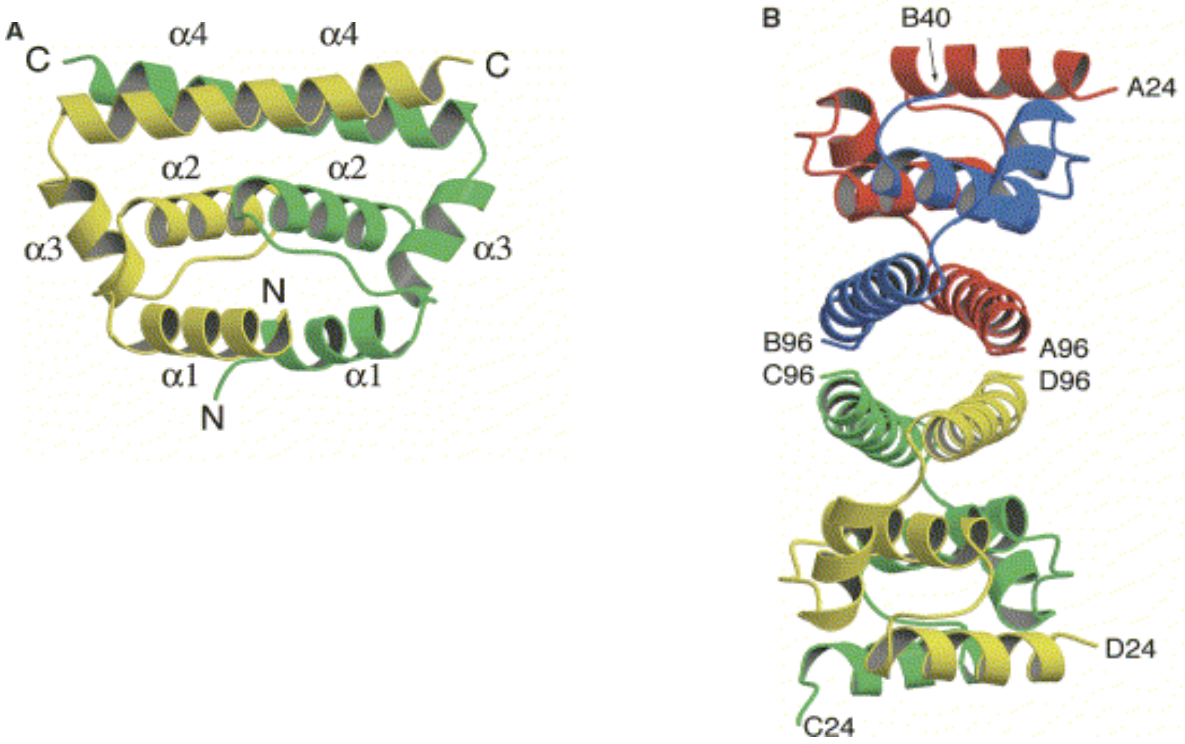

Figure 1-3. Ribbon representation of the capsid protein from Kunjin strain of WNV. Each monomer of capsid protein is given one colour. (A) A ribbon representation of a capsid dimer. (B) A ribbon representation of a capsid tetramer showing the formation of a tunnel between the two dimers (Ma *et al.*, 2004).

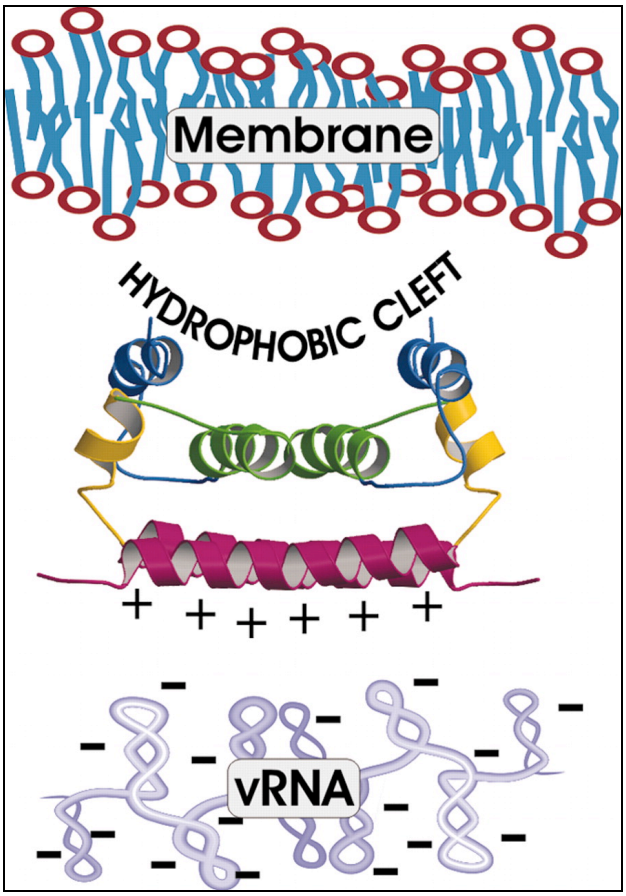

Figure 1-4. Proposed model of flavivirus C protein interaction with the lipid bilayer and viral RNA. The dimerized C protein is shown in between the lipid bilayer membrane and viral RNA (vRNA). The hydrophobic face of the C dimer is facing the membrane while the positively charged surface is facing the RNA. (Ma *et al.,* 2004)

1.6.2 Nucleocapsid dimerization and viral assembly

Nucleocapsid assembly in general involves C-C and C-nuclei acid interactions. Nucleocapsid assembly models of other spherical viruses suggest that when interaction between monomeric C proteins is weak, a nucleic acid scaffold, dimerization and/or oligomerization of C protein could enhance C-C interaction and hence provide the stability to induce assembly (Zlotnick, 2003). The mechanism suggested during nuclei acid-induced assembly is its binding to C protein would elevate local C protein concentration and correctly orientate the protein for dimerization. Mathematical models of assembly based on Hepatitis B virus also show that *in vitro* dimerization of C protein subunits is expected to favour assembly regardless of the role of dimers *in vivo* (Ceres &

Zlotnick, 2002). To understand the effect of dimerization on assembly, analysis of the geometry each subunit of the nucleocapsid is necessary. In general, the more contact points a subunit makes, the more stable the resulting capsid is (Zlotnick, 2003). These models can be extended to the assembly of flavivirus C protein but since structural studies revealed that flavivirus nucleocapsid did not have an ordered density, geometric analysis of the flavivirus nucleocapsid might be irrelevant. Nonetheless, weak interactions between C proteins reduce the likelihood of the formation of kinetic traps but allow for other factors such as nucleation to enhance assembly. This characteristic might be important for viruses where dissociation is part of its life cycle.

A functional study identified the critical residue, Trp69, for C-C self-association. Mutation to this residue abolished or greatly attenuated dimerization of C protein (Bhuvanakantham & Ng, 2005). Evidence suggested that the basic building block of flavivirus nucleocapsid is the dimeric form of the C protein (Kunkel *et al.*, 2001; Patkar *et al.*, 2007). This corroborated with the NMR and crystal structure of the C protein. The importance of a dimeric C proteins in virion morphogenesis is unclear although nucleocapsid-like particles assembled from purified Tick-borne encephalitis virus C protein suggested that C dimers functioned as the building block of nucleocapsid assembly (Kiermayr *et al.*, 2004)

Though it was demonstrated that the amino- and carboxyl- terminal of Kunjin virus is responsible for RNA binding, mutational and deletion studies with Yellow fever, Tick-borne encephalitis and Hepatitis C viruses C proteins suggested that the positively-charged clusters of amino acid residues at the amino- and carboxyl-terminal interacted with nucleic acids cooperatively because they are functionally redundant. Conversely, it

was demonstrated that deletions or mutations to the internal hydrophobic region of C protein was tolerated to a lesser extent (Kofler *et al.*, 2002; Patkar *et al.*, 2007). This is consistent with the notion that hydrophobic interactions between C protein and the lipid membrane is important for assembly of mature virus since the cytoplasmic domains of both M and E proteins are very short, thus they are not likely to provide the necessary interactions for association with the nucleocapsid (Markoff *et al.*, 1997). In any case, the mechanism pertaining to the envelopment of the nucleocapsid is still unclear.

Studies from *in vitro* nucleocapsid assembly of Hepatitis C virus and Tick-borne encephalitis virus C proteins suggested that C-C self-association and the eventual formation of a nucleocapsid was a spontaneous process in the presence of nucleic acid (Kiermayr *et al.*, 2004; Kunkel *et al.*, 2001). It is proposed that nuclei acids form nucleation points for the subsequent oligomerization of the C protein. Indeed *in vitro* study on the assembly of the alphaviruses, a genus of the *Togaviridae*, was also consistent with this idea (Tellinghuisen *et al.*, 1999). In these *in vitro* systems, either viral RNA or short DNA oligonucleotides were sufficient to assemble nucleocapsid-like particles, suggesting that encapsidation may not be specific.

1.6.3 Flavivirus capsid (C) nuclear localisation

Many flavivirus as well as Hepatitis C virus C proteins have been demonstrated to localize in the nucleus of many infected cells (Falcon *et al.*, 2003; Falcon *et al.*, 2005; Mori *et al.*, 2005; Sangiambut *et al.*, 2008; Suzuki *et al.*, 1995; Suzuki *et al.*, 2005; Wang *et al.*, 2002; Westaway *et al.*, 1997). However, the functions of C protein in the nucleus are unclear since positive-stranded RNA viruses are thought to utilize cellular

components in the cytoplasm for replication. Nuclear localisation of C protein for positive-stranded RNA virus is not unusual though. The C protein of coronavirus, another group of positive-stranded RNA viruses, was also reported to localize in the nuclei of infected cells and interacted with nucleolin protein (Chen *et al.*, 2002). Nucleolin is a multifunctional protein involved in activities such as cytokinesis, cell proliferation, ribosome biogenesis, chromatin decondensation and transcription regulation (Tuteja & Tuteja, 1998). This suggested that coronavirus C protein could be involved in any of the above-mentioned functions by interacting with nucleolin protein.

More recently, it has been suggested that coronavirus C protein was involved in cycle arrest during an infection through its interaction with cyclin-cyclin dependent kinase complex (Li *et al.*, 2007; Surjit *et al.*, 2006). Hence, flavivirus C protein could have similar roles in the nucleus during an infection. In addition, flavivirus C protein may also play an important role in transcriptional and translational regulation of host proteins since Dengue virus C protein was reported to interact with the heterogeneous nuclear ribonucleoprotein K (Chang *et al.*, 2001).

Nuclear localisation of C protein is mediated by the nuclear localizing signal (NLS) motif on the protein. The NLS motif is a region on a protein that is rich in basic residues, like lysine and arginine. There are two types of NLS motifs – monopartite NLS, namely one short region of about four to five basic residues, and bipartite NLS which consists of two regions of basic residues separated by 10 to 12 random residues sequence (Dingwall & Laskey, 1991). Thus far, the nuclear localisation signal has been identified on Dengue, West Nile and Hepatitis C viruses C protein (Falcon *et al.*, 2003; Wang *et al.*, 2002). The NLS motif on a protein has been widely demonstrated to interact with

importin-α/β to mediate nuclear localisation of the protein (Friedrich *et al.*, 2006; Reguly & Wrana, 2003; Whittaker & Helenius, 1998). The mechanism of nuclear translocation necessitates the binding of importin-α to the NLS motif. Importin-α then acts as an adaptor for the binding of importin-β. Subsequently, importin-β docks the entire complex of the NLS-bearing protein and importin- α/β complex at the nuclear pore complex for translocation (Gorlich & Mattaj, 1996). Thus far it has been shown that the NLS motif of C protein of Hepatitis C, Dengue and West Nile viruses interacted directly with importin-α to mediate nuclear localisation (Bhuvanakantham *et al.*, 2009; Suzuki *et al.*, 2005) The disruption of nuclear localisation has been shown to be detrimental for viral replication, suggesting that this phase of the viral life cycle is important for viral replication (Bhuvanakantham *et al.*, 2009).

1.6.4 Capsid (C) protein and RNA interaction

The RNA binding properties of the flavivirus C protein was first demonstrated with the Kunjin strain of WNV (Khromykh & Westaway, 1996). Similar reports of C protein RNA binding property were also reported in Hepatitis C virus (Santolini *et al.*, 1994). All these report indicated that the positively charged clusters of the C protein at the amino- and carboxyl- terminals were involved in RNA binding. This property can be extended to all flavivirus C proteins since they all have positively charged clusters.

Although the region on C protein involved in RNA binding was defined in Kunjin virus, the encapsidation signal of the RNA was never defined. However, it was demonstrated that the 5'UTR and 3'UTR of the Kunjin genomic RNA bind to Kunjin C protein (Khromykh and Westaway, 1996). In contrast, encapsidation for Sindbis and

Rubella viruses, both of which belong to the family of *Togaviridae*, were defined (Geigenmuller-Gnirke *et al.*, 1993; Liu *et al.*, 1996). In the Sindbis virus nucleocapsid assembly model, it was proposed that binding of C protein to the virus RNA encapsidation signal, promoted a conformational change in the RNA and this resulted in the dimerization of the capsid protein (Geigenmuller-Gnirke *et al.*, 1993). In addition, encapsidation signal were also found for many other animal viruses like the *Retroviridae* (Banks & Linial, 2000; Beasley & Hu, 2002; McBride & Panganiban, 1996), *Coronaviridae* (Cologna *et al.*, 2000; Narayanan & Makino, 2001), *Bunyaviridae* (Ng, 1987; Severson *et al.*, 2001; Xu *et al.*, 2002) and *Orthomyxoviridae* (Tchatalbachev *et al.*, 2001).

In all these examples, RNA encapsidation signal contains complex secondary confirmation such as stem loops and bulges. Analysis of 5' and 3'UTR of flavivirus genome also reveal such structures. However, the difference between flavivirus genome and the rest of the animal viruses genomes listed above is that flavivirus genome undergoes cyclization whereby the conserved regions on the 5' and 3'UTR are paired (Khromykh *et al.*, 2001). It is therefore, not surprising that the 5' and 3' UTR of Kunjin virus displays C protein binding properties. It is however unknown if other regions of the flavivirus RNA has specific C protein binding properties as well.

In addition to packaging of the RNA, it was recently proposed that the flavivirus C protein acted as an RNA chaperone to assist in the correct folding of RNA molecules (Ivanyi-Nagy *et al.*, 2008). The authors suggested that the intrinsically disordered segments of the C protein induce RNA structural rearrangements and this could have implications for RNA encapsidation and replication.

1.6.5 Phosphorylation of capsid (C) protein and its effects

Phosphorylation of C protein in many plant and animal viruses has been demonstrated to be an integral and important process for viral replication. The disruption of this process is detrimental to viral replication. For example, prevention of phosphorylation of C protein of the Cauliflower Mosaic virus resulted in reduced levels of virus accumulation (Leclerc *et al.*, 1999). Similarly, when C protein of rabies virus was not phosphorylated, both replication and transcription were reduced (Wu *et al.*, 2002). In another study, phosphorylation of C protein of the Potato X virus induced co-translational virion disassembly (Atabekov *et al.*, 2001). In general, the direct effects of phosphorylation on a viral C protein function can be broadly divided into three categories – i) RNA binding, ii) protein-protein interaction including iii) self-oligomerization and iv) nuclear localisation.

Studies from both animal and plant viruses provided evidence that phosphorylation affect the RNA binding ability of C protein. In the case of Hepatitis B and Rubella virus, it was shown that phosphorylation of the C protein attenuated its binding to viral genomic RNA (Gazina *et al.*, 2000; Law *et al.*, 2003). In addition, studies on the potivirus, Potato virus A corroborated with earlier results from the Hepatitis B and Rubella virus whereby phosphorylation of C protein also attenuated its binding to viral RNA (Ivanov *et al.*, 2001). These are not a surprising finding since phosphorylation of a nucleic binding protein would neutralize the positive charges on the protein, thus reducing its affinity for the negatively charged nucleic acids. Besides modulating the charge on the protein, there exists the possibility that phosphorylation might effect a

conformational change in the protein which attenuates its RNA binding (Yu & Summers, 1994).

Conversely, it should also be noted that phosphorylation of some proteins can enhance its RNA binding ability. For example, in Human T-cell leukemia virus type 2, phosphorylation of the Rex protein enhanced RNA binding activity (Green *et al.*, 1992). Evidence also suggested that the RNA binding activity of C protein of the Mouse hepatitis virus, a member of the *Coronaviridae*, was also enhanced by phosphorylation because dephosphorylation is required for the disassembly and release of the viral RNA (Kalicharran *et al.*, 1996). Thus it can be surmised that phosphorylation plays a modulating role in a protein's RNA binding activity.

Protein-protein interaction, including dimerization, can also be affected by phosphorylation and this was demonstrated by studies from various viruses. For example, in the Herpes simplex virus type 1, phosphorylation of its structural proteins promoted tegument dissociation (Morrison *et al.*, 1998). A mutation at the phosphorylation site of the capsid protein, VP1 in polyomavirus was associated with a defect in virion assembly (Li & Garcea, 1994). It has also been proposed that phosphorylation of Cap24 structural protein of Human immunodeficiency virus type 1 is necessary for the disassembly of the virus (Cartier *et al.*, 1999). All these suggested that phosphorylation affect protein-protein interaction and therefore the oligomerization and assembly/disassembly C protein.

Finally, phosphorylation can also affect nuclear localisation of phosphorylated protein. It has been reported that phosphorylation of residues around the NLS motif enhanced nuclear localisation of a protein as is the case for the Simian-Virus-40 large T-antigen (Fontes *et al.*, 2003). Evidence from mutational analyses of the phosphorylation

sites on Hepatitis C virus C protein also showed enhancement of nuclear localisation (Lu & Ou, 2002).

Certainly, it is not only the properties of C protein that could be modulated by phosphorylation. Recent studies with the Hepatitis C virus C protein revealed that phosphorylation targeted the protein for degradation. Hence phosphorylation acts as a modulator of C protein level in infected cells (Majeau *et al.*, 2007).

Therefore, it can be surmised from the above studies that phosphorylation of the C protein plays an important role in modulating its function. Since phosphorylation of the C protein occurs across a variety of viruses, it is possible that this mechanism is conserved for regulating virion assembly and viral replication.

Thus far, there are currently no direct evidence that the flavivirus C is a phosphoprotein although it has been shown that the Hepatitis C virus C protein is a phosphoprotein (Lu & Ou, 2002). Consequently, it is likely that the WNV C protein is functions as phosphoprotein and that its activities such as RNA binding, oligomerization and nuclear localisation are modulated by phosphorylation.

1.7 Objectives

Encapsidation of viral RNA is a critical process in WNV assembly. Characterization of the C protein and RNA interaction would give an insight on how encapsidation might occur. In addition, mounting evidence from other viruses suggests that viral assembly is a regulated process involving host factors. Phosphorylation has often been cited to regulate protein functions. Therefore this study aims to elucidate how phosphorylation might modulate the functions of the C protein, with regards to the processes of nucleocapsid assembly and ultimately viral replication.

Since nucleocapsid assembly would involve the interaction between viral RNA and C protein, the specific objectives of this study are as follows:

(a) To develop assays to study WNV C protein-RNA interaction.

(b) Determine if C protein is a phosphoprotein.

(c) Characterize C protein and RNA interaction and determine how phosphorylation affects this interaction.

(d) Determine how phosphorylation might affect other functions of the C protein with regards to nucleocapsid assemnbly and ultimately viral replication.