RECOMBINANT NS3 SERINE PROTEASE FROM DENGUE VIRUS 2 AS A SCREEN FOR SMALL MOLECULES

NUROHAIDA BINTI AB AZIZ

UNIVERSITI SAINS MALAYSIA
2011

RECOMBINANT NS3 SERINE PROTEASE FROM DENGUE VIRUS 2 AS A SCREEN FOR SMALL MOLECULES

by

NUROHAIDA BINTI AB AZIZ

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

August 2011

ACKNOWLEDGMENTS

First and foremost I would like to express my deepest gratitude to my main supervisor, Professor Maqsudul Alam, who had contributed fruitful ideas, comments, continuous motivation and excellent technical assistance towards the completion of this thesis. I would like to thank my co-supervisor, Dr. Jennifer Saito for her assistance, guidance and supervision throughout this study. I would also like to thank my collaborator, Dr. Irene Newhouse from University of Hawaii for the collaboration work in *in silico* drug design. I would also like to thank Dr. Azat Mukhametov from the Centre for Chemical Biology, Universiti Sains Malaysia for performing additional structural analysis.

I am thankful to Professor Nazalan Najimudin and Dr. Rashidah for their guidance and support. I would like to express my sincere gratitude to my lab members especially Dr. Teh Aik Hong, Dr. Masaomi Kanbe, Suria, Su Yean, Luqman, Beng Soon, Patrick, Lingsze, Sheri, Bee Feong and friends for the motivation and help during my Master of science program. I would also like to thank the Centre for Chemical Biology management team including Mr. Larry, Ms. Roslina, Ms. Nithi and Ms. Komala for all of their help throughout the years.

Most of all, I want to thank my husband, Mohd Aatif for his love, encouragement and support in completing this thesis. Finally, I am grateful to my parents and family members for their constant encouragement and support during this endeavor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

Å	angstrom
AMC	7-amino-4-methyl coumarin
bp	base pair (s)
cDNA	complementary DNA
cm	centimetre
C-terminal	carboxyl terminal
CV	column volume
DENV	dengue virus
DENV1, DENV2, DENV3, and DENV4	dengue virus serotype 1, 2, 3, and 4
DF	dengue fever
DHF	dengue hemorrhagic fever
DNA	deoxyribonucleic acid
DSS	dengue shock syndrome
E. coli	*Escherichia coli*
ER	endoplasmic reticulum
FcγR	Fcγ receptors
g	gram
GRR-AMC	Boc-Glycine-Arginine-Arginine-4-methylcoumaryl-7-amide
HCV	hepatitis C virus
HTS	high-throughput screening
IMAC	Immobilized metal affinity chromatography

IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase pair (s)
kDa	kilodalton (s)
L	litre
M	molar
mg	milligram
min	minute (s)
ml	millilitre
mM	millimolar
mm	millimetre
ng	nanogram
NGC	New Guinea C
nm	nanometer (s)
nM	nanomolar
NS2B	nonstructural protein 2B
NS3pro	nonstructural protein 3 protease domain
N-terminal	amino terminal
ºC	degree Celsius
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
PDB	Protein Data Bank
pmol	picomolar
PVDF	PolyVinyliDene Fluoride
RdRp	RNA-dependent RNA polymerase

RNA	ribonucleic acid
s	second (s)
SAM	S-adenosyl methyltransferase
SAR	structure-activity relationship
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TCA	Trichloroacetic Acid
TGN	trans-Golgi network
TM	transmembrane regions
U	unit (s)
UTR	untranslated region
UV	ultraviolet
V	voltage
v/v	volume/volume
VP	vesicle packets
WNV	West Nile virus
x g	g-force
µg	microgram
µl	microlitre
µM	Micromolar

REKOMBINAN PROTEASE SERINA NS3 DARIPADA VIRUS DENGGI 2 SEBAGAI PENYARING UNTUK MOLEKUL KECIL

ABSTRAK

Jangkitan denggi adalah muncul semula sebagai satu penyakit utama dunia dan diklasifikasikan sebagai patogen utama kategori A. Setiap tahun, dianggarkan 50-100 juta manusia dijangkiti virus denggi dan dianggap sebagai penyebab kepada salah satu penyakit virus bawaan arthropoda paling penting dari segi kematian dan kemorbidan manusia. Penjangkitan virus berlaku melalui gigitan nyamuk *Aedes aegypti* dan separuh daripada populasi dunia berisiko kepada jangkitan. Sehingga sekarang masih tiada lagi drug antivirus atau vaksin diluluskan yang berkesan untuk menetang virus denggi. Fokus tesis ini adalah untuk menggabungkan diantara kuasa pengkomputeran berprestasi tinggi dengan eksperimen makmal dimana rekombinan protease serina NS3 daripada virus denggi 2 sebagai penyaring molekul kecil antivirus yang boleh digunakan untuk menghalang atau merawat jangkitan virus denggi.

Kerjasama dengan Dr. Irene Newhouse, Advance Studies for Genomics, Proteomics and Bioinformatics (ASGPB), University of Hawaii, model molekul dan penyaringan secara *in silico* telah dijalankan keatas perpustakaan sebatian molekul kecil daripada pangkalan data National Cancer Institute (NCI) dan ZINC untuk sebatian kecil yang mengedok ke dalam tapak ikatan protease DENV2 NS2B-NS3. Daripada himpunan calon-calon yang menunjukkan suaian terbaik (53 perencat-perencat molekul kecil yang berpotensi), 4 sebatian larut air yang menunjukkan skor tertinggi, boleh didapati secara komersial telah dipilih untuk penilaian secara *in vitro*.

Gen protease serina NS2B-NS3 daripada virus denggi serotip 2 telah diklon dan diekspresi dalam *E. coli* sebagai protein rekombinan berpenanda heksahistidina. Protease NS2B-NS3 telah ditulen menggunakan kromatografi affinity dan penurasan gel. Asai *in vitro* menunjukkan aktiviti protease terhadap substrat peptida fluorogenik yang mengandungi dua residual berbes. Kesemua 4 sebatian larut air yang menunjukkan skor tertinggi diuji dan mempamerkan aktiviti perencatan secara *in vitro* terhadap rekombinan protease serina NS2B-NS3. Sebatian 4 didapati memberi kesan perencatan paling aktif dimana kadar perencatan sebanyak 64% pada kepekatan 100 µM.

Sebagai kesimpulan, kajian tesis ini membuktikan bahawa rekombinan protease serina NS3 daripada virus denggi 2 boleh digunakan sebagai penyaring molekul kecil antivirus secara *in silico* dan *in vitro*. Sebatian 4 adalah penemuan berharapan dan berpotensi untuk dibangunkan sebagai drug anti-denggi.

RECOMBINANT NS3 SERINE PROTEASE FROM DENGUE VIRUS 2 AS A SCREEN FOR SMALL MOLECULES

ABSTRACT

Dengue infection is re-emerging as a major global disease and is classified as a Category A priority pathogen. Dengue viruses are estimated to infect 50-100 million people annually and are considered to cause one of the most important arthropod-borne viral diseases in terms of human morbidity and mortality. Virus transmission occurs through the bite of the *Aedes aegypti* mosquito and half the world's population is at risk for infection. There is presently no approved vaccine or antiviral drug that is effective against dengue viruses. The focus of this thesis is to combine the power of high performance computing with wet lab experiments for the recombinant NS3 serine protease from dengue virus type 2 as a screen for antiviral small molecules that can be used either to prevent or treat dengue virus infections.

In collaboration with Dr. Irene Newhouse, Advance Studies for Genomics, Proteomics and Bioinformatics (ASGPB), University of Hawaii, molecular modelling and *in silico* screening of small molecule compound libraries from the National Cancer Institute (NCI) and ZINC databases that dock into the DENV2 NS2B-NS3 protease binding site was carried out. From the pool of best-fit candidates (53 potential small molecule inhibitors), the 4 high-scoring water-soluble, commercially available compounds were selected for *in vitro* assessment.

The NS2B-NS3 serine protease gene from dengue virus serotype 2 was cloned and expressed in *E. coli* as a recombinant hexahistidine tagged protein. The NS2B-NS3 protease was purified using affinity and gel filtration chromatography. *In*

vitro assay revealed protease activity toward a fluorogenic peptide substrate containing two basic amino acid residues. All 4 high-scoring water-soluble compounds were tested and exhibited *in vitro* inhibition activity on the recombinant NS2B-NS3 serine protease. Compound 4 was found to be most active inhibitor with 64% inhibition at 100 μM concentration.

In summary, this thesis project has established that the purified recombinant NS3 serine protease from dengue virus type 2 can be used to screen antiviral small molecules *in silico* and *in vitro*. Compound 4 is a promising finding for further development as an anti-dengue drug.

CHAPTER 1

Introduction

1.1 Dengue virus infection

Dengue virus (DENV) is the most important human viral disease transmitted by an arthropod vector, with an estimated annual infection rate in excess of 50 million. The majority of infections are silent with no obvious clinical symptoms. Nevertheless, a significant minority of infected individuals develop a mild febrile illness, dengue fever (DF), or even life-threatening dengue hemorrhagic fever (DHF)/dengue shock syndrome (DSS) which has an increasing incidence in tropical and subtropical countries.

The first reported epidemics of DF occurred in 1779-1780 in Asia, Africa, and North America. During that time, DF was considered a benign, nonfatal disease of visitors to the tropics. The disease was confined to relatively small geographic regions and the four different serotypes of DENV remained isolated. On the contrary, the global prevalence of DENV is now increasing dramatically and DENV epidemics caused by multiple serotypes (hyperendemicity) are more frequent (Gubler, 1998; Rigau-Perez *et al.*, 1998; Gubler, 2002).

The disease is caused by four antigenically related but distinct serotypes of DENV: DENV1, DENV2, DENV3, and DENV4. Despite being an age-old disease, there is no effective treatment for DENV infection. Researchers have endeavored to develop a vaccine for many years with very little success. The reason is that an effective vaccine would have to protect against all four serotypes of DENV.

Considerable efforts are now contributed to the development of antiviral compounds (Sampath and Padmanabhan, 2009).

Efficient and accurate diagnosis of DENV is of primary importance for clinical care. It includes epidemiological consideration (season of the year, travel history), physical examination (high body temperature, blood pressure, evidence of bleeding in the skin or other sites, hydration status, evidence of increased vascular permeability, and tourniquet test), and clinical laboratory tests (virus isolation, nucleic acid detection, detection of antigens, serological tests, and haematological tests) (WHO, 2009).

1.2 Transmission, prevalence, and consequences

Mosquitoes, humans, and lower primates such as chimpanzees, gibbons, and macaques are all considered to be the natural hosts for DENV infections. However, humans are the main amplifying host of the virus (Henchal and Putnak, 1990). DENV is transmitted to humans through the bite of infected female mosquitoes, either *Aedes aegypti* or *Aedes albopitus*, which can usually be found near or in human dwellings. The species is day-active, with most biting activity occurring in the early morning or late afternoon.

The transmission cycle of DENV by the mosquito begins with a DENV-infected person (Gubler, 1998). The person will have virus circulating in the blood for approximately 4 to 7 days. This beginning state is called viremia. During this period, if other uninfected female mosquitoes bite the ill person, those mosquitoes may become infected and becomes infective after an obligatory extrinsic incubation period of 10 to 12 days. After the mosquito becomes infective, it may transmit

DENV by taking a blood meal, or by simply probing the skin of a susceptible person (Rigau-Perez *et al.*, 1998). Symptoms that are caused by DENV infection may last 3 to 10 days after the onset of symptoms as illustrated in Figure 1.1.

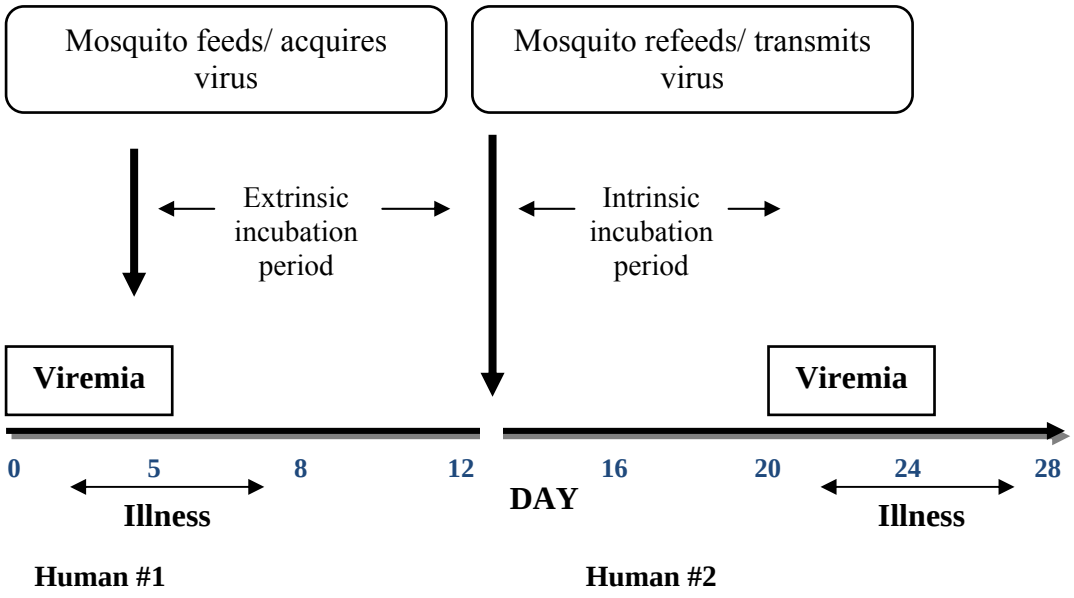


Figure 1.1 Transmission of DENV by *Aedes aegypti* and/or *Aedes albopitus* (WHO 2009. *Dengue: guidelines for diagnosis, treatment, prevention and control -- New edition.* Geneva, World Health Organization).

Currently, 2.5 billion people are at risk for DENV infection. DENV is endemic in more than 100 countries in Africa, the Americas, Southeast Asia, and the Western Pacific as illustrated in Figure 1.2. The most seriously affected areas are Southeast Asia and the Western Pacific region. DENV epidemics occurred sporadically in the Americas from the 18th to the mid-20th century. DF has become an endemic disease since the 1970s. In Southeast Asia, DHF has been recognized for approximately 40 years. Most of the cases are reported from Thailand, Indonesia, and Vietnam.

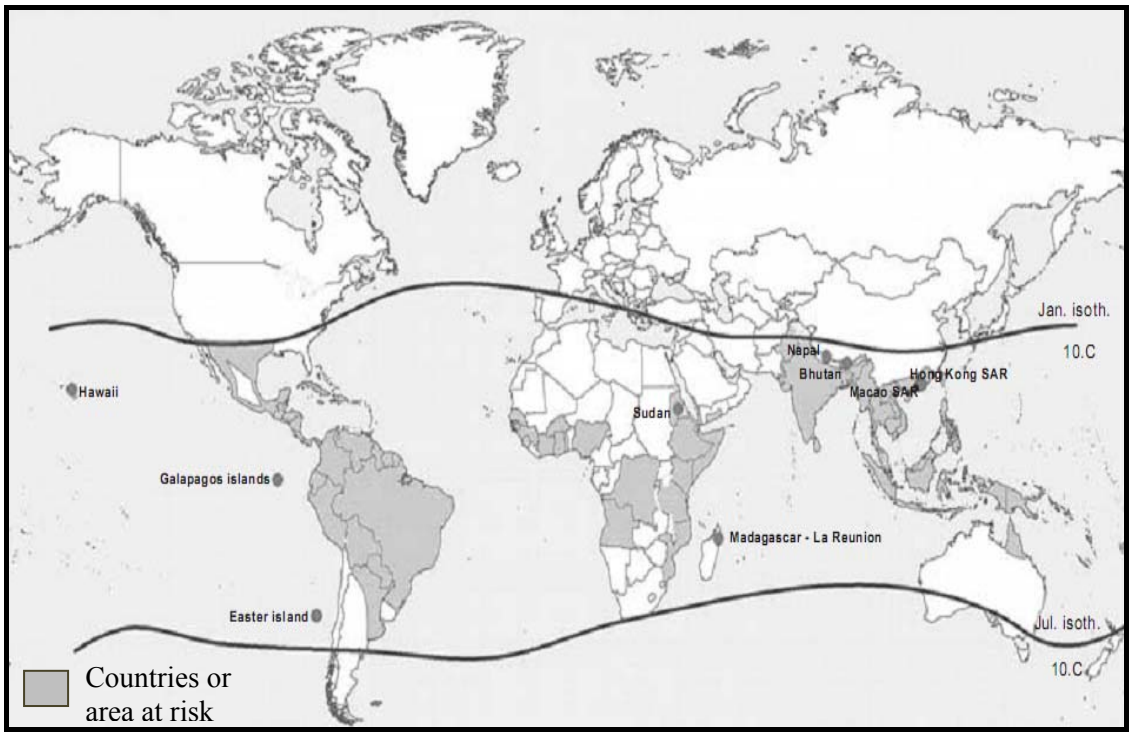


Figure 1.2 Approximate distributions of dengue cases in 2008. Cases were largely confined to subtropical and tropical regions of the world (Tomlinson, S. M., Malmstrom, R. D. & Watowich, S. J. 2009b. New approaches to structure-based discovery of dengue protease inhibitors. *Infect. Disord. Drug Targets,* 9**,** 327-43. Figure 2, page 328).

The global prevalence of DENV is increasing dramatically as a result of the rapid rise of urban populations, expansion of mosquito breeding, and migration of infected people. As a consequence, epidemics caused by multiple serotypes (hyperendemicity) are more frequent and DHF has become a leading cause of hospitalization and death among children in many countries throughout the world (WHO, 2009). As an infectious disease, DENV inflicts a significant health, economic, and social burden on the populations of endemic areas.

In general, most of the DF cases are self-limited in their course and rarely progress to fatal DHF. There are many risk factors associated with the occurrence of DHF such as the virulence of different virus strains, host genetic factors, and age. However, immune response to DENV appears to be a major factor in the pathogenesis of DHF and DSS (Halstead and O'Rourke, 1977). A strong association

of severe disease in humans undergoing a heterotypic secondary infection has been established (Halstead *et al.*, 1970; Vaughn *et al.*, 1997).

Several hypotheses have been proposed and one of these is 'antibody-dependent enhancement'. According to this hypothesis, the enhanced disease severity that is observed after secondary infection by a different DENV serotype is believed to be mediated primarily by pre-existing, non-neutralizing heterotypic antibodies that enhance access of DENV to FcγR-bearing cells. The antibody-dependent enhancement hypothesis as illustrated in Figure 1.3 resulted in an increase in both the total number of FcγR-bearing cells infected and the total amount of virus produced. The infected cells then release vasoactive mediators resulting in the increased vascular permeability and hemorrhagic manifestations (Whitehead *et al.*, 2007). In addition, the target tissue for viral infection could be both mononuclear cells and megakaryocytes in the bone marrow (Halstead, 1989).

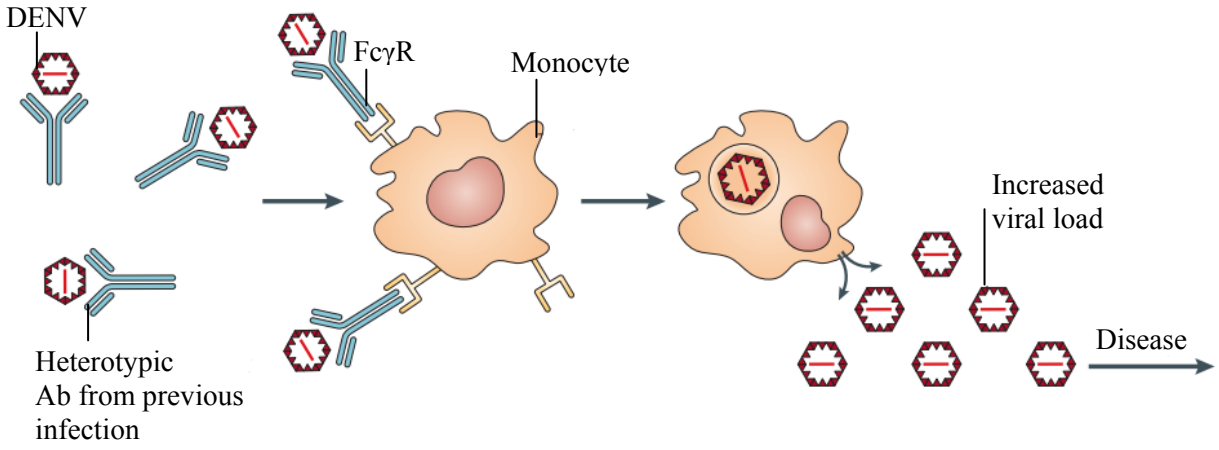


Figure 1.3 Model for antibody-dependent enhancement of DENV replication. Antibody (Ab)-dependent enhancement of virus replication occurs when heterotypic, non-neutralizing Ab present in the host from a primary DENV infection binds to an infecting DENV particle during a subsequent heterotypic infection but cannot neutralize the virus. Instead, the Ab-virus complex attaches to the Fcγ receptors (FcγR) on circulating monocytes, thereby facilitating the infection of FcγR cell types in the body not readily infected in the absence of antibody. The overall outcome is an increase in the overall replication of virus, leading to the potential for more severe disease (Whitehead, S. S., Blaney, J. E., Durbin, A. P. & Murphy, B. R. 2007. Prospects for a dengue virus vaccine. *Nat. Rev. Microbiol.,* 5**,** 518-28. Figure 3, page 524).

There are 3 criteria that must be fulfilled in order to meet the case definition for severe dengue (WHO, 2009). The criteria are severe plasma leakage leading to shock, accumulation of fluid and respiratory distress, severe hemorrhagic manifestations, and severe organ impairment which mainly involve the liver, heart, and central nervous system.

1.3 Status of dengue therapy

There is no specific treatment for DENV infection. The only treatment available is symptomatic treatment with careful clinical management by experienced physicians and nurses. This can often save the lives of DHF patients (WHO, 2009).

Despite considerable work over the years, a licensed vaccine against DENV is still elusive and even today there are only candidate DENV vaccines. A successful vaccine must be tetravalent, capable of simultaneously inducing a high level of long-lasting immunity to all four DENV serotypes (Ray and Shi, 2006). The immune enhancement phenomenon underlying disease pathogenesis and the lack of suitable animal models to evaluate candidate DENV vaccines are the major challenges to vaccine development (Johnson and Roehrig, 1999; Lei *et al.*, 2001).

Various strategies have been used to develop DENV vaccines: live attenuated viruses, chimeric live attenuated viruses, inactivated or sub-unit vaccines, and nucleic acid-based vaccines (Halstead and Deen, 2002). However, efforts to develop a DENV vaccine have focused mainly on live attenuated virus vaccines, inactivated virus vaccine, and subunit virus vaccines (Table 1.1).

Table 1.1 Status of DENV vaccine developments

Vaccine	Developer	Phase
Live attenuated tetravalent vaccine (LAV[a])	WRAIR, GSK*	1/II
Intertypic chimeric vaccine (Δ30[a])	NIAID, NIH*	II
Chimeric vaccine (ChimeriVax[a])	Acambis, Sanofi-Aventis	II
Chimeric vaccine	CDC*	I
Flavivirus-based recombinant DNA vaccine	Navy Medical Research Center	I
Non-flavivirus based recombinant DNA vaccine (E (ecto) protein[b])	Hawaii Biotech	1

(Whitehead, S. S., Blaney, J. E., Durbin, A. P. & Murphy, B. R. 2007. Prospects for a dengue virus vaccine. *Nat. Rev. Microbiol.,* 5**,** 518-28. Table 1, page 525)

* Walter Reed Army Institute of Research (WRAIR); GlaxoSmithKline Biologicals (GSK); National Institute of Allergy and Infectious Diseases (NIAID); National Institutes of Health (NIH); Centers for Disease control and Prevention (CDC)

[a] LAV, Δ30 and ChimeriVax are being tested as tetravalent vaccines (targeting all four DENV serotypes simultaneously)

[b] Recombinant DENV1 E protein ectodomain (N-terminal 80%) formulated in alum

In all these instances, the vaccine viruses are monovalent in that each one is specific to one DENV serotypes. A tetravalent DENV vaccine is based on producing vaccine formulations by mixing all four monovalent vaccine viruses. Studies show that both the live attenuated and the ChimeriVax tetravalent DENV vaccine formulations elicit unbalanced immune response due to viral interference (Swaminathan and Khanna, 2010). The occurrence of this phenomenon, which tends to skew the immune response predominantly towards one serotype, emphasizes the limitations and the risks associated with mixing four monovalent vaccine viruses to create a tetravalent vaccine. However, tetravalent formulations and immunization

schedules are being optimized, so as to confer similar levels of protection against all four DENV serotypes (Ray and Shi, 2006).

In the absence of vaccines, drugs for specific therapy are needed, but no antiviral medications are approved for use against DENV. The proteins required for the fitness of the virus provide several potential targets against which to develop antiviral drugs. Strategies for DENV antiviral drug discovery include structure-based approaches, modulating the host immune response, high-throughput screening (HTS) using virus replication cell-based assays, or HTS specifically targeting viral morphogenesis, the 3' UTR, viral absorption, or assembly and maturation (Tomlinson *et al.*, 2009b).

A challenge for inhibitors discovered with virus replication cell-based HTS is determining the mechanism of inhibition. Testing natural products identified several lead compounds that inhibit DENV replication in cell-culture (Parida *et al.*, 2002; Kiat *et al.*, 2006; Jain *et al.*, 2008), with the components of fingerroot (*Boesenbergia rotunda*) reported to inhibit the DENV protease with a μM inhibition constant (Kiat *et al.*, 2006). A very general strategy utilizes compounds identified from other viral studies and tests them for inhibitory activity against DENV replication. There have been a few discoveries utilizing these various strategies, however, no leads have progressed to clinical trials.

In recent years, health authorities have emphasized disease prevention and mosquito control through community-based programs. Such programs are proper solid waste disposal, improved water storage practices, covering containers to prevent access by egg laying female mosquitoes, and the use of chemical and biological insecticides (WHO, 2009). These programs are very demanding in terms

of time, expertise, and financial resources. Therefore, there are only of limited usefulness for the control of DENV diseases.

1.4 Molecular biology of dengue virus

DENV is a vector borne member of the genus *Flavivirus* and the family *Flaviviridae* (Westaway *et al.*, 1985). The genus *Flavivirus* contains more than 70 members, including yellow fever virus (YFV), Japanese encephalitis virus (JEV), tick-borne encephalitis virus (TBEV), and West Nile virus (WNV). The complete nucleotide sequences of several flaviviruses have been reported and sequence comparison among the flavivirus polyproteins suggested that despite divergences in amino acid sequence, their hydrophobicity profiles are highly conserved, especially within the NS1, NS3, and NS5 proteins (Chambers *et al.*, 1990a).

DENV displays four antigenically related but distinct serotypes: DENV1, DENV2, DENV3, and DENV4. The four serotypes are almost indistinguishable in terms of clinical and pathological symptoms they cause, but they can be identified by neutralization tests, monoclonal antibodies, and polymerase chain reaction (PCR) (Morita *et al.*, 1991). These serotypes also vary in their degree of virulence and infection with one DENV serotype provides lifelong immunity to that virus, but there is no cross-protective immunity to the other serotypes. Each of the four serotypes can cause severe and fatal disease (Rigau-Perez *et al.*, 1998).

DENV is a smooth and spherical enveloped virus with a diameter of 500 Å. The virus contains a single-stranded positive-sense RNA genome of 10,723 nucleotides. The genome is enclosed in the viral capsid which is surrounded by a

host-derived lipid bilayer envelope (Kuhn *et al.*, 2002). The RNA genome has a type 1 cap at the 5' end, but is lacking a poly(A) tract at the 3' end.

The genome is organized into a single open reading frame (ORF) encoding a single polyprotein of 3,391 amino acids. The polyprotein precursor is processed into three structural proteins (C, prM, and E) and seven nonstructural proteins (NS1 to NS5) as illustrated in Figure 1.4. They are arranged in the order NH_2-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-COOH, where C is nucleocapsid, prM is precursor to membrane protein, E is envelope protein, and NS are the non-structural proteins (Westaway *et al.*, 1985; Chambers *et al.*, 1990a; Henchal and Putnak, 1990). These proteins are required for replication and assembly of new virions.

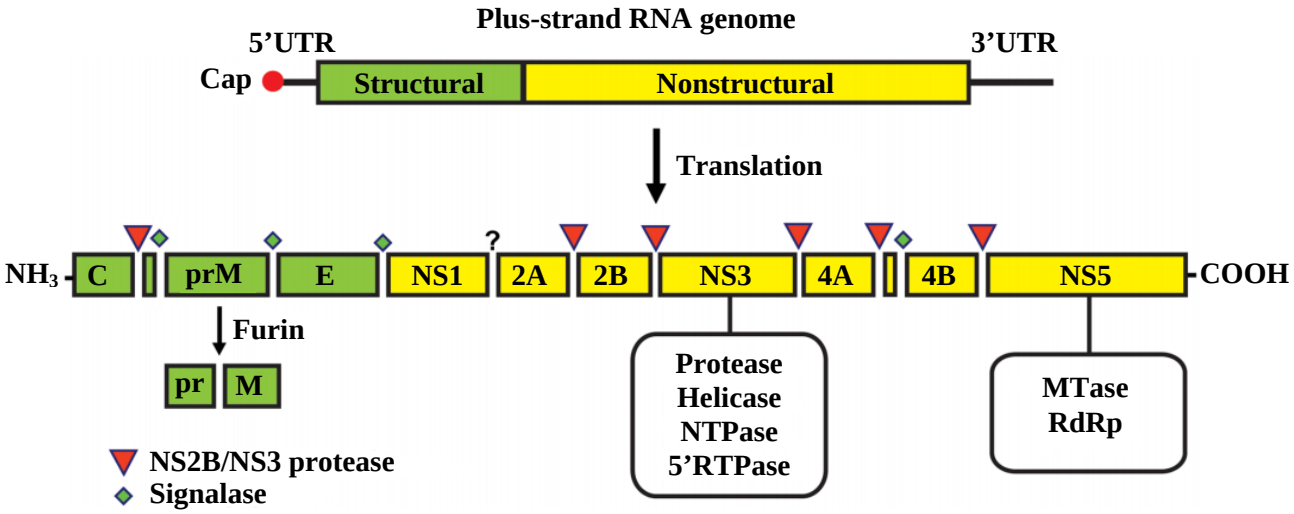


Figure 1.4 Schematic representations of DENV genome organization and polyprotein processing. The 11 kb positive-sense, single-stranded RNA genome contains a single open reading frame which encodes 3 structural proteins (capsid (C), precursor membrane (prM), and envelope (E)) and 7 non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5). The open reading frame is flanked by untranslated regions. Sites of polyprotein cleavage mediated by the viral NS2B-NS3 and by host signalase and furin are shown, and the enzymatic activities of NS3 and NS5 are also indicated (Sampath, A. & Padmanabhan, R. 2009. Molecular targets for flavivirus drug discovery. *Antiviral Res.*, 81, 6-15. Figure 1, page 20).

The C protein consists of ~120 amino acids and is involved with packaging of the viral genome and forming the nucleocapsid (NC) core. This protein is the first viral polyprotein synthesized during translation, has a molecular weight of about 13.5 kDa, and is rich in lysine and arginine residues. This highly basic character probably enables it to interact with the virion RNA (Henchal and Putnak, 1990).

prM (~165 amino acids) and E (~495 amino acids) are glycoproteins, each of which contains two transmembrane helices. Before it is cleaved during particle maturation to yield the pr peptide and the M protein (~75 amino acids), the prM protein might function as a chaperone for folding and assembly of the E protein.

The E protein contains a cellular receptor-binding site(s) and a fusion peptide (Mukhopadhyay *et al.*, 2005). E is associated with viral hemagglutination and neutralization activity, and interacts with cellular receptors to mediate viral attachment and entry (Crill and Roehrig, 2001).

NS1 (46 kDa) is required for flavivirus replication and is presumably involved in negative-strand synthesis by an unknown mechanism. A large deletion in YFV NS1 abolished viral replication but can be complemented in *trans* by functional expression from Sindbis virus vector (Lindenbach and Rice, 1997).

NS2A (22 kDa) is a small hydrophobic transmembrane protein that is involved in production of virus particles and in generation of virus-induced membranes during virus assembly (Leung *et al.*, 2008)

NS3 (70 kDa) and NS5 (104 kDa) are the best characterized nonstructural proteins, with multiple enzyme activities that are required for viral replication. NS3 has three distinct activities: serine protease together with the cofactor NS2B, required for polyprotein processing; helicase/NTPase activity, required for unwinding the double-stranded replicative form of RNA; and RNA triphosphatase, required for

capping nascent viral RNA (Falgout *et al.*, 1991; Zhang *et al.*, 1992; Li *et al.*, 1999). Mutations that affect each activity impair viral replication (Matusan *et al.*, 2001).

NS5 is the largest and most highly conserved flaviviral protein, with greater than 75% sequence identity across all DENV serotypes. It contains two distinct enzymatic activities, separated by an inter domain region: an S-adenosyl methyltransferase (SAM) (Egloff *et al.*, 2002) and an RNA-dependent RNA polymerase (RdRp) (Guyatt *et al.*, 2001). NS4A (16 kDa) is an integral membrane nprotein which may induce membrane rearrangements to form the viral replication complex. NS4B (27 kDa) inhibits the type I interferon response of host cells, and may modulate viral replication via its interaction with NS3 (Sampath and Padmanabhan, 2009).

DENV replicates in the cytoplasm of susceptible host cells, including monocytes, macrophages, and dendritic cells. A specific receptor for internalization of DENV into the host cell has not yet been identified. Several cellular molecules capable of mediating virus attachment are known, but none have been conclusively shown to function as virus receptors (Tassaneetrithep *et al.*, 2003; Lozach *et al.*, 2005; Krishnan *et al.*, 2007; Miller *et al.*, 2008).

As illustrated in Figure 1.5, during virus entry, E proteins forming the glycoprotein shell bind to cell surface receptors that assist in the internalization of the virus through clathrin-mediated endocytosis. Following internalization, the acidic environment of the endosome triggers an irreversible trimerization of the E protein that results in fusion of the viral and cell membrane. This leads to the release of the viral RNA into the cytoplasm (Mukhopadhyay *et al.*, 2005; Krishnan *et al.*, 2007; van der Schaar *et al.*, 2007).

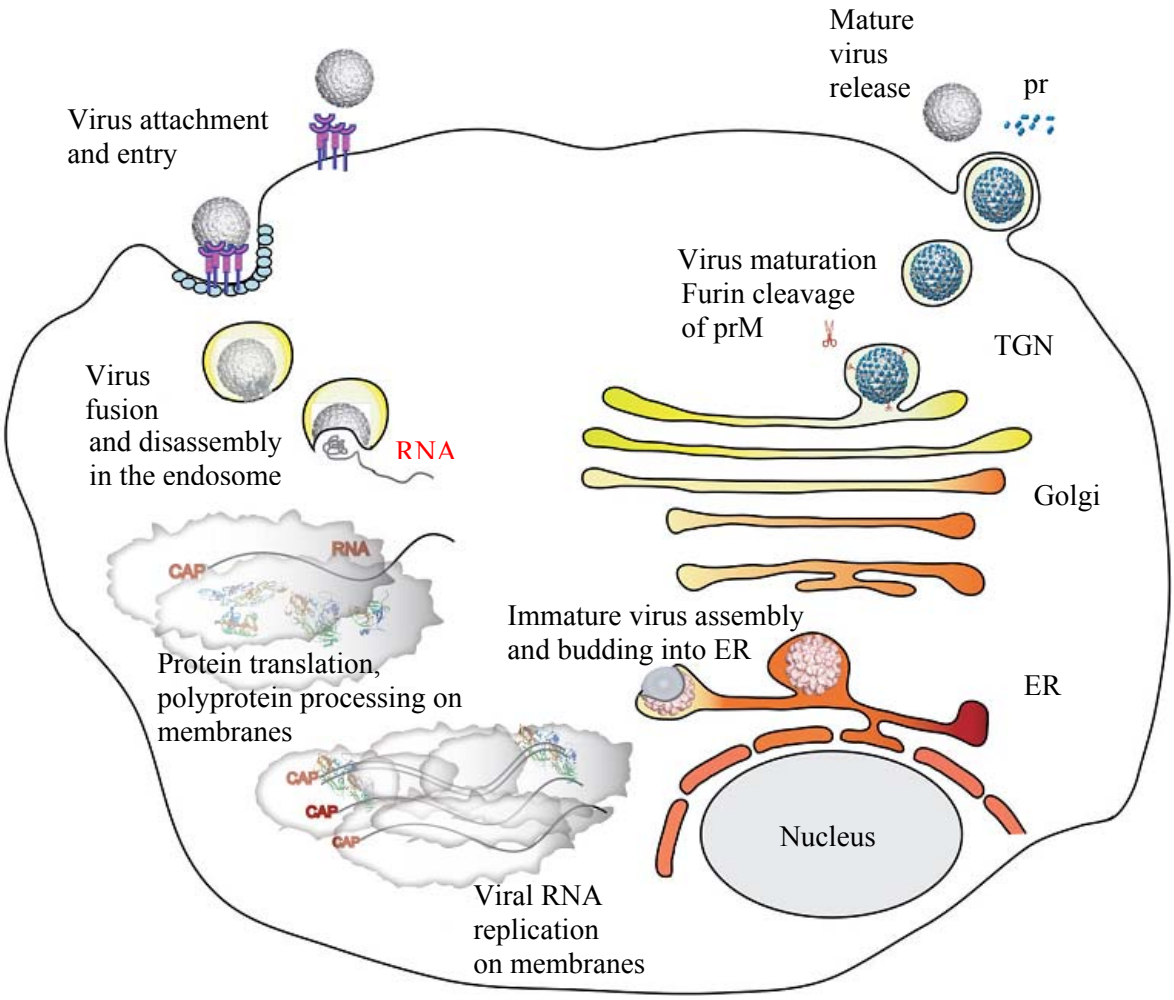


Figure 1.5 The DENV replication cycle. Virions bind to cell-surface attachment molecules/ receptors and are internalized through endocytosis. In the low pH of the endosome, viral glycoproteins mediate fusion of the viral and cellular membranes, allowing disassembly of the virion and release of its RNA into the cytoplasm. The viral RNA is translated into a polyprotein that is processed by viral and cellular proteases. Viral non-structural proteins then replicate the genomic RNA. Virion assembly occurs at the ER membrane. Capsid protein and viral RNA are enveloped by the membrane and its embedded glycoproteins to form immature virus particles, which are then transported through the secretory pathway. In the low pH of the trans-Golgi network (TGN), prM is cleaved by furin. Mature virions are then released into the cytoplasm (Sampath, A. & Padmanabhan, R. 2009. Molecular targets for flavivirus drug discovery. *Antiviral Res.,* 81, 6-15. Figure 2, page 21).

The viral RNA is directly translated into a single polyprotein by the host's translational machinery. The processing of the polyprotein precursor occurs both cotranslationally and post-translationally by host cell and virus-encoded proteases.

Host cell signalase located in the luminal side of the endoplasmic reticulum (ER) is responsible for the cleavages at the C-prM, prM-E, E-NS1, and NS4A-NS4B junctions (Chambers *et al.*, 1990a; Henchal and Putnak, 1990). Previous work suggest that NS1-NS2A cleavage occurs in the ER and NS2A is required to permit a host ER-resident protease, possibly signalase to effect cleavage (Falgout and Markoff, 1995).

The virus-encoded trypsin-like serine protease, a complex of NS2B and NS3, cleaves at a number of sites including the NS2A-NS2B, NS2B-NS3, NS3-NS4A, and NA4B-NS5 junctions (Preugschat *et al.*, 1990) (Figure 1.4). In addition, it is also responsible for the cleavage within the viral protein C, NS4A, and within NS3 itself (Teo and Wright, 1997). The viral RNA replication is catalyzed by a replication complex which is composed of NS5, the RNA-dependent RNA polymerase, and other viral and host factors in the rough ER and in Golgi-derived membranes called vesicle packets (VP) (Mackenzie, 2005).

Newly synthesized RNA encapsulated by C protein is then enveloped by glycoproteins prM and E to assemble immature virus particles that bud into the ER. These immature particles are transported through the secretory pathway to the Golgi apparatus. In the low pH environment of the trans-Golgi, furin-mediated cleavage of prM to M drives maturation of the virus. prM processing destabilizes the prM-E interaction and promotes the formation of E homodimers present in mature infectious virions. Finally, progeny virus particles are released from the cell by exocytosis (Henchal and Putnak, 1990; Perera *et al.*, 2008; Sampath and Padmanabhan, 2009).

1.5 NS2B-NS3: The two-component protease of dengue virus

The NS2B-NS3 protease is a two-component protease. This heterodimeric complex of NS2B and NS3 is responsible for cleavage of the newly translated DENV polyprotein at the NS2A-NS2B, NS2B-NS3, NS3-NS4A, and NS4B-NS5 sites as well as internal sites within the viral protein C, NS2A, NS3, and NS4A (Chambers *et al.*, 1990a; Chambers *et al.*, 1990b; Preugschat *et al.*, 1990; Falgout *et al.*, 1991; Zhang *et al.*, 1992; Clum *et al.*, 1997; Yusof *et al.*, 2000; Bera *et al.*, 2007).

NS3 is a multifunctional protein as illustrated in Figure 1.6. The virus encoded protease lies within the N-terminal 180 amino acid residues of the 618 residue protein. The C-terminal region comprises the RNA-stimulated nucleoside triphosphatase (NTPase) and RNA helicase activities. The functional domains of the protease and NTPase overlap within a region of 20 amino acid residues (residues 160 to 180) (Li *et al.*, 1999).

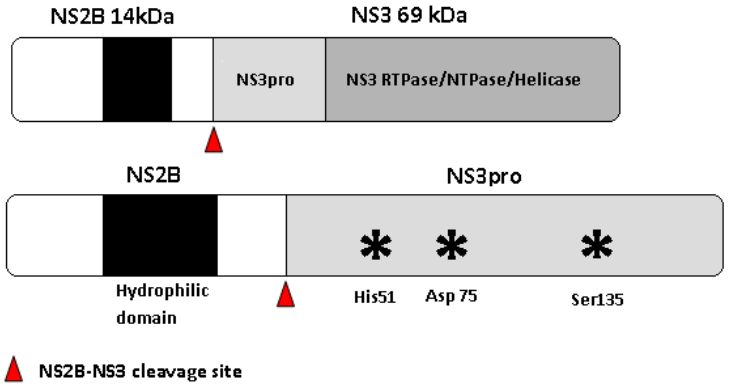


Figure 1.6 Structure and organization of the DENV NS2B-NS3 protease. (A) Scheme for viral enzymes, showing the cofactor domain of NS2B in black and the NS3pro in gray with the catalytic triad: His51, Asp75, andSer135. (Nall, T. A., Chappell, K. J., Stoermer, M. J., Fang, N. X., Tyndall, J. D., Young, P. R. & Fairlie, D. P. 2004. Enzymatic characterization and homology model of a catalytically active recombinant West Nile virus NS3 protease. *J. Biol. Chem.,* 279**,** 48535-42. Figure 1, page 48536).

The serine protease domain of NS3 was identified based on sequence homology to known serine proteases (Bazan and Fletterick, 1989). Four separate regions of significant conservation were identified (boxes 1, 2, 3, and 4). Boxes 1, 2, and 3 encompass the catalytic triad (His51, Asp75, and Ser135) and boxes 3 and 4 contain residues that are involved in substrate binding and recognition (Figure 1.7).

Subsequent biochemical studies confirmed the protease activity within the N-terminal 180 amino acid residues of NS3 (Chambers *et al.*, 1990b; Preugschat *et al.*, 1990). Site-directed mutagenesis experiments performed with YFV showed that replacement of the putative catalytic triad residues abolished protease activity *in vitro*, and when the changes were incorporated into the infectious full-length cDNA clone, virus was not recovered (Chambers *et al.*, 1990b).

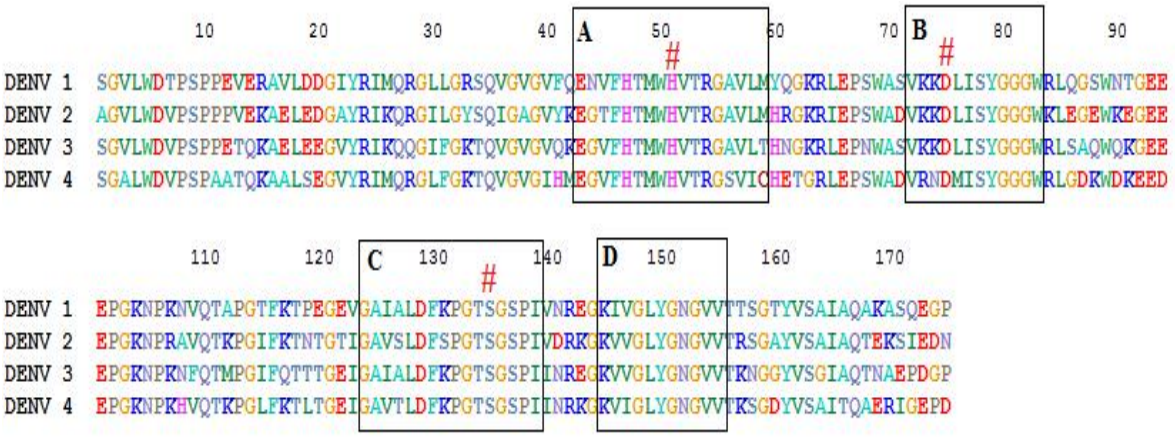


Figure 1.7 Sequences of DENV NS3 serine protease domain. Multiple sequence alignment of DENV protease domain from 4 serotypes. The catalytic triad residues: His51, Asp75, and Ser135, are labeled with the symbol # are found in boxes labeled A, B, and C. The boxes labeled A, B, C, and D identifies regions of significant similarity surrounding the catalytic triad residues and residues that might form the substrate-binding pocket. (Aleshin, A. E., Shiryaev, S. A., Strongin, A. Y. & Liddington, R. C. 2007. Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. *Protein Sci.,* 16**,** 795-806. Figure 1, page 796).

NS2B is an ER-resident integral membrane protein. The protein contains 130 amino acid residues with the molecular mass of 14 kDa. The hydrophobicity plot of NS2B shows that NS2B contains a central hydrophilic domain flanked by two hydrophobic domains at the N-terminus (I and II) and C-terminus (III and IV) (Figure 1.8) (Clum *et al.*, 1997). The hydrophobic sequences are essential for co-translational insertion of the protease cofactor into ER membranes for efficient cleavage of the NS2B/NS3 junction (Clum *et al.*, 1997) but they are dispensable for protease activity. The central hydrophilic region contains 40 amino acids which are conserved among flaviviruses. The individual NS3pro domain, lacking the NS2B part, is catalytically inert (Murthy *et al.*, 1999; Murthy *et al.*, 2000).

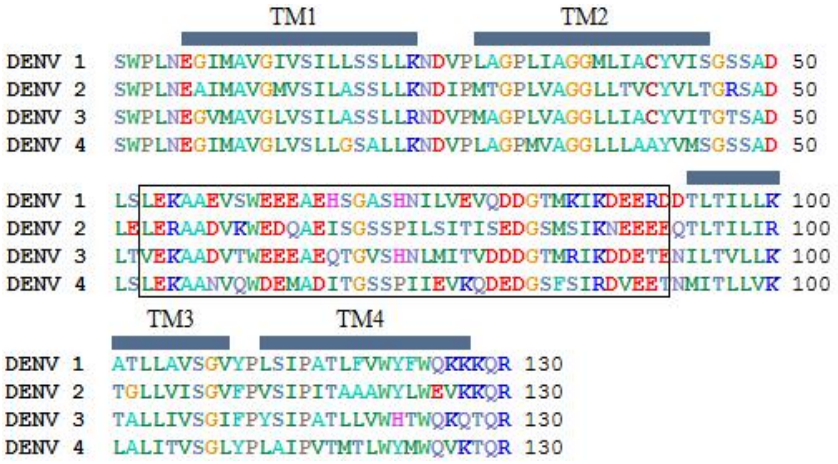


Figure 1.8 Sequences of DENV NS2B. The box indicates the minimal cofactor segment required for activation of the NS3 protease *in vitro*. TM1-TM4 are predicted transmembrane regions (Aleshin, A. E., Shiryaev, S. A., Strongin, A. Y. & Liddington, R. C. 2007. Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. *Protein Sci.,* 16**,** 795-806. Figure 1, page 796).

Results from co-expression studies showed that the proteolytic activity of the NS3 protease was critically dependent upon the presence of its cofactor, NS2B protein (Falgout *et al.*, 1991). Truncation studies in DENV2 showed that the central 40 amino acid hydrophilic domain is sufficient for protease activity (Chambers *et al.*,

1993; Falgout *et al.*, 1993; Clum *et al.*, 1997; Niyomrattanakit *et al.*, 2004). The presence of NS2B resulted in a several thousand-fold activation of the NS3 protease towards dibasic peptide substrates (Yusof *et al.*, 2000). The flanking hydrophobic domains within NS2B are likely to function in promoting membrane association of NS2B-NS3 (Clum *et al.*, 1997).

The kinetic parameters and substrate specificity of DENV protease were reported (Yusof *et al.*, 2000; Leung *et al.*, 2001; Khumthong *et al.*, 2002; Chanprapaph *et al.*, 2005; Shiryaev *et al.*, 2007a; Iempridee *et al.*, 2008). The precursor devoid of the hydrophobic regions but containing the conserved NS2B hydrophilic domain linked to the NS3 protease domain through a carboxy terminal region of NS2B containing the NS2B-NS3 cleavage site was expressed in *E.coli* (Yusof *et al.*, 2000). The precursor, expressed as insoluble inclusion bodies, was purified by denaturation and refolding.

The expression of soluble and active protease was achieved when the hydrophilic portion of the NS2B viral cofactor spanning residues 49-95 (hereafter named CF40) of either WNV or DENV2 was fused to residues 1-169 of the NS3 protein via a flexible (Gly_4-Ser-Gly_4) linker, thus obviating the denaturation and refolding steps in the purification of the protease (Leung *et al.*, 2001).

A number of *in vitro* assays for the viral proteases have been described in several studies (Clum *et al.*, 1997; Yusof *et al.*, 2000; Leung *et al.*, 2001; Walker and Lynas, 2001; Khumthong *et al.*, 2002; Tong, 2002). Either virus-encoded polyprotein or synthetic peptides have been utilized as the substrates. Important information on the regulation and requirements for the viral polyprotein processing were obtained from the assay with virus-encoded polyprotein. The assay with the synthetic peptides

would provide the information on the substrate specificity of the enzymes and were used in the inhibitor screening.

The viral protease has a preference for two basic amino acid residues (Arg-Arg, Arg-Lys, Lys-Arg, or occasionally Gln-Arg) at the P2 and P1 positions preceding the cleavage sites, followed by Gly, Ala, or Ser at the P1' position. The earliest report for the DENV protease *in vitro* assay had used commercially available fluorogenic peptides as the substrates.

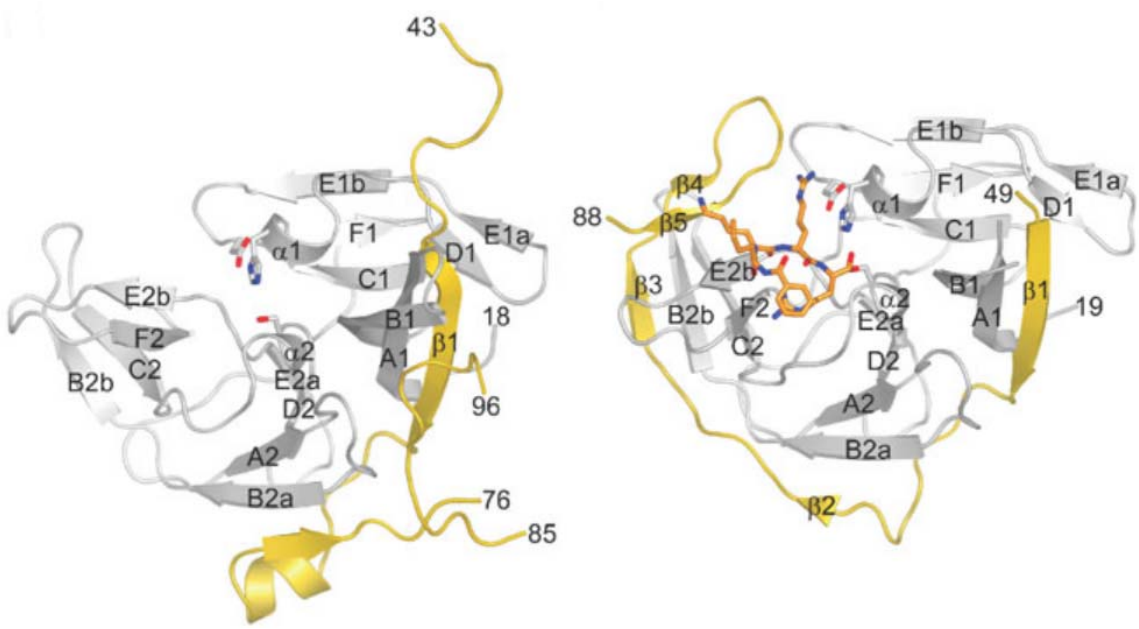
All of these peptides contain two basic amino acid residues (Arg-Arg, Arg-Lys, Lys-Arg) at the P1 and P2 positions preceding the cleavage site. None of the peptides contain an amino acid residue at the P1' position, but rather the P1' residue is replaced by a fluorogenic moiety. Their result revealed that the substrate Gly-Arg-Arg-MCA, which contains a Gly residue at the P3 position, is the most active of the four substrates tested (Yusof *et al.*, 2000).

Li *et al*. (2005) cloned and expressed the protease from all four DENV serotypes (DENV1-4 CF40-Gly_4-Ser-Gly_4-NS3pro) and adapted the *in vitro* assay described by Yusof *et al.* (2000) to screen tetrapeptide and octapeptide libraries comprising ~13,000 substrates.

The tetrapeptide benzoyl-norleucine (P4)-lysine (P3)-arginine (P2)-arginine (P1)-ACMC (Bz-Nle-Lys-Arg-Arg-ACMC) was identified as the optimal substrate with the steady state kinetics parameter k_{cat}/K_m of 51,800 $M^{-1}s^{-1}$ which is >150-fold more sensitive than other published peptides. The sensitivity enabled miniaturization of the assay for high-throughput screening (Keller *et al.*, 2006). Moreover, this ideal tetrapeptide sequence formed the basis for the peptidomimetic approach for finding potent substrate-based inhibitors (Yin *et al.*, 2006a; Yin *et al.*, 2006b).

1.6 Structure of DENV NS3 serine protease

The high-quality crystal structure of active DENV NS2B-NS3 protease (1.5 Å; PDB identifier 2FOM) and WNV NS2B-NS3 protease in the complex with the substrate-based inhibitor Bz-Nle-Lys-Arg-Arg-H (1.68 Å; PDB identifier 2FP7) were resolved (D'Arcy *et al.*, 2006; Erbel *et al.*, 2006). The NS3 protease domains in both structures adopt chymotrypsin-like serine protease folds with two β-barrels, each formed by six β-strands, and the catalytic triad (His51, Asp75, and Ser135) located at the cleft between the two β-barrels (Figure 1.9). Sharing 50% sequence identity, the two NS2B-NS3 protease structures have close structural similarity.



DENV2 NS2B-NS3pro **WNV NS2B-NS3pro**

Figure 1.9 Schematic representation of the structures NS2B-NS3 protease. (Left) The apo-enzyme from DENV2 with the NS2B cofactor in yellow and the catalytic triad represented as sticks. (Right) Complex of WNV NS2B-NS3 protease with the KKRR tetrapeptide. The N-terminal part of the NS2B cofactor is sufficient to stabilize the enzyme (Erbel, P., Schiering, N., D'arcy, A., Renatus, M., Kroemer, M., Lim, S. P., Yin, Z., Keller, T. H., Vasudevan, S. G. & Hommel, U. 2006. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat. Struct. Mol. Biol.,* 13**,** 372-3. Figure 1, page 372).

The structures of WNV and DENV NS2B-NS3 proteases reveal residues of NS2B that are important for the stabilization of the NS3 protease fold. Similarly to the HCV NS4A-NS3 protease, the N-terminal part of the cofactor contributes one β-strand (β-strand 1, NS2B residues 51-57 in DENV) to the N-terminal β-barrel of the protease, which conceals hydrophobic residues from the solvent and provides stabilization to this domain.

This explains the observed strong tendency for NS3 protease and full-length NS3 to aggregate when the strand contributed by NS2B is absent in synthetic constructs. In this respect the N-terminal of NS2B has a chaperone-like role in stabilizing NS3.

On the other hand, the fold adopted by the C-terminal part of the NS2B cofactor shows marked differences between the unliganded DENV NS2B-NS3pro and inhibitor-bound WNV NS2B-NS3pro complexes. In the inhibitor-bound protease complex, a large rearrangement brings residues 67-88 of NS2B in close proximity to the substrate-like inhibitor, forming a belt around the NS3 protease. Residues Arg78-Leu87 of the NS2B cofactor forms a β-loop which interacts with the N-terminal barrel of the NS3 protease, affecting the formation of the active site and substrate recognition.

The contribution of the NS2B cofactor to stabilize both the N- and C-terminal barrels and complete the substrate-binding site is indeed unique to flaviviruses. It differs substantially from those observed with other cofactor-activated viral proteases such as HCV NS4A-NS3pro which requires a short fragment of NS4A to form the active enzyme (Erbel *et al.*, 2006; Lescar *et al.*, 2008).

The unprecedented way in which the NS2B cofactor region forms a belt around the protease domain was confirmed in a second structure that was reported

for the WNV enzyme as a complex with the aprotinin/BPTI inhibitor (Aleshin *et al.*, 2007). The aprotinin occupied all the specificity pockets of the protease and induced a fully formed oxyanion hole, which allowed Aleshin *et al*. (2007) to provide a complete view of the enzyme substrate Michaelis complex for a flavivirus protease. These structures open up new opportunities for discovering flavivirus-specific drugs that could function by interfering with protein-protein interactions that are needed for the activation of the protease in addition to active site directed competitive inhibitors.

The DENV NS3 protease structure in the absence of the NS2B cofactor deviates substantially from DENV NS2B-NS3 protease structures. These differences are observed throughout the entire enzyme and affect the length and location of secondary structure elements (Erbel *et al.*, 2006). Although the protease catalytic site residues (His51, Asp75, and Ser135) were arranged similarly in the NS3 and NS2B-NS3 protease crystal structures, numerous large conformational differences were evident.

For instance, overlaying the catalytic regions of the two structures resulted in position differences of 14Å and 35 Å for Leu31 and Asn119, respectively (Figure 1.10). Of relevance to DENV protease inhibitor design are the large differences in the substrate-binding region between the two structures. The S1 site within the NS3 substrate-binding region formed a deep pocket that could accommodate long positively charged P1 side chains of the substrate. However, in the NS2B-NS3 protease structure, the S1 site forms only a shallow depression. Structure-based drug discovery approaches must consider the differences between the NS3 and NS2B-NS3 structures since small molecules may interact differently with the active sites of NS3 and NS2B-NS3 (Tomlinson *et al.*, 2009b).

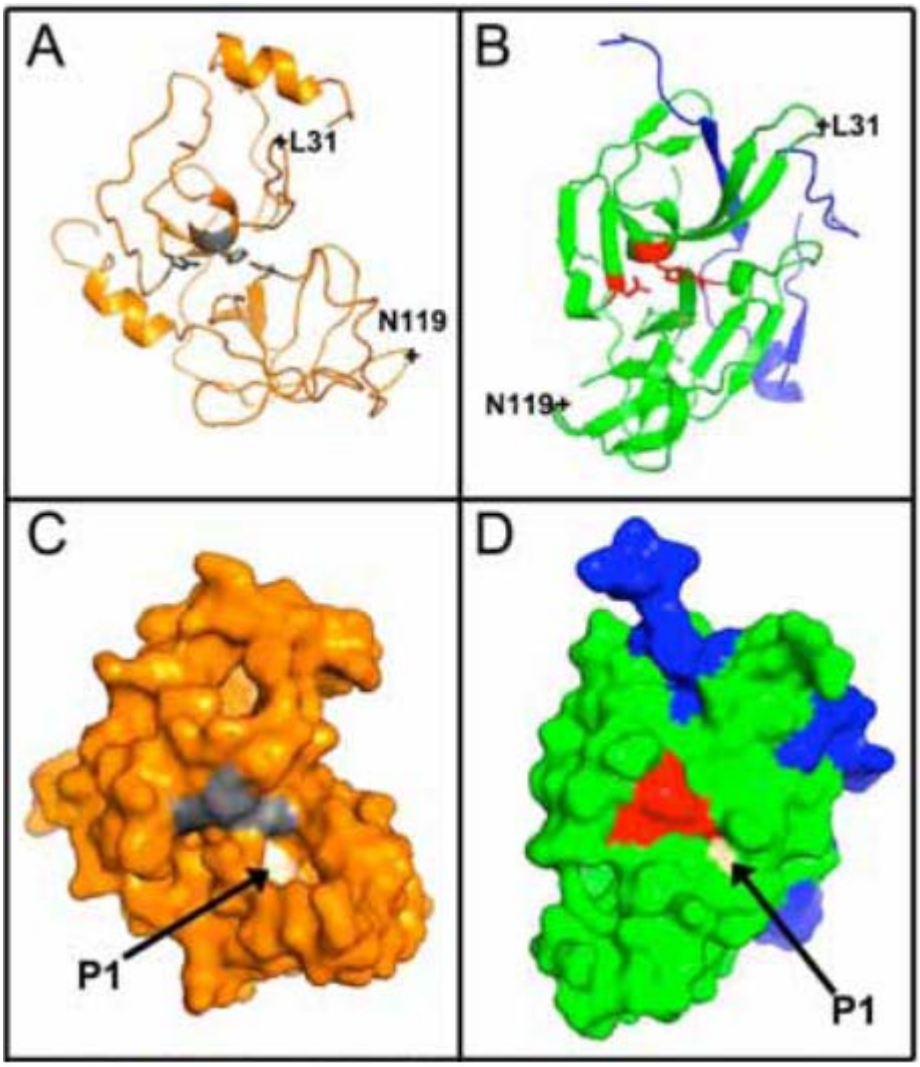


Figure 1.10 Comparison of the DENV NS3 protease (PDB identifier 1BEF) and the DENV NS2B-NS3 protease (PDB identifier 2FOM). Ribbon diagram representation of (A) NS3 and (B) NS2B-NS3 proteins. The conformational changes between the two structures shifts leucine 31 (L31) and asparagine 119 (N119) by 14 Å and 35 Å, respectively. Molecular surface diagrams for (C) NS3 and (D) NS2B-NS3. Arrows point to the substrate binding pocket in the two structures (Tomlinson, S. M., Malmstrom, R. D. & Watowich, S. J. 2009b. New approaches to structure-based discovery of dengue protease inhibitors. *Infect. Disord. Drug Targets,* 9**,** 327-43. Figure 6, page 335).

1.7 Serine protease

Proteases have long been recognized as attractive targets in the drug discovery processes. Serine proteases are the most widely studied group of proteins in biology (Walker and Lynas, 2001). The important role of serine proteases has been elucidated in the pathology of viral infections (Tong, 2002). Many crystal structures and their complexes with either substrates or inhibitors have been resolved (Kim *et al.*, 1996; Erbel *et al.*, 2006; Aleshin *et al.*, 2007).

In humans, serine proteases are involved in many important physiological processes such as inflammation, fibrinolysis, immune response, digestion, blood coagulation, and fertilization. Hence, it is critical that whenever an inhibitor of viral serine proteases is utilized for the treatment of disease in humans, such inhibitor must have a high selectively for the viral protease to minimize the risk of any adverse effects.

Proteases are enzymes that selectively catalyze the hydrolysis of peptide bonds. Proteases are classified into five major classes based on their mechanism of action. These classes are serine proteases, cysteine proteases, aspartic proteases, threonine proteases, and metallo proteases. The classification is made owing to the critical residues used in catalysis.

Serine proteases are characterized chiefly by the presence of an active site serine (Ser) residue, the γ hydroxyl group of which acts as a nucleophile during the hydrolytic process. Two other amino acid residues that are directly involved in the catalytic mechanism are histidine (His) and aspartate (Asp) that together form the catalytic triad (Walker and Lynas, 2001; Hedstrom, 2002). In addition, the enzyme possesses an oxyanion binding site that is made from the backbone amide NH groups