LIGAND BASED DRUG DISCOVERY OF NOVEL DENGUE-2 NS2B-NS3 PROTEASE INHIBITORS

MOHAMED SUFIAN BIN MOHD NAWI

UNIVERSITI SAINS MALAYSIA

2015

LIGAND BASED DRUG DISCOVERY OF NOVEL DENGUE-2 NS2B-NS3 PROTEASE INHIBITORS

by

MOHAMED SUFIAN BIN MOHD NAWI

Thesis submitted in fulfilment of the requirements

for the degree of

Doctor of Philosophy

August 2015

ACKNOWLEDGEMENT

All praise is due to Allah S.W.T because of His bounty I can complete this research towards fulfilling the requirements for my degree of Doctor of Philosophy.

Foremost, I would like to express my sincere gratitude to my supervisor Professor Habibah A. Wahab for the continuous support of my PhD study and research, for her patience, motivation, enthusiasm and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis.

My utmost gratitude to both my field supervisors, Professor Noorsaadah Abd. Rahman and Associate Professor Shafida Abdul Hamid. Their scientific advice, knowledge, many insightful discussions and suggestions make my way to complete this thesis. Special thanks to Prof Mutasem for teaching me insightful knowledge on preparing pharmacopore model and Professor Rohana Yusof for providing space, instrumentation and general lab equipment together with helpful suggestions for my enzyme preparation.

I thank my fellow lab mates in PhDs lab: Choi Sy Bing, Yam Wai Kit, Maywan, Yusuf, Razip, Belal, Muchtaridi, Rosfatihah, Saera, Lee, Lim, Ban Hong, Nadia and Azizairul for the endless helps, stimulating discussions, encouragement, insightful comments and for all the fun we have had in the last four and half years.

My sincere thanks also go to the lab members of Prof. Rohana Yusof Research Laboratory, Department of Molecular Medicine, UM: Dr. Shatrah, Dr. Heh Choon

Han, Tan, Christina, Hani, Mimi, Amar and Dr. Husin for offering me the opportunities to work in their group and helping me on enzyme expression, purification and *in-vitro* works. I am grateful to Dr. Narendra Babu from USM, Dr. Anjan Patail from Kulliyyah of Science IIUM, Dr Lee Yen Kee from Chemistry Department UM and Maywan from PhDs lab for helping me with the synthesis works.

I also thank my friends and supporting staff at the School of Pharmaceutical Sciences USM; Department of Molecular Medicine and Department of Chemistry UM; as well as Kulliyyah of Pharmacy and Kulliyyah of Science of IIUM.

To IIUM scholarship and study leave committee, Ministry of Higher Education and Public Service Department of Malaysia, thank you for the financial support that was awarded to me throughout my PhD studies. I acknowledge Ministry of Science, Technology and Innovations (MOSTI) for the funding of the research.

I would like to express my utmost gratitude to my wife, Siti Hadijah Shamsudin and my children: Afifah, Amirul and Amirah for their moral support that gave me the confidence, will and strength to endure the pressure and tension in pursuing my ambition and to them I dedicate this thesis.

Finally but not least, thank you to my mother, my late father and mother-in-law for their prayers, understanding and inspiration throughout my study. May Allah give His blessing in their lives in this world and in the hereafter.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

(p-Me)Phe	:	Para methyl phenyl
µL	:	Microliter
µM	:	Micromolar
µm	:	Micrometer
2D	:	Two-dimensional
3D	:	Three-dimensional
3D-QSAR	:	Three dimension quantitative structure activity relationship
AMC	:	7-Amino-4-methylcoumarin
APS	:	Ammonium persulphate
BSA	:	Bovine serum albumin
C	:	Capsid
CBB	:	Coomasie Brilliant blue
CBB R-250	:	Coomasie Brilliant blue R-250
CFR	:	Case fatality rate
CHARMm	:	Chemistry at HARvard Macromolecular Mechanics
COSY	:	Correlation spectroscopy
DEN-1	:	Dengue Virus Type 1
DEN-2	:	Dengue Virus Type 2
DEN-3	:	Dengue Virus Type 3
DEN-4	:	Dengue Virus Type 4
DENV	:	Dengue virus
DHF	:	Dengue haemorrhagic fever
DMSO	:	Dimethyl sulfoxide
DS	:	Discovery Studio
DSS	:	Dengue shock syndrome
E	:	Envelope
EI	:	Enzyme-inhibitor complex
ER	:	Endoplasmic reticulum
FV	:	Fit value
g	:	Gram
GFA	:	Genetic function algorithm

HBA	:	Hydrogen bond acceptor
HBD	:	Hydrogen bond donor
Hbic	:	Hydrophobic
HCl	:	Hydrochloric acid
HEPES	:	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	:	Heteronuclear single quantum correlation
IC_{50}	:	Half maximal inhibitory concentration
IPTG	:	Isopropyl–β–D–thiogalactopyranose
JEV	:	Japanese Encephalitis Virus
kDA	:	Kilo Dalton
K_i	:	Dissociation constant for the *EI* complex
K_m	:	Michaelis constant
LB	:	Lysogeny broth
M	:	Membrane
m	:	Meta
mM	:	Millimolar
mm	:	Millimeter
mmol	:	Millimole
MAct	:	Most Active Compound
MD	:	Molecular dinamics
MLR	:	Multiple linear regression
mRNA	:	Messenger ribonucleic acid
MS	:	Mass spectroscopy
NaCl	:	Sodium chloride
NADI	:	Natural Product Discovery System
NC	:	Nucleocapsid core
NCI	:	National Cancer Institute, USA
Ni^{2+}-NTA	:	Nickel-nitrilotriacetic acid
nm	:	Nanometer
NOESY	:	Nuclear Overhause effect spectroscopy
NS2B	:	Non-Structural Protein 2B
NS3	:	Non-Structural Protein 3
NS5	:	Non-Structural Protein 5

NTPase	:	RNA nucleoside triphosphatases
o	:	Orto
OD	:	Optical density
ORF	:	Open reading frame
p	:	Para
PDB	:	Protein Data Bank
prM	:	Pre membrane protein
prM-E	:	Pre membrane envelope
QTOF	:	Quadrupole time-of-flight
RNA	:	Ribonucleic acid
rpm	:	Round per minute
RTPase	:	RNA 5'Triphosphatase
SAR	:	Structure activity relationship
SDS	:	Sodium dodecyl sulphate
SDS-PAGE	:	Sodium dodecyl sulfate polyacrylmide gel electrophoresis
TBEV	:	Tick-borne encephalitis virus
TEMED	:	N,N,N,M'-tetramethylethyldiamine
TLC	:	Thin layer chromatography
TMS	:	Tetramethylsilane
Tris	:	Tris(hydroxymethyl)aminomethane
Tris-HCl	:	Tris(hydroxymethyl)aminomethane – HCl buffer
Trp	:	Tryptophan
Tyr	:	Tyrosine
UTR	:	Untranslated region
V_{max}	:	Maximum velocity
WHO	:	World Health Organisation
WNV	:	West Nile virus
YFV	:	Yellow fever virus
α	:	Alpha
π-π	:	Pi-pi
π-cation	:	Pi-cation
π-σ	:	Pi-sigma

LIST OF PUBLICATIONS

Mohamed Sufian M. Nawi, Noorsaadah Abd. Rahman, Shafida Abd. Hamid and Habibah A. Wahab (2010). Pharmacophore Modeling and in Silico Screening of NADI Database for the Discovery of Dengue Virus NS2B-NS3 Protease Inhibitors. *International Seminar and Expo on Jamu 2010*. Universitas Padjadjaran, Bandung, Indonesia, 5-6th November 2010.

Habibah A Wahab, Suriyati Muhamad, **Mohamed Sufian Mohd Nawi**, Mohd Razip Asarudin, Maywan Hariyono and Ezatul Ezleen Kamarulzaman (2013). Nature, the Source of Lead Compounds for Drug Discovery. *International Conference on Medicinal Chemistry and Timmerman Award 2013*. Universitas Indonesia, Jakarta, Indonesia, 28-29th October 2013.

Habibah A Wahab, Suriyati Muhamad, **Mohamed Sufian Mohd Nawi**, Nur Kusaira Ikram, Maywan Hariyono and Ezatul Ezleen Kamarulzaman (2013). Interoging Nature for Lead Compounds for Drug Discovery Using Computational Techniques. *The 5th Seminar and Workshop on Computer-Aided Drug Design*. Bandung, Indonesia, 9-13th December 2013.

Muhammad Asyraf Abduraman, **Mohamed Sufian Mohd Nawi**, Habibah A. Wahab, Rohana Yusof and Tan Mei Lan (2013). The Development of Dengue Virus NS2B/NS3 Protease Inhibition Assay for Lead Compound Optimization In Drug Discovery Process. *International Conference on Medical and Health Science*. Kota Bharu, 22-24th May 2013.

Muhammad Asyraf Abduraman, **Mohamed Sufian Mohd Nawi**, Habibah A. Wahab, Rohana Yusof and Tan Mei Lan (2013). The Development of Specific NS2B/NS3 Protease Inhibition Assay Using Alphascreen^{TM} Technology. *National Biotechnology Seminar*. Penang, Malaysia, 6-8th June 2013.

PENEMUAN UBAT BERASASKAN LIGAN UNTUK PERENCAT NOVEL PROTEASE NS2B-NS3 DENGGI-2

ABSTRAK

Kes demam denggi dilaporkan meningkat setiap tahun, namun sehingga kini belum ada ubat anti-denggi boleh didapati di pasaran. Oleh itu, pencarian ejen anti-denggi adalah kritikal. Pemotongan poliprotein pelopor oleh enzim protease NS2B-NS3 merupakan proses yang sangat penting untuk replikasi flavivirus. Oleh itu enzim protease NS2B-NS3 sesuai dijadikan sasaran untuk membangunkan ubat anti-denggi. Di Malaysia, Denggi-2 adalah jenis sera yang paling lazim. Dalam kajian ini, pendekatan berasaskan ligan telah dilaksanakan dalam mencari perencat enzim protease NS2B-NS3 Denggi-2 baharu yang berpotensi. Model farmakofor telah dihasilkan daripada pelbagai struktur perencat enzim protease NS2B-NS3 Denggi-2 yang pernah dilaporkan, yang terdiri daripada molekul peptida dan bukan peptida. Model farmakofor terpilih telah digunakan untuk menyaring senarai sebatian dalam pengkalan data *National Cancer Institute, AS* (NCI) untuk mencari perencat protease NS2B-NS3 Denggi-2 yang baharu secara maya. Produk semulajadi dalam pengkalan data Natural Product Discovery System (NADI) juga disaring dan struktur yang aktif secara maya telah dieksploitasi untuk merekabentuk molekul perencat baru, yang kemudiannya disintesis. Aktiviti-aktiviti perencatan sebatian yang aktif secara maya dari pengkalan data NCI dan yang disintesis telah diuji secara biocerakinan *in-vitro* terhadap enzim protease NS2B-NS3 Denggi-2 menggunakan substrat peptida Boc-Gly-Arg-Arg-MCA. Kewajaran pemilihan model farmakofor S5T5H06 dan S2T3H01 telah disahkan secara ujikaji dengan penemuan 3 perencat enzim protease NS2B-NS3 Denggi-2 baharu dari pengkalan data NCI dengan nilai-nilai K_i 17 ± 6

μM (perencat tidak kompetitif), 26 ± 4 μM (perencat kompetitif) dan 95 ± 35 μM (perencat kompetitif). Disamping itu, 1 perencat kompetitif (K_i = 78 ± 41 μM) dan 2 perencat tidak kompetitif (K_i = 29 ± 6 μM dan K_i = 97 ± 49 μM) telah diperolehi dari molekul yang disintesis berasaskan struktur sebatian daripada pengkalan data NADI. Penemuan menarik ini mendedahkan bahawa produk semulajadi dari kepelbagaian-bio Malaysia yang terdapat dalam pengkalan data NADI adalah sumber yang berharga untuk pembangunan ejen anti-denggi. Daripada keputusan ujikaji ini juga, satu andaian dapat dibuat iaitu ligan yang mempunyai farmakofor seperti model S5T5H06 boleh terikat kepada tapak ikatan aktif atau tapak ikatan alosterik dengan mod ikatan yang sama. Daripada kajian "SAR", satu struktur pemula yang berpotensi untuk membangunkan perencat tidak kompetitif terhadap protease NS2B-NS3 Denggi-2 telah dicadangkan. Struktur ini mempunyai nukleus pusat yang terdiri daripada di-fenilmetana. Nukleus pusat ini dihubungkan oleh kumpulan imina di kedudukan 1 dan 1' ke dua tangan fenil, R1 dan R2. Selain itu, struktur pemula ini perlu mempunyai kumpulan hidrofobik dan penerima ikatan hidrogen (HBA) sebagai rantaian sampingan kepada R1 dan penderma ikatan hidrogen (HBD) sebagai rantaian sampingan kepada R2. Berdasarkan keputusan-keputusan yang diperolehi, dapat disimpulkan bahawa farmakofor model S5T5H06 dan S2T3H01 boleh digunakan untuk pencarian perencat enzim protease NS2B-NS3 Denggi-2 yang berpotensi daripada pengkalan data kimia. Farmakofor model S5T5H06 dan S2T3H01 juga boleh dijadikan panduan untuk merekabentuk perencat enzim protease NS2B-NS3 Denggi-2 berpotensi yang baru berdasarkan struktur sasaran yang aktif.

LIGAND BASED DRUG DISCOVERY OF NOVEL DENGUE-2 NS2B-NS3 PROTEASE INHIBITORS

ABSTRACT

The reported dengue cases are increasing yearly, yet no anti-dengue agent is available in the market. Therefore, the search for anti-dengue is critical. In Malaysia, Dengue-2 (DEN-2) is the most prevalent serotype. NS2B-NS3 protease is the enzyme for the cleavage of polyprotein precursor, which is crucial for the flavivirus replications. This makes it a potential target for the development of therapeutics against the dengue virus. In this study, ligand-based approach was implemented in searching for the new potential DEN-2 NS2B-NS3 protease inhibitors. Pharmacophore models were developed from diverse reported structures of DEN-2 NS2B-NS3 protease inhibitors, comprising peptide and non-peptide molecules. The selected pharmacophore models were employed to screen the US National Cancer Institute (NCI) list of compounds to search for new DEN-2 NS2B-NS3 protease inhibitors. The list of natural products in Natural Product Discovery System (NADI) database was also screened and the structures of active hits were exploited for the design of new inhibitor molecules. The designed molecules were then synthesised. The *in-vitro* assay were conducted to examine the inhibitory activities of the selected hits from NCI database and the synthesised compounds towards recombinant DEN-2 NS2B-NS3 protease by using fluorogenic peptide substrate Boc-Gly-Arg-Arg-MCA. The validity of selected pharmacophore models, S5T5H06 and S2T3H01 was experimentally established by the identification of three new DEN-2 protease inhibitors retrieved from NCI database with K_i values of 17 ± 6 μM (non-competitive), 26 ± 4 μM (competitive) and 95 ± 35 μM (competitive). One competitive inhibitor (K_i = 78 ± 41 μM) and two non-competitive inhibitors (K_i = 29

± 6 μM and K_i = 97 ± 49 μM) were found from the NADI guided compounds. This interesting finding revealed that the natural products from Malaysian biodiversity, which are available in NADI database, are valuable resources of active compounds/active fragments for the development of anti-dengue. It was postulated that the active site and the allosteric binding site on DEN-2 protease share common binding mode to the ligand that carry the pharmacophores of the model S5T5H06. From SAR study, a potential lead structure for DEN-2 NS2B-NS3 protease non-competitive inhibitor was suggested. The structure should has a central nucleus made up of diphenylmethane connected by imines at the position 1,1' to two phenyl lateral arms, R1 and R2. Additionally, the structure should have a hydrophobic and a hydrogen bond acceptor (HBA) as side chains to the phenyl group R1 and a hydrogen bond donor (HBD) as a side chain to the phenyl group R2. The findings concluded that pharmacophoric models S5T5H06 and S2T3H01 can be useful for pharmacophoric exploration of chemical databases in searching for the potential DEN-2 NS2B-NS3 protease inhibitors as well as a guide to design new potential inhibitors from the active hits.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 PROBLEM STATEMENT

Dengue is the most prevalent arthropod-borne viral infection in human. It is estimated that dengue virus infection is 50 – 100 million per year globally in over 100 tropic and sub-tropic countries where over 2.5 billion people (over 40% of world's population) are at risk. Out of these 2.5 billion people at risk, 1.8 billion (> 70%) are in Asia Pacific countries (Kumar et al., 2012; WHO, 2009, 2013a, 2013b). The trend of reported dengue cases in Western Asia Pacific region varies from region to region for example from 2006 to 2012, Australia, Cambodia, Lao PDR, Malaysia, Philippines and Vietnam have reported more cases in 2012 than in 2011. Malaysia in particular reported 21,900 of Dengue cases with 35 deaths (CFR 0.2%) for the year 2012 compared to 19,884 cases (36 deaths, CFR 0.2%) in 2011 (WHO, 2013a). As at December 2013, Australia, Lao PDR, Malaysia, New Caledonia and Singapore have reported higher cumulative cases in 2013 compared to 2012 for the same time period (WHO, 2013c). In Malaysia, the cases rapidly increased (43346 cases with 92 death) by the end of 2013, which is two-fold increase relative to 2012 (WHO, 2014). Recent statistics of dengue cases and fatalities in Malaysia are ringing the alarm bells across the country. From 1st to 22nd January 2014, there were 6,155 dengue cases reported, compared to 1,792 cases within the same period in 2013 (six-fold increase). The number of deaths has doubled from five deaths in 2013 to 10 deaths in 2014, all within the same period (Ismail, 2014).

The continuous prevalent of dengue cases drive the researchers to search and develop the vaccine and anti-dengue drugs for the treatment of dengue. The search for lead anti-dengue compounds has been done from many resources. In traditional medicines plants extracts, several species have been reported with anti-dengue activity (Reis et al., 2008). A sulphated polysaccharide, named fucoidan, from marine algae was found as a potential anti-dengue agent (Hidari et al., 2008) while there have been many discoveries of potential anti-dengue from synthetic compounds using many approaches such as *in silico* screening from compounds libraries (Kampmanna et al., 2009; Wang et al., 2009), high throughput *in vitro* screening and structure-based drug design (Tomlinson et al., 2009).

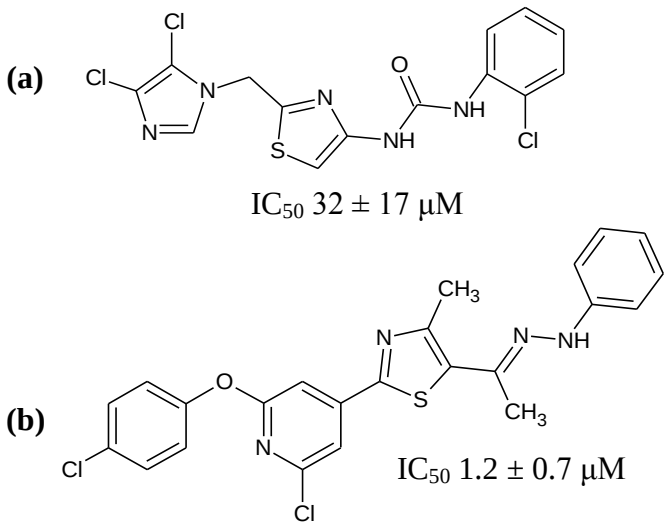
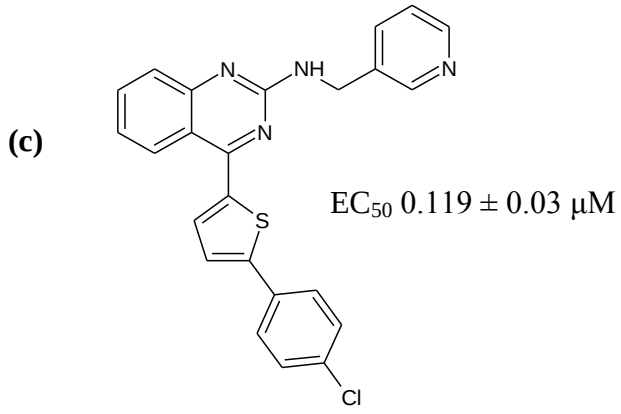


Figure 1.1: Several structures of potential anti-dengue compounds reported by Kampmanna et al., 2009 (**(a)** and (**b))**, Wang et al., 2009 **(c)** and Tomlinson et al., 2009 (**(d)** and **(e)**).



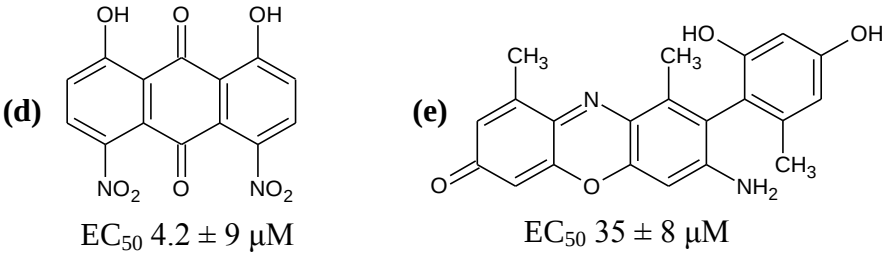


Figure 1.1: Continued.

From the understanding of the viral life cycle, the virus structural proteins (capsid protein, C, membrane-associated protein, prM and envelope protein, E) as well as non-structural proteins (NS1, NS2B, NS3 and NS5) have been identified as potential targets for the development of vaccine and anti-dengue drug. The development of imaging techniques, high resolution x-ray crystallography and computer simulation technology enable the proteins structure models of the virus to be available (Erbel et al., 2006) to assist researchers to understand the interactions of the target proteins and the drugs. This availability helps the researchers to speed up the finding of new drugs at lower cost.

Nevertheless, the development of vaccines and antiviral therapy has seen little success where no licensed antiviral therapy is currently available (Sampath &

Padmanabhan, 2009). The control of disease is focused on the control of mosquito vector, which is costly and often met with limited success (Gubler, 1998). Therefore, the search for the lead compounds and derivatives with their information on the structures and activities is needed in discovering novel inhibitor for the development of anti-dengue.

1.2 DENGUE

Dengue is a disease caused by the infection of the dengue virus. There are four antigenically related but genetically distinct serotypes of dengue viruses; DEN-1, DEN-2, DEN-3 and DEN-4, which are transmitted to human by female *Aedes aegypti* and *Aedes albopictus* mosquitoes. The symptom of a disease similar to dengue was reported since the end of 18^{th} century. In 1780, Benjamin Rush, a physician of Philadelphia, provided the first detailed description of what we know now as dengue fever. However, the first extensively documented outbreak of dengue haemorrhagic fever (DHF) occured in the Philippines in 1953 followed by Thailand in 1958. Since 1960s dengue has spread to over 60 countries (Howard, 2005).

Dengue is most prevalent in tropical Asia, Latin America and Caribbean (Deen et al., 2006; Kyle & Harris, 2008; Mukhopadhyay, et al., 2005). From the estimated number of dengue virus infection of 50 – 100 million per year globally, 250,000 – 500,000 patients develop more severe diseases of dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS) with 1-5% of cases result in fatality (Rawlinson et al., 2006). The incidence of dengue has increased dramatically in the past 50 years due to the population growth and uncontrolled urbanization in tropical and subtropical countries that become breeding sites for the mosquitoes and

problematic vector control (Kyle & Harris, 2008). In addition, the increased global transports networks such as air travel has enabled infected humans to spread viruses (Gibbons & Vaughn, 2002).

Dengue symptoms appear after five to eight days following the bite by an infected mosquito (Smith, 1995). The person infected by dengue virus may manifest ranges of symptoms from silent infection (no symptom) to fatal dengue heamorrhagic fever (DHF)/ dengue shock syndrome (DSS) (Gibbons & Vaughn, 2002). The WHO scheme has classified dengue virus infection into three categories; dengue fever, DHF and DSS.

Dengue fever is clinically defined as an acute febrile illness with two or more manifestations (headache, retro-orbital pain, myalgia, arthralgia, rash, haemorrhagic manifestations, leucopenia) and the occurrence should be at the same location and time as other cases of dengue fever. DHF is defined as a case that must meet all four of the following criteria: fever or history of fever lasting 2-7 days, a haemorrhagic tendency shown by a positive tourniquet test or spontaneous bleeding, thrombocytopenia and evidence of plasma leakage. DHF is further classified into four severity grades according to the presence or absence of spontaneous bleeding and the severity of plasma leakage. The dengue shock syndrome (DSS) refers to DHF grades III and IV, in which shock is present as well as all four DHF defining criteria (Deen et al., 2006; Gubler, 1998). Schematic of WHO classification of symptomatic dengue infection and WHO classification of dengue haemorrhagic fever are shown in Figure 1.2 and Table 1.1, respectively.

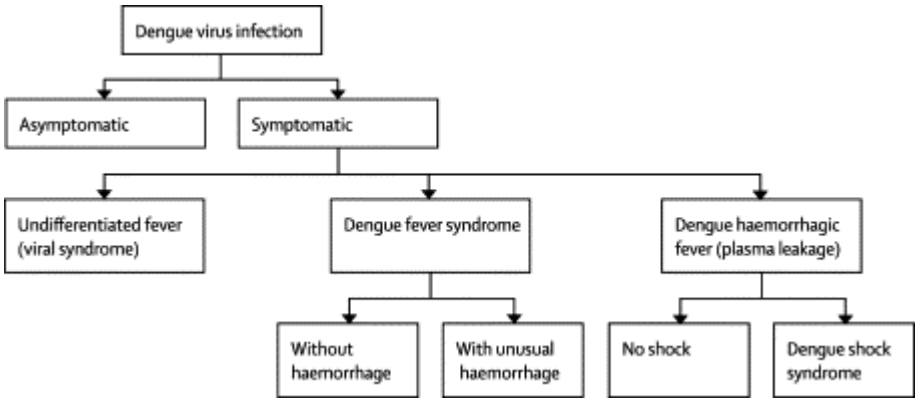


Figure 1.2: WHO classification of symptomatic dengue infection (Source: Deen et al., 2006).

Among the infected person, less than 1% developed DHF/DSS. The development of DHF/DSS was influenced by many factors. The patience with sequential infection of distinct dengue virus serotype from the primary infection showed to have 15-80-fold likely to develop DHF. However, patience that has not undergone a previous infection can also develop DHF. The trigger factors to the development of DHF could be host related such as age, sex, race, nutritional status and genetic predisposition. Children under 15 years old are more likely to develop DHF. However, in the case of DEN-2 outbreak in Santiago De Cuba in 1997, all DHF are in adults. Female are more likely to develop a more severe disease. In the case of race factor, in Cuban epidemic in 1981, white Cubans were 3.5-fold more hospitalised compared to the black Cubans (Howard, 2005; Guzman, et al., 2006; Rawlinson et al., 2006).

Table 1.1: WHO classification of dengue haemorrhagic fever. Source: (Howard, 2005).

Grade	Clinical description
I	Fever with non-specific constitutional symptoms and the only haemorrhagic manifestations being a positive tourniquet test
II	As for Grade I, but accompanied by more extensive haemorrhagic manifestations
III	Sign of circulatory failure or hypertension
IV	Profound shock with pulse and blood pressure being undetectable

Another trigger factor is the viral factor. The differences in genotype between strains also influence the severity of the disease. This phenomenon is shown by the capability of DEN-2 strain from Southeast Asia to cause DHF while the same strain did not cause DHF in epidemic event in Peru in 1995 even though the population had undergone an epidemic of DEN-1 five years before. This is due to the fact that DEN-2 from Southeast Asia and Peru have differences in six amino acid in structural and non-structural proteins as well as at the 5' and 3' un-translated region of virus RNA. However, the exact factors for the minor cases of DHF/DSS development is still not clear (Rawlinson et al., 2006).

1.3 DENGUE VIRUS

Dengue virus is a member of *Flaviviridae* family from *Flavivirus* genera. The *Flaviviridae* family consists of three genera; *Flavivirus, Pestivirus* and *Hepacivirus*. Genus *Flavivirus* contains more than 70 viruses including dengue virus (DENV), Japanese encephalitis virus (JEV), tick-borne encephalitis virus (TBEV), West Nile virus (WNV) and yellow fever virus (YFV). Dengue viruses are further divided into four serotypes 1, 2, 3 and 4 (DEN-1, DEN-2, DEN-3 and DEN-4).

1.3.1 The structure and genome of dengue virus

Dengue virion is a spherical enveloped virus with diameter approximately 500Å. It comprises a single, positive-strand RNA genome of ~10,700 nucleotides that is packed by three structural proteins and a lipid bilayer. The structural proteins are capsid, C; membrane-associated (M, which is expressed as prM, precursor to M) and envelope, E (Kuhn et al., 2002; Mukhopadhyay et al., 2005; Rey, 2003). The RNA of dengue virus consist of a 5'-untranslated region (UTR), a single open reading frame (ORF) and 3'-UTR (Dong et al., 2008). The RNA has type 1 cap at the 5'-terminus (m^7GpppAmp) and polyadenylate [Poly(A)] tail at 3'-terminus (Wahab et al., 2007). The three structural proteins (C, prM and E) are encoded by 5'-terminus of the ORF of the genome. The remainder of the ORF of the genome encodes seven nonstructural proteins i.e. NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (Figure 1.3).

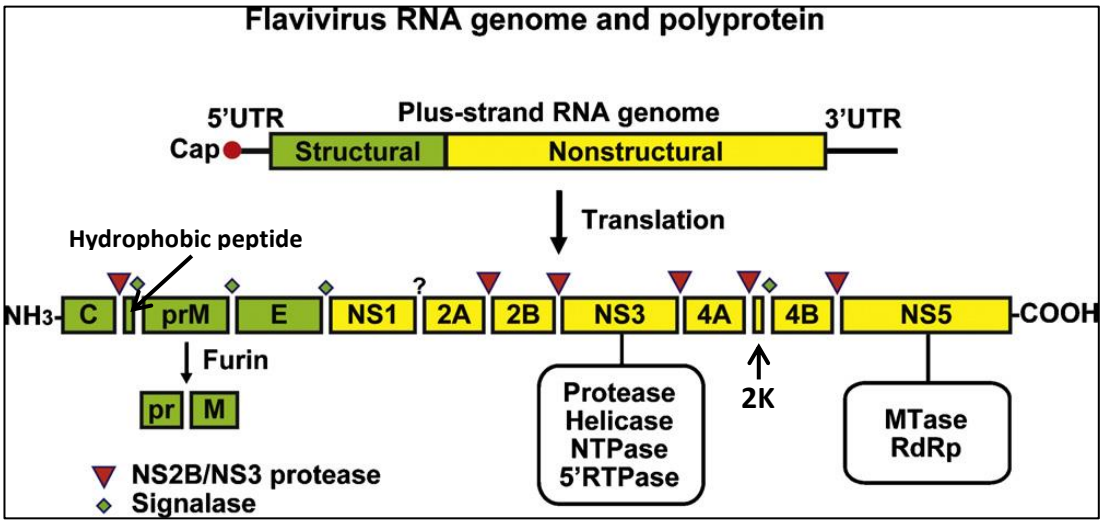


Figure 1.3: Schematic representation of flavivirus genome organization and polyprotein processing. Sites of polyprotein cleavage mediated by the viral NS2B-NS3 and by host signalase are shown. The enzymatic activities of NS3 and NS5 are also indicated. Source: (Sampath & Padmanabhan, 2009)

1.3.2 Life cycle of Flavivirus

The primary mammals' cells infected by dengue virus are monocytes, macrophages and dendritic cells (Rawlinson et al., 2006; Sampath & Padmanabhan, 2009). The life cycle of dengue virus and flavivirus in general, started with the attachment and binding of the virions on the host cell surface, followed by the entry into the cell by the process of endocytosis to form endosome. Endosome then undergoes the process of fusion where the RNA is released into the cytoplasm. The viral RNA is next translated into a polyprotein, which is cleaved into viral proteins (structural and non-structural) by polyprotein processing. Then, the replication of RNA and virus assembly occur. Finally, budding occurs where the new mature virions are released from the host cell by exocytosis (Guzman, et al., 2006; Lescar et al., 2008; Mukhopadhyay et al., 2005). The schematic diagram of the flavivirus life cycle is illustrated in Figure 1.4.

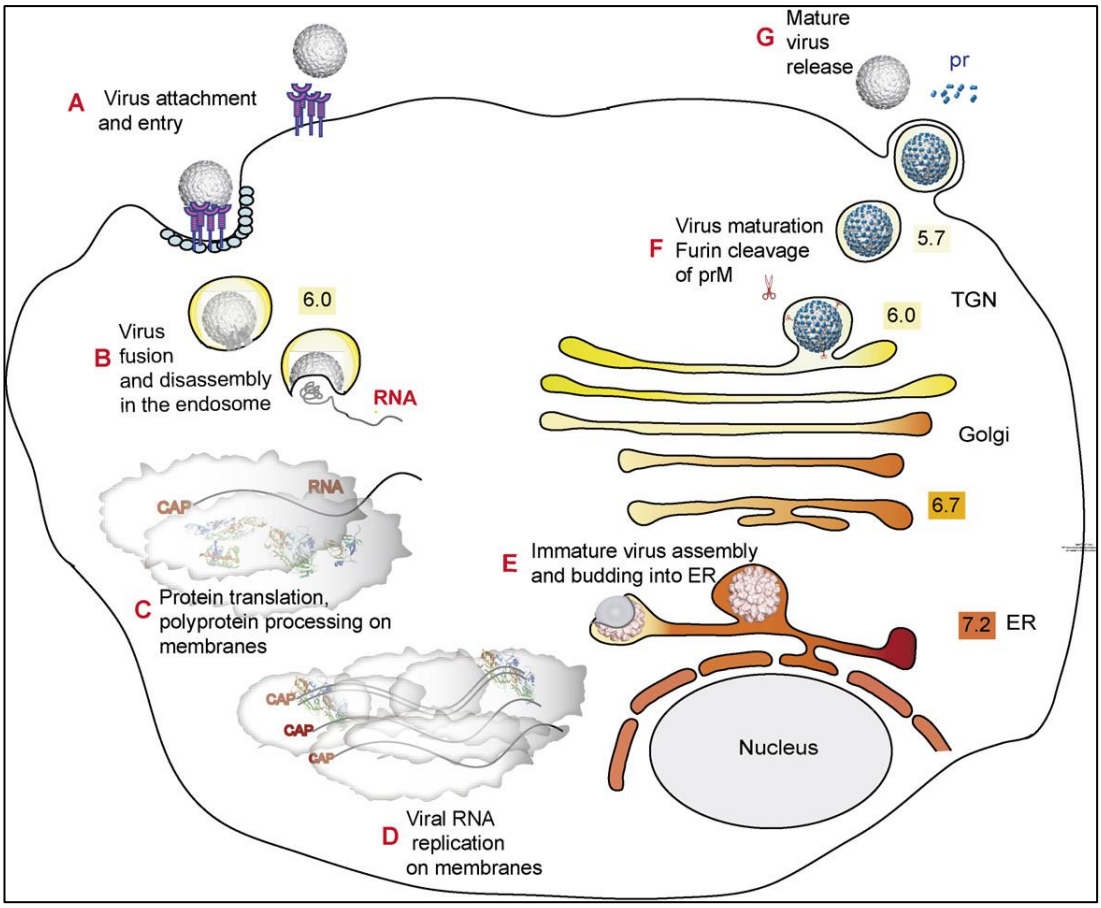


Figure 1.4: Flavivirus life cycle. Numbers shown in boxes refer to the pH of the respective compartments. Source: (Perera et al., 2008).

1.3.2.1 Entry to the cell

In the initial dengue virus infection process, the virion attaches to the host cell by the binding of E protein onto the receptor on the host cell surface (Rawlinson et al., 2006) . Then, the virion enters the cell by receptor-mediated endocytosis to form endosome (Sampath & Padmanabhan, 2009). There are different receptors for dengue virus on different cell types for the attachment. Fcɣ-R is the receptor for macrophages and monocytes, while the C-type lectine DC-SIGN(CD209) is the receptor for dendritic cell for the infection of all four dengue serotypes (Rawlinson et al., 2006).

Following endocytosis, the RNA genome is released into the cytoplasm of the host cell through the fusion of the viral membrane and the host-cell endosomal membrane. Acidic environment (low pH) in endosome mediates the structural reorganization of E protein (Sampath & Padmanabhan, 2009) from homodimers into homotrimers (Rawlinson et al., 2006). The trimeric conformation of the E protein forms more open structure, exposing the fusion peptide, thereby bringing the viral and endosomal membrane into close proximity and allowing insertion of the virus into the host cell membrane. As a result, fusion of the viral and host-cell endosomal membranes occurs. Upon fusion, the nucleocapsid, which carries virus RNA is released into the cytoplasm of the host cell (Lescar et al., 2008; Mukhopadhyay et al., 2005; Rawlinson et al., 2006). The capsid protein and viral RNA dissociation is probably a spontaneous process that occurred upon the release of nuclecapsid (Rawlinson et al., 2006).

1.3.2.2 Translation and polyprotein processing

In cytoplasm, the viral RNA served as mRNA that is directly translated into a long single polyprotein precursor of approximately 370 kDa. The polyprotein precursor is arranged in the order of NH2-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-COOH (Wahab et al., 2007). The polyprotein is then cleaved into the three viral structural (C-prM-E) and seven non-structural (NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5) proteins by viral-encoded serine protease (NS2B-NS3), host cell proteases (signalase and furin) as well as other unknown protease (Lescar et al., 2008; Mukhopadhyay et al., 2005; Rawlinson et al., 2006; Sampath & Padmanabhan, 2009).

In addition to the ten proteins, two small hydrophobic peptides are released from the polyprotein (Figure 1.3). One is derived from the C-terminus of the anchored capsid protein; after the cleavage of this fragment, the mature capsid is released. The second one is a small fragment between the NS4A and the NS4B that is called 2K (Guzman, et al., 2006).

The viral NS2B-NS3 protease is involved in the cleavage on the cytoplasmic side of the endoplasmic reticulum (ER) membrane whilst host cellular protease takes part in the cleavage in the ER lumen (Lescar et al., 2008). The viral NS2B-NS3 protease is responsible for the cleavage at the NS2A/2B, NS2B/NS3, NS3/NS4A, NS4A/2K and NS4B/5. Host protease cleaves at the C/prM, prM/E, E/NS1 and 2K/NS4B junctions (Figure 1.3) that release prM, E and NS1 proteins in the ER lumen (Guzman, et al., 2006; Rawlinson et al., 2006).

1.3.2.3 RNA replication

Genome replication occurs in intracellular membranes (Mukhopadhyay et al., 2005) and Golgi-derived membranes called vesicle packed (Sampath & Padmanabhan, 2009). Non-structural proteins mediate the transcription of viral RNA. Thus, transcription started after the non-structural proteins have been synthesised and released from the polyprotein precursor (Lescar et al., 2008). Starting from the genomic plus-strand RNA, the complimentary negative-strand RNA is synthesised, which then served as the template for the production of an additional genomic plus-strand RNA (Guzman, et al., 2006).

1.3.2.4 Viral assembly and release of new progeny virion

Virus assembly occurs on the surface of the endoplasmic reticulum (ER) (Mukhopadhyay et al., 2005). The newly synthesised plus stranded RNA is coated with C protein to form nuclecapsid. The mechanism of interaction of the C protein in nucleocapsid is still not clear (Sampath & Padmanabhan, 2009). Then, E and prM envelop the nucleocapsid resulting in the formation of non-infectious, immature viral and subviral particles in the lumen of ER. Immature viral particle contains E and prM proteins, lipid membrane and nucleocapsid. On the other hand, subviral particle does not contain nucleocapsid. These particles are non-infectious as they cannot induce host-cell fusion. They are transported through the trans-Golgi network where the maturation of the virus particles occurs. In the trans-Golgi network, the prM is cleaved to M by the host protease furin and rearrangements of E protein occur, resulting in the formation of mature, infectious particles (Perera et al., 2008). Subviral particles are also cleaved by furin. Mature and subviral particles are subsequently released from the cell by exocytosis (Mukhopadhyay et al., 2005; Sampath & Padmanabhan, 2009).

1.4 PROTEINS OF DENGUE VIRUS

1.4.1 Structural proteins and functions

The capsid protein (~100 amino acids) is involved in the packaging of viral genome/form nucleocapsid core (NC). The prM protein (~165 amino acids) might function as a chaperone for the folding and assembly of E protein during particle maturation. The E protein (~495 amino acids) constructs the envelope structure of the virus.

1.4.1.1 Capsid protein

Dengue-2 virus capsid protein exists in dimer form and has been suggested to be the building block for nucleocapsid assembly. Nuclecapsid core formed in the early stage of assembly process, consists of one copy of genomic RNA and multiple copies of the capsid protein. However, NCs are rarely found in flavivirus-infected cells, which indicated that the particle formation is a coordinated process between the membrane-associated capsid protein and the prM-E heterodimers in the ER (Mukhopadhyay et al., 2005). The cryoEM reconstruction of mature and immature dengue virus show that the nucleocapsid is not a well-formed protein shell. It has been suggested that the capsid protein facilitates the binding of the nucleocapsid to the lipid membrane (Ma et al., 2004).

1.4.1.2 E and M proteins

The E protein, the external structural protein of dengue virus is responsible for the binding with cellular receptor and interacts with the host cell membrane for fusion and penetration, as well as viral assembly and budding. E protein also possesses antigenic determinants that elicit neutralizing antibodies (Huang et al., 2008; Kuhn et al., 2002; Modis et al., 2004, 2005; Mukhopadhyay et al., 2005; Zhang et al., 2004). The ability of monoclonal antibody to neutralise dengue virus serotype 1, 2 and 3, is primarily by inhibiting the attachment to the host. Therefore, the change in viral surface structure was assumed to be responsible for inhibiting viral attachment to cell (Lok et al., 2008).

During the flavivirus life cycle, the virion exists in three main conformational states: immature, mature and fusion-activated. The conformation of the E protein differs in these three states. The conformational rearrangement occur from prM–E heterodimers in immature particle to E homodimers in the mature particle, and finally to E homotrimers in the fusion-activated particle (Perera et al., 2008).

Following the assembly process, an immature virus particle is produced in ER, which have pre-membrane protein (prM) that must be proteolytically processed during virion maturation. There are two types of immature virus; "spiky" and "smooth" virus particles. Immature virus particles have a diameter of ~600 Å (Figure 1.5A and 1.5C) with a spiky surface protein shell consisting of prM and E proteins, which form 180 heterodimers that are arranged as 60 trimeric spikes. Within each spike there are three prM-E heterodimers (Zhang et al., 2003; Zhang et al., 2004). The prM protein forms a cap-like structure that protects the fusion peptide on E protein and prevents the fusion of the virus with the host membrane.

In Golgi apparatus, low pH environment trigger rotational and translational movement of E protein resulting in the transition of E protein from prM heterodimer (Figure 1.5C) to anti-parallel E homodimers (Figure 1.5D). The "smooth" viral morphology is formed but is still considered immature due to the presence of prM protecting the fusion peptide on E. Maturation process of flavivirus is directed by proteolitic cleavage of the precursor membrane protein (prM) (Li et al., 2008). Then, furin cleaves the prM into pr and M. The cleaved pr portion remains associated with E and is only released from the virus particle following the exit into neutral pH environment. Following maturation, the particle is released into a neutral

environment and pr peptide is released from the virion (Perera et al., 2008).

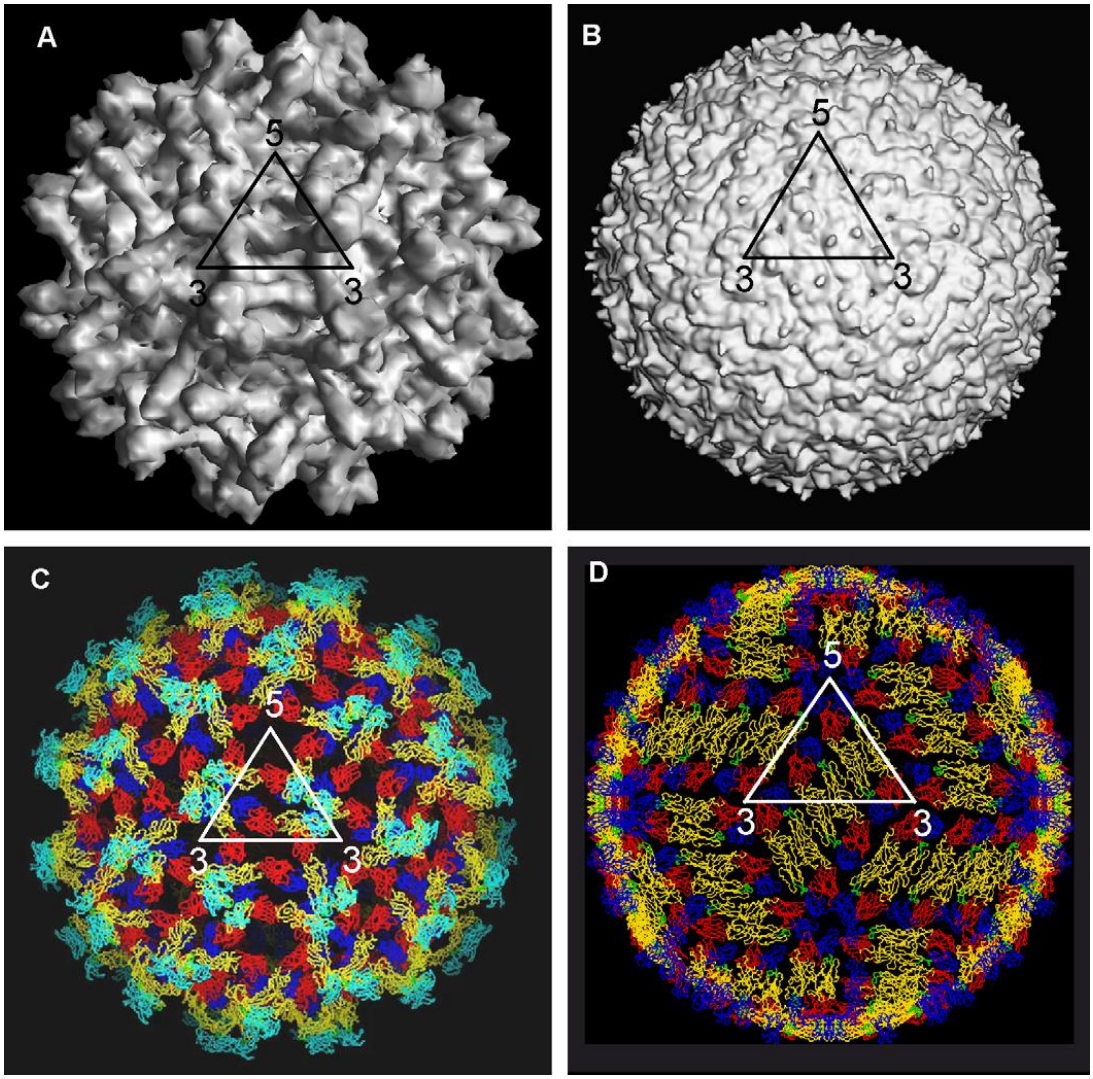


Figure 1.5: Structure of flaviviruses. (A) A surface shaded view of the cryo-EM reconstruction of immature DENV-2 showing the spiky surface features. (B) A surface shaded view of the cryo-EM reconstruction of mature DENV-2 showing the relatively smooth surface features (C) Fit of the atomic coordinates of the E protein into the immature virus and (D) Mature virus showing the arrangement of the E proteins on the surface of the virion. Source: (Perera et al., 2008)

Mature flavivirus particles (diameter ~500 Å) have a relatively smooth surface with the lipid bilayer membrane completely covered by the envelope (E) and membrane (M) protein shell (Figure1.5B and 1.5D). This shell consists of 180 copies of the E protein arranged as 90 homodimers forming a herringbone pattern or so-

called protein rafts that lie flat on the viral surface. Following the attachment of mature virus particle, specific cell-surface receptors mediate endocytosis of the virus. In the endosome, in the low pH environment, the E homodimers within the mature virion dissociate and re-arrange into fusion-active homotrimers. The fusion peptide of the virus becomes exposed at the distal end of the E protein and is inserted into the host membrane, which promotes fusion of the viral and host membranes (Modis et al., 2004; Perera et al., 2008).

1.4.2 Non-structural proteins and functions

Nonstructural proteins are essential for viral replication. NS1 involves in early steps of viral replication (Lindenbach & Rice, 1999), while NS3 shows enzymatic activities, which is involved in the viral polyprotein processing and genome replication. NS3 functions as a serine protease (NS2B-NS3 complex mediate proteolytic processing of polyprotein), 5'-RNA triphosphatase (RTPase), nucleoside triphosphatase (NTPase) and helicase (Luo et al., 2008). NS5 exhibits two enzymatic activities which are involved in the methylation of 5'-cap structure of genomic RNA (methyltransferase) and RNA-dependent RNA polymer (Dong et al., 2008; Geiss et al., 2009; Zhou et al., 2007). Due to the improved knowledge on the structures and functions of the proteins, the development of vaccine and therapeutic design currently target E, NS1, NS3 and NS5 proteins (Wahab et al., 2007).

1.4.2.1 Dengue virus NS2B-NS3 protease

Proteins are made up of a long chains of amino acids that bonded by carbon-nitrogen bonds called amide bonds or peptide bonds. Chemical hydrolysis of amide bond requires activation energy of ~25 kcal/mol. Therefore, under normal physiological conditions (pH~7, T = 37 °C), hydrolysis of amide bond would be very slow. However, catalysis of amide bond hydrolysis under physiological condition is accomplished at much faster rate by enzymes called proteases or peptidase.

The protease or proteolytic enzymes not only catalyze protein hydrolysis at a fast rate but also specific and selective in order to prevent uncontrolled proteolysis of the organism. Proteases can be divided into classes based on their mechanism of action, which are serine, cysteine, aspartic, threonine and metallo proteases. As an example, a serine protease is an enzyme in which one of the amino acids at the active site is serine and the catalytic nucleophile in serine protease is hydroxyl group of the active site serine.

Protease binds with the substrate in a specific manner that is represented by Schechter and Berger system of nomenclature. According to this system, the amino acid residues (or side chains) of substrate are labeled from the N to C terminus as P_n,.....,P_3,P_2,P_1,P_1',P_2',P_3',.....P_n' and the corresponding binding sub sites of the enzyme are labelled as S_n,.....,S_3,S_2,S_1,S_1',S_2',S_3',.....S_n'. The peptide bond hydrolysis is carried out between P_1,P_1' and the bond is called the scissile bond (Figure 1.6). The P_1 residue of the substrate is called the primary specificity residue. The specificity and selectivity of a protease depend on the nature of the P_1 residue. As an example trypsin only cleaves at lysine or arginine residues (positively charged

residues/basic residues at P_1). Serine proteases are classified as elastase-like (small hydrophobic residues at P_1), trypsin-like (positively charged residues/basic residues at P_1) or chymotripsin-like (large hydrophobic residues at P_1).

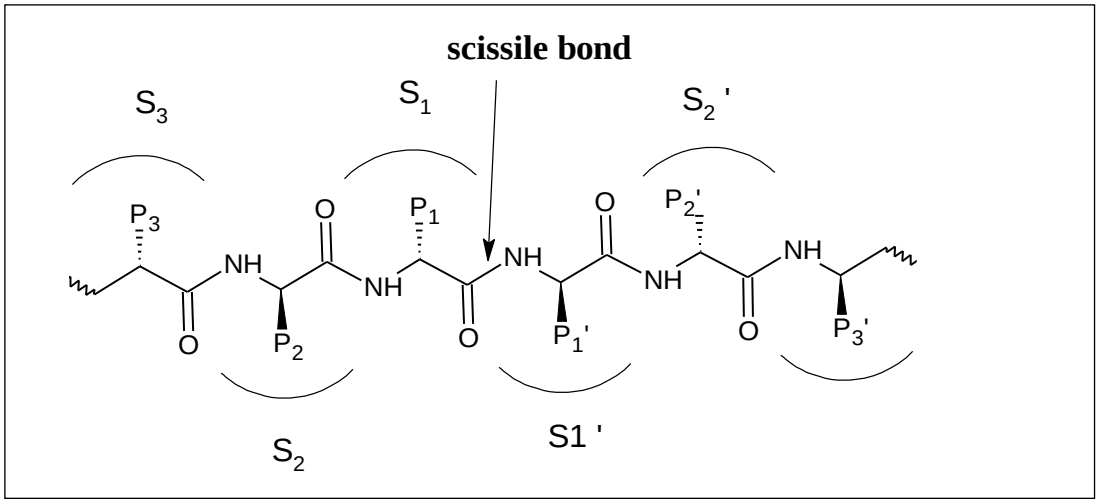


Figure 1.6: Schehter and Berger System of Nomenclature.

NS2B-NS3 protease is a serine protease. Serine protease utilises a combination of mechanisms that are common to enzyme catalysis. First, enzyme binds to the substrate to form an enzyme-substrate Michaelis-Menten complex (E-S complex) utilizing non-covalent bonding interaction such as ionic interactions, dipole-dipole interactions, hydrophobic interactions, hydrogen bonding and Van der Waal's interactions. When the substrate is bound to the active site of the enzyme, the carbonyl group of the scissile amide bond is exposed for catalysis by enzyme.

The active site of serine protease consists of three conserved amino acid residue; serine, histidine and aspartic acid that together form the "catalytic triad". Serine protease utilises an active site serine residue in a covalent catalytic cleavage of peptide bonds. Nucleophilic group of serine is hydroxyl, which is a poor nucleophile. However, the serine hydroxyl is activated by general base catalysis

through hydrogen bonding network involving aspartic acid, histidine and serine residues nearby. This process converts serine to alkoxide (ionised serine) through a mechanism called the charge relay system (Silverman, 2004). The active site of serine (activated nucleophile) then attacks the carbonyl group of the scissile amide bond (Figure 1.7).

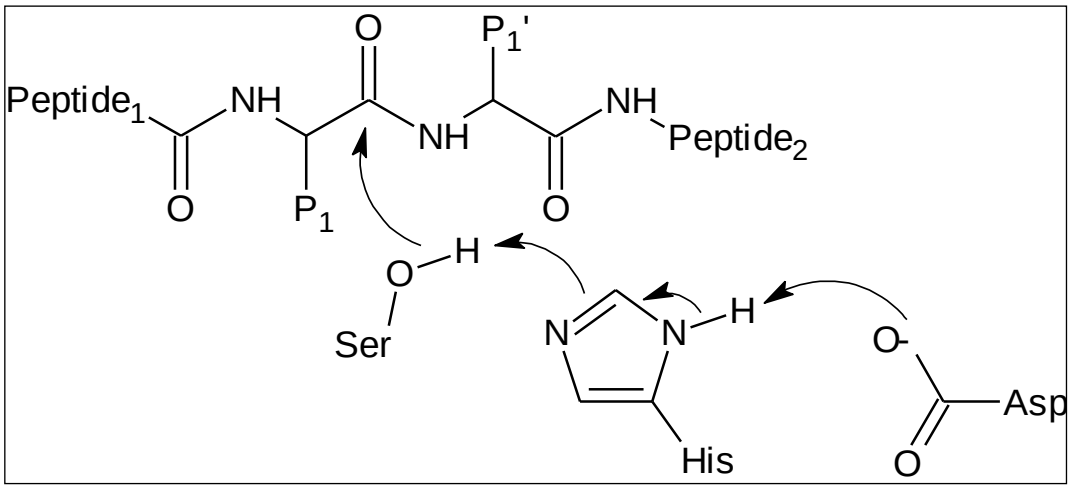


Figure 1.7: Charge relay system for activation of an active site serine residue.

Dengue NS3 protein is made up of 618 amino acids. It consists of two domains that possess enzymatic activity: a trypsin-like serine protease located in the N-terminal domain of the protein and a C-terminal domain that possess RNA helicase and NTPase activity (Murthy et al., 1999). The NS3 protease domain adopts a chymotrypsin-like fold with two β-barrels, each formed by six β-strands (Figure1.8), with the catalytic triad (His51-Asp75-Ser135) located at the cleft between the two β-barrels.

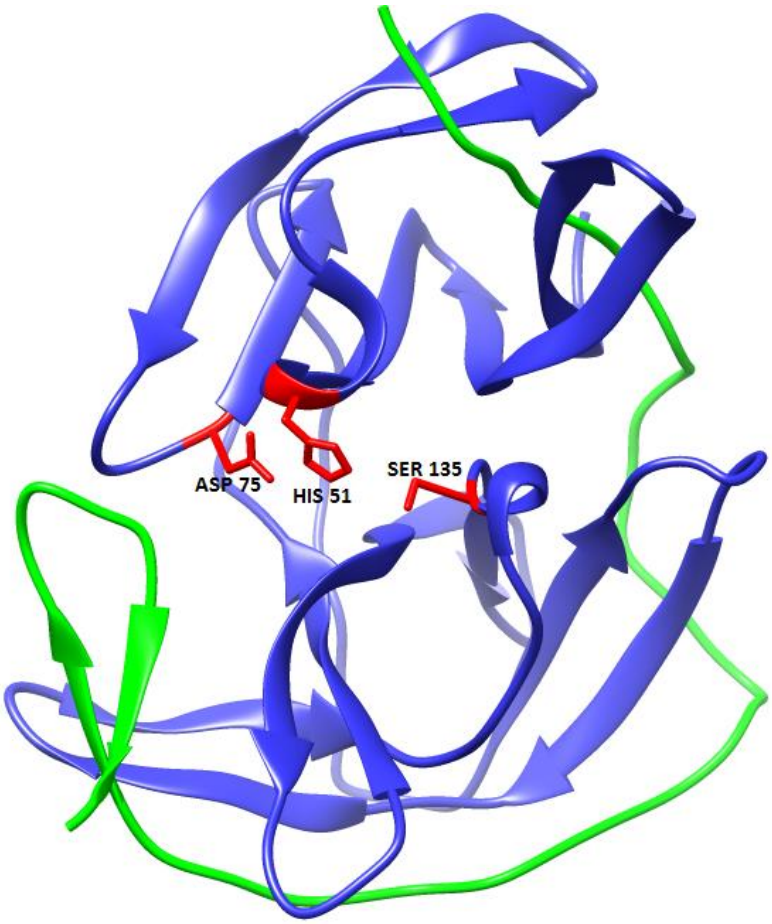


Figure 1.8: Structures of NS2B–NS3 protease (Blue, NS3pro; Green, NS2B). Residues of the catalytic triad Histidine 51 (HIS 51), Aspartate 75 (ASP 75) and Serine 135 (SER 135) are shown as red-colored stick models. Source: (Wichapong et al., 2009).

It has been reported that Dengue virus NS3 protease alone is not active for proteolysis. The enzymatic activity of NS3 protease is enhanced by its interaction with NS2B protein (130 amino acid) that functions as a co-factor. Both NS2B and NS3 are required for proteolytic processing of dengue virus polyprotein (Falgout et al., 1991). From the X-ray crystal structure of NS2B (PDB ID: 2FOM) (Erbel et al., 2006), residues 51-57 of NS2B form a β-strand which associates with the NS3 protease. Erbel et al. (2006) also expressed a construct containing residues 49–66 of NS2B in a soluble form, indicating that the N-terminal part of NS2B is sufficient to stabilise the enzyme. However, the truncated NS2B-NS3 protein is catalytically

inactive and cannot bind a substrate-based inhibitor, which suggests that the C-terminal part of NS2B directly interacts with the substrate-binding site. This is in contrast to HCV NS3 protease, where a short fragment of NS4A, including residues corresponding to β-strand 1, is sufficient to yield a fully active enzyme (Erbel et al., 2006).

It is also suggested that NS2B functions as a chaperone in assisting the folding of NS3 protease to an active conformation (Leung et al., 2001). A 47-residues central hydrophobic region of NS2B is found to be essential for the activation of NS3 protease where the absence of NS2B has adverse effects on both the stability and the catalytic activity of NS3 protease (Murthy et al., 1999; Yusof et al., 2000). Leung et al. (2001) also reported that the enzyme is inactive on its own or after the addition of 13-residues cofactor but is active when fused to the 40-residues cofactor. The first homology model with complete structure of the protease including the complete cofactor NS2B in the productive form was presented by Wichapong et al. (2009).

The active site of NS2B-NS3 protease consists of four pockets, which are S1, S2, S3 and S4 that interact with the substrate residues P1, P2, P3 and P4, respectively. The sequence alignment of Dengue 2 Virus (PDB ID: 2FOM) and West Nile Virus (2PF7) of NS2B and NS3 domains (Erbel et al., 2006) showed that all residues in the pockets are either identical or similar for these two viruses (Wichapong et al., 2009) as shown in the Figure 1.9.

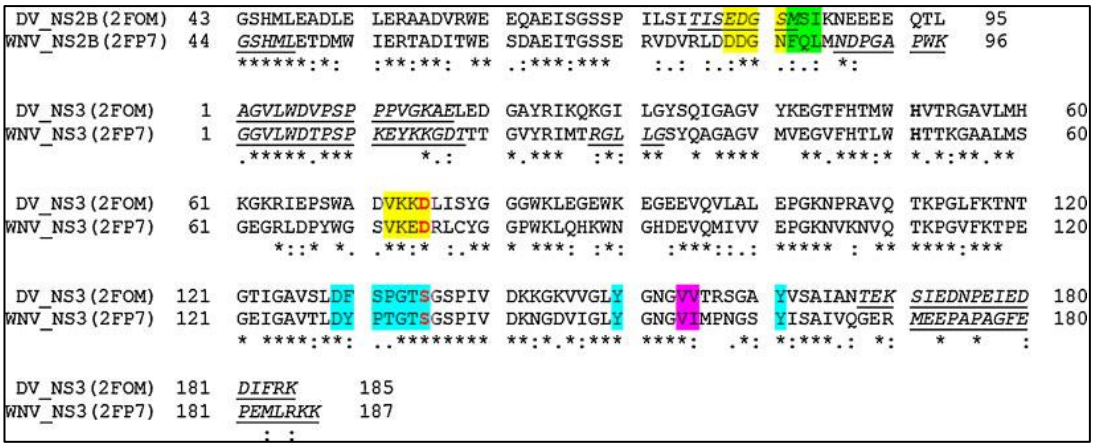


Figure 1.9: Sequence alignment of DENV and WNV NS2B and NS3 domains. Stars indicate residues that are identical whereas dots and colons indicate similar residues. Residues which are drawn italic and underlined are missing residues in the X-ray structures. Residues located in the S1, S2, S3, and S4 pockets are colored cyan, yellow, green, and magenta, respectively, and catalytic triads are displayed in bold red. Source: (Wichapong et al., 2009).

Previous alignment studies indicated that NS3 protease has structurally similar to trypsin-like serine protease (Murthy et al., 1999). However, unlike trypsin, NS3 protease has a marked preference for a pair of basic residues (either Arg or Lys) at P2 and P1 in its peptide substrate followed by a small, unbranch amino acid (Gly or Ser) at P1' (Chambers et al., 1991; Chambers et al., 1990)

Wichapong and co-workers (2009) had constructed homology model of DEN-2 Virus NS2B-NS3 protease complexed with the peptidic inhibitor (Bz-Nle-Lys-Arg-Arg-H). Based on the MD simulation of this homology model, the interactions between the enzyme active sites and inhibitors were revealed. It was suggested that His51 and Asp129 of NS3 protease play significant role in stabilizing the basic inhibitor P1 residue Arg or Lys (Li et al., 2005; Wichapong et al., 2009). From the simulation, the interaction of Arg-P2 and residues Asp75 (C-terminus NS2B), Asn152 (NS3) and Gly182 (NS2B) were observed. This result suggested that C-terminus of NS2B not only play an important role for stabilizing and reorganizing the

NS3 protease but is also essential for the interaction with P2 residue of the inhibitor.

All residues in S3 pocket are made up from NS2B domain. MD simulation showed that the H-bond was only observed between Lys-P3 and Tyr161. The S4 pocket of DEN-2 protease is formed by Val154, Val155, Thr156 and Arg157 suggested that it is dominated by hydrophobic residues. Thus the preferred residue at the P4 position of peptidic inhibitor is norleucine or leucine (Li et al., 2005). Wichapong and co-workers (2009) also revealed that the hydrogen bond between Nle-P4 of inhibitor and the S4 pocket is very low.

1.4.2.2 Dengue virus NS2B-NS3 protease as a potential drug target

The replication of flaviviruses requires the correct processing of their polyprotein by the viral NS2B-NS3 protease. Therefore, NS2B-NS3 protease is a potential target for the development of therapeutics against the dengue virus (Tomlinson et al., 2009). Moreover, protease inhibitors have a successful history as being developed into antiviral drugs (Hsu et al., 2006; Malcolm et al., 2006; Wlodawer & Vondrasek, 1998). A few protease inhibitors have been approved by US Food and Drug Administration (FDA) for HIV treatement such as saquinavir mesylate (1995), ritonavir (1996) and indinavir (1996), amprenavir (1999) and atazanavir sulphate (2003) (Hsu et al., 2006). For dengue virus, the potential peptides and small molecules as protease inhibitors have been studied by a few reserchers focusing on NS2B-NS3 protease (Chanprapaph et al., 2005; Ganesh et al., 2005; Kiat et al., 2006; Yin et al., 2006a).