VIRTUAL SCREENING AND *IN VITRO* ASSAY OF POTENTIAL INHIBITORS AGAINST DENGUE-2 NS2B-NS3 PROTEASE

VANEE A/P GANESH

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

September 2016

ACKNOWLEDGEMENT

All praises goes to the Almighty for giving me the strength and courage to accomplish this research work and put it in writing.

Foremost, my deepest gratitude and appreciation to my supervisor, Prof. Dr. Habibah A. Wahab, for the guidance, encouragement, patience and support provided throughout the research period. I am gratified by her understanding during critical situations in life, when I need to be away from laboratory. I thank Dr. Ezatul for her support and mentoring as well.

I would like to thank my seniors, friends and lab mates from PhDS. They have been supportive and helpful. I thank Dr. Yusuf, Dr. Maywan, Adila and Stella, with whom I have grown my friendship with. Their presence has definitely made this journey easier to go through.

I thank the support staff from Pharmaceutical Chemistry (J03), School of Pharmaceutical Sciences, USM, for their assistance during my work routine over there. I thank Ministry of Higher Education (MOHE) for providing me financial assistance to fund my tuition fees.

I thank my parents, especially my Mother, who understands my research journey, for motivating me and for being supportive.

TABLE OF CONTENTS

CHAPTER 4 - CONCLUSION

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
ADT	AutoDockTools
ASN	Asparagine
ASP	Aspartic acid
C	Capsid
C	Carbon
d	Doublet
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
DEPT	Distortionless enhancement of polarisation transfer
DHF	Dengue haemorrhagic fever
DMSO	Dimethyl sulfoxide
DSS	Dengue Shock Syndrome
E	Envelope
ER	Endoplasmic reticulum
FTIR	Fourier transform infrared spectroscopy
GLU	Glutamic acid
GLY	Glycine
HCl	Hydrochloric acid
HIS	Histidine
HMBC	Heteronuclear Multiple-Bond Correlation
HPLC	High pressureliquid chromatography
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
IC_{50}	Half maximal inhibitory concentration
ILE	Isoleucine
IUPAC	International Union for Pure and Applied Chemistry
kcal/mol	kilocalorie per mol
kDA	Kilo Dalton

LCMS	Liquid Chromatography - Mass Spectrometry
LYS	Lysine
M	Membrane
m	multiplet
MeOH	Methanol
MET	Methionine
mRNA	Messenger Ribonucleic acid
NADI	Natural Product Discovery System
NCI	National Cancer Institute, USA
NMR	Nuclear Magnetic Resonance
NS2B	Non Structural Protein 2B
NS3	Non Structural Protein 3
NS4A	Non Structural Protein 4A
NS4B	Non Structural Protein 4B
NS5	Non Structural Protein 5
PDB	Protein Data Bank
PHE	Phenylalanine
prM	Pre membrane protein
PRO	Proline
RMSD	Root mean square deviation
RNA	Ribonucleic acid
s	Singlet
SER	Serine
THR	Threonine
TYR	Tyrosine
v/v	Volume/volume
VAL	Valine
µg	Micro gram
µl	Micro litre
µM	Micro molar
π	Pi
%	Percentage
°C	Degree Celcius

SARINGAN MAYA DAN *IN VITRO* UNTUK PERENCAT BERPOTENSI TERHADAP PROTEASE NS2B-NS3 DENGGI- 2

Abstrak

Walaupun penyakit denggi merupakan beban global semasa yang tinggi, namun sehingga kini tidak ada penawar yang pasti untuk denggi. Walaupun terdapat usaha-usaha pembangunan vaksin yang dijalankan , cabaran imunisasi yang sukar diatasi adalah perlindungan lengkap terhadap kesemua empat serotype di mana perlindungan imunisasi yang tidak lengkap boleh menyebabkan pesakit yang mempunyai risiko untuk menghidapi Demam Hemoragik Denggi (DHF) dan Sindrom Kejutan Denggi (DSS). Berdasarkan faktor-faktor ini , kepentingan terapi antivirus masih amat diperlukan.Namun begitu, proses penemuan dan pembangunan ubat yang memakan masa menambah lagi kepada beban ini. NS2B / NS3 enzim protease mempunyai peranan penting dalam pembelahan pelopor poliprotein - satu proses yang penting untuk replikasi flavivirus - menjadikannya sasaran terapeutik yang sesuai. Kajian ini menggunakan kaedah bantuan komputer, melalui penggunaan AutoDock Vina untuk melakukan penyaringan maya terhadap sebatian dari NCI Kepelbagaian Set Data dan juga dari Sistem Penemuan Produk Semulajadi pangkalan data (NADI) terhadap protein sasaran, NS2B / NS3 denggi jenis 2. Keputusan penyaringan maya telah di analisa untuk mendapatkan maklumat mengenai interaksi yang menyumbang kepada setiap pertalian mengikat. Ujian *in vitro* telah dilakukan untuk menentukan aktiviti perencatan daripada empat puluh sebatian NCI terhadap enzim protease dan tujuh ekstrak tumbuhan terhadap DEN-2 NS2B / NS3 dengan menggunakan substrat peptida Boc-Gly- Arg-Arg-MCA. Dua sebatian NCI di kodkan NSC127133 dan NSC 343256 merencatkan protease pada 5.73µM dan 30µM, masing-masing. Americanin A, sejenis sebatian neo-lignan yang

di asingkan daripada buah *Morinda citrifolia* menunjukkan aktiviti perencatan dengan nilai IC_{50} pada 167 μM.. Kajian ini juga mengandaikan bahawa tapak alosterik juga boleh memainkan peranan dalam aktiviti perencatan NS2B-NS3pro.

VIRTUAL SCREENING AND *IN VITRO* ASSAY OF POTENTIAL INHIBITORS AGAINST DENGUE-2 NS2B-NS3 PROTEASE

Abstract

Despite the current global burden, there has been no definite cure for dengue. Although efforts on vaccine development are ongoing, the strategy faces challenges of constant immunization, where by incomplete protection against all four serotypes may lead to patients at risk of progressing to dengue haemmorhagic fever (DHF) and dengue shock syndrome (DSS). Considering these factors, antiviral therapy is still in significant need. However, the time consuming process of drug discovery and development is adding to this burden. NS2B-NS3 protease plays crucial role in the cleavage of polyprotein precursor - an important process for flavivirus replication, making it a suitable therapeutic target. This study employed computer-aided approach, with the use of AutoDock Vina to virtually screen compounds from National Cancer Institute (NCI) Diversity Data Set as well as from in-house Natural Product Discovery System database (NADI) against the target protein, NS2B-NS3 protease of dengue virus type 2 (DEN-2). Virtual screening results were analyzed to obtain information on interactions contributing to each binding affinity. The *in vitro* assay was then carried out to determine inhibitory activities of forty NCI compounds and seven plant extracts towards DEN-2 NS2B-NS3 protease by using fluorogenic peptide substrate Boc-Gly- Arg-Arg-MCA. Two NCI compounds coded NSC127133 and NSC343256 inhibited protease at 5.73 µM and 30 µM, respectively. Americanin A, a neo-lignan compound isolated from the fruit of *Morinda citrifolia* showed inhibitory activity with the IC_{50} of 167µM. It is postulated in this study, that the allosteric site of NS2B-NS3 could play a role in the inhibitory activity of the NS2B-NS3pro. .

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Statement of the problem

The mosquito-borne dengue virus, is an emerging pathogen, belonging to the *Flaviviridae* family and *Flavivirus* genus, continues to be a constant threat to children and adults worldwide (Tomlinson et al., 2009). Dengue has been ranked as the most critical form of mosquito-borne viral disease, as reported by the World Health Organization in 2012 (World Health Organization, 2012). Dengue virus serotypes (DEN-1, DEN-2, DEN-3, and DEN-4) have been identified to be the main causative agents triggering dengue fever, dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). WHO's statistics also estimated that 50 to 100 million infections occur annually in 100 countries already endemic to dengue, with the spreading of the disease to the previously unaffected areas (World Health Organization, 2012). As of December 2015, it was reported that Malaysia is experiencing 67.6% and 16.3 % increase in dengue cases compared to number of cases reported in year 2013 and 2014 respectively (Western Pacific Regional Office, 2013, 2015). A number of factors including massive urbanization, overpopulation, inconsistent *Aedes aegyptii* eradication programme, poor living conditions, and mutating strains (Edelman, 2007), poor waste management and lack of basic infrastructure (interruptive water supply which prompts public to collect and store water at their homes) have caused dengue epidemic to be a major challenge to tackle. (World Health Organization, 2002).

Despite current global burden, there has been no definite cure for dengue (Lam, 2013). Although efforts on vaccine development are ongoing, the strategy faces challenges of constant immunization, where by incomplete protection against all four serotypes may lead to patients at risk of progressing to DHF and DSS (Noble et al., 2010). Considering these factors, the importance of antiviral therapy is still in significant need (Chawla et al., 2014). Targeted antiviral approach to dengue has a more promising approach, by exploiting viral machineries critical to the viral replication before onset of the disease itself (Noble et al., 2010). The general idea is to discover an inhibitor which is able to bind to any part of this viral machinery to halt further development. The search for inhibitor could be performed through multiple methods;- enzyme based screening, viral replication based screening, structure based rational design, virtual screening, and fragment-based screening (Noble et al., 2010). There have been notable efforts from Malaysian researchers in the search for dengue NS2B-NS3 inhibitors with similar approaches (Heh et al., 2013; Kiat et al., 2007).

1.3 Dengue

1.3.1 Overview

Dengue is the most prevalent arthropod-borne virus causing more human morbidity and mortality compared to other arthropod-borne viruses today (Alen & Schols, 2011). The virus which remains to be major public concern in the tropical region depends on vectors namely *Aedes aegypti* and *Aedes albopictus* to infect living organisms including humans and non-humans (Gubler, 1998). Four antigenically distinct serotypes of the virus have been determined; - DEN-1, DEN-2, DEN-3 and DEN-4; with DEN-2 and DEN-3 being the most prevalent serotypes

(Panhuis et al., 2010; Raheel et al., 1943). It has been shown that infection with one dengue serotype does not provide complete immunization to other serotypes (Gubler, 1998), and hence this provides the complicated challenge of producing a cure against all four serotypes.

Infections by dengue virus can be asymptomatic for most cases or may trigger a benign syndrome, dengue fever (DF) and more severe syndromes such as dengue hemorrhagic fever (DHS) and dengue Shock Syndrome (DSS) (Chawla et al., 2014; Libraty et al., 2002). The classic DF is characterized by self-limited dengue fever which is accompanied by non specific symptoms such as rashes, headache, nausea/vomiting, malaise, myalgia, retro-orbital pain, and arthrolgia, with the last three symptoms are also displayed in DHF/DSS conditions (Kalayanarooj, 2011). Other signs of DHF/DSS include systemic capillary leakage, thrombocytopaenia and hypovolaemic shock which may progress to death with improper or absence of treatment (Martina et al., 2009).

The exact mechanism of DHF/DSS remains unclear, although secondary infection with different serotype is believed to be the main factor (Thisyakorn et al., 2014). The prevalence of clinical manifestation of dengue is age-specific, with infants at greater risk being affected severely by DHF/DSS followed by children and adults (Hammond et al., 2005).

1.3.2 History of Dengue

Dengue disease occurrences increased dramatically following the ending of World War II and the urbanization that followed after (Sun et al., 2013). However, evidences suggested much earlier existence of interaction between dengue viruses and humans in the third century. A Chinese medical encyclopedia from Jin Dynasty

(265–420 AD) records a condition called “water poison” linked to flying insects, which is the first record of possible dengue case. Other records having similar descriptions were made during the 7th and 10th Century [Tang Dynasty (CE 610) and Northern Sung Dynasty (CE 992), respectively]. Clinical symptoms described in those reports including rash, fever, myalgia, and hemorrhagic manifestations (Weaver et al., 2013) .

Few centuries later, which coincide with traders travelling through sea, conditions mimicking dengue, were reported in other places like, French West Indies (1635), and Panama (1699) (Weaver & Vasilakis, 2013). A century later, the disease reached pandemic level by spreading to Batavia (present day Jakarta), Cairo, Philadelphia, and Cadiz and Seville, Spain. Shipping vessels allowed breeding and transportation of humans from one place to another, thus allowing for slow but progressive development of dengue viruses globally (Gubler et al., 2002), along with causing endemic intervals of 10-40 years (Murray et al., 2013).

Beginning of World War 2 brought significant changes in the way DENV spreads, so much that it called for scientific studies on the disease, its etiologic agents and development of diagnostic tests (Weaver & Vasilakis, 2013). The ending of World War 2 leads to uncontrolled urbanization and improper sewage management which contributed to *Aedes aegyptii* mosquitoes active breeding and spread of hyperendemicity within Southeast Asia (Anker & Schaaf, 2000). Early 80’s in America saw an increased DENV activity due to abandonment of *Aedes aegyptii* eradication programme. In Africa, dengue prevalence was recorded happening in 19^{th} century (Weaver et al., 2013), and was not detected thereafter until the year 1964 due to poor surveillance system (Causey et al., 1970).

1.3.3 Epidemiology of Dengue

Epidemiology of a disease is the study of the distribution and factors triggering a particular health issue and applying the knowledge on disease control strategies. Today, 40% of world's population (about 2.5 million people) in 100 countries around Asia, Americas, the Caribbean and Africa are at high risk of developing dengue (Hanafusa et al., 2008). The World Health Organization (WHO) estimated 50 to 100 million infections occur annually, with 500,000 cases of dengue hemorrhagic fever and 25, 000 deaths (Bentsi-enchill et al., 2013; Chokephaibulkit et al., 2013). In Australia, case fatality rate (%) has remained nil as reported within the year 2007 to 2011, with the lowest number of dengue cases as compared to Cambodia, Lao People's Democratic Republic, Malaysia, Philippines, Singapore and Vietnam within the same period. Among these 8 countries, Philippines recorded highest level of death cases (921 cases) with 187, 031 dengue cases in the year 2012 alone (Arima, Chiew, & Matsui, 2015).

For the year 2014 and 2015, Malaysia recorded 108,698 and 120,836 reported dengue cases respectively (Ghani, 2016). All four DENV serotypes were prevalent variably, at a given dengue endemic period in Malaysia. For example, in the year 2004, DEN-1 accounted to 73.4 % of the reported dengue cases and 58.6 % in 2005. DEN-2 was the predominant serotype in the year 2006 and 2007, at 36.4 % and 53.0 % respectively. The least common serotype which gives rise to more or less than 5 % of dengue virus isolated was DEN-4 (Mia et al., 2013; Mohd-Zaki et al., 2014). Recent report suggests serotype shifts from DEN-3 and DEN-4 to DEN-2 has caused surge in dengue outbreaks in the year 2013 (Ng et al., 2015).

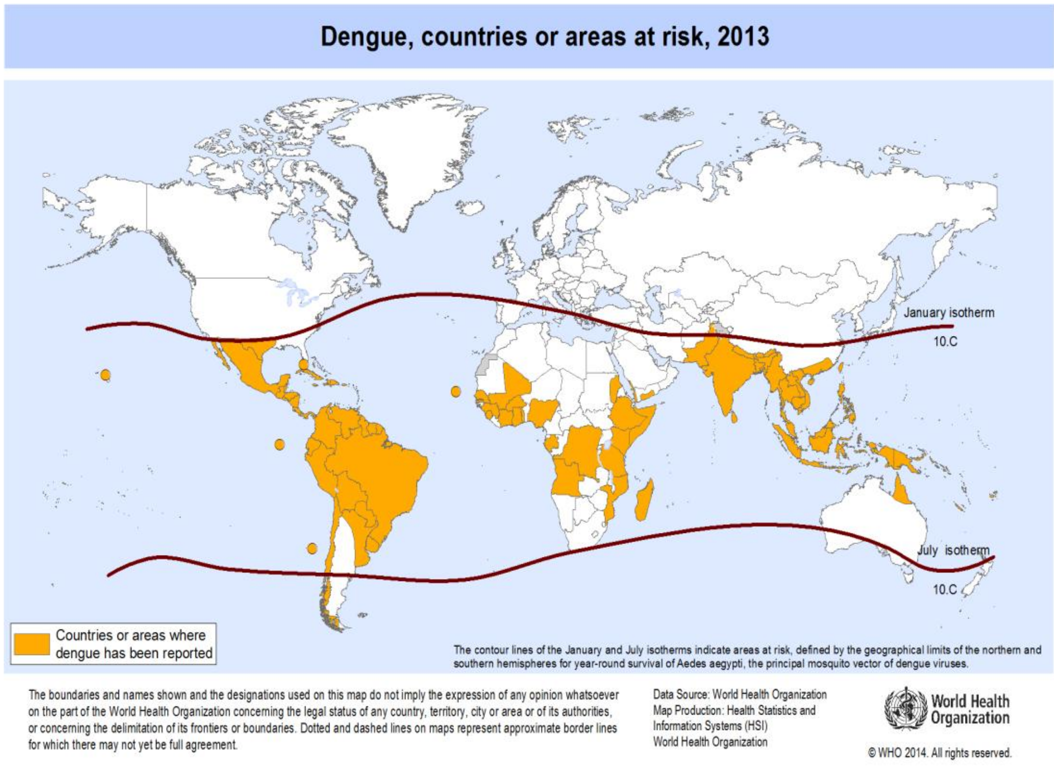

Figure 1.1: Countries at risk for dengue for the year 2013. Source: [(World Health Organization, 2014) http://www.who.int/ith/en/)]

By majority, dengue infections are caused by vector bites and in rare cases could be caused by transplant of organs and blood of infected donors ("Epidemiology," 2015). Seasonal increase of dengue cases occur at areas of tropics and subtropics, where heavy rainfall promotes optimal breeding sites for mosquitoes. Poor waste management and unreliable water supplies which prompt civilian to store water in containers further facilitated mosquito breeding (Monath, 1994). Coincidences of high density mosquito populations with high number of people not immune to one of the four serotypes (DEN-1, DEN-2, DEN-3 and DEN-4) contribute to dengue endemics at a particular region. Dengue cases remained restricted until middle of 20^{th} century before becoming a global threat (Murray et al., 2013).

1.3.4 Morphology and Life Cycle of Dengue Virus

1.3.4.1 Overview

A mature dengue virus is roughly spherical in shape at a diameter of about 500 Å (Zhang et al., 2003). The family of *Flaviviridae* consists of three genera including flavivirus, pestivirus and hepaciviruses, with dengue belonging to genus flavivirus. Other viruses belonging to genus flavivirus are West Nile virus (WNV), yellow fever virus (YFV), tick-borne encephalitis virus (TBEV), Murray Valley encephalitis virus (MVEV), Kadam Virus (KADV), and Ngoye virus (NGOV) to name but a few (Bollati et al., 2010; Mukhopadhyay et al., 2005)

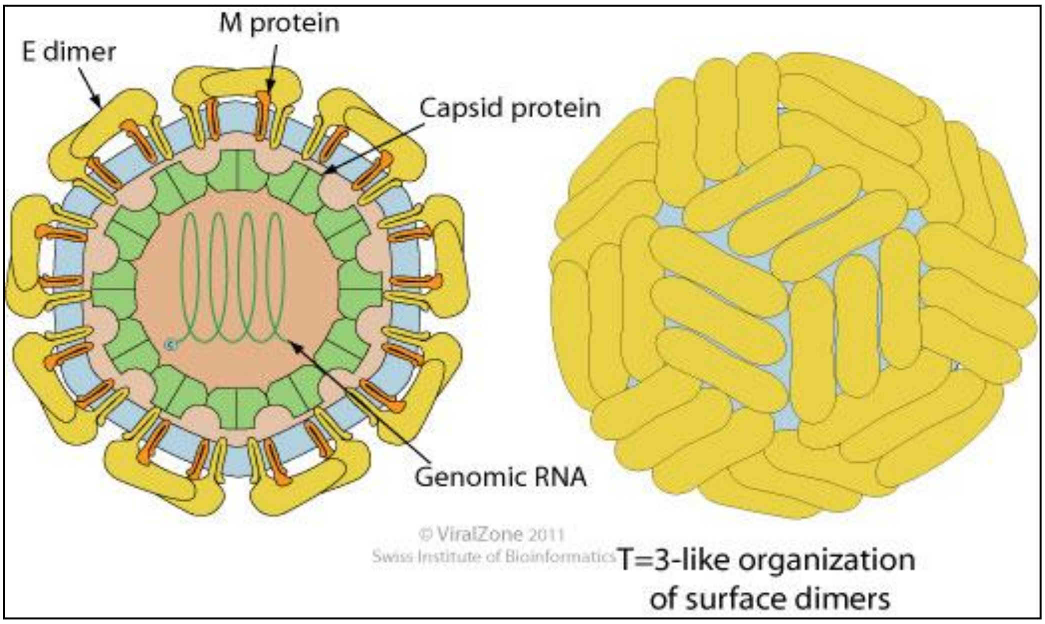

Figure 1.2: Cross-section of diagram of a flavivirus virion. Three proteins (E, M and C) make up the virus's membrane. E-protein assumes a herring-bone arrangement as depicted in the diagram on the right. Source: (Swiss Institute of Bioinformatics, SIB [http://www.expasy.org/viralzone])

1.3.4.2 Virus Genomic RNA

The virus genome is composed of a single strand, of an approximate 11 kb of positive sense ribonucleic acid (RNA) molecule. The RNA genome is a single open reading frame encoding 3,391 amino acid residues which make up for the three structural proteins (C, prM, and E) and non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (Zuo et al., 2009).

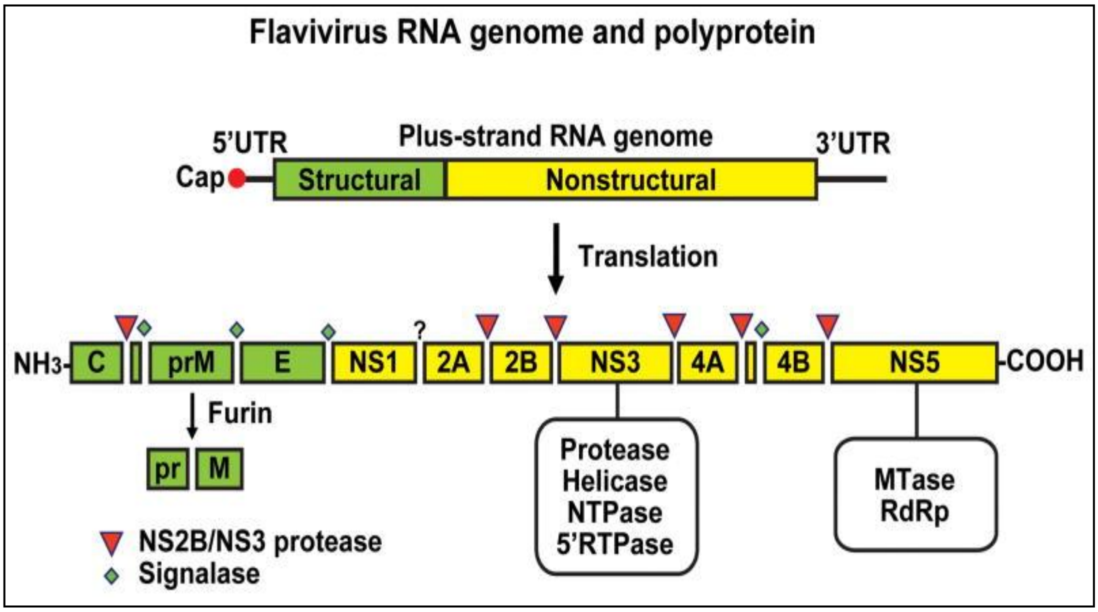

Figure 1.3: Schematic representation of flavivirus genome organization and polyprotein processing. Source: (Sampath et al., 2009)

1.3.4.3 Structural Proteins

The three structural proteins ; capsid (C), envelope (E) and membrane (M) along with lipid bilayer encloses the interior genomic RNA of the flavivirus (Modis et al., 2003). A mature virus particle is enveloped by 180 envelope (E) glycoprotein molecules attached to an equal number of lipid membrane (M) protein layer (Lok et al., 2012). The E protein crystal structure reveals three domains; the structurally central N-terminal domain I, dimerization domain II, and C-terminal, immunoglobulin-like domain III (Pokidysheva et al., 2006; Zhang et al., 2003). Both

E and M proteins are associated to the host-derived lipid bilayer. Interior of the lipid bilayer is the nucleocapsid core consisting of capsid (C) proteins that encompasses the flavivirus RNA genome (Jones et al., 2003). It is suggested that the nucleocapsid core bears a lower density as compared to that located at the outer glycoprotein shell, which further suggest that the structure of nucleocapsid core is poorly ordered or has a variable orientation in relative to the glycoprotein shell (Kuhn et al., 2002). Each component of the structural proteins in the flavivirus is critical for propagation. The E protein, which is considered as class II fusion protein, mediates viral attachment via cellular receptors and fusion with the endosomal membrane, thus enabling virus entry (Crill et al., 2001). The creation of nucleocapsid following the association of genomic RNA and capsid proteins is not clearly understood. However, it is shown that in the absence of capsid protein, virus like particles (VLPs) which are produced lack the RNA, rendering them non-infectious. Thus, nucleocapsid core, in some way is critical to the propagation of infectious flaviviral particles, while suggesting early interaction of the C proteins with the genome RNA, during the viral assembly process (Jones et al., 2003).

The membrane (M) protein is a product of a polyprotein which was first cleaved into precursor membrane (prM) and E proteins. Immature virus bears the prM linked to E proteins while in a neutral pH environment within the endoplasmic reticulum. During maturation, within the trans Golgi network, the precursor (pr) would dissociate from the membrane (M), along with dimerisation of E proteins (Zhou et al., 2014) The pr portion of the prM, helps masks the E from premature fusion while it is going through the acidic trans Golgi network (Stadler et al., 1997). Removal of the pr is done by furin cleavage activity; an event that directs rearrangement of the E proteins and induces virus infectious process (Zhang et al., 2004).

1.3.4.4 Non Structural Proteins

The polyprotein precursor can be processed either co-translationally or post-translationally, mediated by host signalase located within the cellular endoplasmic reticulum membranes or by virus encoded proteases (Zhang et al., 1992). Apart from structural proteins, the single polyprotein is also processed into seven non-structural proteins known as ; NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Chambers et al., 1990) .

The NS1 is a glycoprotein with a mass of 43-48 kDa, which is expressed intracellularly and has been shown to play a role in the flaviviral RNA replication (Amorim et al., 2014; Flamand et al., 1999; Mackenzie et al., 1996; Rice et al., 1997). The carboxyl (C) - terminal region of the envelope glycoprotein encodes a hydrophobic signal sequence that prompts the translocation of NS1 into the endoplasmic reticulum, where it then dimerizes rapidly (Falgout et al., 1989). The NSI protein would then proceed to cell surface where it becomes membrane associated (Amorim et al., 2014; Jacobs et al., 2000).

Through immunofluorescence and cryo-immuno electron microscopy studies, it was revealed that NS1 colocalize with the dsRNA and other components of the replication complexes (Mackenzie et al., 1996; Westaway et al., 1997). This observation supports the claim of NS1 protein as a cofactor in viral replication (Khromykh et al., 2000).

Several hypotheses noted the importance of NS1 protein in causing autoimmune processes and disruption of circulatory system, due to cross reactive antibodies. This event leads to decrease of platelet count, endothelial cell apoptosis, complement activation and then host cell damage (Chen et al., 2009; Falconar et al., 2011; Kurosu et al., 2007; Lin et al., 2002; Martina et al., 2009).

Following NS1 is the flaviviral NS2 protein which is made up of NS2A and NS2B. NS2A is a 22-kDa hydrophobic protein (Xie et al., 2013) and is made up of 224 amino acids from the cleavage of NS1-NS2A and NS2A-NS2B. Its N and C termini produced within the ER catalyzed by host signalases and within the cytoplasm by viral proteases, respectively. The internal cleavage by NS2B-NS3 serine protease generates a truncated form of NSA, known as, NS2α (Kümmerer et al., 2002). The NS2A also co-localizes with replication complexes, suggesting its role in viral RNA synthesis (Mackenzie et al., 1998). This process is still not well understood by researchers.

The flaviviral NS2B complexes with NS3 (Cahour et al., 1992; Chambers et al., 1991). The NS2B-NS3 has been given much attention as a suitable drug target for the past few decades. The NS2B domain, is of approximately 14kDa (Chambers et al., 1991), bearing a central conserved hydrophilic domain which is flanked by two hydrophobic domains at the N-terminus and one hydrophobic domain at the C terminus (Clum et al., 1997; Yusof et al., 2000).

It was discovered by Clum that the central hydrophilic domain of NS2B consisting of 40 amino acids was the most optimal and sufficient for the activation of NS3 protease (Clum et al., 1997; Noble & Shi, 2012). Even though hydrophobic domains of NS2B are dispensable for protease activity, it was indicated by NS2B hydrophobic domains deletion analysis that these domains play a role in cotranslational membrane insertion of the full NS2B protein, in order for NS3pro activation (Clum et al., 1997).

NS3 is a multi-functional protein (69 kDa), which is also as crucial in polyprotein processing and RNA replication. In its N terminus domain, the NS3 bears a trypsin-like serine protease domain (180 amino acid residues), of which its activity is contributed by non-covalent interaction with the 40 amino acid hydrophilic domain of the membrane-bound NS2B. The C terminal of the NS3 is made up of nucleotides and RNA binding motifs with RNA helicase, 5-nucleoside triphosphatase (NTPase), and RNA 5-triphosphatase (RTPase) activities (Li et al., 2005; Luo et al., 2008; Xu et al., 2005; Yusof et al., 2000).

NS2B–NS3 has been shown to cleave at the cleavage junctions between NS2A/2B, NS2B-NS3, NS3/NS4A, NS4A/NS4B, and NS4B/NS5, in addition to producing C termini of mature Capsid (Arias et al., 1993; Chambers et al., 1990; Falgout et al., 1989; Preugschat et al., 1990; Wengler et al., 1991; Zhang et al., 1992b).

It was only of recent years that research studies have been focused on the actual role of NS4A protein in viral replication. In a paper published in 1998 by MacKenzie and others, it was revealed through observation on cells infected by flavivirus named Kunjin virus (KUNV), that NS4A colocalizes within the vesicular packets (VP), suggesting its role in replication by targeting or anchoring within replication complex (RC). However, detailed information to understand this process was lacking during that period (Mackenzie et al., 1998). Apart from this, it was indicated that viral replication also owes to the interaction between NS1 and NS4A (Lindenbach et al., 1999)

NS4A is a 16 kDa hydrophobic protein, with its initial residues (residues 1 to 49) function as the cofactor for NS3 helicase (Shiryaev et al., 2009). Meanwhile, the subsequent regions (residues 50 to 73, residues 76 to 89, and residues 101 to 127) possess hydrophobicity, are membrane associated and do not interact with NS3. Also present within the NS4 is a small loop that exposes NS4-2k cleavage site, along with the C-terminal segment known as 2k, which acts as signal sequence that would direct translocation of NS4B towards the ER lumen (Miller et al., 2007; Shiryaev et al., 2009). NS4A in association with the other viral and host proteins triggers membrane rearrangements needed during viral replication (McLean et al., 2011; Roosendaal et al., 2006).

The NS4A has also been recently proven as a stronger determinant in viral replication by inducing autophagy, thereby protecting host cell death – a requirement for successful infection process. However, the mechanism involved in the regulation of autophagy by NS4A protein is still yet to be determined (McLean et al., 2011)

NS5, is the largest flaviviral protein (100 kDa) , multifunctional and bearing well conserved domain(Bollati et al., 2010). At its N terminus, the NS5 bears the S-adenosyl-L-methionine-dependent methyltransferases, whilst at its C terminus places the RNA-dependent RNA polymerase (RdRp) domain functioning in mRNA capping – a process vital for viral replaication (Botting et al., 2012; Egloff et al., 2002).

1.3.4.5 Life Cycle of Flavivirus

Flavivirus depends on mosquitoes as the primary vectors in transmission (Lindenbach et al., 2007). Many studies have been done to understand the interaction between flavivirus and mosquitoes. Mutagenesis study has proposed few residues within the hinge region of the DENV which are critical in its infection process with the vector. In another study, it was found that the loop motif between F and G beta strands (FG loop) within the domain 11 of E protein is a determinant in the binding with mosquito cells while the binding of the virus to mammalian cells is suspected to be independent of this FG loop (Hung et al., 2004). However, in a later study, it was revealed that the FG loop is as important in infection with mammalian cells as well (Erb et al., 2010). As an infected mosquito bites a human host, the virus is orally transmitted and enters the cell via receptor mediated endocytosis (Stiasny et al., 2006). A change in environmental pH within the cell encourages the release of genomic RNA in the cell's cytoplasm (Clyde et al., 2006). Following this, the RNA is translated into polyprotein precursor which would be cleaved in to its structural (C, E, prM) and non-structural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) proteins (Henchal et al., 1990). Virions are assembled and pass through the ER and merge with its cleaved structural proteins (prM and E) to mature and bud off from the cell via exocytosis. Sometimes, immature viral particles that are lacking nucleocapsid could escape as well as normal by products during viral assembly (Mukhopadhyay et al., 2005)

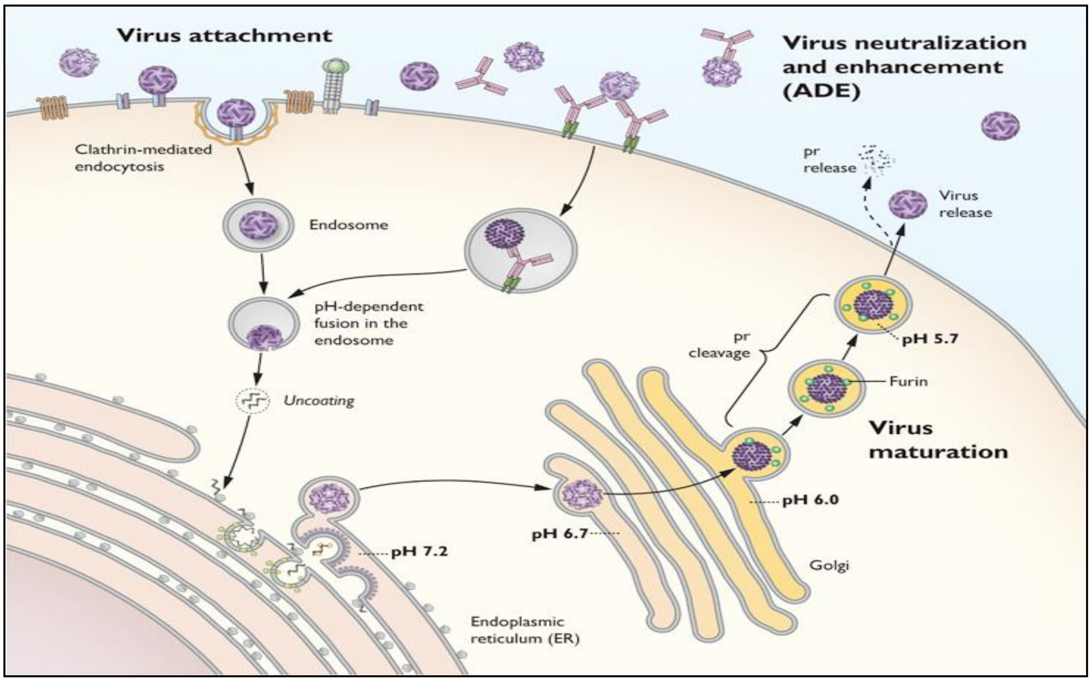

Figure 1.4: Illustration of flavivirus replication pathway at different pH conditions. Source : (Pierson et al., 2012)

Virus Entry to Cell

A number of human host cells including the macrophages, monocytes, Langerhans cell and dendritic cells are target for DENV infection. DENV has been proposed to attach to host cells through the commonly expressed glycosaminoglycan heparan sulfate (Chen et al., 1997; Crance et al.,, 2002), and Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN) for immature dendritic cells (Tassaneetrithep et al., 2003). The virus then enters via receptor-mediated endocytosis and proceeds to endosomes. Acidic pH condition in the endosome causes structural change of E protein in which the homodimeric form of E proteins start to dissociate and its monomer rearranges in a way that promotes the fusion of viral membrane with the endosomal membrane (Zaitseva et al., 2010).

Translation and Polyprotein processing

The RNA molecule which has been released is then translated into a single polyprotein within the ER-derived membranes (Clyde et al., 2006). Proteases derived from host and virus aid the processing of the single polyprotein into ten proteins inclusive of structural and non structural proteins (Perera et al., 2008). Host derived enzymes known as peptidases are responsible for the cleaving of structural proteins while virus-derived serine protease aids in cleavage between non-structural proteins (Lindenbach et al., 2007).

RNA Replication

With the release of NS5 protein, the viral RNA is transcribed from 3' end resulting in minus strand RNA (Henchal et al, 1990; Lindenbach et al., 2007). This minus strand RNA is transcribed back to plus strands RNA, resulting in transient intermediate dsRNA. The dsRNA is separated to allow NS5 polymerase to bind and initiate RNA synthesis (Lescar et al., 2008). The separation or unwinding of intermediate dsRNA is believed to be triggered by RNA helicase activity of NS3 protein (Sampath et al., 2006). The NS5 associates with promoter region located within 5' end of genome, there by initiating RNA synthesis at 3' end through long range RNA-RNA interactions (Filomatori et al., 2006).

Viral assembly and Release

The synthesized viral RNA translocates to cytoplasm and thereafter assembled with other virus particles within rough ER lumen (Uchil et al., 2003). Prior to assembly, viral RNA is encapsulated with C protein (Perera et al., 2008) . This is followed by E and prM proteins arrangement around nucleocapsid, forming an immature virus particle (Mackenzie et al., 2001). This particle then exits from the rough ER lumen and enters the Golgi, where the virus particles mature. Virus maturation is performed by furin which cleaves prM to M along with structural re-arrangements of E protein. The mature virus particles then exits the host cell by exocytosis (Mukhopadhyay et al., 2005).

1.4 NS2B-NS3 protease and its active site

NS2B-NS3 protease is a key virus-encoded domain crucial in processing polyprotein precursor, evidently stressing the role of NS2B-NS3 protease in the viral replication. Hence, it is a reliable and promising therapeutic target in drug discovery efforts (Jang et al., 2015). Active site of NS2B-NS3 lies within the NS3 protease domain (Salaemae et al., 2010). In similarity with other flaviviral systems, three residues (His51, Asp75 and Ser135) forming the catalytic triad have been proven to be crucial in conferring protease activity of the serine protease, of which when removed in *in vitro* experimental studies, diminished the functionality of the enzyme (Falgout et al., 1998). Noble and his team have reported the event of NS2B forming a β hairpin structure which folds around the NS3 protease, leading to the formation of active, closed conformation (Noble et al., 2012). The hydrophilic domain of NS2B (residues 49-95) is said to be fused to NS3 protease via a Gly4-Ser4-Gly4 linker, leading to active protease (Leung et al., 2001). Some of the residues residing within the C-terminus of NS2B region were implicated by mutagenesis study to contribute to proteolytic activity which includes L74, I76 and I78 (Niyomrattanakit, et. al., 2004).

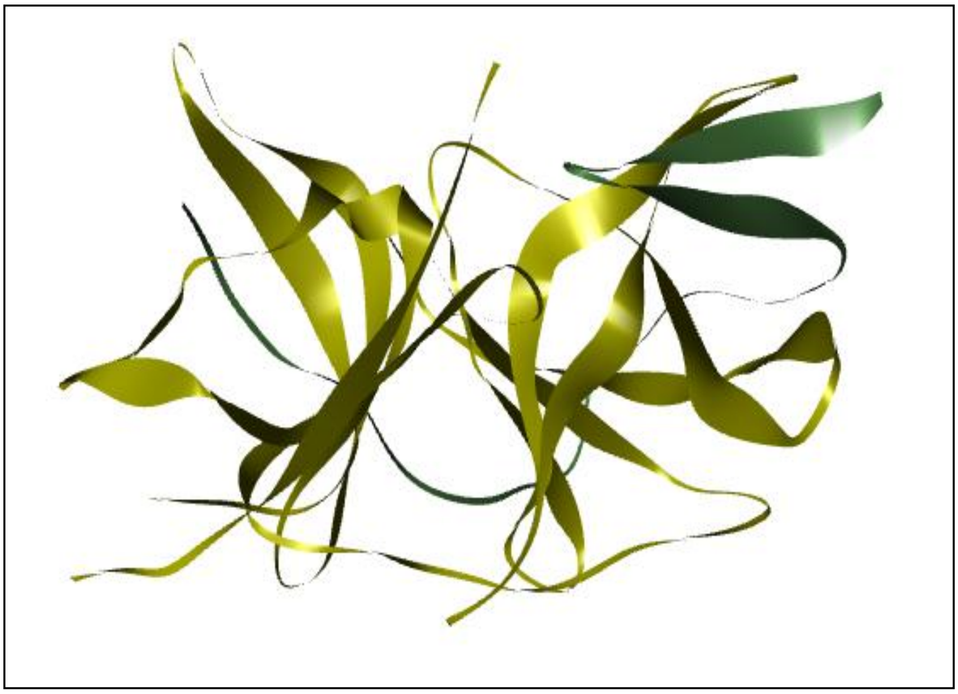

Figure 1.5: Ribbon representation of NS2B-NS3pro.(Noble et al., 2012)

Li and team have demonstrated that all the four serotypes of dengue proteases share similar substrate specificities, by incorporating tetrapeptide substrate, benzoyl-norleucine (P4)-lysine (P3)-arginine (P2)-arginine (P1)-ACMC (Bz-Nle-Lys-Arg-Arg-ACMC). It was concluded that at P1 and P2 positions, dibasic residues were preferred while at P3 and P4 positions, basic or aliphatic residues were preferred. The substrate binding pockets in NS3 protease is lined by highly conserved residues which spans within the S1 to S4 region (Li et al., 2005) .

Mutagenesis studies revealed role of other residues apart from the catalytic triad being crucial in the protease activity. GLY151 was suggested to aid in stabilizing the tetrahedral position formed at Ser135 along with that of the E2-F2 strands in the protease fold (Salaemae et al., 2010). ASN152 is located at the S2 subsite and it forms hydrogen bonding with the side chain of the P2. GLY133 is found to be an important part of NS3 sequences, which determines ideal conformation for substrate binding within the oxyanion hole. TYR150, on the other hand, is said to stabilize placement of the P1 via pi-cation interaction along with stabilizing E2 strand of the C-terminal β-barrel in NS3 protease. Mutation at SER163 too inactivates the enzyme (Chappell et al., 2005), and is proposed to line with GLY153 forming bulky entry at the binding site, and was also suggested to stabilize substrate binding via a hydrogen bond with P1 arginine (Salaemae, et. al., 2010).

Proteases have been aimed as therapeutic target, and such effort has produced success stories in search for HIV-1 protease inhibitors. Hence, it is believed that aiming dengue protease to inhibit DENV replication can be considered as a valid therapeutic target (Salvesen et al., 2010).

1.5 Computer-aided drug discovery

The whole process of drug discovery and development is often synonymed with searching for a needle in a haystack. It takes as long as 17 years and cost nearly 800 million US dollars from lead identification to clinical trials (Cerqueira et al., 2015). Given the limited amount of drugs reaching clinical trials compared to a huge amount at the initial stages, the resources spent on the whole drug discovery cycle is monumental. Prior to lead optimization, a myriad of stages supersede which includes chemical synthesis, extractions, compound isolations and *in vitro* screenings to identify hits against a target protein. Hits identification alone consumes so much to time, money and human capitals.

Early 1980s saw an interest in computer-aided drug discovery (CADD) as exposed by a cover article of Fortune magazine titled “The Next Industrial Revolution: Designing drugs by computer at Merck” (Drie, 2007). It was also mentioned by Green from GlaxoSmithKline, “The future is bright. The future is virtual”; implying the prominence and growing importance of computational tools in

R&D of pharmaceutical industries. (Kapetanovic, 2008). Post genomic period witnessed abundance of information of small molecules and protein crystal structures, which enables a wide application of CADD. This caused an inevitable integration of CADD as part of drug discovery pipeline. (Jorgensen et al., 2004)

In silico filters such as those function to eliminate redundant compounds (poor absorption, distribution, metabolism, excretion and toxicity, ADMET) are available that aids focusing on the more promising drug targets (Tan et al., 2006). Rational design of drugs that could bind to a target protein is also another application within CADD; an approach which significantly saves more time.

1.5.1 Virtual screening of ligand libraries

The objective of receptor-based virtual screening is to search for ligands from libraries added with prediction of respective binding affinities and conformation against the protein of interest (Lyne, 2002). A number of programs which can execute such calculations include DOCK (Ewing et al., 2001), FlexX (Rarey et al., 1996), GOLD (Jones et al., 1997) and AutoDock (Morris et al., 2009; Trott et al., 2011). This study focuses on usage of AutoDock for virtual screening, since it has been established as a reliable tool since 2010 for docking predictions with an added advantage of being a free source program.

1.5.2 AutoDock

AutoDock program is suite software developed by Morris and team to facilitate prediction of binding modes between macromolecules and drug-like ligands, by employing semi-empirical free energy force field. This technique of computational calculation allows for the prediction of free binding energies along with binding constants for the docked ligand (Morris et al., 1998). The model which was applied in the calculation of free binding energies is as below;

$$\Delta G = \Delta G\ vdw + \Delta G\ hbond + \Delta G\ elec + \Delta G\ conform + \Delta G\ tor + \Delta G\ sol$$

The first four pairwise calculations relate to dispersion/repulsion, hydrogen bonding, electrostatics and deviation from original conformation, torsional entropy and desolvation respectively (Morris et al., 1998). AutoDock uses grid-based method, in which rapid evaluation of binding energies of trial conformations are calculated and stored in a grid file to be used as a look up table by AutoDock during docking simulation. AutoDockTools (ADT) was created as part of graphical user interface, which enables user to prepare coordinate files, perform experimental design and perform data analysis (Morris et al., 2012).

1.5.3 AutoDock Vina

AutoDock Vina was introduced as the new software for molecular docking and virtual screening by The Scripps Research Institute. Its speed is of two orders magnitude in comparison to AutoDock 4, by utilizing multithreading in multi-core machines. Vina is also compatible with AutoDock tools, and utilizes similar input file formats as required by AutoDock 4. However, users would not need to perform grid map calculations, as Vina performs this task along with the clustering and ranking of results in a way not visible to the users (Trott & Olson, 2011) .

It is believed that Vina is able to comprehend more complicated ligands (large ligands, ligands having larger number of rotatable bonds) in comparison to AutoDock 4 (Chang et al., 2010). In regard to this, this study makes use of Vina to perform molecular docking and virtual screening