THEOBROMINE-BASED SYNTHETIC DERIVATIVES AS POTENTIAL INHIBITORS OF DENGUE VIRUS II SERINE PROTEASE

FADI GHASSAN MOHAMMED SAQALLAH

UNIVERSITI SAINS MALAYSIA

2023

THEOBROMINE-BASED SYNTHETIC DERIVATIVES AS POTENTIAL INHIBITORS OF DENGUE VIRUS II SERINE PROTEASE

by

FADI GHASSAN MOHAMMED SAQALLAH

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

February 2023

To whom used to call me their soulmate ...
To whom I spent hours talking to while they weren't really there ...
To whom I wished them to be here ...
To pray for me ...
Instead, I am praying for my soul's second half ...
Then, now, and forever ...

To
Mom & Dad

and to
Maywan Hariono
you were loved

ACKNOWLEDGEMENT

First and foremost, I would like to thank *God* for the patience *He* has given me throughout my entire life, and especially through my study. Without your mercy, I would have been lost. Without your blessings, I would not have been the man I am today. Thank you, *Allah*, for everything. Give thanks to *Allah*, for *He* is good. To *Him* who alone does great wonders. Who by *His* understanding made the heavens. Who spread out the earth upon the waters. Who made the great lights, and the sun to govern the day. Who made the moon and stars to govern the night; *His* love endures forever.

I would like to thank my father and mother for their continuous support. You both raised me well when I was little and always encouraged and loved me as I am older, without boundaries. I hope I will be able to return even if the smallest part of what you offered me.

I am deeply grateful to my supervisor, Professor Dr. Habibah A. Wahab, for giving me the opportunity to start my doctorate journey. With you, I have learned a lot, and I am still learning every day from the ocean of experience and wisdom you got. You taught me how to survive difficulties and supported me whenever needed. I also would like to thank my co-advisors, Dr. Belal O. Alnajjar and Dr. Yap Beow Keat, for the patient guidance, continuous monitoring of my progress, and encouragement they have provided in the past years.

I would also like to thank Mrs. Qamar Atari Al-Masri whom supported me through the beginning of my journey. I will always remember your support, and I pray to *God* that I will be able to offer the same to anyone whom deserves it in the future. You taught me what giving without asking means.

Finally, I would like to thank Mr. Sami Karam, Mr. Abdulsalam Q. Almashhadani,

Dr. Raghdaa Al-Zarzour, Mr. Salah Al-Shehade, and Dr. Anas Al-Shishani (from the *School of Chemical Sciences, USM*) for their support and guidance throughout the synthesis part of this study. Mrs. Nurul Hanim Salin, Dr. Iffah Izzati Binti Zakaria (both from the *National Institutes of Biotechnology Malaysia (NIBM)*), and Dr. Nadeem A. Abdalrazaq for their help and support throughout the *in-vitro* investigation. Finally, I would like to thank Dr. Kesevan R. Kumaran, Dr. Maywan Hariono, Dr. Selestin Rathnasamy, Dr. Ahmad Naoras Bitar, and Mr. Kinan Keshkeh for their support, encouragement, and being true friends throughout my journey. You all are life saviours, and you were there whenever I needed your help.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS AND UNITS

Å	Angstrom
°C	Celesius degrees
°K	Kelvin degrees
Da	Dalton
g	Gram
$g.kg^{-1}$	Gram per kilogram
$g.L^{-1}$	Gram per litre
$g.mol^{-1}$	Gram per mol
hr	Hour(s)
Hz	Hertz
kb	Kilobase
kg	Kilogram
mg	Milligram
MHz	Mega Hertz
min	Minute(s)
mL	Millilitre
mol	Mole
N	Normal (normality) concentration
ns	Nanosecond
µM	Micromolar
w/v	Weight per volume per cent concentration
w/w	Weight per weight per cent concentration
α	Alpha
β	Beta
γ	Gamma
λ_{em}	Emission swavelength
λ_{ex}	Excitation wavelength
δ_C	Carbon shift in the NMR spectrum
δ_H	Proton shift in the NMR spectrum

LIST OF ABBREVIATIONS

[*x*]	Concentration of *x*
1/[S]	Reciprocal of substrate concentration
1/*V*	Reciprocal of reaction velocity
^{13}C	Carbon 13 isotope
^{1}H	Proton
ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity pharmacokinetic descriptors
ALogP	Ghose-Crippen-Viswanadhan octanol-water partition coefficient
AM1-BCC	AM1-bond charge correction charge method
AMBER MD	Assisted Model Building with Energy Refinement molecular dynamics package
AMC	7-amino-4-methylcoumarin
AMU	Atomic mass unit
ATPase	Adenosine triphosphatase
BBB	Blood-brain barrier
BOC	Butyloxycarbonyl
BSA	Bovine serum albumin
CAC	Chloroacetyl chloride
CADD	Computer-Aided Drug Design
CDC	Centers for Disease Control and Prevention
CNS	Central nervous system
COX-1	Cyclooxygenase-1
CYP2D6	Cytochrome P450 2D6
D_2O	Deuterium oxide
DCM	Dichloromethane
DENV	Dengue / dengue virus
DF	Dengue fever
DHF	Dengue haemorrhagic fever
DMF	Dimethylformamide

DMSO	Dimethyl sulphoxide
DMSO-d_6	Deuterated dimethyl suphoxide
DSC	Differential scanning calorimetry
DSS	Dengue shock syndrome
DW	Distilled water
E	Envelope protein of DENV
EC_{50}	Half maximal effective concentration
ESP	Electrostatic potential
Et_3N	Triethylamine
Et_3N.HCl	Triethylamine hydrochloride salt
EtOAc	Ethylacetate
EtOH	Ethanol
ff14SB	force field 14 Stony Brook
FT-IR	Fourier-transform infrared spectroscopy
GAA	Glacial acetic acid
GAFF	General AMBER Force Field
GPU	Graphics processing unit
H_2O	water / aqueous
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
HDL	High-density lipoprotein
IC_{50}	Half maximal inhibitory concentration
J coupling	Scalar coupling
K_i	Inhibition constant
K_m	Michaelis-Menten constant
LC-MS	Liquid Chromatography-Mass Spectrometry
LD_{50}	Median lethal dose
LDL	Low-density lipoprotein
LEaP	Link, Edit, and Parm programs
LEB	Lowest energy of binding
M	Membrane protein of DENV
M.wt.	Molecular weight
m/z	mass-to-charge ratio

MeOH	Methanol
MeOH-d_4	Deuterated methanol
$MgSO_4$	Magnesium sulphate
MM-GBSA	Molecular mechanics with generalised Born and surface area
MM-PBSA	Molecular mechanics Poisson-Boltzmann surface area
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NMR	Nuclear magnetic resonance
NPT	Isothermal-isobaric ensemble
NS1	Non-structural protein 1
NS2A	Non-structural protein 2A
NS2B	Non-structural protein 2B
NS2B/NS3pro	The NS2B/NS3 protease receptor
NS3pro	Non-structural protein 3 protease
NS4A	Non-structural protein 4A
NS4B	Non-structural protein 4B
NS5	Non-structural protein 5
NSAIDs	Non-steroidal anti-inflammatory drugs
NVE	Microcanonical ensemble
NVT	Canonical ensemble
ORF	Open reading frame
PDB	Protein Data Bank
PMEMD	Particle Mesh Ewald Molecular Dynamics
POVME	POcket Volume MEasurer
prM	Membrane precursor
PSA	Polar surface area
RadGyr	Radius of gyration
R_f	Retention factor
RFU	Relative fluorescence unit
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
rpm	Round per minute

RT	Room temperature
RTPase	RNA tri-phosphatase
SANDER	Simulated Annealing with NMR-Derived Energy Restraints
SASA	Solvent-accessible surface area
SDF	Structure-data file format
TBDMS-Cl	*tert*-butyldimethylsilyl chloride
TBDMS	*tert*-butyldimethylsilyl
TLC	Thin-layer chromatography
Tris	Trisaminomethane
Tris-HCl	Trisaminomethane-hydrochloric acid buffer solution
TxA_2	Thromboxane A_2
vs.	versus
WHO	World Health Organization
$ZnBr_2$	Zinc bromide
ΔE_{EEL}	Electrostatic energy
ΔE_{GB}	Polar Solvation Energy using generalised Born approach
ΔE_{nplr}	Non-polar solvation energy using generalised Born approach
ΔE_{PB}	Polar Solvation Energy using Poisson–Boltzmann approach
ΔE_{surf}	Non-polar solvation energy using Poisson–Boltzmann approach
ΔE_{vdW}	van der Waals energy
ΔG_{bind}	Total energy
ΔG_{gas}	Total gas phase free energy
ΔG_{solv}	Total solvation free energy

LIST OF APPENDICES

TERBITAN SINTETIK BERASAKAN TEOBROMINA SEBAGAI PERENCAT BERPOTENSI TERHADAP PROTEASE SERINA VIRUS DENGGI II

ABSTRAK

Denggi merupakan penyakit viral RNA yang tersebar di seluruh daerah perbandaran tropika dan subtropika. Sehingga Jun 2022, terdapat lebih daripada 17,497 kes jangkitan denggi telah dilaporkan di Malaysia sejak awal tahun ini, yang dianggap peningkatan sebanyak 57.6% berbanding enam bulan pertama 2021. Sehingga kini, tiada rawatan berlesen penuh untuk pengurusan jangkitan denggi. Virus denggi (DENV) menggunakan protease NS2B/NS3 untuk membelah poliprotein sepanjang replikasi dan kitaran hayat virus. Oleh itu, perencatan protease virus boleh menyekat replikasi DENV. Sebelum ini, beberapa data yang tidak diterbitkan oleh kumpulan kami menunjukkan bahawa *Theobroma cacao* menghalang aktiviti protease DENV2. Theobromine ialah alkaloid semulajadi, yang biasa ditemui dalam *T. cacao*. Theobromine mempunyai perancah kimia purin yang menyerupai 19 antivirus terkenal yang sama ada tersedia secara komersial atau di bawah penyiasatan klinikal, termasuk asiklovir. Beberapa kajian telah melaporkan sintesis theobromine dan derivatifnya, tetapi sehingga kini tiada yang berkaitan sebagai agen antivirus.. Dalam kajian ini, sekumpulan 137 molekul hibrid berasaskan teobromina telah direka bentuk dengan menggabungkan teobromina (pada kedudukan °1; amina sekunder) dengan entiti kimia dan sebatian-sebatian lain yang boleh didapati secara komersial. Konjugat yang direka bentuk telah dikaji secara *in-siliko* untuk pengikatan dan interaksinya di tapak pengikatan pemangkin NS2B/NS3pro menggunakan dok molekul dan dinamik molekul. Molekul-molekul yang menunjukkan pengikatan dan interaksi yang tertinggi

telah disintesis, dicirikan, dan diuji secara *in-vitro* untuk aktiviti perencatan dan jenis perencatan terhadap protease. Hasil kajian menunjukkan bahawa konjugasi teobromina dengan moeiti kimia lain melalui amina sekundernya telah meningkatkan pengikatan afiniti, kestabilan, dan interaksi teobromina pada tapak pemangkin protease secara signifikan. Pemprofilan ADMET menunjukkan bahawa kesemua molekul yang direka dijangkakan selamat, bukan hepatotoksik, pengikat non-CYP2D6 dan pengikat protein non-plasma. Pengiraan tenaga MM-PB(GB)SA menunjukkan bahawa sebatian C30a dan C78f mempunyai tenaga bebas pengikatan terendah iaitu -30.13 dan -37.14 kcal.mol^{-1}, manakala teobromina pada -0.90 kcal.mol^{-1}. Seterusnya, tiga belas sebatian telah disintesis menggunakan tindak balas berbilang langkah dan dicirikan menggunakan DSC, FT-IR, LC-MS, dan ^{1}H & ^{13}C NMR. Penemuan kajian dinamik molekul adalah bersetuju dengan keputusan *in-vitro*, di mana C30a merencat protease dengan IC_{50} sebanyak 90.19 μM secara kompetitif. Namun, C78f mampu menghalang protease virus dengan IC_{50} sebanyak 22.54 μM tetapi dengan cara yang tidak kompetitif. Manakala, IC_{50} teobromina tidak dapat ditentukan kerana ia hanya menunjukkan 19.46% perencatan terhadap protease pada 500 μM. Kesimpulannya, konjugasi teobromina dengan moeiti kimia lain melalui penghubung amida boleh meningkatkan aktivitinya terhadap DENV NS2B/NS3 protease *in-siliko* dan *in-vitro*. Jenis moeiti kimia yang juga dikonjugasi mempengaruhi keseluruhan aktiviti molekul yang direka bentuk dan jenis perencatan yang ditunjukkannya.

THEOBROMINE-BASED SYNTHETIC DERIVATIVES AS POTENTIAL INHIBITORS OF DENGUE VIRUS II SERINE PROTEASE

ABSTRACT

Dengue is an RNA viral disease that is spread across tropical and subtropical urbanised districts. Up to June 2022, more than 17,497 cases of dengue infections have been reported in Malaysia since the beginning of the year, which is considered an increment of 57.6% compared to the first six months of 2021. Moreover, up to date, there is no licensed drug treatment for the management of dengue infections. Dengue virus (DENV) uses the NS2B/NS3 protease to cleave its polyprotein throughout the virus's replication and life cycle. Hence, inhibiting the viral protease can suppress the replication of DENV. Previously, some unpublished data by our group showed that *Theobroma cacao* inhibited the DENV2 protease. Theobromine is a naturally occurring alkaloid commonly found in *Theobroma cacao*. Theobromine has a purine chemical scaffold which resembles that of 19 famous antivirals, which are either commercially available or under clinical investigations, including acyclovir. Moreover, only a few studies have reported the synthesis of theobromine and its derivatives, but to date, none have been related to antiviral agents. In this study, a group of 137 theobromine-based hybrid molecules were designed by conjugating theobromine (at position °1; the secondary amine) with other commercially available chemical entities and compounds. The designed conjugates were studied *in-silico*, for their binding affinities and interactions at the catalytic binding site of NS2B/NS3pro using molecular docking and molecular dynamics. The top-ranked designed molecules were synthesised, characterised, and *in-vitro* tested for their inhibitory activities and type of inhibition towards the protease. Results show that conjugating theobromine

with other chemical moieties through its secondary amine has significantly enhanced the binding affinity, stability, and interactions of theobromine at the catalytic site of the protease. ADMET profiling showed that all the designed molecules are predicted to be safe, non-hepatotoxic, non-CYP2D6 binders, and non-plasma protein binders. MM-PB(GB)SA energy calculations showed that compounds C30a and C78f possessed the lowest free energies of binding of -30.13 and -37.14 kcal.mol^{-1}, whereas theobromine was found at -0.90 kcal.mol^{-1}. Next, thirteen compounds were synthesised using a multistep reaction and characterised using DSC, FT-IR, LC-MS, and ^{1}H & ^{13}C NMR. The findings of the molecular dynamics study were in agreement with the *in-vitro* results, where C30a inhibited the protease with an IC_{50} of 90.19 μM in a competitive manner. Interestingly, C78f was able to inhibit the viral protease with an IC_{50} of 22.54 μM, but in a non-competitive manner. In comparison, the IC_{50} of theobromine could not be determined as it showed only 19.46% inhibition towards the protease at 500 μM. To conclude, conjugating theobromine with other chemical moieties through an amide linker can significantly improve its activity towards DENV NS2B/NS3 protease *in-silico* and *in-vitro*. The conjugated chemical moiety type also influenced the overall activity of the designed molecule and the type of inhibition it exhibits.

CHAPTER 1

INTRODUCTION

1.1 Background

Dengue is a mosquito-borne viral disease caused by the dengue virus (DENV) (Harapan *et al.*, 2020). It is a tropical and sub-tropical communicable disease. DENV is a member of the *Flavivirus* genus, derived from the Latin word *"flavus"*, meaning yellow, in accordance to jaundice caused by the yellow fever virus (Kuhn *et al.*, 2002). Members of the Flaviviridae family, including DENV, consist of a single enveloped positive RNA strand (Simmonds *et al.*, 2017; Fishburn *et al.*, 2022). DENV has five distinct serotypes (DENV1 through DENV5), with DENV2 being the most prevalent among the other serotypes in Malaysia and Southeast Asia (Lee *et al.*, 2007). Those serotypes are classified according to their biological and immunological criteria (Mustafa *et al.*, 2015), and they share approximately 65% of their genomes (Sasmono *et al.*, 2015). A DENV serotype can be identified through the detection of its RNA genome using reverse-transcription polymerase chain reaction (RT-PCR) (Tian *et al.*, 2022) or by its non-structural protein 1 (NS1) antigen (Prommool *et al.*, 2021). DENV is considered a prototype to recognise and study the *Flavivirus* genus structure and replication cycle (Lescar *et al.*, 2018). DENV genome is around 11 kb in length and has a long open reading frame (ORF). This ORF encodes a polyprotein that lies between two untranslated regions (UTR); the 5'-UTR and the 3'-UTR (Zeng *et al.*, 2018).

Globally, it is estimated that there are about 284-528 million cases of dengue infections each year, of which around 24% of these cases require hospitalisation (Bhatt *et al.*, 2013). Seventy per cent (70%) of the reported dengue cases and the estimated burden remain in Asia (Bhatt *et al.*, 2013). In 2000, there were 505,430 dengue cases

reported globally to the World Health Organization (WHO). However, these numbers have manifested enormously in the last few years; for example, in 2019 alone, there were 5.2 million DENV infection cases (WHO, 2021).

According to the Centers for Disease Control and Prevention (CDC), Malaysia is rated as a country with a continuous and frequent risk of dengue infections (CDC, 2020a). In 2019, there were more than 124,777 cases and 174 mortality events in Malaysia (WHO, 2019). Up to June 2022, more than 17,497 cases were reported with ten related death cases. The latter numbers are considered an increment of 57.6% compared to the first six months of 2021, which had 11,100 cases and six related death incidents (WHO, 2022a). Figure 1.1 represents the number of reported DENV cases, DENV haemorrhagic fever (DHF), and the total number of related death cases between 2000 and 2019, as published by the Department of Statistics of Malaysia (DOSM, 2020). DENV infections create a substantial economic burden on the affected countries of around USD 950 million in Southeast Asian countries (Ng *et al.*, 2022), where in 2009/2010 Malaysia alone spent on average a total of USD 178 million (Arumugam, 2023).

DENV is considered an arbovirus. Arbovirus is an acronym for ARthropod-BOrne VIRUS. This type of virus life cycle is very complex because it requires more than one host to complete it. This unique characteristic is the major challenge in monitoring the spread of arboviruses. Therefore, it is important to study their life cycle, prophylaxis, and treatment (Sukhralia *et al.*, 2018). Three species of the mosquito genus *Aedes* are the main transmitting vectors. These are *Aedes aegypti, Aedes polynesiensis*, and *Aedes albopictus*, where *A. aegypti* is considered the main causative agent for the infection (Figure 1.2) (Malavige *et al.*, 2004). *A. aegypti* first originated

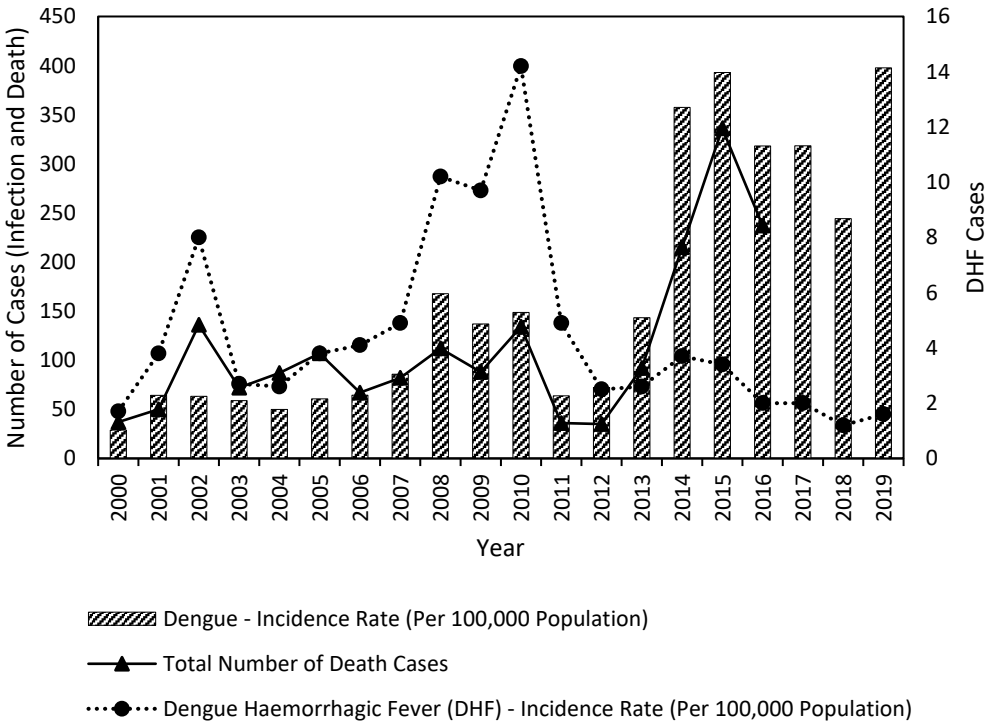
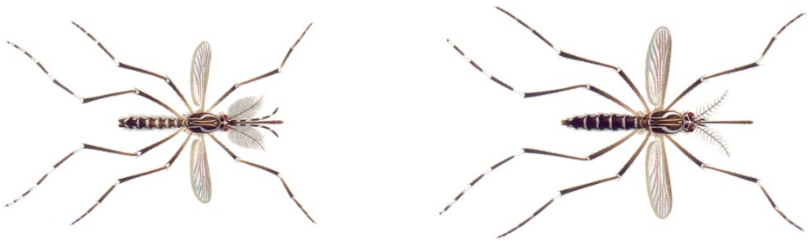


Figure 1.1 Incidence rate of reported DENV infections, Dengue Haemorrhagic Fever (DHF) manifestations (per 100,000 population for both statistics), and the total related death cases reported between 2000 and 2019 according to the Department of Statistics of Malaysia (DOSM).

in Africa. It is assumed that during the ages of slavery, combined with harsh conditions, *A. aegypti* was introduced into other parts of the world. That is where it consequently spread to tropical and subtropical districts around the globe (Kraemer *et al.*, 2015). *A. albopictus* is a zoophilic species from Asia encompassing the Indian and Pacific Oceans islands. During the 1980s, it rapidly expanded its reach to Europe, the United States and Latin America. Nowadays, both *A. aegypti* and *A. albopictus* exist in nearly all Asian countries and the massive lands of the Americas (Kraemer *et al.*, 2015). Even though *A. aegypti* is considered a tropical mosquito, its distribution appears to be affected by climate changes in tepid regions of the globe (Jansen & Beebe, 2010).



(a) **(b)**

Figure 1.2 *Aedes aegypti*, a dengue transmitting vector. **(a)** the male, and **(b)** the female.

Adapted from (Sanjad, 2003)

The mosquito vectors are anthropophilic, meaning they flourish close to human accommodations (Jansen & Beebe, 2010; Crawford *et al.*, 2017). Mature females favourably feed on humans, whereas other vertebrate species constitute merely a minor fraction of their blood meals. Unlike many other mosquito species, *A. aegypti* is a day-biting mosquito that feeds on numerous hosts in a single gonotrophic cycle. Females prefer to lay eggs in man-made or artificial containers like pot plant bases, flower vases, water tanks, or any other vessel usually found around or inside the houses. Eggs are placed on or close to the water surface in vessels. After embryonation, they can survive dehydration for up to one year (Jansen & Beebe, 2010).

Infection with DENV can result in a variant of complications and symptoms. These complications mainly vary between asymptomatic dengue fever (DF), mild DF, dengue haemorrhagic fever (DHF), or dengue shock syndrome (DSS) (Khetarpal & Khanna, 2016). The severity of these complications generally varies according to the infecting serotype. In 2015, a meta-analysis reported that infection with DENV3 has the greatest severity as a primary infection among other serotypes in infected patients from Southeast Asia. However, DENV2 was found with the highest severity when it comes to secondary infections in Southeast Asian and non-Southeast Asian regions

(Soo *et al.*, 2016). Moreover, it has been reported that DENV2 infections are associated with DHF and DSS (Vicente *et al.*, 2016; Wang *et al.*, 2020). Hence, DENV2 can be considered the most clinically severe serotype (Kyle & Harris, 2008). Secondary infections might manifest during primary infection of DENV. This manifestation comes as a result of a vulnerable immune system or an accessible route of entry for another pathogen which might be viral, bacterial, or mycotic (Manohar *et al.*, 2020).

DF usually resolves within five to seven days. It might vary between being asymptomatic to mild fever. Symptoms might include pyrexia, nausea, vomiting, cephalgia, arthralgia, myalgia, and rash (Kularatne, 2015; Khetarpal & Khanna, 2016; Stewart-Ibarra *et al.*, 2018; CDC, 2020b). Children might suffer febrile seizures and delirium due to fever. Severe retro-orbital ache on eye movement or by applying some pressure to the eyeball is also usual (Kularatne, 2015). No direct or symptomatic treatment is usually provided (CDC, 2020b). Patients whom persist ill in spite of their temperature subsiding are more likely to progress to DHF (Malavige *et al.*, 2004).

DHF, on the other hand, is typically a result of secondary DENV infections. Nonetheless, it may follow primary infections on occasions, particularly in infants. DHF is characterised by the aforementioned symptoms of DF, in addition to epistaxis haemorrhage (gums bleeding, petechiae and ecchymosis, hematemesis, melaena, and menorrhagia in females), features of circulatory failure (hypotension, tachycardia, narrow pulse pressure, and poor capillary refill-time), pleural effusions, ascites, glomeruli injury and pericarditis (Malavige *et al.*, 2004; Kularatne, 2015; Pagliari *et al.*, 2016).

Additionally, DHF can be divided into three stages; febrile, leakage, and convalescent. The febrile stage starts with abrupt onset of intermittent pyrexia accompanied by rigours and facial flush. Epigastric distress, myalgia, vomiting, abdominal pain, sore throat, and convulsions are common. Hepatomegaly is perceived in nearly all patients, while splenomegaly might be perceived in some. Plasma leakage is usually characterised by tachycardia and hypotension. Intense plasma leakage could develop further signs of circulatory disturbance (Malavige *et al.*, 2004; Malavige & Ogg, 2017). Insufficient or improper treatment often leads to an intense shock. In DHF, haemorrhage is not associated with depletion in thrombocyte count and could occur from any site in the patient's body. The third stage in DHF, convalescence, is typically brief and uneventful. It is characterised by appetite return, bradycardia, petechial rash, erythema, islands of pallor, and itching (recovery rash) (Malavige *et al.*, 2004; Muller *et al.*, 2017). Other complications might manifest in worse conditions which might lead to DSS. These complications include but are not restricted to thrombocytopenia ($<100\times10^9$ L^{-1}), encephalopathy, encephalitis, hepatic failure, myocarditis, and disseminated intravascular coagulation (DIC) leading to immense haemorrhage (Malavige *et al.*, 2004).

1.2 Problem Statement

Since DENV is an endemic viral infection caused by dengue virus (DENV), it is present in more than 120 countries around the world. The transmission of the disease is already high, but modern urbanisation throughout tropical regions such as Southeast Asia and Latin America has accelerated the spread of the disease. Each year, an estimated 390 million DENV infections occur around the world. Although DENV is responsible for up to 25,000 deaths annually worldwide (WHO, 2021), up to this day,

there is no treatment for dengue, and only one licensed vaccine is available in the market. Dengvaxia® is the only licensed vaccine approved to be used for children 9-16 years old who are in areas at risk of DENV in 19 countries and have been DENV-infected previously (Almas *et al.*, 2022; WHO, 2022b). Furthermore, the enormous vector control efforts, which cost more than USD 178 million (in Malaysia) annually, were not able to stop its rapid emergence and global spread (Arumugam, 2023).

Many researchers took the initiative across decades of hard work trying to discover a treatment for DENV infections. However, none of these has seen the light to be licensed and used clinically (Troost & Smit, 2020). Of the taken initiatives, peptides, crude extracts, phytochemicals, synthesised molecules, and repurposing of certain antivirals were reported to act on various targets of the DENV particulate, on certain host's receptors to prevent the virus entry to the cells, or by improving the immune system of the hosts (Malabadi *et al.*, 2011; Chew *et al.*, 2017; Troost & Smit, 2020; Wellekens *et al.*, 2022).

Among the targeted proteins in DENV is the NS2B/NS3 protease. This protease plays a crucial role in the virus life cycle that inhibiting it would prevent DENV replication and minimise symptoms and mortality rate, not to mention minimising the cost of health care systems (Bollati *et al.*, 2010). Hence, this study investigates the potentials of certain theobromine derivatives which share a similar chemical scaffold with approved antiviral agents as inhibitors of the NS2B/NS3 protease. It is hoped that this study will further contribute to the discovery and development of future potent NS2B/NS3 proteases, as well as anti-dengue therapeutics.

1.3 Study Objectives

The overarching objective of this study is to develop and evaluate a series of novel and effective anti-DENV protease agents based on theobromine through a comprehensive approach. This can further be achieved as shown in the flowchart presented in Figure 1.3 and through the following defined goals:

a. To design various theobromine conjugates which have higher affinities towards DENV protease assessed via molecular docking approach.
b. To assess the dynamic evolution of theobromine derivatives in complex with NS2B/NS3pro using molecular dynamics and molecular mechanics energy calculations.
c. To synthesise the best-evaluated candidates which are able to prove their theoretical activity against the target enzyme according to their affinity and the stability of the complex.
d. To measure the *in-vitro* activity of the synthesised theobromine conjugates through their percentage of inhibition, half-maximal inhibitory concentration (IC_{50}), and the mechanism of inhibition they offer towards the protease.

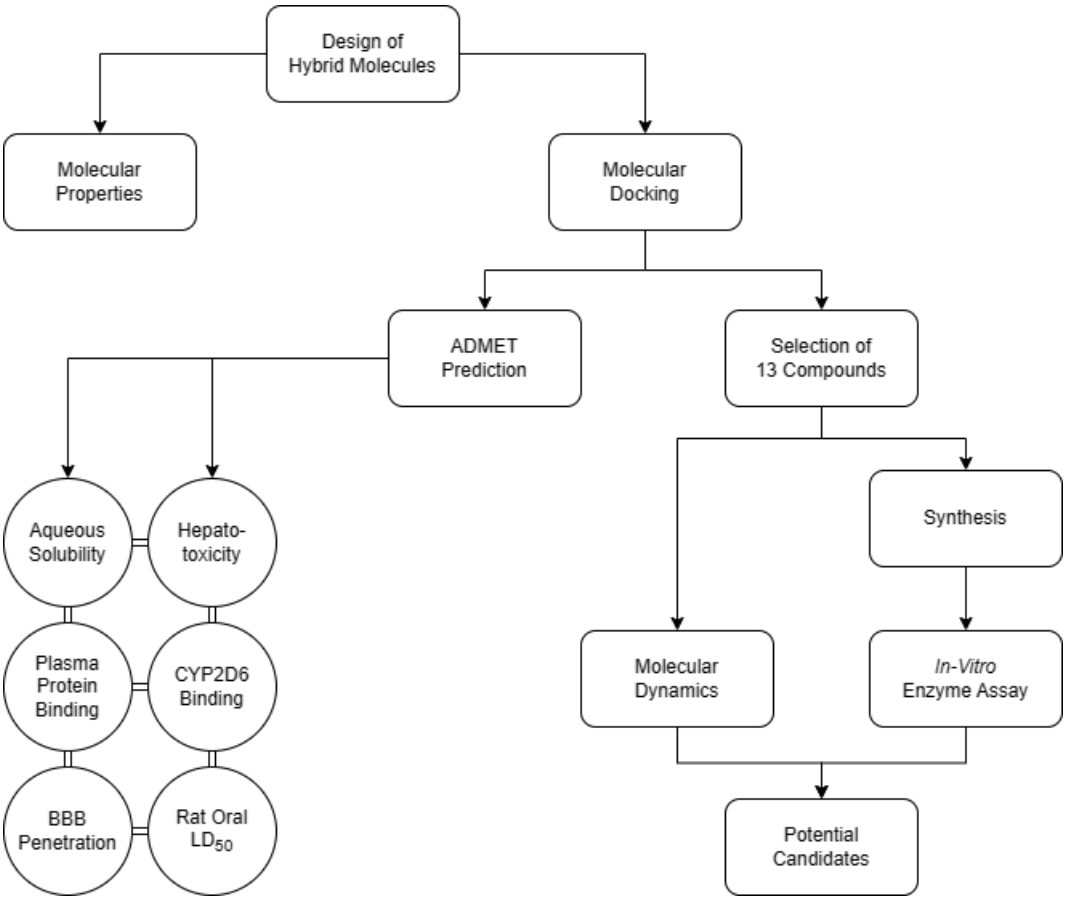


Figure 1.3 Flowchart of the conducted study.

CHAPTER 2

LITERATURE REVIEW

2.1 Dengue Virus Structure and Replication Cycle

DENV is a spherical enveloped virus characterised by a smooth surface and a diameter of around 500 Å (Burke & Monath, 2001). DENV envelope consists of an outer lipid bilayer which consists mainly of palmitoyloleoylphosphatidylcholine (POPC), dilinoleylphosphatidylcholine (DUPC), dipalmitoylphosphatidyl ethanolamine (DPPE), dioleoylphosphatidylserine (DOPS), palmitoylsphingomyelin with a choline headgroup (PPCS), palmitoylsphingomyelin with an ethanolamine headgroup (PPCE), and ceramide (Reddy & Sansom, 2016), and two structural proteins; the membrane (M) and the envelope (E) (Kuhn *et al.*, 2002). The lipid bilayer is derived from the host during the virion assembly inside the infected cell's cytoplasm (Reddy & Sansom, 2016). The M and E proteins are embedded in the lipid bilayer with stoichiometric amounts; 180 M and 180 E (Zhang *et al.*, 2003). The envelope glycoprotein (E) lies as dimers at the surface of the virion and is responsible for binding and fusion with the host cell's receptor. On the other hand, the M protein is mainly responsible for the conformational changes of the E protein during the process of maturation of the virion and the smooth appearance (Nasar *et al.*, 2020).

The most common receptors which DENV utilises to internalise inside the cells belong to the dendritic cells that include the human mannose-binding receptor (MR) (Miller *et al.*, 2008), the dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN; CD209) (Wu *et al.*, 2000), T-cell immunoglobulin and mucin-domain containing-3 and -4 (TIM3 and TIM4), langerins receptors, and the Fc portion of immunoglobulin (Ig)G (FcγR) (Boonnak *et al.*, 2008; Begum *et al.*, 2019).

Underneath the DENV envelope comes the capsid, which consists of the structural capsid protein (C) arranged asymmetrically and not as ordered as the earlier proteins (Uno & Ross, 2018). Its main function is to form a nucleocapsid that embraces the viral ribonucleic acid (RNA) (Figure 2.1) (Christopher *et al.*, 2003; Nasar *et al.*, 2020).

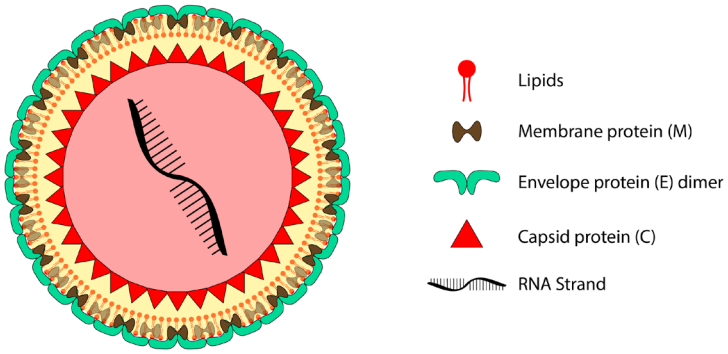


Figure 2.1 An illustration of dengue virus structure showing the lipid bilayer, the structural proteins, and the positive RNA strand.
Illustration was made using Adobe Illustrator.

Flavivirus positive RNA genome is a 5'-capped RNA ((+)ssRNA). Upon infection, the virus gets internalised via receptor-bound endocytosis, creating an endosome within the cytoplasm of the host cell (Heinz & Allison, 2000). The acidic environment of the endosome disrupts the E protein leading to fusion with the endosome (Mukhopadhyay *et al.*, 2005). Once the nucleocapsid is released into the cytoplasm, it dissociates, leaving the RNA, which would serve as a messenger-RNA (mRNA). The ORF migrates to the rough endoplasmic reticulum and is translated into a large polyprotein. This polyprotein is cleaved co- and post-translationally into ten proteins (Welsch *et al.*, 2009). This cleavage is hosted by two proteases; the host furin-type (or any other Golgi-localized proteases) and the viral serine protease embedded in the N-terminal domain of non-structural protein 3 (NS3pro), which requires non-

structural protein 2b (NS2B) for its activity (Bollati *et al.*, 2010). Upon cleavage, the N-terminus of the polyprotein yields the three structural proteins; E, membrane precursor (prM), and C, while the rest encodes seven non-structural (NS) proteins; NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Figure 2.2) (Welsch *et al.*, 2009; Bollati *et al.*, 2010). DENV proteins' functions are summarised in Table 2.1.

The replication process begins when a complementary negative strand of the RNA ((-)ssRNA) that is created to serve as a template for the replication of the positive strand via an enzymatic process by NS5, NS3, other viral NS proteins, and some host factors. Upon replication, the newly created positive RNA strands participate in translating more viral proteins, or they form new nucleocapsids as they associate with C proteins. The newly formed nucleocapsids remain at the lumen of the endoplasmic reticulum, where they acquire their lipid bilayer in association with the structural proteins; E and prM. Next, proteolysis of prM takes place in the trans-Golgi network, which triggers rearrangement and dimerisation of E. These mature viral particles are released via an exocytosis mechanism to infect other cells and repeat the cycle (Figure 2.2) (Mukhopadhyay *et al.*, 2005; Murugesan & Manoharan, 2020).

Table 2.1 Dengue virus translated proteins and their functions.

Protein	Function	Ref.
Structural		
E	a. Recognition and binding to the host cell. b. Involved in uncoating of virus by enabling fusion of viral and endosomal membranes. c. Neutralises antibodies.	(Modis *et al.*, 2004)
prM/M	prM: a. Drives the assembly of viruses. b. Acts as a chaperone for the folding of the E. M: forms ion channels.	(Premkumar *et al.*, 2005; Zhang *et al.*, 2012)
C	a. RNA binding. b. Nucleocapsid formation.	(Christopher *et al.*, 2003; Byk & Gamarnik, 2016)
Non-structural		
NS1	a. Viral RNA replication. b. Can be used as a marker for early diagnosis. c. It activates macrophages, platelets, and mononuclear peripheral cells. d. Inhibits cellular autophagy during infection.	(Flamand *et al.*, 1999; Kassim *et al.*, 2011; Modhiran *et al.*, 2015)
NS2A	Viral replication and assembly.	(Xie *et al.*, 2013)
NS2B	Acts as a NS3-protease co-factor.	(Falgout *et al.*, 1991)
NS3	a. Serine protease cleaves viral polyprotein. b. Acts as an adenosine triphosphatase (ATPase), RNA triphosphatase (RTPase), and RNA helicase.	(Xu *et al.*, 2005; Luo *et al.*, 2008b)
NS4A	a. Promotes membrane rearrangement. b. Inhibits cellular autophagy during infection.	(Stern *et al.*, 2013; Echavarria-Consuegra *et al.*, 2019)
NS4B	a. Acts as a negative modulator of the NS3 helicase function. b. Blocks IFN-α/β-induced signal transduction and helps virus to escape host's innate immune response.	(Nemésio *et al.*, 2012)
NS5	a. Methyl transferase domain (MTase). b. RNA-dependent RNA polymerase (RdRp).	(Yap Thai *et al.*, 2007; Liu *et al.*, 2010)

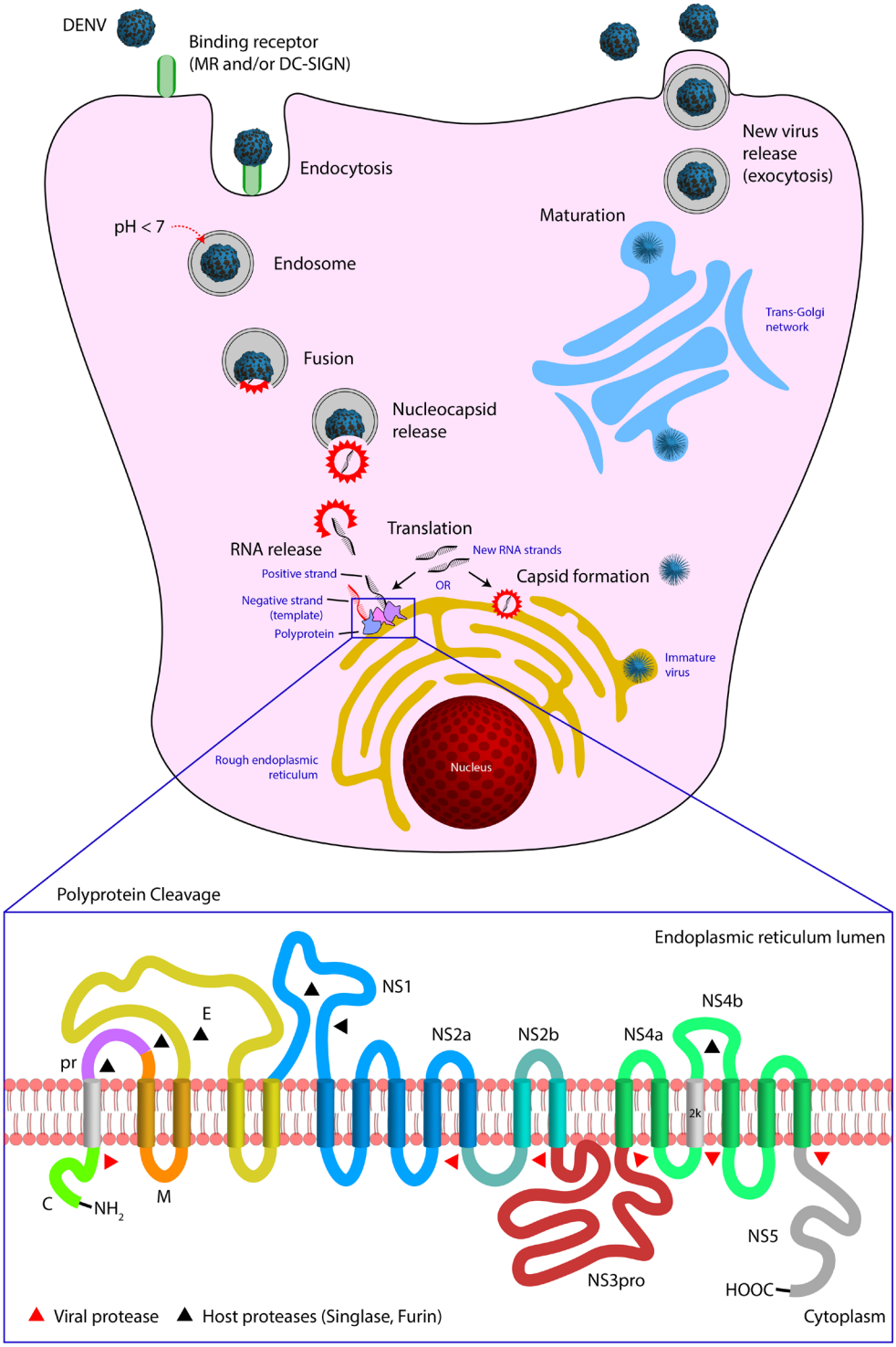


Figure 2.2 *Flavivirus* life cycle and the cleavage of the translated polyprotein into the structural (C; capsid, M; membrane, and E; envelope) and the non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5).

Illustration was made using Adobe Illustrator based on (Welsch *et al.*, 2009).

2.2 DENV NS2B/NS3pro

As aforementioned in Table 2.1, NS3 protein is a multifunctional protein that can act as a trypsin-like serine protease, ATPase, RTPase, and RNA helicase (Xu *et al.*, 2005; Luo *et al.*, 2008b). The N-terminus of NS3 (residues 1-180) is considered the protease domain, while the rest of the protein (residues 180-618; the C-terminus) functions as the RNA helicase. Like any other trypsin-like serine protease, NS3pro has a catalytic triad made of His51, Asp75, and Ser135 (Figure 2.3) (Bazan & Fletterick, 1989; Erbel *et al.*, 2006). NS3pro was proven to require a co-factor, NS2B, for its functionality (Yusof *et al.*, 2000; Leung *et al.*, 2001), where NS2B provides structural stability for NS3pro and permits protein folding for a successful proteolytic activity (Leung *et al.*, 2001). NS2B has a hydrophilic central region, which behaves as a chaperone to stabilise the hydrophobic termini of NS3pro by surrounding it like a belt (Figure 2.3) (Tomar *et al.*, 2017). Both domains, the NS3pro and NS2B, are connected through a Gly_4-Ser-Gly_4 linker (Leung *et al.*, 2001; Luo *et al.*, 2008a). This NS2B/NS3 is responsible for the cleavage of multiple proteins at the junctions of C-prM, NS2A-NS2B, NS2B-NS3, NS3-NS4A, and NS4B-NS5 (Tomar *et al.*, 2017). Thus, making it a crucial enzyme complex for the proteolysis and cleavage activities of the translated polyprotein during the virus life cycle. Another part of the catalytic triad is the oxyanion hole. The oxyanion hole is a positively charged pocket formed by the NH groups of Gly133, Thr134, and Ser135 (Figure 2.4) (Hedstrom, 2002; Noble *et al.*, 2012). The role of the oxyanion hole is discussed further in the catalysis mechanism part (see Section 2.3).

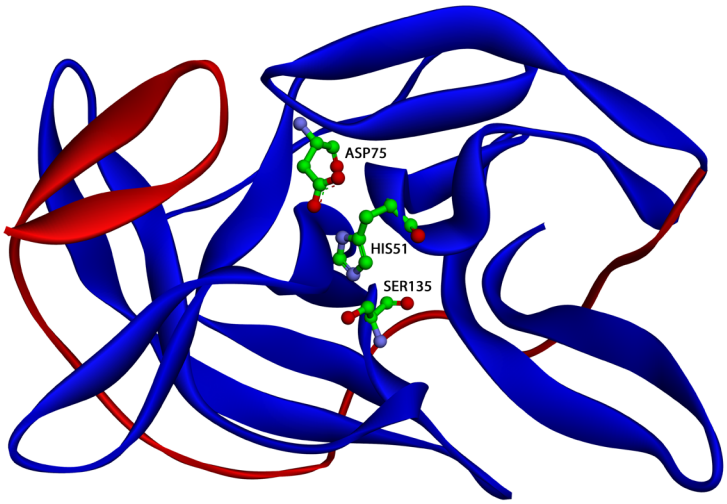


Figure 2.3 A ribbon representation of DENV2 NS2B/NS3pro showing the NS3pro domain in blue, the NS2B co-factor in red, and the catalytic triad amino acid residues; His51, Asp75, and Ser135, as green-coloured carbons sticks and balls. Illustration was made using BIOVIA Discovery Studio.

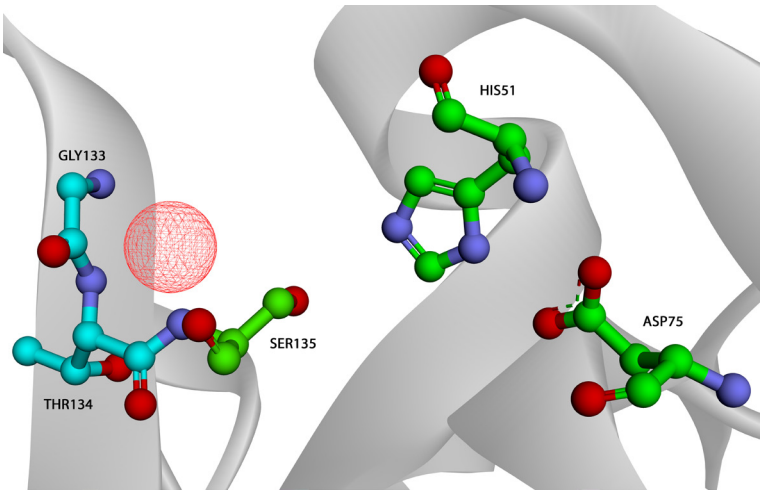


Figure 2.4 A ball and stick 3D representation of the catalytic triad amino acid residues; His51, Asp75, and Ser135 (green-coloured backbone atoms) and the oxyanion hole (red mesh sphere) formed by Gly133, Thr134 (cyan-coloured backbone atoms), and Ser135 residues.

Illustration was made using BIOVIA Discovery Studio.

2.3 Catalytic Mechanism of NS2B/NS3pro

DENV NS2B/NS3pro is a trypsin-like serine protease complex that is well-studied and understood for its catalytic mechanism (Kraut, 1977; Warshel *et al.*, 1989; Hedstrom, 2002; Radisky *et al.*, 2006; Di Cera, 2009). In the case of *Flavivirus*, the catalytic mechanism relies on three amino acid residues, the catalytic triad, which includes His51, Asp75, and Ser135. Each residue of the catalytic triad has a defined role where Ser135 interacts with the carbonyl of the substrate, His51 has the ability to donate and accept protons throughout the catalytic process, and Asp75 shows the ability to accept protons as the final step of the cleavage (Hedstrom, 2002).

The catalysis process starts with activating Ser135 residue via a general acid. This activation allows Ser135 to deprotonate, giving the H^+ to His51. Once Ser135 is activated, an acylation reaction initiates as it interacts with the carbon of the carbonyl group at the peptide scissile region and disrupts the resonance by allowing the carbon of the carbonyl to form five transitional bonds. This transition state creates a tetrahedral intermediate which allows the amide bond to break where the C-terminus of the peptide is now free to leave. The C-terminal region receives a proton from His51, and His51 recompensates this proton by deprotonating Asp75. In the next step, a water molecule protonates, allowing a diacylation reaction by giving the carbonyl's carbon a hydroxyl group (^-OH) and the H^+ to His51 after returning the proton it took from Asp75. This second transition state creates another tetrahedral intermediate that permits the cleavage of the peptide N-terminus (Figure 2.5) (Hedstrom, 2002; Radisky *et al.*, 2006).

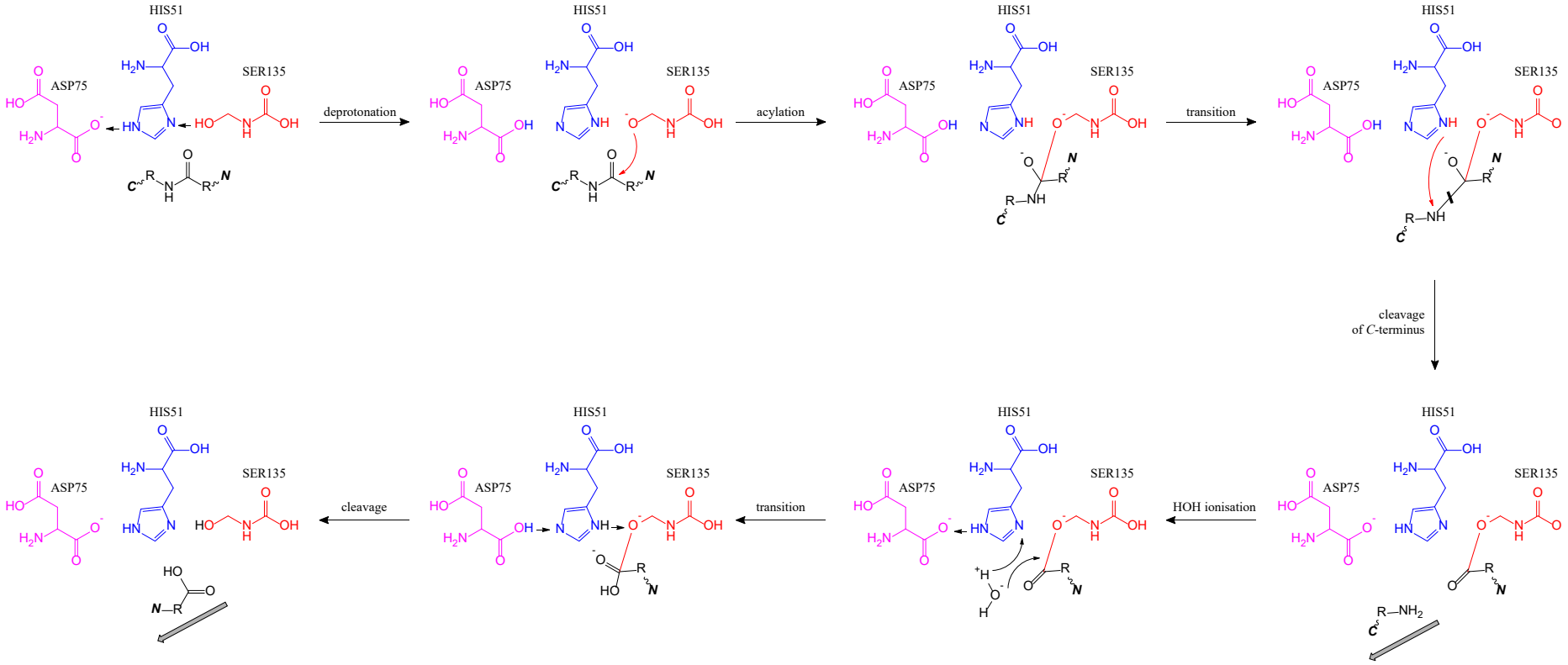


Figure 2.5 The catalytic mechanism of NS2B/NS3pro, starting with the activation of serine residue which interacts with the peptide carbonyl creating a transition state of a tetrahedral intermediate. This allows the breakage of the peptide bond, and the C-terminus (*C*) of the peptide leaves. Next, a water molecule gets ionised and a second transition state creates another tetrahedral intermediate. Finally, the N-terminus (*N*) of the cleaved peptide leaves.

The oxyanion hole activates the carbonyl of the scissile peptide bond. It allows it to interact with the NH's of Gly133, Thr134, and Ser135 to stabilise the oxyanion (O^-) of the peptide carbonyl during the transition state (Hedstrom, 2002; Radisky *et al.*, 2006; Noble *et al.*, 2012). It is noteworthy that even when the catalytic triad is disabled, a remaining activity of the protease is still there, which was attributed to the oxyanion hole's contribution (Hedstrom, 2002).

2.4 Substrate Specificity of NS2B/NS3pro

The binding site of *Flavivirus* NS2B/NS3pro enzyme is divided into five sub-pockets (S); S1', S1, S2, S3, and S4 (Aleshin *et al.*, 2007). This terminology was first introduced by Schechter and Berger (Schechter & Berger, 1967). These sub-pockets are located on both sides of the catalytic region and are labelled from the N- to C-termini. These sub-pockets correspond to the positions (P) of the amino acid residues of the bound substrate denoted as P1', P1, … etc. The numbering system starts from the scissile bond, where that substrate amino acid is referred to as P1, and the corresponding sub-pocket is S1 (Schechter & Berger, 1967). Each S on the protease accommodates one amino acid P of the substrate (McDonald, 1985). The cleavage point is set to lie between S1 and S1', and the numbering system continues from that point onwards (Figure 2.6).

It has been found that DENV NS2B/NS3pro tend to prefer dibasic residues (Arg-Arg, Arg-Lys, and Lys-Arg) at P1 and P2 of the bound substrate in order to achieve the catalytic motif (Niyomrattanakit *et al.*, 2006), while it can tolerate bulky residues (e.g., Trp, Phe, or Tyr) at P1' and P2' (Shiryaev *et al.*, 2007b) considering

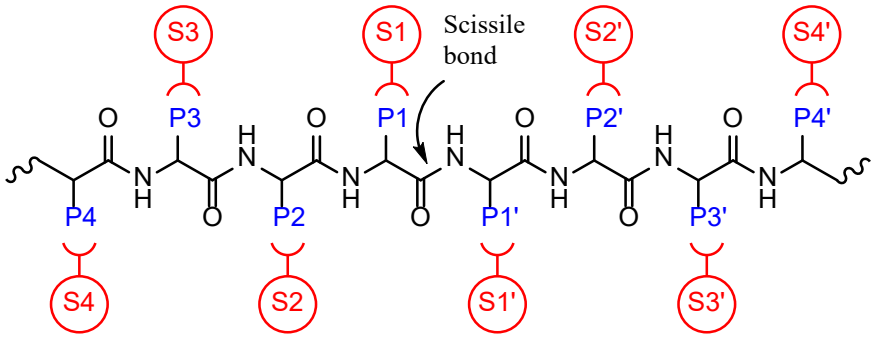


Figure 2.6 A schematic representation of the numbering system introduced by Schechter and Berger of the protease sub-pockets (S) and the corresponding positions (P) of the bound substrates' amino acid residues.

Illustration was made using PerkinElmer ChemDraw Professional.

that Ser and Gly are the preferred residues at P1' (Li *et al.*, 2005; Niyomrattanakit *et al.*, 2006; Shiryaev *et al.*, 2007a) and Arg at P2' (Lin *et al.*, 2016; Lin *et al.*, 2017). However, P3 and P4 were found to be largely exposed to solvent as both corresponding sub-pockets are small in size (Aleshin *et al.*, 2007). Studying the substrate preferences of the protease sub-pockets can help in the proper designing and synthesis of the enzyme inhibitors.

2.5 Mechanisms of NS2B/NS3pro Inhibition

Serine proteases and corresponding inhibitors have been studied thoroughly for their impact on coagulation, digestion, inflammation, wound healing, viral replication, and disease manifestation (Farady & Craik, 2010; Almonte & Sweatt, 2011). Proteolytic inhibitors of the proteases are meant to block their activity via a conformational change mechanism (Almonte & Sweatt, 2011). Inhibitors can mainly be classified into three main categories; competitive, irreversible, and allosteric (Farady & Craik, 2010).

Competitive inhibitors act in a lock-and-key fashion where they interact with the sub-pockets of the protease binding site. Their interaction is competent, not as strict as the specified substrate, and in a non-catalytic manner. These inhibitors are built to mimic the peptide-substrate sequence. They bind through an extended β-sheet with the enzyme in a way that resembles the substrate. Their scissile bond gets hydrolysed slowly, but no fragments are released, and the amide bond is re-ligated (Farady & Craik, 2010). On the other hand, irreversible inhibitors are candidates which require catalysis by the binding site of the enzyme. They interact with the active site's sub-pockets in a substrate-like manner, utilise the proteolytic activity of the enzyme to trap, and inhibit the enzyme activity. Alternatively, allosteric inhibitors bind to an exo-site of the protein. Allosteric or exo-sites are secondary binding sites which are distinct from the main binding pocket, and their interaction with an inhibitor can have a critical effect on the enzyme activity. The inhibition mechanism of the allosteric inhibitors is usually through affecting the surface area of protein-protein interaction and by altering the main site's sub-pockets specificity towards the substrate amino acids, thus inhibiting its catalytic abilities (Farady & Craik, 2010).

Up to date, four allosteric sites have been discovered for DENV NS2B/NS3pro (Figure 2.7). The first allosteric site is formed by Lys73, Lys74, Leu85, Glu86, Gly87, Glu88, Trp89, Thr120, Ile123, Ala125, Val 147, Asnl52, Val162, Ala164, Ala166, Asn167, Thr168, and Glu169 residues and is located at the back of the catalytic binding site and can be inhibited by curcumin, dabrafenib, idelalisib, and nintedanib (Lim *et al.*, 2020a; Uday *et al.*, 2021). The second allosteric active site at DENV NS2B/NS3pro is located at the side of the main pocket between the 120s loop and the 150s loop. This allosteric site has been studied with a mutation of Ala125 to Cys and found to be inhibited by aldrithiol, 5,5'-dithiobis-(2-nitrobenzoic acid), and

biarylchloromethylketone (Yildiz *et al.*, 2013). The third active site also lies on the back side of the main pocket of the protease, overlapping with some shared amino acids with the catalytic pocket. The third allosteric site is lined mainly by Met49, His51, Lys73, Lys74, Asp75, Leu76, Trp83, Thr120, Gly148, Leu149, Gly151, Ile165, Ala166, Asn167, and is inhibited by a pyazine derivative (Yao *et al.*, 2019). The fourth and final discovered allosteric site lies at the other side of the catalytic pocket and is constituted mostly by the NS3 residues Lys15-Asp20 in loop, and Gly21-Arg24 in β-strand, as well as the NS2B residues Leu54-Val59. This allosteric site of DENV protease was found to be inhibited by myricetin (Dang *et al.*, 2022).

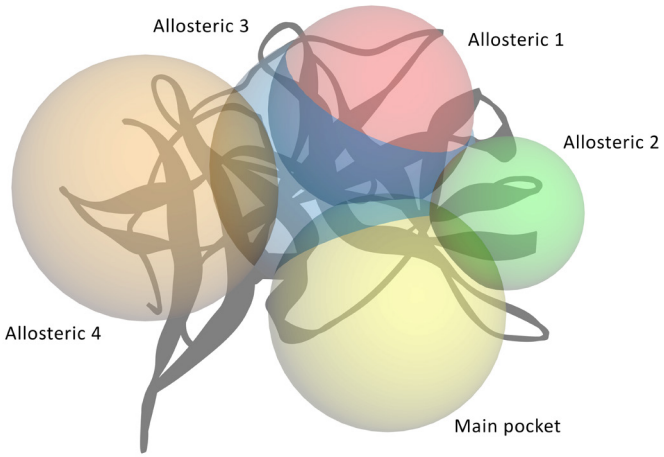


Figure 2.7 An illustration representing DENV NS2B/NS3pro (grey ribbon), its catalytic binding pocket (yellow sphere), and the four discovered allosteric sites 1-4 (red, green, blue, and orange spheres, respectively).
Illustration was made using BIOVIA Discovery Studio.

Although the main binding pocket of DENV NS2B/NS3pro utilises the polyprotein precursor for its catalytic activity, studies have also shown that it can catalyse the commercial tetrapeptide substrate Nle-Lys-Arg-Arg (norleucine-lysine-arginine-arginine) (Ulanday *et al.*, 2016). Hence, an aldehyde-based (-COH) inhibitor (Bz-Nle-Lys$_{(P4)}$-Arg$_{(P3)}$-Arg$_{(P2)}$-H$_{(P1)}$) was designed, which contains benzoyl (Bz) as a

protection group (Erbel *et al.*, 2006). It has been found that Ser135 binds covalently with the aldehyde-carbonyl carbon. The resulting hydroxyl group interacts with NS3pro at His51, which creates a tetrahedral hemiacetal. P1 residue side chain forms a salt-bridge with Asp129 of NS3 and a hydrogen bond with the carbonyl of Phe130. P2 and P3, nonetheless, stabilise the β-hairpin of NS2B in a closed conformation by interacting with the carbonyls of Gly82 and Met84 of NS2B, respectively. Moreover, P2 and P3 also interact with Gly151 and Asn152 of NS3, respectively, through a single backbone interaction (Noble *et al.*, 2012).

2.6 Dengue Prophylactic Measures

Vector control represents a principal method in controlling vector-borne diseases, i.e., DENV (Wilson *et al.*, 2020). Vector control methods of *Aedes sp*. mosquitoes can be distinguished into three approaches which include environmental, biological, and chemical (Buhler *et al.*, 2019). The environmental approach to control the spread of DENV is through waste management, the use of water containers' covers, and the elimination of breeding spots (Buhler *et al.*, 2019). On the other hand, biological measures include the use of *Bti* (*Bacillus thuringiensis israelensis*), which produces certain toxins that can kill the larval habits of insects and fungi in water containers while expressing no effects on other organisms. Other reported biological measures include the use of copepods (*Mesocyclops sp.*) and larvivorous fish (*Poecilia reticulata* and *Gambusia affinis*), where they feed on the larvae of the mosquito before they turn into adults (Walton, 2007; Horstick *et al.*, 2017).

In contrast, chemical methods are the most commonly used measures for vector control (Alkuriji *et al.*, 2020). They include sprayable insecticides and fumigation through the application of synthetic pyrethroids, deltamethrin, *S*-bioallethrin,

piperonyl butoxide, pyriproxyfen, and permethrin. Temephos (an organophosphate larvicide) and malathion (an organophosphate insecticide) are other chemical agents which have been used to treat water and control the larval habits of the mosquito (Horstick *et al.*, 2017).

A greener approach for controlling the DENV vector is the *Wolbachia*-based strategy (Sim *et al.*, 2020). *Wolbachia sp.* is a genus of Gram-negative bacteria that is naturally present in some invertebrates and can prevent RNA viruses from replication. Hence, no offspring can develop if an infected male *Aedes* mosquito fertilises an uninfected female. On the contrary, if an uninfected male fertilises an infected female, the offspring will be bearing *Wolbachia,* and the number of laid eggs will be less. This way, the *Wolbachia* strategy can offer a safer and more economical approach to controlling the spread of mosquito-borne diseases (Alkuriji *et al.*, 2020). Similarly, green pesticides from the plant kingdom were seen to prove their efficacy against *Aedes sp.* and their larvae. These include extracts of *Ocimum americanum* (family Lamiaceae, leaf) (Murugan *et al.*, 2007; Gbolade & Lockwood, 2008), *Jatropha curcas* (family Euphorbiaceae, leaf) (Njom, 2022), *Citrus limon* (family Rutaceae, fruit peel) (Gomes *et al.*, 2019), *Artocarpus blancoi* (family Moraceae, leaf) (Pineda-Cortel *et al.*, 2019), *Allium sativum* (family Amaryllidaceae) (Jarial, 2001), in addition to some essential oils from ginger, oranges, lemongrass, rosemary, and thyme, among many others (Waliwitiya *et al.*, 2009; Ery Agus *et al.*, 2016; Maia *et al.*, 2019; Kurniasih *et al.*, 2021).

2.7 Dengue Management

Despite the aforementioned prophylactic measures, up to date, there is no definite effective antiviral treatment for DENV infections. Supportive care is the