SEARCH OF DENGUE PROTEASE INHIBITORS FROM NATURAL PRODUCT DATABASE USING ENSEMBLE-BASED VIRTUAL SCREENING

NOR FARRAH WAHIDAH BINTI RIDZWAN

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

SEARCH OF DENGUE PROTEASE INHIBITORS FROM NATURAL PRODUCT DATABASE USING ENSEMBLE-BASED VIRTUAL SCREENING

NOR FARRAH WAHIDAH BINTI RIDZWAN

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF BIOTECHNOLOGY

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

SEARCH OF DENGUE PROTEASE INHIBITORS FROM NATURAL PRODUCT DATABASE USING ENSEMBLE-BASED VIRTUAL SCREENING

ABSTRACT

As the number of reported dengue cases appears to increase each year, it has become a great concern globally that there has yet to be any specific treatment for it. Nevertheless, previous studies have identified several of the dengue virus proteins as the potential target for the development of antiviral whereby, the viral protease has become the most favourable target. Virtual screening (VS) provides an alternative option from the conventional screening in drug discovery and has frequently been used by researchers worldwide mainly due to its cost-effectiveness. An extension of the standard VS called the ensemble-based incorporates the flexibility of receptor and ligand in the VS. Therefore, this approach that takes into account the existing multiple conformations of the target's structure would ultimately enhance the obtained VS findings. In general, this study aimed at identifying natural-based compounds that can serve as dengue protease inhibitors via the ensemble-based approach and to further evaluate the generated protein-ligand interactions. The collections of natural-based ligands were retrieved from the Super Natural II database utilising adapted criteria of the Lipinski's rule of five. Prior to the screening procedure, the three-dimensional structure of the protease was first obtained from the protein data bank (PDB), followed by model generation via MODELLER to address the missing residues and were further evaluated using PROCHECK. Next, molecular dynamic (MD) simulations of the protease were conducted and assessed via GROMACS. Several clustered structures of the protease were then generated and selected from the MD simulations' trajectories representing as the macromolecule or receptor in the subsequent virtual screening or molecular docking via AutoDock. The molecular docking simulations were conducted twice with the initial

simulation aimed to identify the ligands that showed a strong affinity towards the protease whereas the second simulation was done to verify and strengthening the initial results. Based on the findings, several ligands showed promising potential as the dengue protease inhibitors that bears strong binding energy and interaction with at least one of the catalytic triads of the dengue protease structure. Thus, the listed ligands would provide a good starting point for extending the utilisation and confirmation of this studies and may be used as lead-compounds in generating an antiviral for the dengue virus.

Keywords: dengue protease (NS2B/NS3), natural-based antiviral, protease inhibitors, ensemble-based virtual screening, dengue fever.

PENCARIAN PERENCAT PROTEASE DENGGI DARI PANGKALAN DATA BAHAN SEMULAJADI MENGGUNAKAN PENYARINGAN MAYA ENSEMBLE-BASED

ABSTRAK

Ketiadaan rawatan khusus untuk merawat demam denggi sehingga kini telah membimbangkan semua pihak apabila laporan kes demam denggi semakin meningkat pada setiap tahun. Namun begitu, terdapat kajian terdahulu yang telah berjaya mengenal pasti beberapa protein pada virus dengue yang berpotensi untuk digunakan dalam perkembangan antivirus dan salah satu protein yang sering menjadi sasaran utama adalah protease. Penyaringan maya (PM) merupakan kaedah alternatif daripada penyaringan konvensional yang digunakan dalam pengenalpastian ubat dan ia semakin kerap digunakan oleh para penyelidik di seluruh dunia kerana ia merupakan kaedah yang lebih murah apabila dibandingkan dengan kaedah konvensional. Kaedah ensemble-based merupakan salah satu kaedah lanjutan dari PM yang melingkupi fleksibiliti reseptor dan ligand semasa PM dijalankan. Oleh itu, kaedah ini yang mengambil kira pelbagai konformasi yang wujud pada reseptor akan menambahkan lagi kejituan kepada keputusan yang diperolehi. Tujuan kajian ini adalah untuk mengenal pasti sebatian berasaskan bahan semulajadi yang boleh bertindak sebagai perencat protease virus denggi dengan menggunakan kaedah ensemble-based dan seterusnya untuk mengenal pasti interaksi antara protein dengan ligand yang terbentuk atau dibentuk. Koleksi ligand yang berasaskan sumber semulajadi diperoleh daripada pangkalan data Super Natural II dengan menggunakan kriteria yang bersesuaian dari peraturan lima Lipinski. Sebelum prosedur PM dijalankan, struktur tiga dimensi protease diperolehi daripada Protein Data Bank (PDB). Kemudian, beberapa model protease dihasilkan dengan menggunakan MODELLER untuk menangani residu yang

dilaporkan hilang diikuti dengan penilaian model-model tersebut dengan menggunakan PROCHECK. Seterusnya, simulasi molekul dinamik (MD) diaksanakan dan dinilai menggunakan GROMACS pada protease. Melalui trajektori yand diperoleh daripada simulasi MD, struktur protease yang terjana dikelompokkan dan dipilih untuk digunakan sebagai makromolekul ataupun reseptor di dalam prosedur berikutnya iaitu PM ataupun dikenali juga sebagai molecular docking dengan menggunakan AutoDock. Prosedur ini dilaksanakan dua kali dimana, ligand yang menunjukkan interaksi yang kuat dengan protrease dikenalpasti pada saringan pertama. Manakala, saringan simulasi yang kedua dijalankan untuk mengesahkan dan memperkuatkan lagi keputusan yang diperolehi. Berdasarkan kaedah yang dinyatakan, kajian ini telah menemui beberapa ligand yang berpotensi untuk dijadikan sebagai perencat protease denggi dimana setiap ligand yang terpilih dapat membentuk interaksi yang stabil bersama protease. Oleh yang demikian, ligand yang tersenarai boleh digunakan untuk pengesahan selanjutnya dan dijadikan sebagai sebatian utama dalam penghasilan dan perkembangan antivirus untuk denggi.

Kata kunci: protease denggi (NS2B/NS3), antivirus berasaskan bahan semulajadi, perencat protease, penyaringan maya ensemble-based, demam denggi.

ACKNOWLEDGEMENTS

First and foremost, I would like to take this opportunity to give my highest gratitude to Dr. Saharuddin bin Mohamad, my supervisor for this study, for providing me ample of opportunities to learn, explore, discover and refine this project from the beginning until its completion. In addition, I would also like to thank Dr. Saha for guiding me and provide advices regarding this project and also regarding life in general. I appreciate all of the time that Dr. Saha was willing to spend with me to give bits of advice and motivation apart from sharing his valuable life experiences.

Secondly, I would like to thank all of my friends especially Khairunnisa binti Abdul Rashid, who have supported me directly and indirectly throughout my study in the University of Malaya. Last but not least, I would also like to thank Mrs. Sugunadevi Rajagopal, the assistant science officer in the Bioinformatics department, whom always provide me with the assistance I need in a courteous and friendly manner from the start until the end and I would also like to thank my parents for their financial support. Thank you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

%	:	Percent
3D	:	Three-dimensional
Å	:	Angstrom
Ala	:	Alanine
Arg	:	Arginine
Asn	:	Asparagine
Asp	:	Aspartic acid
atm	:	Atmospheric pressure
C	:	Capsid
CADD	:	Computer aided drug design
Cα	:	Carbon-alpha
Da	:	Dalton
DENV	:	Dengue virus
DENV-1	:	Dengue virus first serotype
DENV-2	:	Dengue virus second serotype
DENV-3	:	Dengue virus third serotype
DENV-4	:	Dengue virus fourth serotype
DENV-5	:	Dengue virus fifth serotype
DF	:	Dengue fever
DOPE	:	Discrete optimized protein energy
E	:	Envelope
Gly	:	Glycine
GPU	:	Graphics processing unit

HDF	:	Haemorrhagic dengue fever
His	:	Histidine
HPLC	:	High-performance liquid chromatography
HTS	:	High-throughput screening
K	:	Kelvin
kcal/mol	:	Kilocalorie per mole
LBDD	:	Ligand-based drug discovery
Leu	:	Leucine
Lys	:	Lysine
M	:	Membrane
MD	:	Molecular dynamics
Met	:	Methionine
MSA	:	Multiple sequence alignment
NMR	:	Nuclear magnetic resonance
NS	:	Non-structural
ns	:	Nanosecond
PDB	:	Protein Data Bank
pg.	:	Page
Phe	:	Phenylalanine
prM	:	Premembrane
Pro	:	Proline
ps	:	Picosecond
RAM	:	Random access memory
RMSD	:	Root-mean-square deviation
RMSF	:	Root-mean-square fluctuation
RNA	:	Ribonucleic acid

SBDD	:	Structure-based drug discovery
Ser	:	Serine
SO_4	:	Sulphate
ssRNA	:	Single-stranded RNA
Thr	:	Threonine
Tyr	:	Tyrosine
Val	:	Valine
vHTS	:	Virtual high-throughput screening
VS	:	Virtual screening

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

Dengue has been around for centuries and has begun to spread widely across the world with the aid from its vector, the *Aedes* mosquito. As the global travelling rates increases, the chances of disease spread also increases including those that are associated with vectors (Kulcyński et al., 2017). Although precautionary steps and preventions were placed, the continuous rise of reported cases for dengue fever (Hii et al., 2016) showed that the current measures are not sufficient enough to stop the dengue virus (DENV). The symptoms of dengue range from mild fever to a deadlier form that is referred to as haemorrhagic dengue fever (HDF). Most hospitalised patients would eventually recover with the help of medical assistance by managing the corresponding symptoms. However, those who are unable to receive the appropriate treatment may not be so fortunate (Bäck & Lundkvist, 2012). Undeniably, the absence an antiviral against DENV may indirectly contribute to the mortality rate of dengue patients.

One would assume that since the DENV has been identified and studied for more than a century (Murray, 2013), there should be ample of antivirals and vaccines available. However, despite the countless breakthrough made in dengue research, there is yet to be a single effective antiviral available (Nedjadi et al., 2015) and unfortunately, the first approved vaccine (World Health Organization, 2016) has faced a tremendous backslash and currently in a controversial state (Normile, 2016). Therefore, the search for dengue antiviral remains essential to treat those who are infected by it. A high-throughput screening (HTS) procedure that is capable of conducting the conventional screening assays rapidly is a trending method in today's drug discovery process (Symański et al., 2012). Despite its ability to significantly speed up the screening process and reduce errors, the cost of it remains as one of the biggest hurdles. Therefore,

an alternative procedure such as virtual screening (VS), would significantly reduce the cost of screening procedure since the physical form of the compounds are not needed (Kore et al., 2012).

The two primary methods that are commonly employed in VS are molecular docking and molecular dynamics (MD) simulations (Leelananda & Lindert, 2016). In general, molecular docking allows the prediction of binding location of the ligand on the target whereas molecular dynamics simulation mimic the natural dynamics present within the molecule (Meng et al., 2011; Zhao & Caflisch, 2014). A combination of docking and followed by MD simulation is often seen in many studies with the latter focused on the stability of the protein-ligand interaction throughout a stipulated duration. An extension of VS called ensemble-based VS is considered to perform better than the standard VS since it incorporates the flexibility of the ligand and the target as well (Silwoski et al., 2013). In the standard VS, the molecular docking procedure is usually conducted in a rigid-target and flexible-ligand manner. Ignoring the flexibility present on a target especially if the target is a protein may not be ideal since it is a known fact that protein is a structure with its own dynamic (Orozco, 2014). The concept behind the ensemble-based VS is to utilised multiple structural conformations to represent the target flexibility in the docking procedure (Amaro & Li, 2010). Therefore, it gratifies the need for a flexible target during the molecular docking procedure and would evidently strengthen the VS findings for potential inhibitors. With respect to that, the objectives of this study are:

1. To identify natural product compounds with the potential to serve as inhibitors to dengue virus protease via ensemble-based virtual screening.
2. To analyse the protein-ligand interactions between potential inhibitor candidates and dengue virus protease.

CHAPTER 2: LITERATURE REVIEW

2.1 The Dengue Virus

Dengue fever (DF) which is caused by the infection of the notorious dengue virus (DENV) is frequently reported in the tropical and subtropical countries (Nedjadi et al., 2015). Moreover, it is estimated that DENV infections may reach up to few millions each year across various countries (Simmons et al., 2012). This alarming rate of infection raised major concerns worldwide thus, various studies and research are still ongoing to combat dengue.

2.1.1 Background of Dengue

The word 'dengue' has an unclear origin but was stipulated to originate from the phrase "Ka dinga pepo" which means, a disease caused by an evil spirit in Bantu language used by Swahili people (Heilman et al., 2014). Even though the origin of the DENV is not precisely known, current research has proposed that it may be originated from primates located in the Asian forests (Scitable, 2014a). Nevertheless, it was known that the first dengue-like recorded case was found in the Chinese medical encyclopaedia dated back during the Jin dynasty (Gubler, 1998). Prior to the usage of the term 'dengue fever' which only started after the year 1828, the disease was known as 'break-bone fever' and was also called 'bilious remitting fever' (Heilman et al., 2014) in which both represent as the symptoms of the disease.

In the early periods of time, DF was considered as a benign disease and the occurrence between epidemics was sparse (Gubler & Clark, 1995). The earliest recorded epidemics that resembled dengue took place in the West Indies back in the year 1635 (Gubler, 1998). Meanwhile, the first reported dengue epidemic that occurred

in the Southeast Asia region was dated in the 1950s (Holmes et al., 1998). It was inferred that the starting point of DENV distribution throughout the Pacific and Southeast Asia regions might have begun during and after the World War II (Gubler, 1998). Following the world war, the Pan American Health Organization (PAHO) began a campaign to eliminate mosquitos in order to contain the yellow fever disease and dengue but was stopped in the 1970s (Murray et al., 2013). A study has suggested that due to the discontinued campaign, it has allowed the spread of DENV that originated from Asia to America (Messina et al., 2014).

Almost a century later, dengue has now become an endemic disease in over 100 countries (Calisher, 2005) whereas the recent DF epidemic in 2017 occurred in Sri Lanka and Burkina Faso (World Health Organization, 2017). As speculated by researchers after the 2014 dengue outbreak in Japan, these current dengue outbreaks and more future outbreaks are expected especially in the absence of a cure for DENV infections (Quam et al., 2016).

2.1.2 The Vectors of DENV Transmission

The main player for the DENV transmission that acts as a vector is a mosquito of the genus *Aedes* namely *Aedes aegypti* (*A. aegypti*) *and Aedes albopictus* (*A. albopictus*) with the latter being less efficient than the former (Murray et al., 2013). Both of the vectors are reported to be especially active during the daylight hours (Sharma, 2013). Furthermore, the female mosquitos for both of the species are the bearer and transmitter of DENV apart from other viruses since they require blood for their eggs production (Virginia Tech, 2015).

A. aegypti often become the primary concern since it has developed the adaptability to follow the urbanisation trends, enabling it to live and breed in close proximity to

humans (Greenwood, 2016). Moreover, it also prefers human's blood over other animals and is known to obtain their blood source from several individuals (Lafrance, 2016). This particular characteristic would ultimately increase the viral transmission from a single infected mosquito to multiple people within that mosquito's lifetime. Although its origin is still debatable, the spread of *A. aegypti* is significantly broad mainly in the tropical and subtropical countries (Murray et al., 2013). A recent study has also revealed that the mosquito is becoming more resistant towards insecticides (Thi et al., 2016) which are often used as a mean to control and/or eliminate their population. Therefore, with their resilient characteristics, countries with high population of *A. aegypti* often become the central location of dengue outbreaks (Lutomiah et al., 2016).

In the case of *A. albopticus*, it has a preference in biting mammals but not exclusively on humans (Rezza, 2012) in contrast with *A. aegypti*. The common name of *A. albopticus* is 'tiger mosquito' (Paupy et al., 2009) and unlike *A. aegypti*, they obtain their blood source from only one person instead of multiple people (Rezza, 2012). Nevertheless, researchers still emphasise on controlling their population since not only it has the capability to transmit DENV, they also have been continuously spreading across the country (Gratz, 2004). Acknowledging the danger, a recent campaign to eradicate their population was conducted in Australia and has shown a promising outcome (Muzari et al., 2017).

Although those two mosquitos are considered as the primary vector for DENV and their occurrence globally has been thoroughly studied and recorded (Kreamer et al., 2015), there are other potential vectors which include *Aedes japonicas*, *Aedes polynesiensis*, and *Aedes scutellaris* (Carrington & Simmons, 2014). On a side note, a current study has identified and confirmed that there was a high probability of DENV transmission to occur from an infected mother to a newborn child (Arragain et al.,

2016). Moreover, a previous study done by Chen and Wilson (2004) reported that the transmission of DENV could even occur through the exposure of the skin mucous membrane from an infected person. Therefore, it is clear that the transmission of DENV does not solely rely on vectors since direct human to human transmission is also plausible.

2.1.3 Overview of DENV Serotypes, Morphology, and Replication Cycle

The first successful isolation of the first DENV serotype (DENV-1) was achieved in the year 1943 in Japan, centuries after the first reported dengue case (Messina et al., 2014). There are currently four main serotypes of DENV (Bäck & Lundkvist, 2013). Even though a fifth serotype (DENV-5) was discovered in Malaysia (Normile, 2013), there has yet to be any follow up or any recorded genomic sequence on the new serotype. The earliest isolation of the second serotype (DENV-2) was done in Trinidad a decade after the DENV-1 isolation (Anderson, 1956), the third serotype (DENV-3) was isolated in Puerto Rico in 1963 (Russel et al., 1966) whereas the fourth serotype (DENV-4) was identified in Brazil in 1981 (Osanai, 1983). Each of the first four serotypes is serologically different from one another though they shared 60% sequence similarity in their genome (Dwivedi et al., 2017). Among the four serotypes, DENV-2 is considered as the most virulent since its infection would likely lead to a more severe form of DF, known as the HDF (Prommalikit & Thisyakorn, 2015). Moreover, it is also capable of infecting a larger number of mosquitos in comparison to the other serotypes (Rico-Hesse, 2015).

Regarding the DENV structure as shown in Figure 2.1, it is spherical with a diameter of 50 nm for the mature and 60 nm for the immature virus (Perera & Kuhn, 2008). The DENV structure (Figure 2.2A) can generally be divided into two main compartments namely the outer shell and inner core. The shell of DENV is made out of the envelope

(E) proteins and membrane (M/prM) proteins (Prera & Kuhn, 2008). Meanwhile, the core of DENV consists of the capsid (C) proteins that are often found in association with its genome, a positive single-stranded RNA (ssRNA) (Cruz-Oliveira et al., 2015).

The viral genome encodes for structural genes and non-structural genes which are translated into structural and non-structural proteins respectively (Solanki et al., 2017) as shown in Figure 2.2B. Overall, there are three structural proteins (E, M, and C) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) (Yacoub et al., 2016) produced by the genome. In brief, apart from providing structural integrity for DENV, the structural proteins also involve in the virus replication and assembly process (Brault et al., 2011; Nemésio et al., 2011; Byk & Gamarnik, 2016). It is also interesting to note that a study has identified that the M protein is also capable in triggering apoptosis of the virus (Brabant et al., 2009). Meanwhile, the functions carry out by most of the non-structural proteins are still unclear although some, such as NS3 and NS5 have a clear enzymatic functions in DENV (Zeidler et al., 2017).

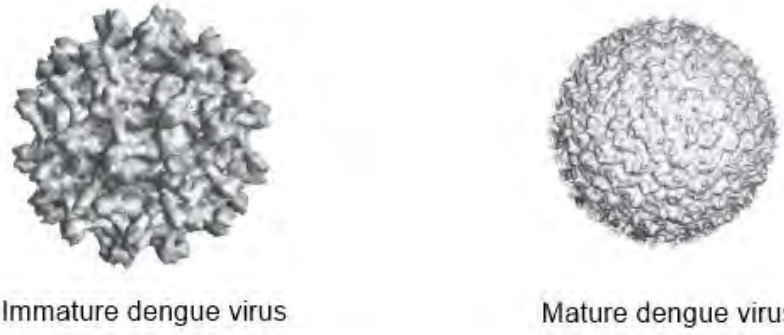


Figure 2.1: Structure of an immature and a mature DENV reconstructed from the cryo-electron microscopy. The immature DENV shows a rough surface making it having a larger diameter (60 nm) than the mature DENV. On the other hand, the mature dengue virus has a smoother surface and acquires a smaller diameter (50 nm). This figure is adapted from Perera and Kuhn (2008).

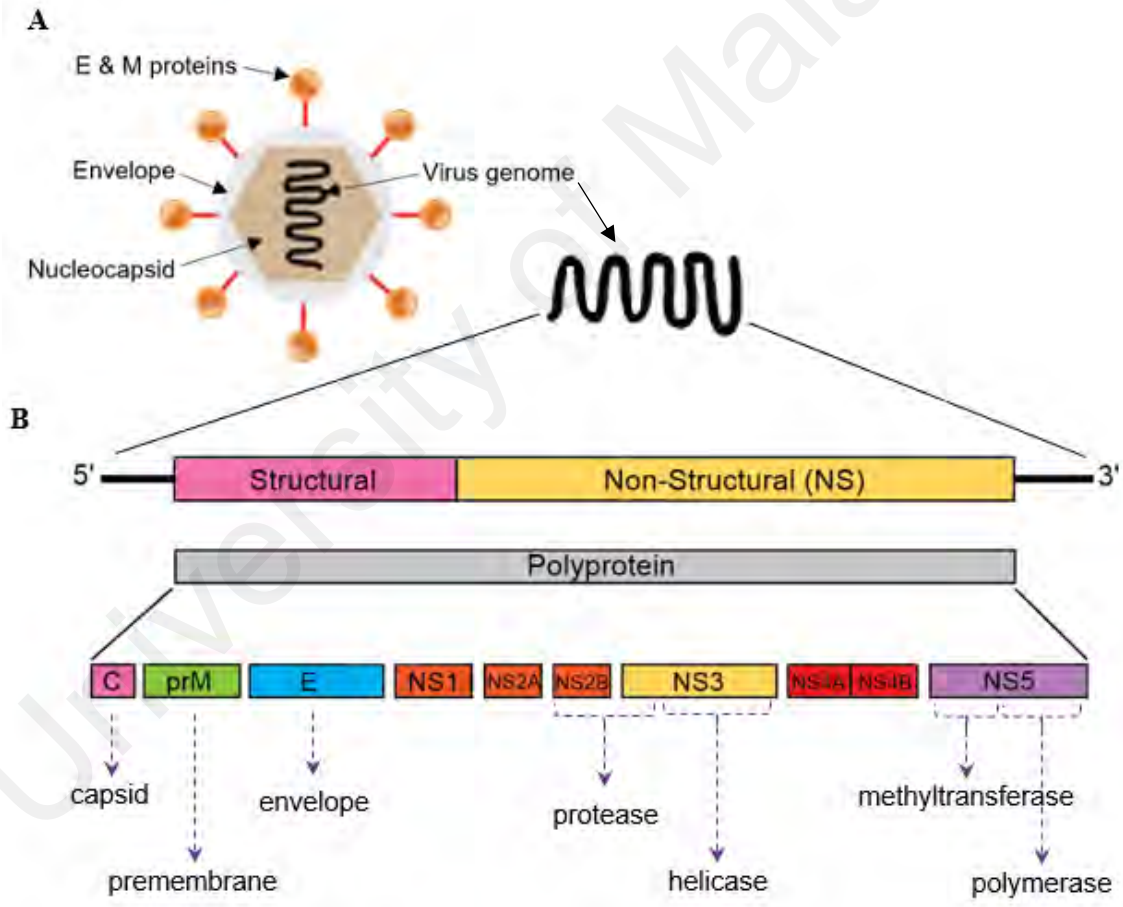


Figure 2.2: The general morphology and genomics of DENV. The basic structure of DENV (**A**) showing the general components that made up the DENV with the combination of capsid and virus genome made up the nucleocapsid. The DENV genome organisation (**B**) is similar with other flaviviruses having a positive ssRNA, translated into a single polyprotein that is further cleaved and processed (purple arrows) to produce the structural proteins (capsid, premembrane, and envelope). Only several of the non-structural proteins (NS2B, NS3, and NS5) undergo further processing and assembly (purple arrow) to produce the protease, helicase, methyltransferase, and polymerase enzyme of the DENV. These images were adapted from Scitable (2014b) and Simmonds et al., (2017).

A complete life cycle of DENV would include its replication within the vector and the host cells. The infected vector does not show any adverse effect or become ill since the vector is capable in generating a quick and efficient immune response during the viral replication period and hinders further amplification of the DENV RNA within it (Sánchez-Vargas et al., 2009). Unfortunately, the vector immunity does not eliminate the DENV entirely from its' system, and thus, the remaining DENV can still be transmitted to others (Sánchez-Vargas et al., 2009). In general, the replication process (Figure 2.3) begins with the viral entry into the cell via endocytosis upon binding with the host cell's receptor (Solanki et al., 2017). The capsid carrying the DENV genome is then released into the cell cytoplasm followed by uncoating and release of the virus genome (Clyde et al., 2006). The translation of the positive ssRNA generates a polyprotein consisting of the virus structural and NS proteins that are further processed in the endoplasmic reticulum (Perera & Kuhn, 2008). The transcription of the positive ssRNA coincides with the translation process to generate more polyproteins apart from a new genome (Clyde et al., 2006). The generated polyproteins are then processed and yield separate proteins used in virus packaging and replication process (Solanki et al., 2017). Finally, a newly assembled virion is later released from the cell via exocytosis (Solanki et al., 2017).

.

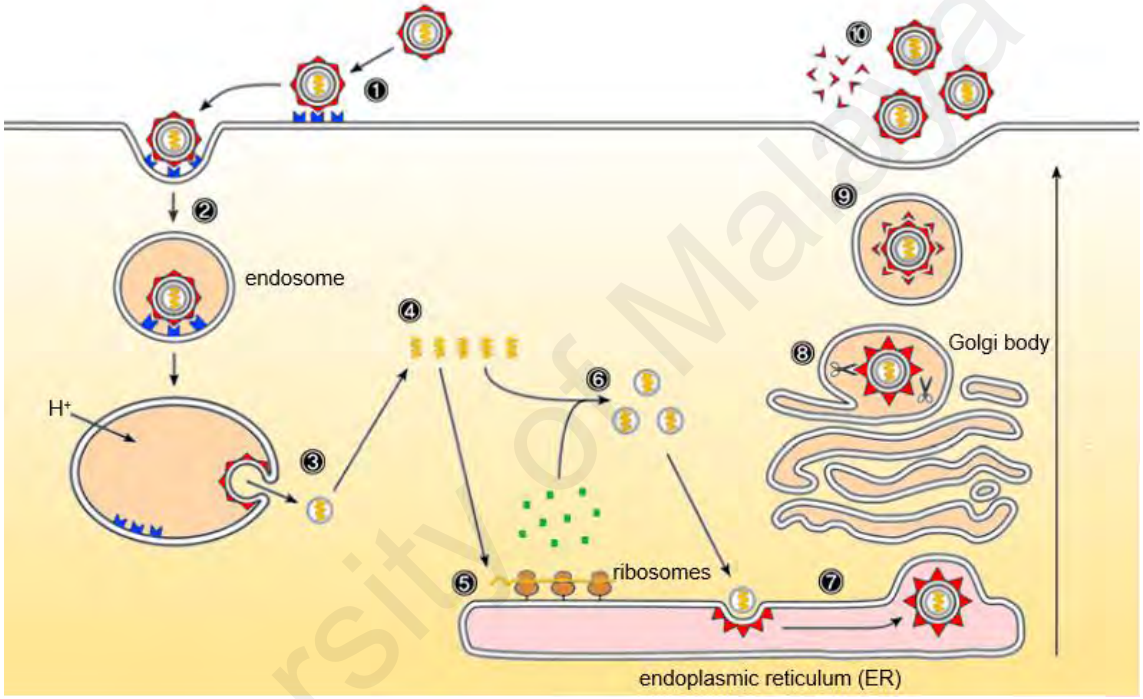


Figure 2.3: Overview of DENV replication process in an infected cell. The initial step in DENV infection starts from the binding of the virus to the cell's receptor (1) followed by endocytosis which forms an endosome carrying the DENV (2). The release of the capsid (3) is triggered when the endosome pH is lowered due to the proton (H^+) influx. Next, the capsid releases the virus RNA in the cytoplasm (4) to allow translation to takes place (5). The newly synthesised DENV protein together with a new copy of DENV RNA are packed (6) and assembled at the ER to form immature DENV (7). The immature DENV is then processed by the furin at the Golgi body (8) forming a mature DENV with its subviral particles (9). Finally, the release of the mature DENV and subviral particles from the infected cell occurred via exocytosis (10). This figure is adapted from Rodenhuis-Zybert et al., (2010).

2.1.4 Potential Targets and Their Current Promising Inhibitors

The comprehensive understanding of the DENV replication process enabled researchers to investigate relevant druggable targets that can be exploited for the antiviral production. The suggested potential targets can be either the virus own protein or the host's proteins that contribute in the virus replication process (Qi et al., 2008; Noble et al., 2010; Solanki et al., 2017). Among the host proteins that are considered as druggable include furin, glucosidase, and kinases (Noble et al., 2010). Meanwhile, the E protein, methyltransferase, helicase, and protease of DENV are among the other suggested targets (Qi et al., 2008; Noble et al., 2010).

Throughout the viral replication process, an immature virus shell is made out of the premembrane (prM) protein. Maturation of the virus involves in the conversion of prM to M protein and is achieved by the cleavage of prM protein by the host furin (Noble et al., 2010). This step is considered as essential in the viral maturation process (Elshuber et al., 2003) and thus, inhibiting the host furin would prevent the virus maturation and eventually prevent further infection. On the other hand, the host glucosidase is responsible in ensuring the correct folding and glycosylation of the viral proteins (Courageot et al., 2000) whereas the host kinases participate in the viral assembly and their release (Chu & Yang, 2007). Without the assistance of these host proteins, the replication of DENV within the host cell would come to a halt and the remaining viral components will eventually be discarded.

The E protein of DENV plays a pivotal role in the virus entry by mediating the binding of the virus to the host cell surface receptor and subsequently in the membrane fusion of the virus (Noble et al., 2010; Modis et al., 2004). Disruption in their interaction would ultimately prevent the DENV infection or hinders further spread of the virus within their host (Solanki et al., 2017). Meanwhile, a study has confirmed that

only the DENV methyltransferase generated from the N-terminal of the NS5 protein is capable of methylating some of the virus RNA substrates involved in RNA capping (Dong et al., 2007). Hence, the DENV methyltransferase appears as one of the attractive targets due to its specific and essential function (Noble et al., 2010) and it is also the most conserved dengue protein structure (Solanki et al., 2017). In addition to the methyltransferase, the DENV RNA helicase is inferred to have several roles during the virus replication process (Noble et al., 2010). Evidently, since DENV produce its very own helicase to carry out their specific function (Noble et al., 2010), it has also become one of the favourable targets in designing drug against dengue (Qi et al., 2008).

Although most researchers usually opt for the viral proteins instead of the host proteins as the drug target to prevent host toxicity, it does not prevent their investigation to be carried out since there has yet to be any effective antiviral against dengue (Solanki et al., 2017). Corresponding to that, a recent study was able to achieve a significant inhibition of DENV replication using a specific furin inhibitors (Kouretova et al., 2017). Moreover, Peng et al. (2017) reported that luteolin was able to act as uncompetitive inhibitors of furin and reduced the DENV count in an *in vivo* experiment. Apart from furin, Sayce et al. (2016) clarified and confirmed the action of iminosugars which inhibit the alpha-glucosidases on the endoplasmic reticulum. In addition, a previous study suggested the use of alpha-glucosidase inhibitor with another broad-spectrum antiviral would be more effective in treating DENV infection (Chang et al., 2011). However, there were limited studies that target the host kinases, one of which mentioned an antiviral that acts on the dengue protease and the host kinases (Vincetti et al., 2015).

There are currently many identified potential inhibitors against the DENV proteins such as butanoate that has the potential to inhibit the E protein (Desai et al., 2015).

Meanwhile, the attractive characteristics of the DENV methyltransferase have invited countless of study throughout the year searching for its inhibitor. A recent study by Benmansour et al. (2017) utilised several techniques which include the fragment-based method and was followed by a thermal-shift assay to validate their findings on the identified methyltransferase inhibitors (Benmansour et al., 2017). On the other hand, a non-metallic compound called suramin (Basavannacharya & Vasudevan, 2014), together with the analogue compounds of benzothiazole and pyrrolone (Sweeney et al., 2015) were reported to have the ability to inhibit the DENV helicase.

Most if not all of the mentioned and several other previous studies that have identified compounds with a promising potential to serve as the drug against dengue has yet to be marketed. The reason being might be that most of them are still in the pre-clinical trial stage (Solanki et al., 2017). However, despite their profound activity against DENV, the existence of multiple serotypes in DENV (Gu & Shi, 2014) often present as a challenge and limitation in their utilisation for the generation of antiviral against DENV.

2.1.5 Dengue Protease as the Favourable Target

The DENV protease is made up of the NS3 protease domain and to achieve optimal enzymatic activity, it requires a cofactor from the NS2B protein (Phong et al., 2011) to form a complex (NS2B/NS3). The protease is a serine protease with the catalytic triads (His-51, Asp-75, and Ser-135) located at its active site (on the protease domain) as the central player for its catalytic activity (Nitsche et al., 2014). Crystal structures of DENV protease complex deposited in the Protein Data Bank (PDB) as shown in Figure 2.4, is often associated with a non-peptide linker (Gly_4-Ser-Gly_4) which connects the cofactor and domain to achieve complex stability (Arakaki et al., 2002). Concerning its function, the complex is involved in the DENV polyprotein processing translated from the

ssRNA (Noble et al., 2010). Thus, targeting the complex would prevent the production of a new dengue protein (non-structural and structural) to be used in the virus assembly step.

Among all of the potential targets, the dengue protease is considered as the most preferred or favoured. One of the main reasons is, there are currently several approved drugs used to treat HIV (Menendez-Arias, 2010) and HCV (Kok, 2016) that targets the virus protease, hence, it is considered as a proven strategy. In addition, the dengue protease structure also shared a high structural similarity in all of the DENV serotypes (Dwivedi et al., 2017). However, discovering a potent compound to act as the protease inhibitor remain as a challenge since the active site has a shallow surface (Solankit et al., 2017) and exhibits a charged (negative) feature (Noble et al., 2010).

Nonetheless, countless of efforts and studies have been made in order to identify compounds that could serve as the inhibitor of the DENV protease. In the year 2017 alone, there were more than ten reported literatures that focus on identifying a potent inhibitor against the dengue protease. Among them was a study conducted by Lin and colleagues (2017) whereby they were able to determine a couple of cyclic peptides that showed inhibitory activity on the dengue protease. Another study also claimed to discover a cyclic peptide that inhibits the protease activity (Takagi et al., 2017). Meanwhile, apart from discovering a competitive inhibitor, other studies also focus on finding a non-competitive or uncompetitive inhibitor of protease (Aguilera-Pesantes et al., 2017). Therefore, as the search continues in the identification of promising compounds for the DENV activity inhibition, it will hopefully lead to the production of an effective antiviral against DENV.

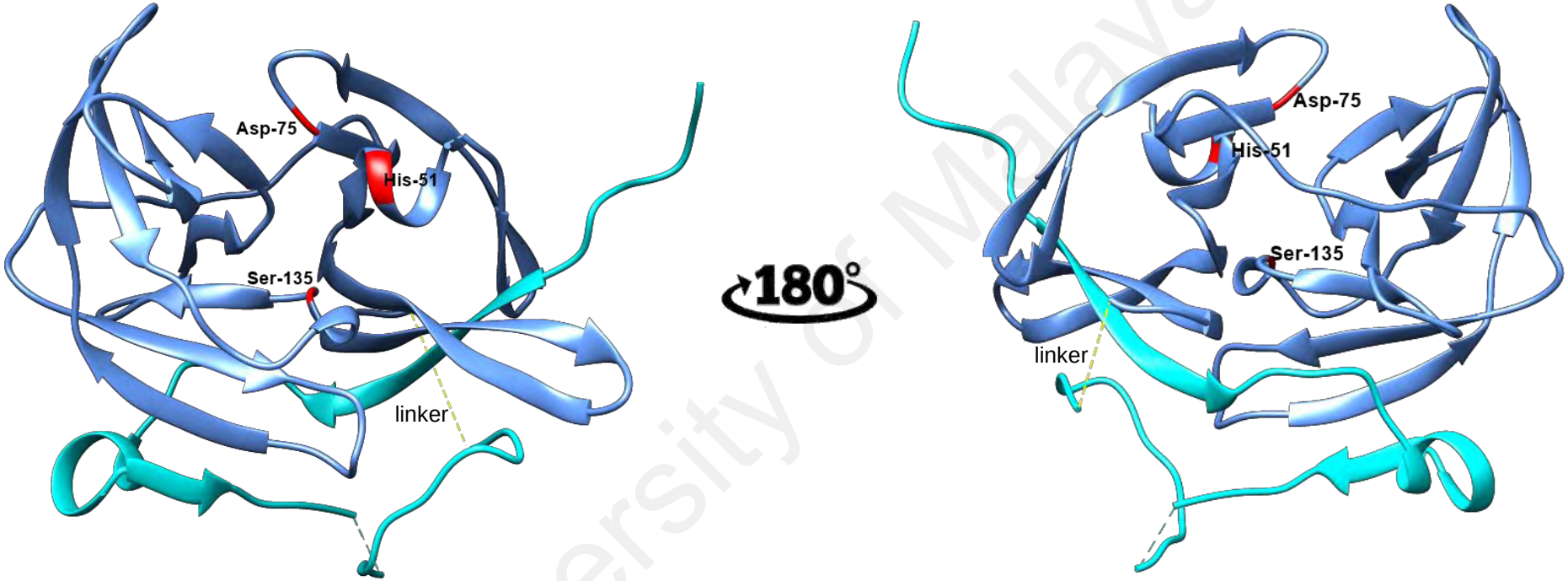


Figure 2.4: The first crystallised structure of dengue protease complex (PDB ID: 2FOM) deposited in the PDB. The NS2B cofactor (cyan) is connected to the NS3 protease domain (blue) via a non-peptide linker (yellow). The dashed lines represent missing residue on the crystallised structure which includes the linker. The catalytic triads (His-51, Asp-75, and Ser-135) that are located at the active site on the protease domain are marked by the red colour. This image was generated via UCSF Chimera.

2.2 Current Effort in Combating Dengue

In the absence of a specific treatment for dengue, most of the attention turns towards the prevention method in order to control and hopefully lower the number of reported dengue cases. The universally applied method is by controlling the breeding of the *Aedes* mosquito. Furthermore, after more than a century of study in DENV, the first approved vaccine is finally available on the market (World Health Organization, 2016). Evidently, the prevention approach mainly aims to reduce the spread of DENV and may not be able to eliminate the dengue threat entirely and thus, the development of a specific antiviral against DENV is still much needed.

2.2.1 Vector Control

Since the vector is capable of not only transmitting the DENV but other viruses as well, controlling or eradicating their population is considered as an attractive solution. Moreover, the uncontrolled urbanisation and the climate changes have allowed the formation of favourable breeding ground for the mosquitoes have shown to contribute towards the increasing number of dengue cases (Hii et al., 2016). Hence, a substantial amount of efforts have been put forward in vector control and there are currently three primary employed methods categorised as the environmental, chemical, and biological means.

The environmental mean in eradicating the mosquito population mainly involving the environmental sanitation and management such as cleaning and removing container that may become their breeding sites (Yacoub et al., 2016). It requires the whole community to work together to make sure there is no pool of stagnant water left for days. Although the effort in raising public awareness especially in ensuring the elimination of stagnant pool of water in the household areas has not stopped, the

numbers of dengue cases are still on the rise (Azam, 2016). Therefore, this method alone is not sufficient in controlling the vector population.

Despite being the fastest mean in eliminating the vector, the usage of synthetic insecticides which is the chemical mean in vector control is often used as the last resort especially during an outbreak (Yacoub et al., 2016). Another approach via chemical mean involves in targeting the larvae instead of the adult mosquito by using larvicides (Yacoub et al., 2016). A study conducted in Cambodia reported that interventions of the mosquito life cycle using the larvicides is considered as cost-effective in curbing the mosquito population (Suaya et al., 2007). Overall, this approach is not the method of choice since it has several deteriorating disadvantages such as it would result in the emergence of highly resistance mosquito against insecticide (Luz et al., 2011) and it also causes environmental and health hazard (Achee et al., 2015).

The biological mean that is currently being widely accepted is by infecting the vector with a bacterium called *Wolbachia* (Yacoub et al., 2016). The considerably high acceptance is due to the fact that utilising *Wolbachia* is considered as a natural mean since it can be commonly found infecting other insect species (Araújo et al., 2015). In general, vector infected with *Wolbachia* would cause the vector to have a shorter lifespan (McMeniman et al., 2009) and also hinders the DENV replication hence, ultimately disrupt the DENV transmission (Moreira et al., 2009). The field trials are currently being held in several countries including Vietnam, Malaysia, and Indonesia (Yacoub et al., 2016) and if proven effective, it may be integrated with another prevention method to significantly reduce the *Aedes* mosquito population.

2.2.2 Immunisation Against DENV

Those who are infected by one of the DENV serotypes may confer immunity against that particular serotype within a particular period (Forshey et al., 2016). However, a subsequent infection, especially by another serotype would increase the risk of developing a more severe form of dengue that may be deadly (Murphy & Whitehead, 2011). Undoubtedly the existence of multiple DENV serotypes complicates the generation of an effective vaccine. Regardless to that, the first approved vaccine named Dengvaxia by Sanofi Pasteur was recently made available in the market (World Health Organization, 2016).

Dengvaxia was first licensed in Mexico in 2015 (Yacoub et al., 2016) and contain a live recombinant tetravalent dengue virus that was aimed to provide immunity against all of the four DENV serotypes (Solanki et al., 2017). However, there has been variation level of effectiveness against each of the serotypes and is only recommended for children above nine years old (Solanki et al., 2017). The promising report together with the recommendation by the WHO (World Health Organization, 2016) has made Dengvaxia as an attractive solution (Godói et al., 2017) especially in the dengue-endemic countries (Gupta, 2016; Halstead, 2016). The Philippines is one of the countries that invested millions of dollars on Dengvaxia and launched a publicly funded vaccination programme for children in order to curb the ever-increasing dengue cases (Normile, 2017).

Nevertheless, a vaccine is not without its danger and as highlighted by Aguiar et al. (2016) on Dengvaxia but was not emphasised upon. As a result, the children in the Philippines experienced the adverse effects the vaccine brings hence raising the controversial issues of Dengvaxia and Sanofi Pasteur's reports (Normile, 2017). This problem would evidently affect other vaccines' status, acceptance, and development

since many more vaccine candidates are currently undergoing clinical trials (Solanki et al., 2017).

2.3 Natural Source for Antiviral Development

Natural compounds, especially from the medicinal plants often provide additional advantages against synthetic products especially with their wide range of pharmacophore properties (Harvey et al., 2015). It was estimated that around 34% medicines approved by the Food and Drug Administration (FDA) consists of natural products together with their derivatives and may show an increasing trend in the future (Newman & Cragg, 2012). Moreover, natural compounds are often considered to be relatively safer than the synthetic compounds.

2.3.1 Traditional Means of Dengue Treatment and Management

Since DF has been around for centuries, home remedies were the primary solution in treating any conditions prior to the availability of today's current drugs and medical treatments. Most, if not all of the home remedies used natural sources such as herbs and other medicinal plants. Although the exact mechanism on how does most of the remedies are able to relieve a particular illness may still be unclear, the fact that it is cheaper and often without any deteriorating side effect (Chin, 2014) may be the main reason that it is still implemented in today's society.

Papaya leave juice is among the frequently used alternatives therapy in DF (Chin, 2014; CureJoy, 2017). The extract helps in increasing the platelet count by preventing platelet destruction in dengue patients (Sarala & Paknikar, 2014). Other than the papaya leave juice, neem juice and oil are also used as a traditional mean in reducing fever and they also help to increase the white blood cell and platelet count (Seal, 2015). Moreover, a study has shown that neem also holds an inhibitory potential for the DENV

replication process (Parida et al., 2002). Meanwhile, in the Philippines, a herb called "tawa-tawa" is commonly used as the treatment for DF (Chin, 2014; CureJoy, 2017) with the capability to reduce DF symptoms (Mir et al., 2012). Other traditional alternatives include kakamachi leaves, coriander, barley grass, and orange juice which mostly helps in increasing platelet count and reducing the body temperature (Seal, 2015).

2.3.2 The Databases of Natural Compounds

Acknowledging the importance and opportunities present in natural products (Koehn & Carter, 2005), several databases dedicated to compiling and storing data, including their structural information are available and are frequently updated (Xie et al., 2015; Harvey et al., 2015). Several of the natural product databases that are currently available are listed in Table 2.1.

Among all of the natural-based databases, Super Natural II appeared to be the most comprehensive by having the current number of entries of more than 325,000 compounds (Banerjee et al., 2015). It is the second updated version and was initially known as the Supernatural database that was published in the year 2006 (Banerjee et al., 2015). Currently, it incorporates 16 vendors with their retrievable information and also provides a brief report of the stored natural compounds (Xie et al., 2015). Apart from providing compound search by properties (name, molecular weight, toxicity class, etc.), it also provides structure search by drawing it via ChemDoodle (Xie et al., 2015). Other exciting features of the Super Natural II database include toxicity prediction and compound clustering (Xie et al., 2015; Banerjee et al., 2015). This open source database undoubtedly would serve as an essential resource for researchers worldwide and is accessible at http://bioinf-applied.charite.de/supernatural_new/index.php?site=home.

Table 2.1: Available databases specialising in storing compounds obtained from natural sources.

Database	Number of current entries	General information	Reference
Super Natural II	325,509	♦ comprises of several other natural databases/ suppliers	Banerjee et al., 2015
Universal Natural Product Database (UNPD)	229,358	♦ assembled molecules from other available traditional Chinese medicines databases ♦ incorporated in Super Natural II database	Gu et al., 2013
TCM Database@ Taiwan	37,170	♦ currently the world largest database for traditional Chinese medicine ♦ incorporated in Super Natural II database	Chen, 2011
NuBBE Database	~ 640	♦ compounds from Brazilian sources ♦ incorporated in Super Natural II database	Valli et al., 2013
Natural Products Alert (NAPRALERT)	NA*	♦ a relational database of all-natural products ♦ a premium database	Loub et al., 1985

*NA: not available

2.4 Drug Discovery for Potential Dengue Antiviral Compounds

Hundreds of thousands if not millions of dollars were spent on dengue research alone in order to understand the dengue replication process and its pathogenesis to enable the identification of potential targets that can be exploited for drug development. With the proposed potential targets mentioned earlier (Qi et al., 2008; Noble et al., 2010; Solanki et al., 2017), further research is required to discover suitable inhibitory compounds that are capable in intervening or disrupting the dengue life cycle.

2.4.1 Conventional Screening Approach

The conventional ways to identify potential compounds can be either via *in vitro* or *in vivo* means. There are a plethora of techniques and approaches involving assays available with their established protocols (Hughes et al., 2011). Although the choice of assays depends on many aspects such as the availability of equipment and infrastructure, the most frequently used technique is via affinity chromatography mainly due to its established methods (Lomenick et al., 2011). Currently, utilisation of a high-throughput screening (HTS) method that uses automated robots and detectors are trending especially in the pharmaceutical industry (Symański et al., 2012).

Nevertheless, be it either automated or manually, both approaches require robust assay protocols to identify the desired compounds successfully. Moreover, apart from acquiring the compounds, generation of a stable targeted macromolecule is also needed prior to the screening procedure. Gong et al. (2013a) reported a method to generate a stable dengue NS5 polymerase to be used as a target molecule including the corresponding biochemical assay for analysis. Other tests in identifying antiviral compound especially for DENV include the plaque reduction assay (Gong et al., 2013b), high-performance liquid chromatography (HPLC) coupled with fluorometry

(Nitsche & Klein, 2013), and by the usage of genetically engineered reporter virus system (Fischl & Bartenschlager, 2013).

The validation obtained from these assays is a necessity to proceed through all of the pre-clinical trials. However, these protocols are often laborious, time-consuming and expensive. Moreover, the cost of the overall drug research and development has shown an increasing trend throughout the years (DiMasi et al., 2016). Availability of a relatively cheaper alternative which is by virtual high-throughput screening (vHTS) or sometimes known as virtual screening (VS), is gaining popularity in this modern day especially with the advance development of sophisticated computer hardware and software (Kore et al., 2012).

2.4.2 *In silico* Approach

Started in the 1980s, computer aided drug design (CADD) is an *in silico* means of rational drug design and discovery that provides multiple strategies for drug design by using the computational software (Kore et al., 2012). Generally, there are two methods in CADD, namely ligand-based and structure-based drug discovery. The ligand-based drug discovery (LBDD) is typically applied when the three-dimensional (3D) structure of the target of interest is unknown or unavailable contrariwise with the structure-based drug discovery (SBDD) (Leelananda & Lindert, 2016). Methods available for LBDD include but not limited to pharmacophore modelling, quantitative structure-activity relationship and molecular similarity approach (Acharya et al., 2012). Although construction of the predictive 3D structure of the target can be done via homology modelling if the structure is absent, it is often a challenge and sometimes unsuitable (Leelananda & Lindert, 2016) thus, making the LBDD as a better option.

The concept behind SBDD is based on a hypothesis that favourable interaction between the target and a molecule (or a ligand) will subsequently affect the target's biological activity (Silwoski et al., 2014). Approved drugs developed via SBDD include the drug that treats glaucoma called Dorzolamide (Talele et al., 2010) and Amprenavir, an HIV protease inhibitor (Kim et al., 1995). Aforementioned, SBDD requires the availability of the potential target in a 3D structure whereby most of it is usually deposited in the PDB (Rose et al., 2015). Nevertheless, in the absence of the 3D configuration, the target structure prediction can be generated from homology modelling, threading or by *ab initio* folding (Leelananda & Lindert, 2016). The two most widely used methods in SBDD are molecular docking and molecular dynamics (MD) simulation.

Molecular docking mainly involves in predicting the best possible orientation of the target and ligand and simultaneously estimates the generated binding affinity (Leelanada & Lindert, 2016; Meng et al., 2012). Ligands obtained from a particular database are readily in 3D configuration (Leelananda & Lindert, 2016) and can even be manually generated using an open source bioinformatics tools such as VegaZZ (Pedretti et al., 2004) and Avogadro (Hanwell et al., 2012). The approach in molecular docking methods includes using both of the receptor and ligand in a rigid or flexible form, or a mixture of both (Meng et al., 2011). However, incorporating flexible ligand and a flexible receptor is computationally demanding and occupied a more extended period of completion (Lavecchia & Di Giovanni, 2013) therefore, most researchers utilised the flexible ligand and rigid receptor approach. Currently there are several of molecular docking programs available such as AutoDock Vina (Trott & Olson, 2011), AutoDock4 (Morris et al., 2009), FlexX (Rarey et al., 1996), Glide (Freisner et al., 2004) and GOLD (Jones et al., 1997). In essence, these programs calculate scores using specified docking

algorithm to rank the poses and binding energy produced between the target and ligand (Leelanda & Lindert, 2016).

Meanwhile, MD simulation aimed to simulate the existing natural molecular forces of the protein structure (Zhao & Caflisch, 2014) within the set duration according to Newton's law of physics (De Vivo et al., 2016). Thus, it provides comprehensive insights of the molecular behaviour and interactions at the atomic level in its natural environment (Durrant & McCammon, 2011). Integrating MD simulation after docking procedure allows a better understanding of the 'after effect' of ligand docking and the induced fit effect towards the targeted protein as well. The popular MD software are AMBER (Case et al., 2007), GROMACS (Abraham et al., 2015), and NAMD (Philips et al., 2008). The current research that applied vHTS via molecular docking and MD simulation reported to identified promising compounds filtered from the ZINC database to be used as inhibitors of DENV helicase and protease (Mirza et al., 2016).

2.4.3 The Ensemble-based Virtual Screening

Aforementioned, to perform docking simulation using a flexible protein conformation is computationally demanding and time-consuming. Moreover, it also requires a reliable tool to successfully conduct and analyse the procedure (Lavecchia & Di Giovanni, 2013). Hence, studies incorporating this approach which would provide valuable data and information are somewhat challenging (Meng et al., 2011). An alternative approach to conduct the mentioned method is via ensemble-based virtual screening.

In nature, macromolecule especially protein does not exist in a static form but exists in multiple conformations as a result of interaction with its molecule and the environment condition (Orozco, 2014). Even though most SBDD studies and

researchers address this knowledge by applying MD simulation after docking, it is somewhat implying that the molecular structure would only have a significant change after the docking of a particular ligand. Nonetheless, this is not the case especially for protein structure. Showed by Jamroz et al. (2016), in the absence of a ligand, the structure showed multiple configurations throughout the timesteps. Therefore, the ensemble-based approach acknowledged the dynamic present in a protein by emphasising on using various conformations obtained from the protein as clusters (Amaro & Li, 2010). The ensemble clusters can be then subjected to molecular docking for further analysis. The multiple conformations can be obtained by either obtaining several X-ray crystallography structures, nuclear magnetic resonance (NMR) structures or from MD trajectories (Amaro & Li, 2010).

Performing ensemble-based virtual screening indirectly gratify the needs of flexible protein configuration in docking process (Leelanda & Lindert, 2016; Silwoski et al., 2013; Amaro & Li, 2010). Moreover, apart from owning the previously stated advantages of conducting vHTS, ensemble-based offers more in-depth understanding and analysis of the protein-ligand interaction within the different target conformations. Recently, a study utilising this approach to better understand the inhibitors' profile against a critical enzyme in the Alzheimer's diseases emphasised on the importance of incorporating the structural flexibility in molecular docking (Sorin et al., 2017). Unfortunately, studies and research using this approach to screen potential lead compounds as the dengue antiviral are still lacking.

CHAPTER 3: MATERIALS AND METHODS

Ensemble-based virtual screening comprised of multiple stages that necessitates the need to utilise variable software and programs. To ensure the robustness of the protocol, generated data or result from majority of the stages were assessed before proceeding to the subsequent step. The main stages include the collection of the targets (protein and ligand), molecular dynamics simulations of the protein, followed by virtual screening as summarised in Figure 3.1.

3.1 Workstation Equipment

A workstation using Ubuntu Linux version 14.04 LTS (64-bit) as the operating system together with 32 gigabytes of Random Access Memory (RAM) was used to conduct this study. Moreover, it was also equipped with 12 Intel® Core™ i7-4930K 3.40 gigahertz (GHz) microprocessor and used GeForce GTX780 Ti as the graphics processing unit (GPU).

3.2 Acquiring and Refining the Dengue Virus Protease Complex

As the year progresses, the number of the deposited three-dimensional (3D) structure of protease for dengue virus (DENV) in the Protein Data Bank (PDB) does not show any tremendous increment with the current total stands at 16 structures. Although there are currently two nuclear magnetic resonance (NMR) structures of dengue protease available in PDB, the corresponding publication is yet to be made available hence were excluded in this study.

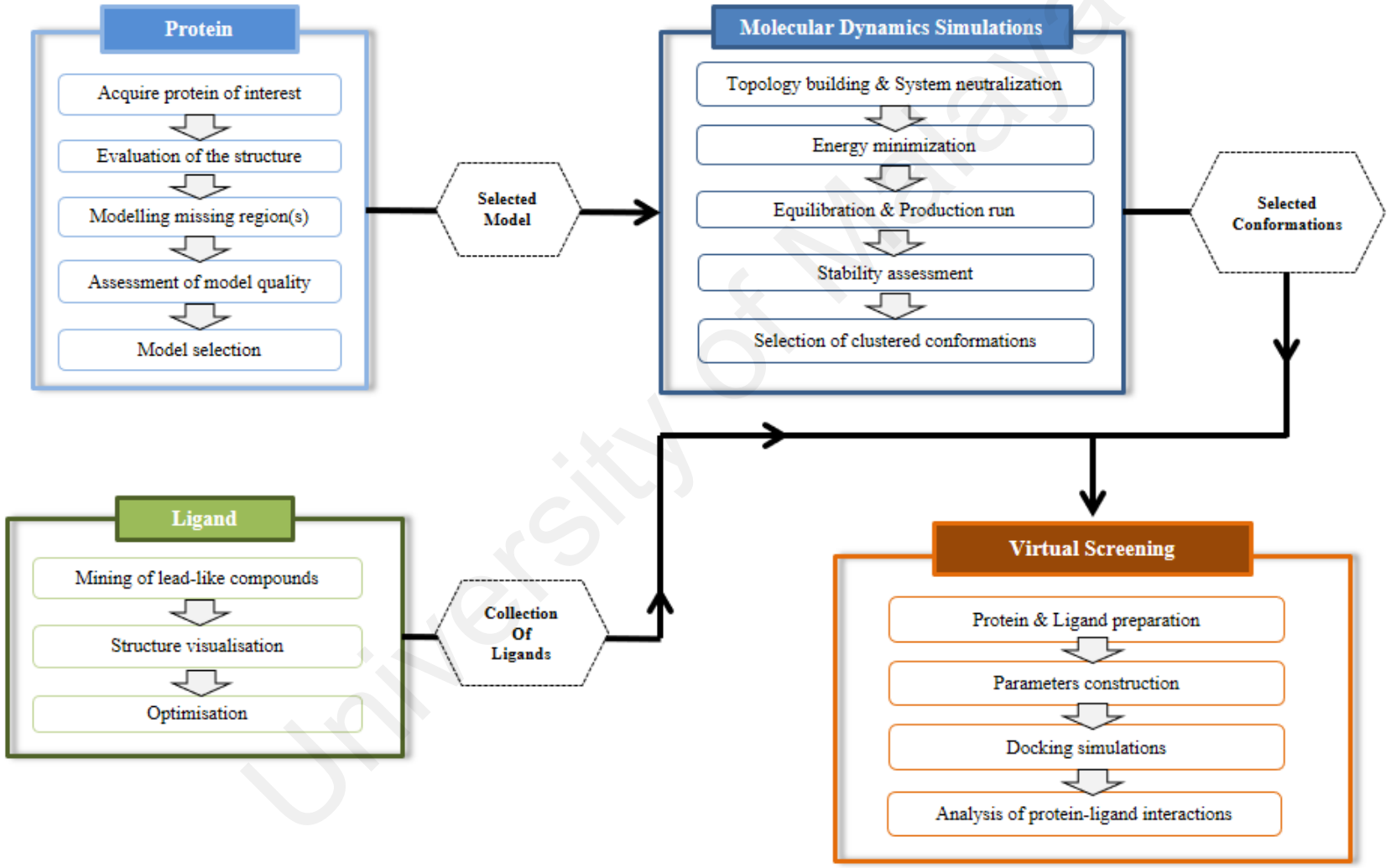


Figure 3.1: The general workflow of ensemble-based virtual screening.

3.2.1 Selection of Dengue Protease Complex Structure

The main target of interest was the dengue protease complex (holoenzyme structure) where the structure comprised of two non-structural (NS) proteins, namely the NS2B cofactor and NS3 protease domain. There were several dengue protease complex structures deposited in PDB showing two distinct conformations, namely the 'open' and 'closed' structure (Nitsche et al., 2014). In general, the 'open' and 'closed' conformations differ in their NS2B cofactor positions whereby in the 'open' conformation, it is located away from the active site whereas it is adjacent to the active site in the 'closed' conformation. In this study, both of the 'open' and 'closed' structural configurations were utilised to ensure maximal coverage of the available structure. All of the available structures were analysed and the selected representative for the 'open' configuration was the protease complex with the PDB ID: 4M9K (Yildiz et al., 2013) whereas 3U1I (Noble et al., 2012) was selected for the 'closed' configuration. These structures were further assessed and refined.

3.2.2 Protease Complex of Different Serotypes

The selected 3D structures were from different serotypes in which the 'open' complex was from the second serotype (DENV-2) whereas the 'closed' complex was from the third serotype (DENV-3). Therefore, to determine the degree of similarity shared between the four main DENV serotypes, multiple sequence alignment (MSA) was conducted via T-COFFEE MSA server (Di Tommaso et al., 2011). The protein sequences of NS2B and NS3 of the four serotypes were obtained from RefSeq (Table 3.1) and MSA was employed together with the protein sequences of the selected 'open' and 'closed' structures. Moreover, a second MSA for the NS3 sequences was carried out using only the protease domain sequence extracted from the full NS3 sequences obtained from RefSeq.

Table 3.1: The RefSeq ID of NS2B, NS3, and protease domain position for each of the dengue virus serotype.

Serotype	RefSeq ID		Position of Protease Domain on NS3
	NS2B	NS3	
DENV-1	NP_733809.1	NP_722463.1	20 – 167
DENV-2	NP_739586.2	NP_739587.2	19 – 167
DENV-3	YP_001531171.3	YP_001531172.2	21 – 167
DENV-4	NP_740320.1	NP_740321.1	19 – 167

3.2.3 Refinement of Dengue Protease Complexes

The protease complex structures were viewed using UCSF Chimera (Pettersen et al., 2004) followed by removal of any artefacts and crystal water present on the structures. Chain A (NS2B) and Chain B (NS3) were chosen in the 'closed' structure to be used as the representative complex. Moreover, the synthetic ligand present was also excluded from the protease structure and was later used as a control ligand. It was noted that both of the selected 'open' and 'closed' protease complex structures have missing atoms and residues within the structure.

The recorded missing atoms were replaced back into the structures using UCSF Chimera whereas MODELLER 9.18 (Sali & Blundell, 1993) was utilised to address the missing residues recorded on the structures. There are two main methods available in MODELLER for loop modelling namely the standard and DOPE-based protocols. The DOPE-based protocol consists of the DOPE and DOPEHR modelling protocols. All of the three protocols (standard, DOPE, and DOPEHR) were employed for the loop modelling in the protease complex where 15 models were generated from each of the protocols. Therefore, a total of 45 models were produced for each of the protease complexes. Throughout the modelling process, restraints were applied on the existing residues to prevent them from moving. The best model from each protocol was then chosen based on the discrete optimised protein energy (DOPE) score.

The quality of the selected models was further evaluated using PROCHECK (Laskowski et al., 1993). The total of six selected models generated from each protocol for both of the protease complexes was submitted to PROCHECK. The stereochemical quality assessment of the generated models was done by analysing the generated Ramachandran plots. Based on the results, the most favoured model for each of the protease complex was selected and utilised for the following procedures.

3.3 Molecular Dynamics (MD) Simulations

In order to simulate the existing natural molecular forces of the protein structure (Zhao & Caflisch, 2014), MD simulations of the modelled 'open' and 'closed' protease complexes were performed using GROMACS version 5.0 (Abraham et al., 2015). The general preparation steps were topology preparation, energy minimisation, and equilibration run which later followed by the final production run. To obtain sufficient sample size, replicates of MD simulation (n = 3) for each of the protease complex was performed. Moreover, both of the protease structures were subjected into a separate MD simulation with similar parameters and conditions.

3.3.1 Protease Complex Topology Preparation

Both of the selected protease complexes ('open' and 'closed') were hydrated in a cubical box with the TIP3P water model (Jorgensen et al., 1983) having a minimum distance of 15 Å apart from the protein. Counter ions were also added replacing the existing water molecules to established system neutralisation.

3.3.2 System Energy Minimisation, Equilibration and Production Run

Prior to MD runs, the system underwent steepest descent energy minimisation utilizing CHARMM22-CMAP force field (MacKerell et al., 2004) with the maximum

number of 50,000 steps. The long long-range electrostatic interactions within the system were calculated and treated by particle-mesh Ewald (PME) method (Darden et al., 1993). Moreover, Verlet neighbour list algorithm (Verlet, 1967) was used for the nearest-neighbour search.

In addition to the mentioned parameters, the simulations' condition for both of the complexes were maintained at a constant temperature of 300 K by using the V-rescale thermostat (Bussi et al., 2007) throughout the equilibration run (200 ps). Whereas the reference pressure for the pressure coupling was set to 1 atm by using Parrinello-Rahman barostat (Parrinello & Rahman, 1981). Furthermore, all bonds involving hydrogen atoms were constrained using LINCS algorithm (Hess et al., 1997). A total duration of 10 ns simulation of the production run using the parameters mentioned previously was performed for both of the protease complexes. The trajectory frames generated from the MD simulation were captured at 10 ps intervals. Finally, the analysis of the MD simulation was then conducted upon each of the completed simulations.

3.3.3 MD Simulations Analysis and Conformations Clustering

Verification of the system stability was analysed via the root-mean-square deviation (RMSD) and total energy of the protein generated throughout the stipulated simulation period. In addition, the root-mean-square fluctuations (RMSF) of the residues were also generated to evaluate the protein flexibility. The generated MD simulation trajectories were also visualized using UCSF Chimera.

Prior to clustering, the replicates of MD simulation trajectories produced from each of the protease complexes were first concatenated, generating a single trajectory with a total duration of 30 ns for each. Next, the RMSD matrices were computed for each of the concatenated trajectories by least square fitting on the carbon-alpha (Cα). The

matrix was then utilised to cluster the generated conformations via gromos clustering algorithm to group conformations with similar configuration together within a fixed RMSD threshold. Several clustering cutoff values were tested (2.0, 2.5, 3.0, and 3.5 Å) on both of the concatenated trajectories. Based on the obtained result, the first 10 generated clusters were then selected from the 'open' protease complex simulation (cutoff = 2.5 Å) whereas, the first 11 clusters were selected from the 'closed' protease complex simulation (cutoff = 3.0 Å). Therefore, a total of 21 selected conformations from the cluster were then extracted and renamed as models accordingly

3.4 Collection of Potential Protease Inhibitors from Natural Resources

In this study, compounds originated from natural resources were used in the search for potential protease complex inhibitors. Several of the searchable databases focusing on natural products are listed in Table 2.1. It is interesting to note that the majority of the listed databases were incorporated within the Super Natural II database making it the most comprehensive natural database and hence, it was utilised in this study to produce a list of ligands for the virtual screening purposes. Moreover, as previously discussed in Section 2.3.2, this database also provides easy compound search and filter using any fixed properties. The home page of the Super Natural II database is as shown in Figure 3.2.

3.4.1 Ligand Search Criteria and Optimisation

In order to ensure the extracted ligands would serve as a promising compound, a set of criteria was fixed. The criteria used (Table 3.2) was mainly based on Lipinski's rule of five (Lipinski, 1997). However, certain criterion such as the range of molecular mass was changed to cater the natural compound properties (Lipinski, 2004; Harvey et al., 2015). Based on the criteria employed, the generated compounds were extracted from

the database, inspected using UCSF Chimera, and renamed accordingly. Each of the ligands was then treated for energy minimisation via Avogadro (Hanwell et al., 2012) by applying the universal force field (Rappe et al., 1992). In addition, the synthetic ligand excluded earlier from the 'closed' protease complex was also included into the collection to act as a control ligand.

3.5 Ensemble-based Virtual Screening

The collection of ligands together with the selected conformations of the complexes was utilised for the ensemble-based virtual screening procedure. The docking simulations of the collected ligands against the selected protease complex models were employed via AutoDock4 version 4.2 (Morris et al., 2009). In addition, the preparation of molecules, analysis, and visualisation of the docking results were carried out using AutoDockTools version 1.5.6 (Morris et al., 2009). The docking procedure was performed in a rigid-protein flexible-ligand manner.

3.5.1 Preparation of Macromolecule Models and Ligands

In the model preparations, hydrogens were first added on each of the selected models followed by the merging of the non-polar hydrogen atoms. Next, Gasteiger partial charges were added onto the models. On the other hand, the collection of ligands was first converted into an accepted format via Open Babel (O'Boyle et al., 2011) followed by setting up the root detection and allowing torsional rotation during docking for each of the ligands. The merging of non-polar hydrogen atoms and addition of Gasteiger charges were also applied on the ligands.

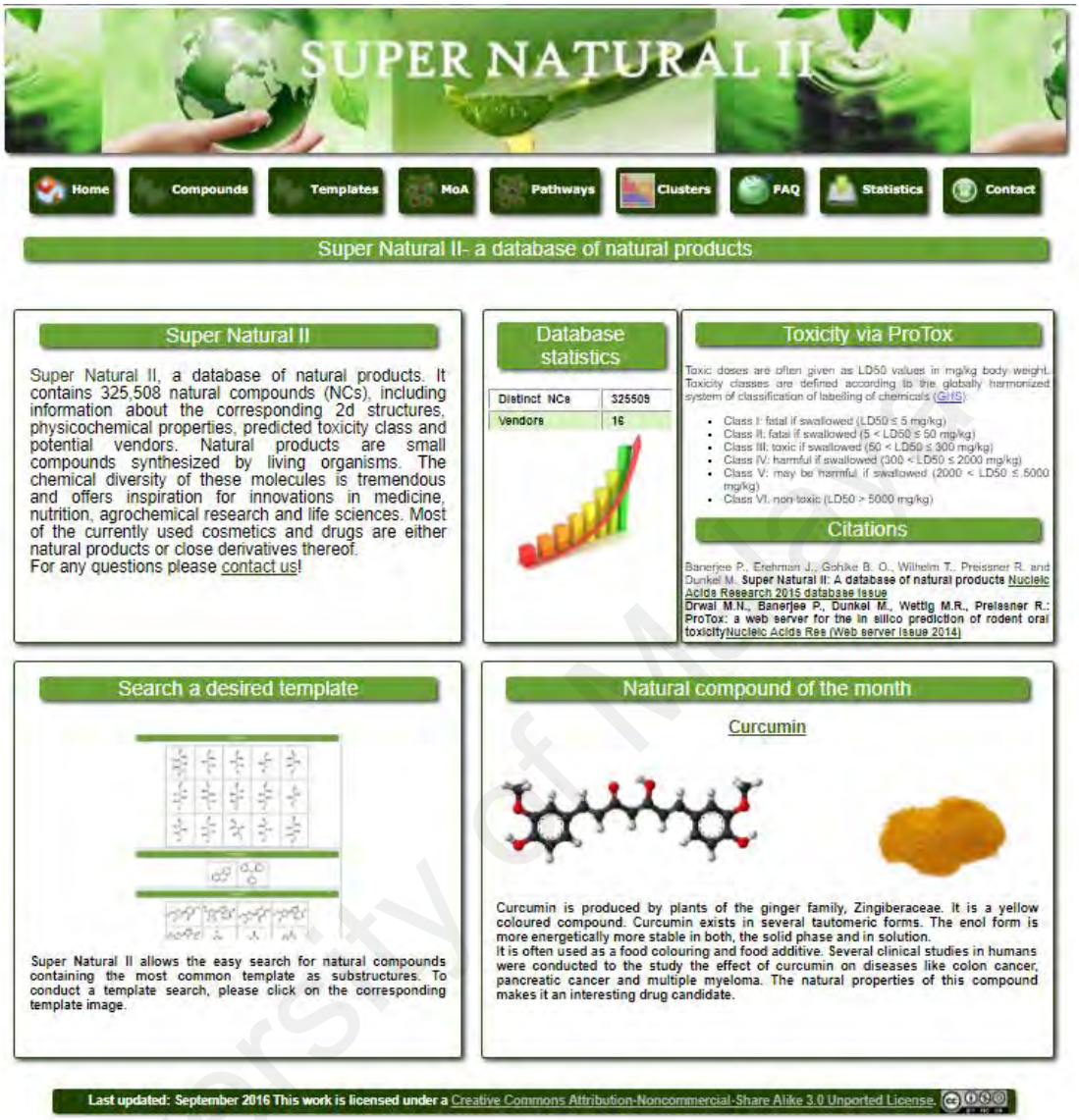


Figure 3.2: The home page of Super Natural II database website.

Table 3.2: Criteria employed in filtering ligands within the Super Natural II database.

Criteria	Selection
Molecular mass	100–349 Da
Lipophilicity (logP)	0.0–2.9
Hydrogen bond acceptors	1–9
Hydrogen bond donors	1–4
Classification	Non-toxic
Availability	Purchasable

3.5.2 Parameters Set Up and Initial Screening

To allow the ligands to move and rotate freely, a grid box with the dimension of 60×60×60 points and 0.375 Å spacing was set up via AutoGrid which covered the active site. Moreover, the grid box coordinate for both of the protease complexes was ensured to be centralised on the catalytic triad on each of the model. The Lamarckian genetic algorithm with its' default parameters was applied and the initial docking simulation was performed having the search run set to 10. The clustering analysis (RMSD < 2 Å) was then conducted upon completion of the docking simulation to identify promising inhibitory ligands.

3.5.3 Redocking and Assessment of Selected Ligands

Based on the initial screening results, ligands that showed favourable binding energy with the control ligand as a reference were chosen for more in-depth analysis. To further confirmed and refined the findings, the docking simulation was conducted again using the selected ligands and the control ligand against all of the models. The redocking simulation was performed with the same parameters used in the initial docking except for the search run was set to 100. Further evaluation of the protein-ligand interaction was later done by observing the interaction existed between the ligand and protein in the first representative model for both of the 'open' and 'closed' proteases using AutoDockTools.

CHAPTER 4: RESULTS

4.1 Assessments of the Dengue Protease Complex Structures

The currently available dengue protease complex (NS2B/NS3) is tabulated in Table 4.1. The X-ray crystallography structure for both of the selected dengue protease complexes obtained from PDB is as shown in Figure 4.1. It was observed that the selected representative for the 'open' protease complex structure (4M9K) does not contain any ligands and appeared to be clean from any artefacts or crystallised water. However, the selected representative for the 'closed' protease complex structure (3U1I) consists of synthetic ligands, crystallised water, sulphate ions (SO_4) and additional chain of protease complex structure which was removed later on.

4.1.1 Morphology of the Selected Protease Complexes

Further assessment showed that the 'open' protease complex structure consists of 247 amino acid residues (48 reported missing) which generated the secondary structure of four helices and 18 beta sheets. On the other hand, the 'closed' protease complex structure made up of 242 amino acid residues (39 reported missing), having one helical and 21 strands of beta sheet conformation in its 3D structure. Despite the differences, both of the structure shared some similarity in the NS3 domain region but showed a significant difference in the NS2B co-factor region visualised in Figure 4.2. The NS2B fragment of the 'closed' protease complex forms a β-hairpin adjacent to the active site hence was given the name 'closed'. Meanwhile in the 'open' protease complex configuration, the NS2B fragment forms a 3_{10}-helix together with a beta strand structure away from the active site.

Table 4.1: X-ray crystallography structure of dengue protease complexes (NS2B/NS3) deposited in PDB (Nitsche et al., 2014).

PDB ID	Resolution (Å)	Name	Serotype	Remarks	Conformation	Reference
4M9K	1.46	NS2B-NS3 protease from dengue virus at pH 5.5	2	• 4M9K and 4M9M structures are similar • 4M9T structure was altered with the presence of the uncaptured electron density probe • 4M9I and 4M9F are mutant protease complexes that share similar structure with the wild protease (4M9K & 4M9M)	Open	Yildiz et al., (2013)
4M9M	1.53	NS2B-NS3 protease from dengue virus at pH 8.5				
4M9T	1.74	NS2B-NS3 protease from dengue virus in the presence of DTNB, a covalent allosteric inhibitor				
4M9I	2.40	A125C NS2B-NS3 protease from dengue virus at pH 5.5				
4M9F	2.70	Dengue virus NS2B-NS3 protease A125C variant at pH 8.5				
3U1J	1.80	Aprotinin bound to Dengue virus protease	3	• NS2B fragment was partially resolved in 3U1J but was fully resolved in 3U1I	Closed	Noble et al., (2012)
3U1I	2.30	Dengue virus protease covalently bound to a peptide				
3L6P	2.20	Crystal Structure of Dengue Virus 1 NS2B/NS3 protease	1	• Both of the structure is similar to one another and consist of mutated NS2B fragments that does not contribute to the active site formation	Open	Chandramouli et al., (2010)
3LKW	2.00	Crystal Structure of Dengue Virus 1 NS2B/NS3 protease active site mutant				
2FOM	1.50	Dengue Virus NS2B/NS3 Protease	2	• Share a highly similar structure to 4M9K and 4M9M	Open	Erbel et al., (2006)

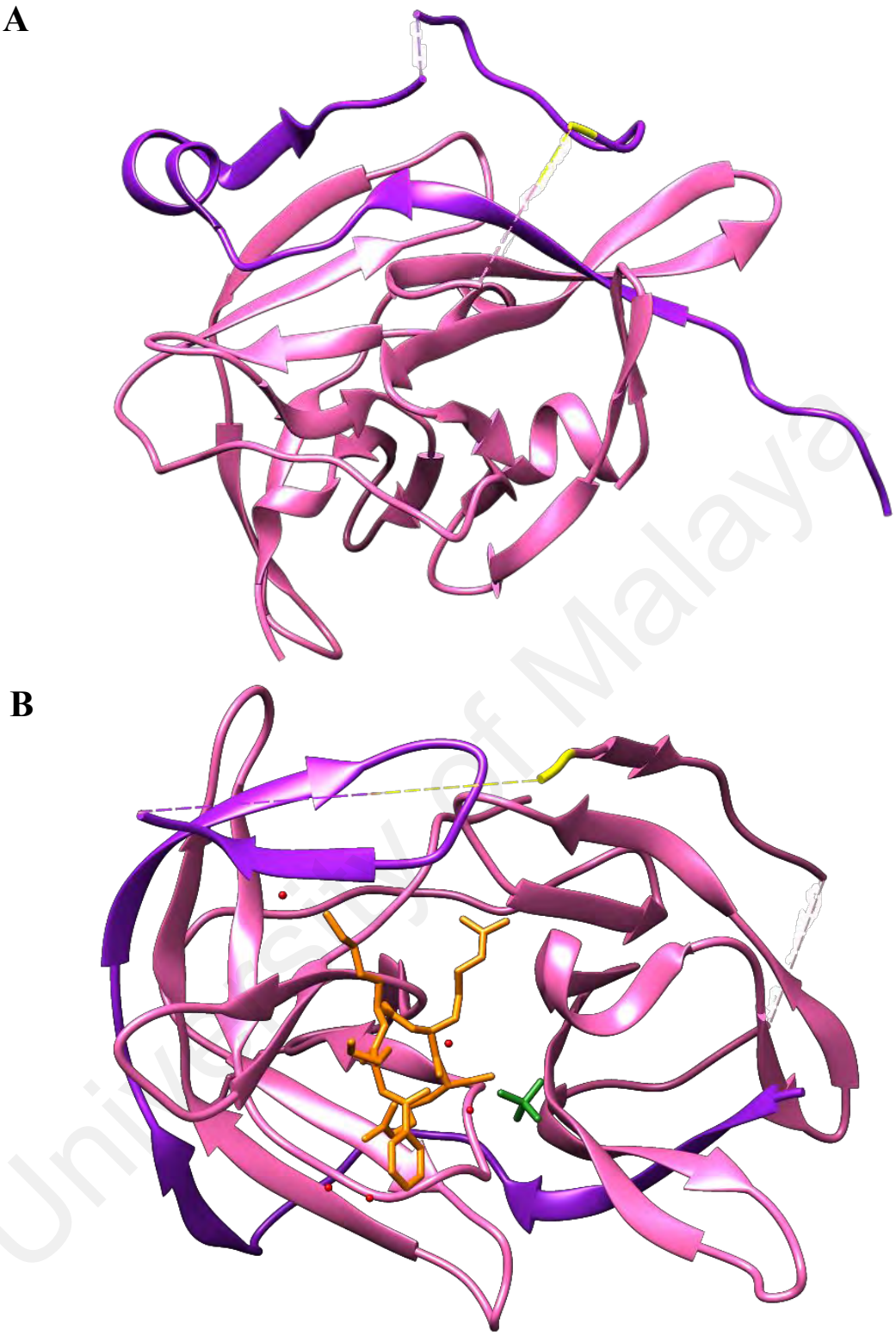


Figure 4.1: The X-ray crystallography of selected structures obtained from the PDB. The 'open' protease complex, PDB ID: 4M9K (**A**) showed a clean model whereas the 'closed' protease complex, PDB ID: 3U1I showed the presence of synthetic ligand (orange), crystallised water (red) and sulphate ion (green) (**B**). The NS2B cofactor and NS3 domain regions on both structures coloured in purple and pink respectively, whereas the non-peptide linker is in yellow. The dashed line represents missing residues. This image was generated via UCSF Chimera.

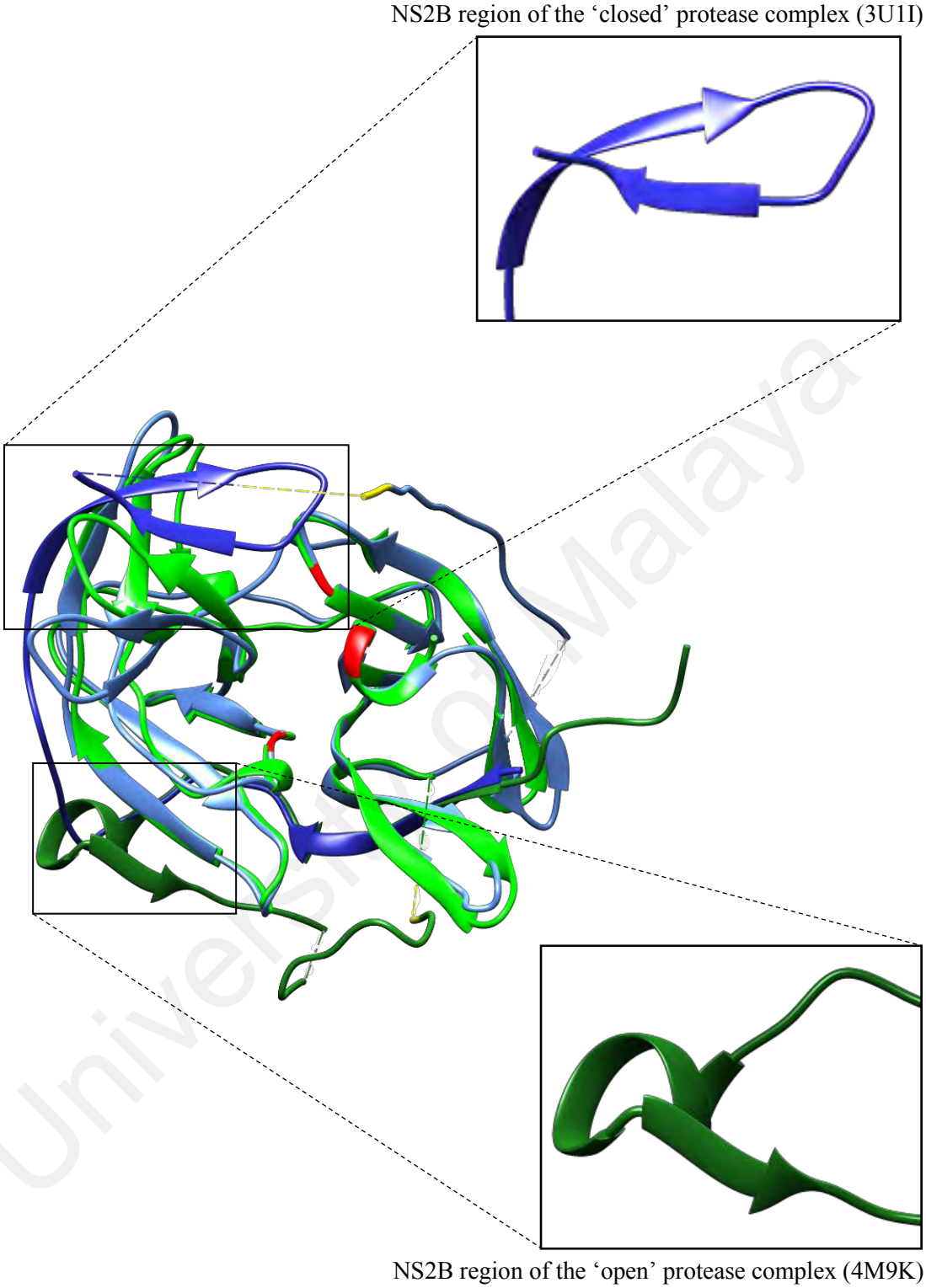


Figure 4.2: The superimposed structure (RMSD ~0.9 Å) of the 'open' protease complex (green) and 'closed' protease complex (blue). The red coloured regions are the catalytic-triad (His-51, Asp-75, and Ser-135) and the linker is in yellow. The dashed line represents missing residues. This image was generated via UCSF Chimera.

4.1.2 MSA of Protease Complex Sequences Between Serotypes

The MSA results revealed a total score (consistency value) of 97 for the NS2B protein sequence alignments (Appendix A). On the other hand, the total score for the entire NS3 protein sequence was 78 (Appendix B) whereas a score of 99 was obtained for the NS3 protease domain sequence alignments (Appendix C). Overall, it appeared that all of the four dengue serotypes shared quite a high degree of protein sequences similarity for the protease. Moreover, based on Figure 4.2, it was clearly visualised that the protease complexes between two different serotypes shared a high degree of structural similarity especially at the NS3 protease domain region.

4.1.3 Loop Modelling and Model Evaluation

Loop modelling was conducted to address the missing residues present on both of the protease complexes. Since there has yet to be any comprehensive study that was done to suggest which of the three protocols (standard, DOPE, and DOPEHR) generate the most reliable model, all of the protocols were utilised. The best model produced from each of the protocols was submitted to PROCHECK to evaluate their quality.

The quality assessment revealed that the standard loop modelling protocol exhibited the most favourable result compared to the other two protocols. The model produced by each of the protease structure via the standard loop modelling protocol is shown in Figure 4.3. Based on the Ramachandran plot, the latter protocols (DOPE and DOPEHR) have less than 90% residues in the most favoured region for both of the generated complexes (Appendix D), hence were considered to be in insufficient quality to be used further. On the other hand, the standard modelled ‘open’ protease complex structure showed that there were 90.9% and 7.6% of residues located in the most favoured and additionally allowed regions respectively. However, 0.5% (Ala-78) and 1% (Lys-45 and Asn-46) of the residues were in the generously allowed and disallowed regions

respectively as shown in Figure 4.4. The two residues that were out of the allowed regions were located in the loop area of the protease complex and away from the active site hence it would not be involved in the interactions with any ligand.

Meanwhile the assessment on the standard modelled 'closed' protease complex structure showed that 90.5% of the residues were in the most favoured regions. There were 8.5% of residues located in the additionally allowed regions and 1.1% residues were in the generously allowed region namely Asp-66 and Lys-75. Fortunately, there were no residues present in the disallowed region within the Ramachandran plot.

Overall, these results indicated that the generated models from the standard modelling procedure acquired a decent quality thus were selected to be further utilised in the subsequent steps. On a side note, the numbering of the catalytic triad (His-51, Asp-75, and Ser-135) was modified in the generated models and was recorded as shown in Table 4.2.

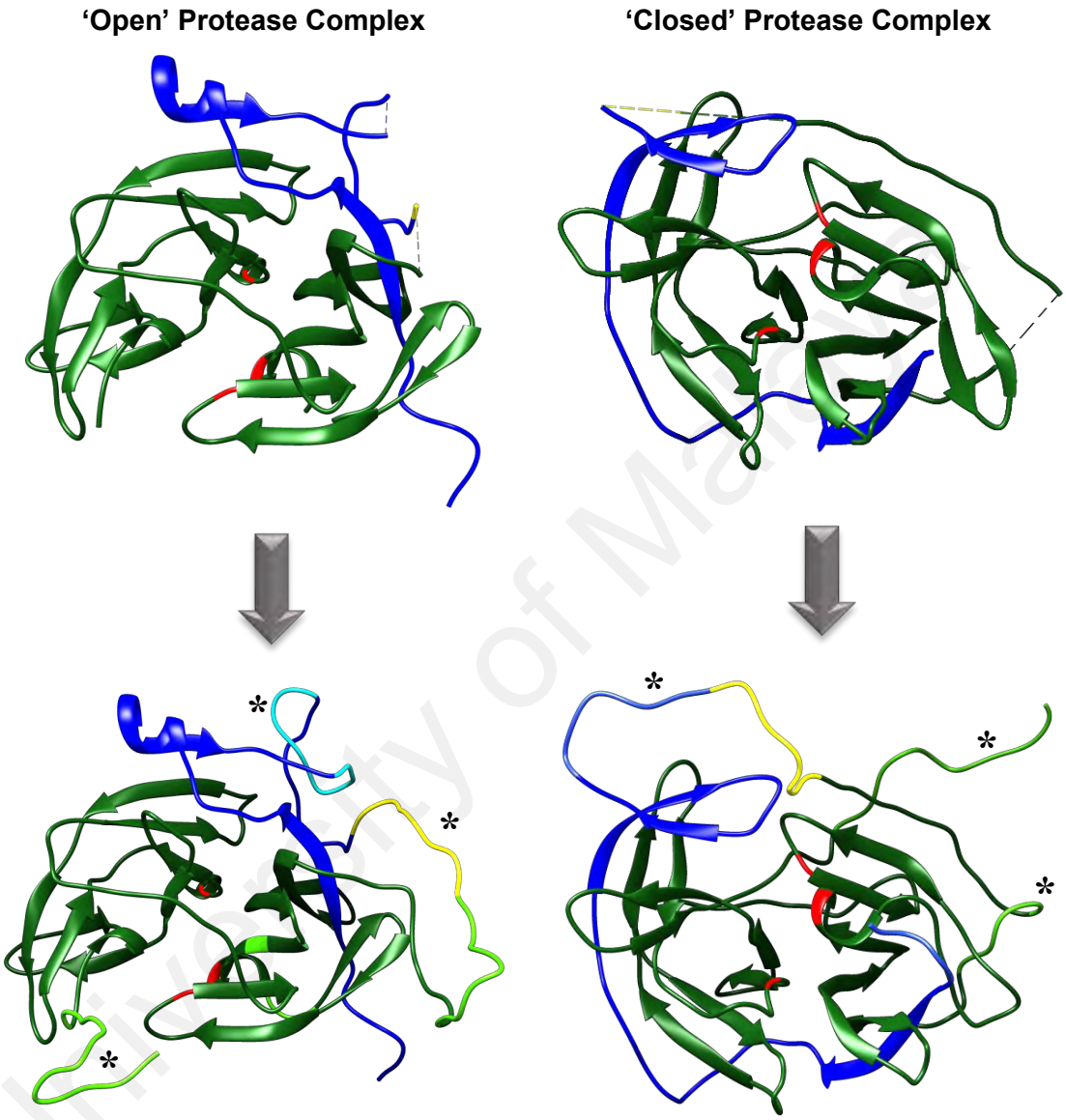


Figure 4.3: Generated protease models based on the standard loop modelling protocol. The top structures were the initial structures obtained from the PDB with missing residues (dashed lines) whereas the bottom structures were the modelled protease structures where the missing residues were resolved. The modelled loop regions in the NS2B cofactor region (dark blue) are coloured in light blue. Meanwhile, the modelled loop regions are coloured in light green in the NS3 domain region (dark green). The non-peptide linker is in yellow and the red regions on the NS3 domain are the catalytic triads. This image was generated via UCSF Chimera. The asterisk symbol(*) indicates modelled region.

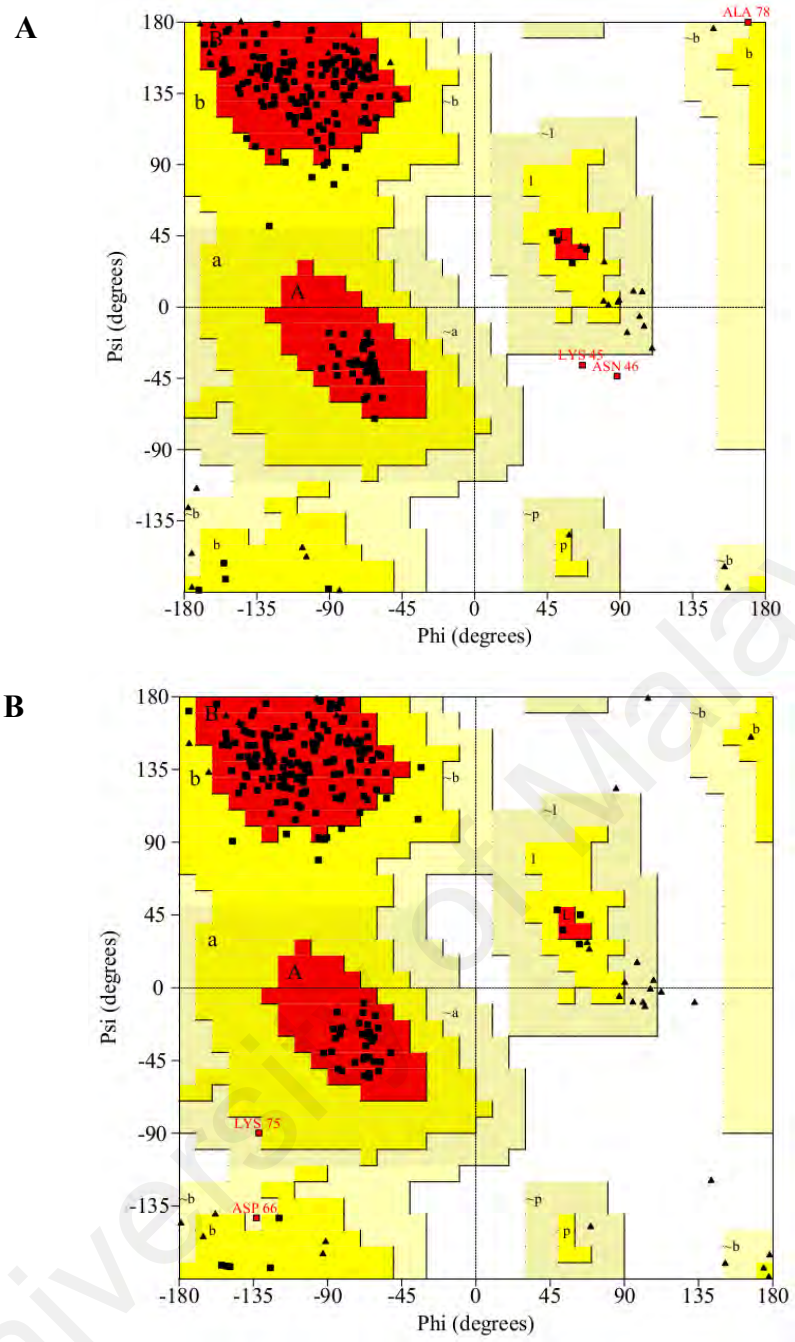


Figure 4.4: Ramachandran plot produced by PROCHECK for the standard loop modelling of the best model generated for the 'open' protease (**A**) and the 'closed' protease complex (**B**).

Table 4.2: The new positions of catalytic triads on the generated models.

Modelled structure	Catalytic Triad		
	His-51	**Asp-75**	**Ser-135**
'Open' protease complex	His-113	Asp-137	Ser-197
'Closed' protease complex	His-111	Asp-135	Ser-195

.

4.2 Stability and Flexibility of Protease Complex throughout MD Simulations

The MD simulations performance of the 'open' and 'closed' protease complexes were analysed and the associated graphs were generated via Grace as shown in Figure 4.5 and Figure 4.6 respectively. Evaluation on the 'open' protease complex RMSD of the carbon-alpha (Cα) atoms (Figure 4.5A) revealed that the initial run (black) showed several continuous stages of a slight RMSD value increment and stabilized after ~4 ns. The second run (red) showed three different stages of convergent, which were from ~1 ns to ~5 ns, ~6 ns to ~8 ns, and ~8 ns to 10 ns. Whereas the third run (green) showed the system convergence was achieved after ~2 ns. Despite the differences in the RMSD values between the replicates, all of the replicates generated a similar total energy pattern from each of the systems (Figure 4.5B) with the average value of –247,679 kcal/mol. Moreover, RMSF analysis also generated similar pattern between the replicates as shown in Figure 4.5C. Apart from the high fluctuation from the terminal regions, there was a large fluctuation shown in the region between the 60^{th} and 80^{th} residue. The catalytic triads however, appeared to be in a stable region of the protease complex.

The RMSD of the Cα atoms for the replicates of 'closed' protease complex also showed different pattern for each (Figure 4.6A). Nevertheless, the initial (black) and second run (red) showed that both of the systems tend to be convergent after ~1 ns. Meanwhile, the third run (green) appeared to reach convergent after ~6 ns of the simulation time. The total energy of all the replicates as shown in Figure 4.6B generated similar pattern and showed that the systems were stable throughout the simulation period bearing the average value of –203,567 kcal/mol. Likewise, similar pattern of RMSF were viewed in all of the replicates (Figure 4.6C) with high fluctuation seen on the region between the 40^{th} and 60^{th} residue whereas the triads were in stable regions.

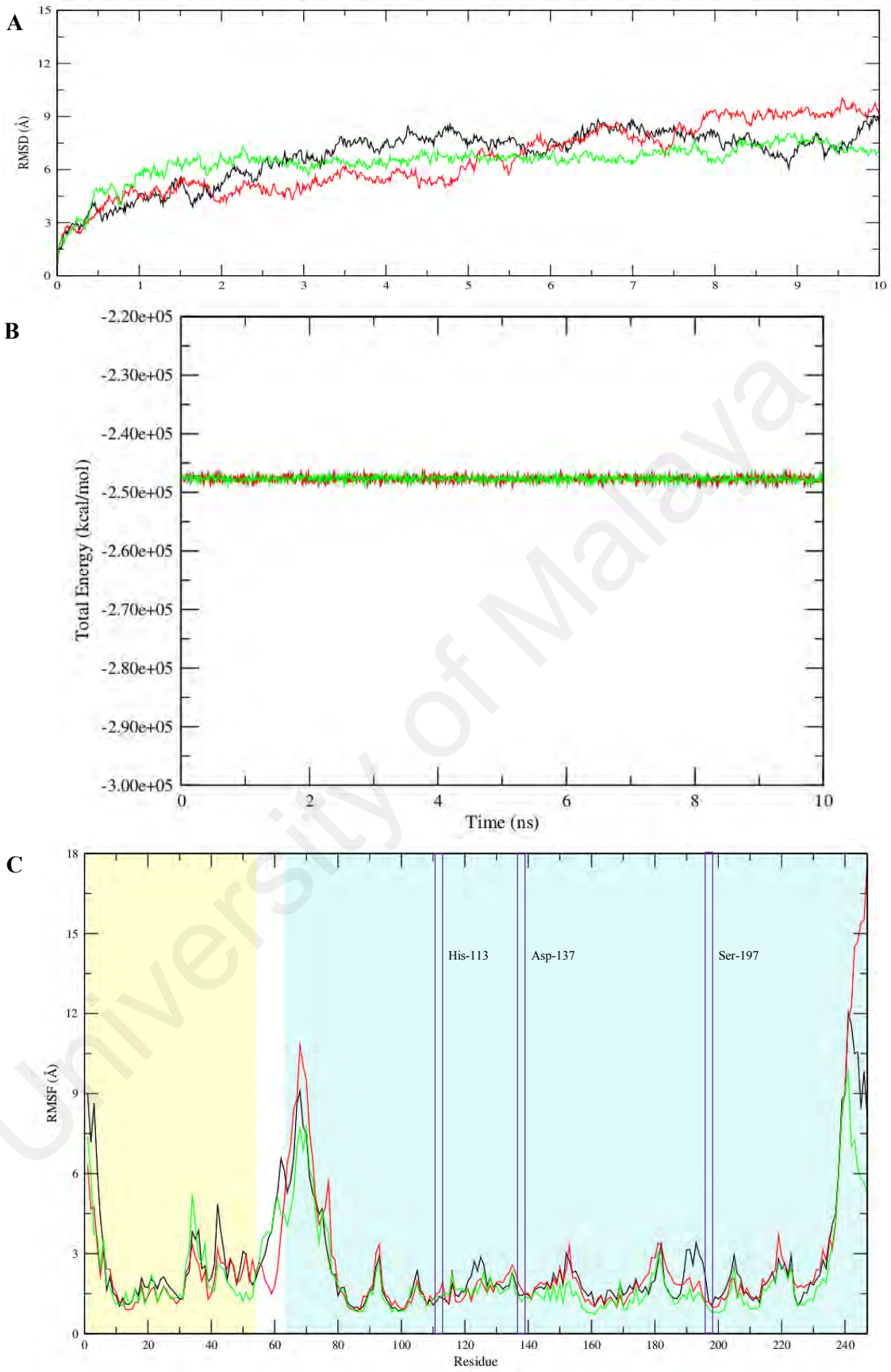


Figure 4.5: Generated graph via Grace for MD simulation analysis of the 'open' protease complex structure. The RMSD of the Cα atoms (**A**), total energy (**B**), and RMSF (**C**) of the protein throughout the simulation for each of the replicates (n = 3), with the first run in black, second in red and third in green lines. The purple bar indicates the catalytic triad positions whereas the light yellow and blue shades designate the NS2B and NS3 region respectively.

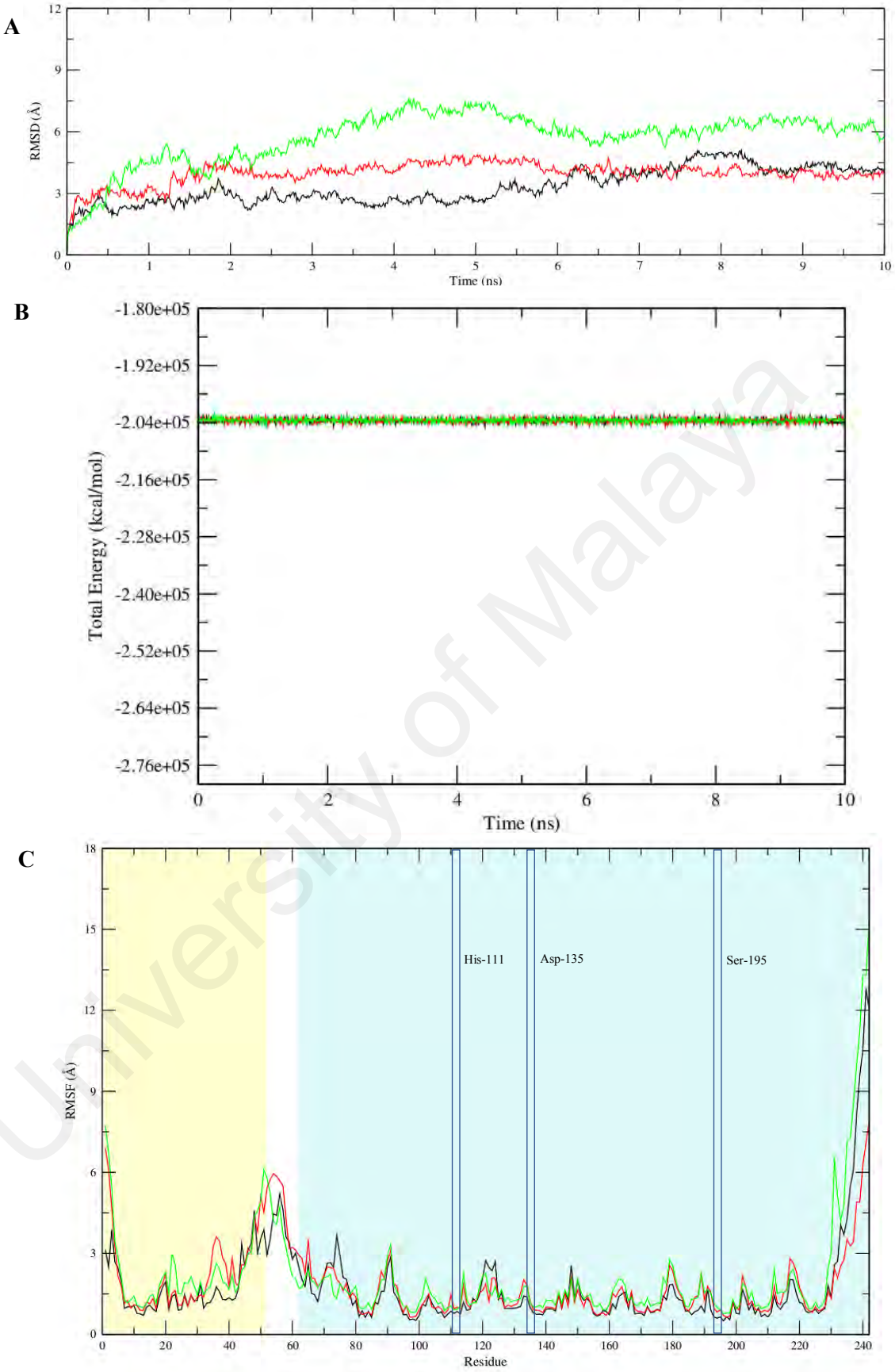


Figure 4.6: Generated graph via Grace for MD simulation analysis of the 'closed' protease complex structure. The RMSD of the Cα atoms (**A**), total energy (**B**), and RMSF (**C**) of the protein throughout the simulation for each of the replicates (n = 3), with the first run in black, second in red and third in green lines. The purple bar indicates the catalytic triad positions whereas the light yellow and blue shades designate the NS2B and NS3 region respectively.

4.3 Clustered Conformations of Protease Complexes

The concatenated trajectories for both of the 'open' and 'closed' protease complexes generated a total of 3001 structures for each. The protease structural conformation for both of the complexes at different time step in the first 10 ns of the concatenated MD simulation is shown in Figure 4.7. Nonetheless, structures that shared similar conformation throughout the simulation were clustered into one group following a specified cutoff value. Based on the tested cutoff values (2.0, 2.5, 3.0, and 3.5 Å), a varied number of clusters was generated in each of the trajectories shown in Table 4.3. However, the most favourable cutoff values for the 'open' and 'closed' trajectories were 3.0 Å and 2.5 Å respectively and were used to obtain the required representative protease structures.

The first 10 cluster groups selected in the 'open' protease complex represents 98% of the total generated structures (Table 4.4) and similarly, the first 11 selected in the 'closed' protease complex represented 98% of the total structures (Table 4.5). Based on the total number of structures within each of the cluster group, an average structure was generated by the GROMACS cluster module as a representative of the cluster group (Figure 4.8). Visualised in Figure 4.9, the notable differences between each of the cluster configuration was within the loop regions of the protein. Moreover, the most significant differences between the groups of clusters (referred as models henceforth) can be seen on the terminals of the structure.

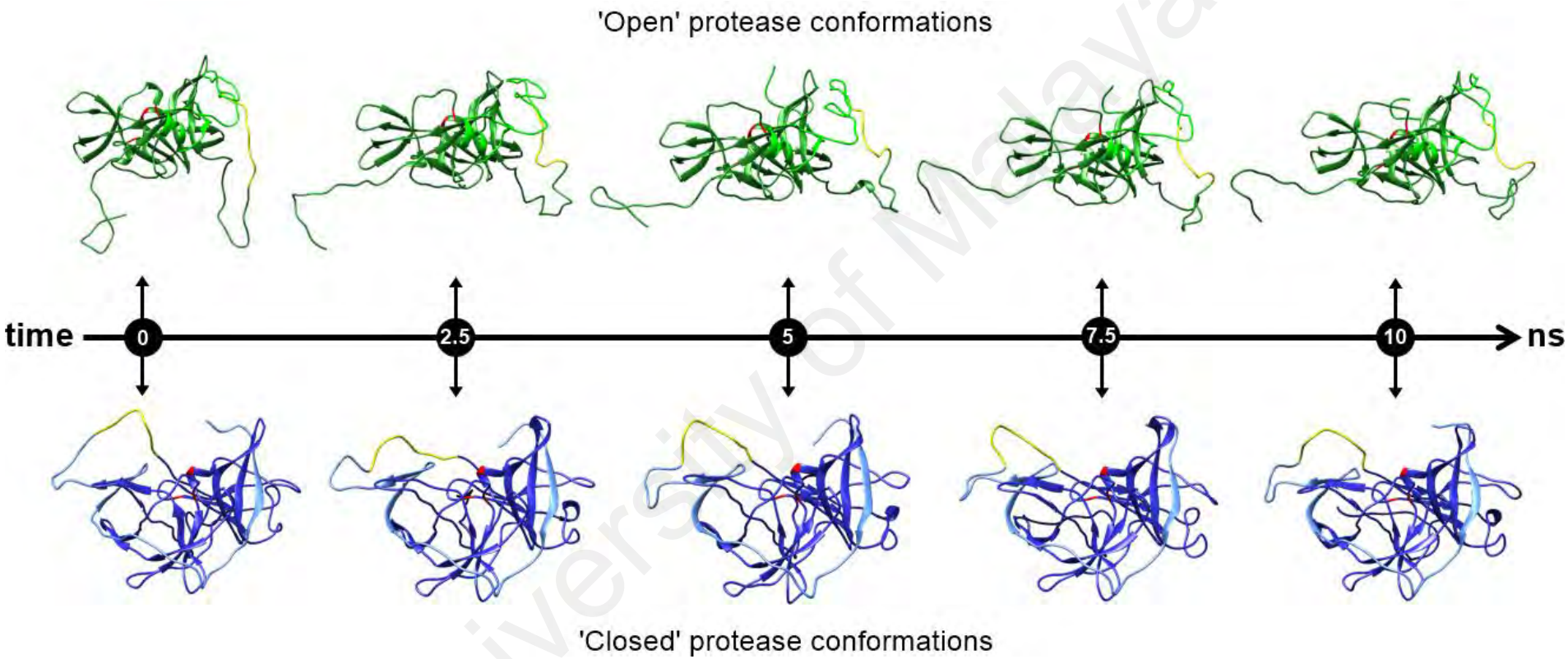


Figure 4.7: The protease complex structure conformational changes at different time stage during the MD simulation. The structural trajectories of the 'open' and 'closed' protease complex were sampled at 0, 2.5, 5, 7.5, and 10 ns from the concatenated MD simulation. The conformational variation can be seen mostly in the loop regions on each of the structure throughout the time step on both of the protease structures ('open' and 'closed'). The NS2B cofactor is shown in light green and light blue on the 'open' and 'closed' protease respectively whereas the NS3 is shown in dark green ('open') and dark blue ('closed'). The non-peptide linker is coloured in yellow meanwhile the catalytic triads are in red colour in both of the protease structures. The structural images were generated via UCSF Chimera.

Table 4.3: Total number of generated clusters in the concatenated trajectories with different cutoff values.

Concatenated trajectory	Cutoff value, RMSD (Å)			
	2.0	**2.5**	**3.0**	**3.5**
'Open' complex	72	31	17	12
'Closed' complex	46	20	11	7

Table 4.4: Generated cluster conformations for the 'open' protease complex (RMSD = 3.0 Å).

Cluster group	Number of structures	Renamed as
1*	783	Model_1
2*	467	Model_2
3*	461	Model_3
4*	364	Model_4
5*	273	Model_5
6*	174	Model_6
7*	124	Model_7
8*	118	Model_8
9*	116	Model_9
10*	49	Model_10
11	35	—
12	18	—
13	12	—
14	3	—
15	2	—
16	1	—
17	1	—
Total	3001	—

*selected group

Table 4.5: Generated cluster conformations for the 'closed' protease complex (RMSD = 2.5 Å).

Cluster group	Number of structures	Renamed as
1*	721	Model_1
2*	675	Model_2
3*	452	Model_3
4*	381	Model_4
5*	232	Model_5
6*	165	Model_6
7*	123	Model_7
8*	61	Model_8
9*	49	Model_9
10*	47	Model_10
11*	24	Model_11
12	20	—
13	14	—
14	9	—
15	8	—
16	7	—
17	5	—
18	3	—
19	3	—
20	2	—
Total	3001	—

*selected group

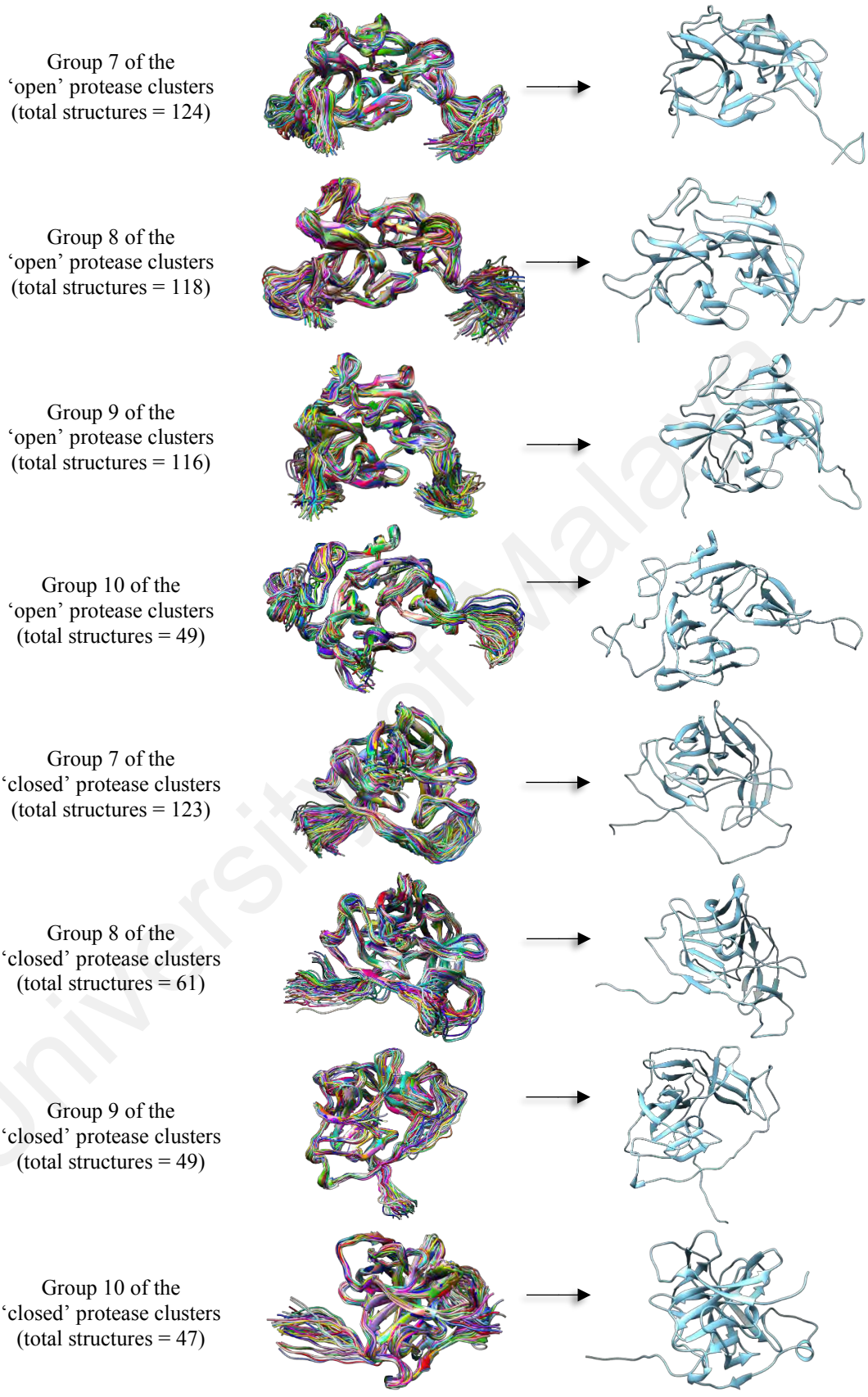


Figure 4.8: The generation of an average structure based on the total structures in each cluster group. The superimposed structures (RMSD ≤ 1.0 Å) on the left consisted of all of the structures within each of the group whereas the structures on the right were the generated structure for each of the group. This image was generated via UCSF Chimera.

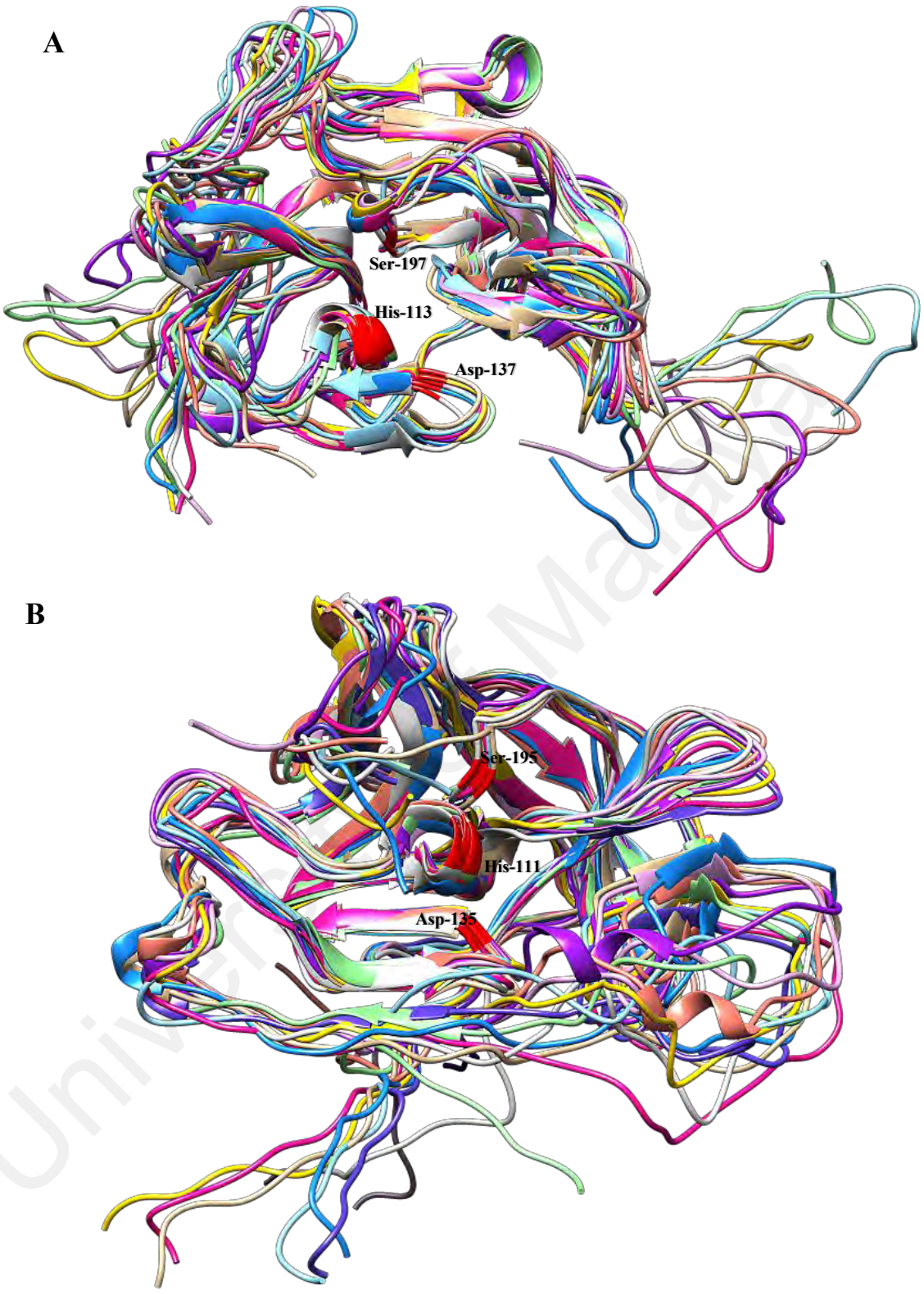


Figure 4.9: Superimposed structures (RMSD ≤ 2.0 Å) of the selected conformations based on RMSD clustering for the 'open' (**A**) and 'closed' protease (**B**) complexes. The 10 averaged models for the 'open' protease originated from a total of 2,929 structures whereas the 11 averaged structures of the 'closed' protease were obtained from a total of 2,930 structures. The major differences between each of the model conformations were seen mostly at the loop and terminal regions of the protein structure. The catalytic triad positions are coloured in red. The image was generated via UCSF Chimera.

4.4 The Collection of Ligands

Based on the search criteria as listed in Table 3.2, a total of 570 hits of natural compounds were obtained (results not shown). The energy minimisation applied on each of the ligands ensured the clashes that may exist among atoms of the ligand were removed apart from allowing it to develop a reasonable starting pose prior to docking. Furthermore, the synthetic peptide inhibitor (Bz-Ne-Lys-Arg-Arg-H) found in the original 'closed' protease complex was also included in the collection to act as a control ligand, totalling the collection of ligands to 571.

4.5 Virtual Screening Evaluation

The initial run comprised of 571 ligands that were docked against 21 protease models (10 'open' and 11 'closed'). In the initial docking simulation analysis, 10 search runs were conducted and generated a preliminary result. The result was then used to further filter the collection of the ligands. Out of the 570 collected ligands, the top 10 ligands that showed satisfactory binding energy values (Table 4.6) with reference to the control ligand were selected for redocking purposes.

4.5.1 Redocking Simulation Analysis

The basic information of the selected ligands is shown in Table 4.7. These ligands together with the control ligand were then subjected to another docking simulation with all of the 21 protease models. Interestingly, the redocking simulations which were performed with 100 search runs revealed eight of the ligands displayed a better binding energy (more negative) on average than the control ligand in the 'open' models (Table 4.8). Meanwhile the redocking of ligands in the 'closed' models showed that all of the selected ligands displayed a better binding energy on average compared to the control ligand (Table 4.9).

Table 4.6: The estimated lowest binding energy (kcal/mol) for the control and several ligands that yield favourable value in the initial search run.

Protease							**Ligands**						
'Open' Protease	Control	**Lig556**	**Lig557**	**Lig558**	Lig560	Lig561	**Lig562**	**Lig563**	**Lig564**	**Lig565**	**Lig566**	**Lig567**	**Lig569**
Model_1	–5.58	–5.80	–6.56	–6.27	–4.96	–4.39	–6.80	–6.88	–4.30	–5.20	–6.24	–7.61	–5.51
Model_2	–6.10	–6.34	–6.64	–6.16	–6.13	–5.34	–6.10	–6.23	–4.82	–5.50	–6.01	–6.96	–6.07
Model_3	–4.12	–5.79	–6.42	–6.29	–4.56	–4.01	–4.93	–5.39	–3.81	–4.39	–6.35	–6.38	–4.52
Model_4	–7.85	–5.92	–5.97	–6.05	–5.09	–4.25	–6.09	–6.02	–4.95	–5.37	–6.17	–6.42	–5.55
Model_5	–4.69	–5.60	–5.91	–5.78	–5.49	–5.03	–5.85	–6.12	–6.71	–6.42	–5.65	–6.33	–5.03
Model_6	–5.36	–7.78	–8.19	–8.33	–6.16	–5.37	–7.69	–7.9	–5.94	–6.45	–7.62	–8.44	–6.41
Model_7	–7.19	–5.88	–5.89	–6.22	–4.93	–3.96	–5.99	–5.86	–3.73	–4.93	–5.70	–6.89	–6.13
Model_8	–5.29	–5.07	–6.32	–5.76	–4.57	–4.06	–5.68	–5.46	–5.03	–4.20	–6.07	–7.22	–5.57
Model_9	–5.25	–5.89	–6.02	–5.50	–4.43	–3.79	–5.53	–5.39	–4.19	–5.21	–5.83	–6.50	–5.21
Model_10	–6.41	–5.78	–6.02	–5.42	–6.14	–5.53	–6.19	–6.11	–5.14	–5.07	–5.98	–7.08	–5.61
'Closed' Protease	Control	**Lig556**	**Lig557**	**Lig558**	Lig560	Lig561	**Lig562**	**Lig563**	**Lig564**	**Lig565**	**Lig566**	**Lig567**	**Lig569**
Model_1	–6.12	–6.54	–6.38	–7.01	–5.42	–5.04	–6.05	–5.97	–5.27	–6.08	–6.31	–7.50	–6.11
Model_2	–5.51	–6.56	–7.46	–7.34	–6.04	–4.93	–7.87	–7.71	–6.70	–6.87	–7.32	–7.94	–7.11
Model_3	–6.66	–6.91	–7.21	–7.06	–5.73	–5.42	–7.21	–7.44	–5.94	–6.39	–7.04	–7.41	–5.40
Model_4	–8.46	–7.21	–7.34	–7.82	–5.78	–5.12	–7.75	–7.39	–6.18	–6.24	–7.35	–7.98	–6.75
Model_5	–7.76	–7.30	–6.79	–7.18	–5.57	–4.84	–6.56	–6.86	–6.02	–6.32	–7.29	–7.87	–5.86
Model_6	–6.37	–6.54	–6.78	–6.47	–4.91	–4.53	–6.37	–6.24	–5.91	–6.12	–7.17	–7.46	–5.19
Model_7	–7.09	–6.40	–7.89	–6.75	–5.29	–4.83	–7.22	–7.09	–6.07	–7.19	–7.40	–8.27	–6.62
Model_8	–5.82	–7.42	–6.95	–6.27	–5.40	–5.17	–6.07	–5.85	–3.99	–4.73	–6.58	–7.66	–5.35
Model_9	–6.20	–7.59	–7.83	–6.77	–5.43	–4.89	–7.20	–7.19	–5.22	–6.29	–7.87	–7.94	–5.25
Model_10	–7.00	–6.57	–7.37	–7.67	–6.24	–5.10	–7.08	–7.14	–5.29	–5.95	–7.38	–8.68	–6.07
Model_11	–7.17	–6.87	–7.64	–6.72	–5.68	–5.23	–6.91	–7.12	–5.80	–6.46	–7.24	–8.25	–6.53

bold ligands were among the selected (top 10) ligands

Table 4.7: The selected ligands chemical formula and their corresponding structure.

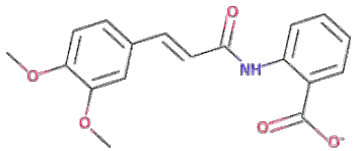
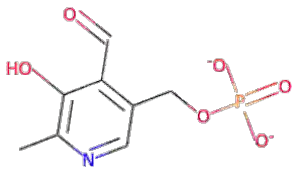
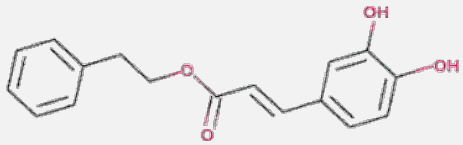
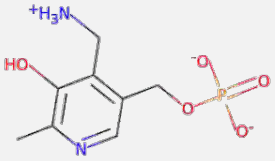
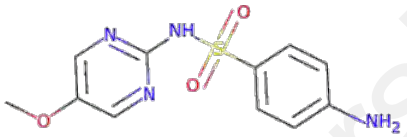
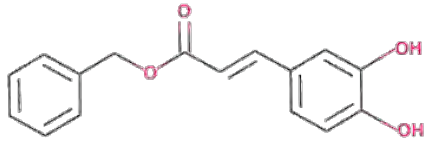
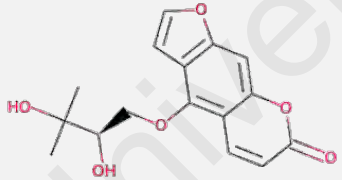
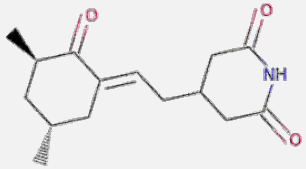
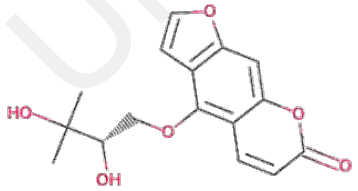
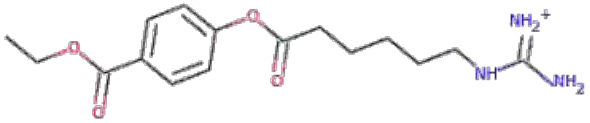
Ligands	Formula	Structure	Ligands	Formula	Structure
Lig556	$C_{18}H_{16}NO_5$		Lig564	$C_8H_8NO_6P$	
Lig557	$C_{17}H_{16}O_4$		Lig565	$C_8H_{12}N_2O_5P$	
Lig558	$C_{11}H_{12}N_4O_3S$		Lig566	$C_{16}H_{14}O_4$	
Lig562	$C_{16}H_{16}O_6$		Lig567	$C_{15}H_{21}NO_3$	
Lig563	$C_{16}H_{16}O_6$		Lig569	$C_{16}H_{24}N_3O_4$	

Table 4.8: The estimated lowest binding energy (kcal/mol) for the control and selected ligands in the 'open' protease models.

Model	Control	Lig556	Lig557	Lig558	Lig562	Lig563	Lig564	Lig565	Lig566	Lig567	Lig569
Model_1	–5.64	–5.66	–6.59	–6.95	–6.78	–6.71	–4.45	–5.18	–6.53	–7.41	–5.95
Model_2	–5.44	–6.59	–6.60	–7.23	–6.61	–6.73	–4.97	–5.97	–6.70	–6.96	–5.90
Model_3	–4.17	–5.67	–6.46	–6.42	–5.19	–5.24	–4.34	–4.49	–6.33	–6.23	–5.41
Model_4	–6.10	–5.79	–6.07	–6.48	–5.69	–5.79	–4.99	–5.36	–6.24	–7.39	–5.29
Model_5	–5.24	–5.29	–5.76	–5.73	–5.75	–5.96	–6.41	–4.97	–6.48	–6.35	–4.77
Model_6	–5.83	–7.83	–8.15	–8.32	–7.39	–7.73	–5.95	–6.45	–7.88	–8.40	–6.25
Model_7	–4.93	–5.69	–5.80	–6.38	–5.95	–5.83	–4.11	–4.93	–5.54	–6.18	–5.92
Model_8	–6.38	–5.48	–5.97	–5.58	–5.63	–5.53	–4.31	–4.88	–5.87	–6.95	–6.05
Model_9	–4.69	–5.85	–5.77	–5.52	–5.51	–5.32	–4.25	–5.21	–5.98	–6.20	–5.47
Model_10	–5.08	–6.98	–6.58	–5.69	–6.06	–6.13	–5.31	–5.58	–7.03	–7.00	–5.82
Average	*–5.35*	*–6.08*	*–6.38*	*–6.43*	*–6.06*	*–6.10*	*–4.91*	*–5.30*	*–6.46*	*–6.91*	*–5.68*

Table 4.9: The estimated lowest binding energy (kcal/mol) for the control selected ligands in the 'closed' protease models.

Model	Control	Lig556	Lig557	Lig558	Lig562	Lig563	Lig564	Lig565	Lig566	Lig567	Lig569
Model_1	–6.02	–7.52	–7.23	–7.04	–7.12	–6.87	–6.24	–6.12	–7.09	–7.57	–6.08
Model_2	–6.31	–6.60	–7.03	–7.36	–7.88	–8.09	–6.69	–6.90	–7.01	–8.10	–6.52
Model_3	–6.26	–6.69	–7.13	–7.09	–7.26	–7.16	–5.64	–5.55	–7.18	–7.54	–5.93
Model_4	–7.23	–7.01	–7.26	–7.82	–7.45	–7.43	–6.29	–6.67	–7.04	–7.83	–6.59
Model_5	–6.10	–7.05	–7.22	–7.23	–6.69	–6.93	–6.05	–6.41	–7.32	–7.61	–5.20
Model_6	–5.13	–6.41	–6.89	–6.40	–6.25	–6.41	–5.88	–6.45	–7.20	–7.29	–4.94
Model_7	–6.76	–6.27	–7.26	–7.73	–6.92	–6.90	–6.44	–7.25	–7.17	–7.88	–7.20
Model_8	–5.32	–7.72	–6.78	–6.38	–6.46	–6.39	–6.82	–5.28	–6.80	–7.32	–5.93
Model_9	–5.89	–7.90	–7.67	–7.13	–7.12	–7.18	–5.81	–6.31	–7.91	–7.36	–6.14
Model_10	–5.90	–7.21	–7.37	–7.60	–7.02	–6.99	–5.79	–6.21	–7.38	–8.52	–6.00
Model_11	–5.10	–6.72	–7.07	–7.17	–6.92	–6.79	–6.20	–6.68	–6.96	–8.17	–6.68
Average	*–6.00*	*–7.01*	*–7.17*	*–7.18*	*–7.01*	*–7.01*	*–6.17*	*–6.35*	*–7.19*	*–7.74*	*–6.11*

4.5.2 Protease-Ligand Interactions Analysis

Since the first model generated from each of the protease complex represents the majority of the generated structural conformations (Table 4.4 and 4.5), it was utilised for further docking evaluation. Analysis in the protein-ligand interactions summarised in Table 4.10 revealed that all 10 of the selected ligands formed hydrogen bond with at least one of the catalytic triad in the first model (Model_1) of the 'open' protease. On the other hand, only eight of the selected ligands showed similar properties in the 'closed' protease as shown in Table 4.11. The results also showed that in the 'open' model, the ligands tend to form hydrogen bonds with the His-113 and Ser-197 residues of the catalytic triads. Whereas the ligands docked in 'closed' model preferred only the Ser-195 residue and formed multiple hydrogen bonds with the other nearby residues. Overall, Lig567 has shown to yield the highest affinity (most negative binding energy value) towards both of the protease complex structures. The visual representation of the protein-ligand interaction between Lig567 with the 'open' and 'closed' protease complex is shown in Figure 4.10 and Figure 4.11 respectively.

Table 4.10: Protein-ligand interaction of the first 'open' protease model (Model_1) with the control and selected ligands.

Ligand	Total number in cluster (out of 100)	Mean binding energy (kcal/mol)	Hydrogen bond interactions		Distance (Å)
			Protein atom	Ligand atom	
Control	2	–4.28	**His-113**: NE2	O	2.105
			Ser-197: HG1	O	1.771
Lig556	64	–5.32	**His-113**: NE2	O	1.940
			Ser-197: HG1	O	1.636
			Gly-215: HN	O	1.943
Lig557	63	–6.27	**His-113**: NE2	O	2.067
			Arg-116: HH21	O	2.039
Lig558	68	–6.12	**His-113**: NE2	O	2.058
			Ser-197: HG1	O	2.167
Lig562	97	–6.56	**His-113**: NE2	O	2.129
			Ser-197: HG1	O	1.968
Lig563	100	–6.44	**His-113**: NE2	O	2.260
			Ser-197: HG1	O, O	2.034
			Gly-215: HN	O	2.198
Lig564	40	–4.22	**His-113**: NE2	O	2.113
			Ser-197: HG1	O	1.683
			Tyr-212: HH	O	1.653
Lig565	31	–5.08	**His-113**: NE2	O	1.995
			Tyr-212: HH	N	1.986
Lig566	54	–6.00	Arg-116: NE	O	1.923
			Ser-197: HG1	O	2.001
Lig567	90	–7.26	**His-113**: NE2	O	2.091
			Gly-215: HN	O	1.978
Lig569	13	–4.92	**His-113**: NE2	O	2.018
			Ser-197: HG1	O	1.749
			Tyr-212: HH	O	1.857

The bold residues are the catalytic triads.

Table 4.11: Protein-ligand interaction of the first 'closed' protease model (Model_1) with the control and selected ligands.

Ligand	Total number in cluster (out of 100)	Mean binding energy (kcal/mol)	Hydrogen bond interactions		Distance (Å)
			Protein atom	Ligand atom	
Control	1	–5.02	**His-111**: NE2	O	2.149
			Asn-212: HD21	N	1.931
			Asn-212: HN	O	2.246
Lig556	53	–6.47	Arg-114: NE	O	1.986
			Gly-193: HN	O	1.849
			Thr-194: HN	O	1.837
			Ser-195: HN	O	2.045
Lig557	8	–6.66	Arg-114: NE	O	2.052
			Arg-114: HH21	O	1.761
			Gly-193: HN	O	1.891
			Thr-194: HN	O	1.984
			Ser-195: HN	O	1.987
Lig558	12	–6.70	Thr-194: HN	O	2.173
			Ser-195: HN	O	1.710
Lig562	3	–6.79	Arg-114: HH21	O	1.824
			Gly-193: HN	O	1.997
			Thr-194: HN	O	1.965
			Ser-195: HN	O	2.196
Lig563*	63	–5.78	Met-40: HN	O	2.063
			Asn-212: HD21	O, O	2.154
Lig564	21	–5.56	Gly-193: HN	O, O	1.997
			Thr-194: HN	O	2.061
			Ser-195: HN	O	2.021
			Tyr-221: HH	O	2.095

*ligand that did not form hydrogen bond with the catalytic triad. The bold residues are the catalytic triads.

Table 4.11 (continued)

Ligand	Total number in cluster (out of 100)	Mean binding energy (kcal/mol)	Hydrogen bond interactions		Distance (Å)
			Protein atom	Ligand atom	
Lig565	75	–5.75	Arg-114: NE	O	2.048
			Arg-114: HH21	O	1.761
			Gly-193: HN	O	1.967
			Ser-195: HN	O	2.161
Lig566	24	–7.09	Arg-114: NE	O	1.952
			Arg-114: HH21	O	2.081
			Gly-193: HN	O, O	2.009
			Ser-195: HN	O	1.774
Lig567	33	–7.54	Arg-114: HH21	O	2.127
			Gly-193: HN	O	2.033
			Thr-194: HN	O	1.638
			Ser-195: HN	O	1.835
Lig569*	6	–5.32	Met-40: HN	N	1.982
			Asn-212: HN	O	1.906

*ligand that did not form hydrogen bond with the catalytic triad. The bold residues are the catalytic triads.

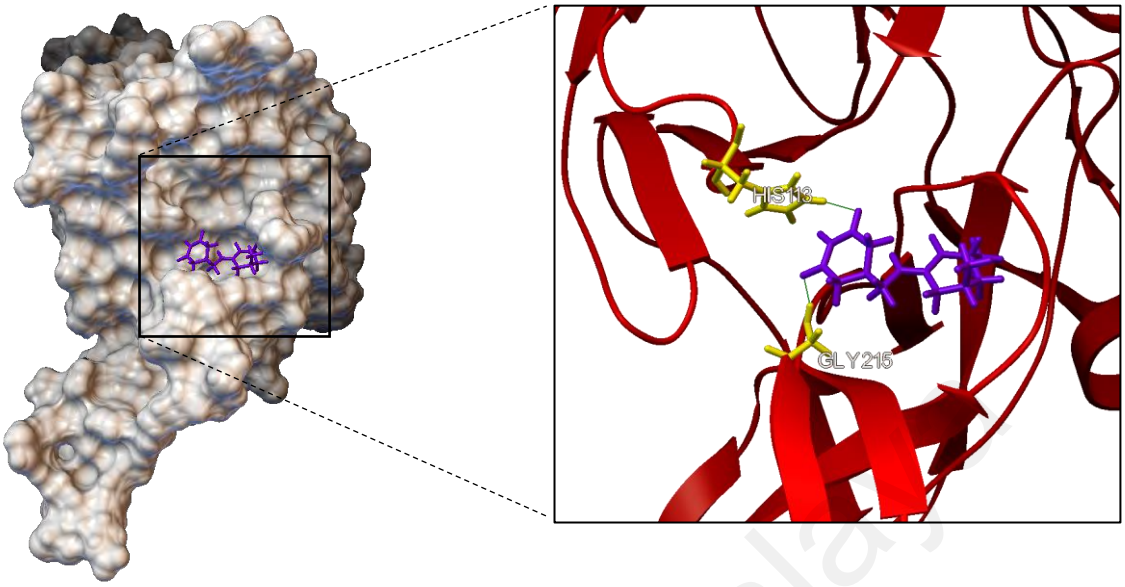


Figure 4.10: The predicted orientation of Lig567 based on its lowest binding energy to the active site of the 'open' protease complex (Model_1). The zoomed-in image shows the hydrogen bonds interaction (green lines) between the protease residues (yellow) with Lig567 (purple). This image was generated via AutoDockTools.

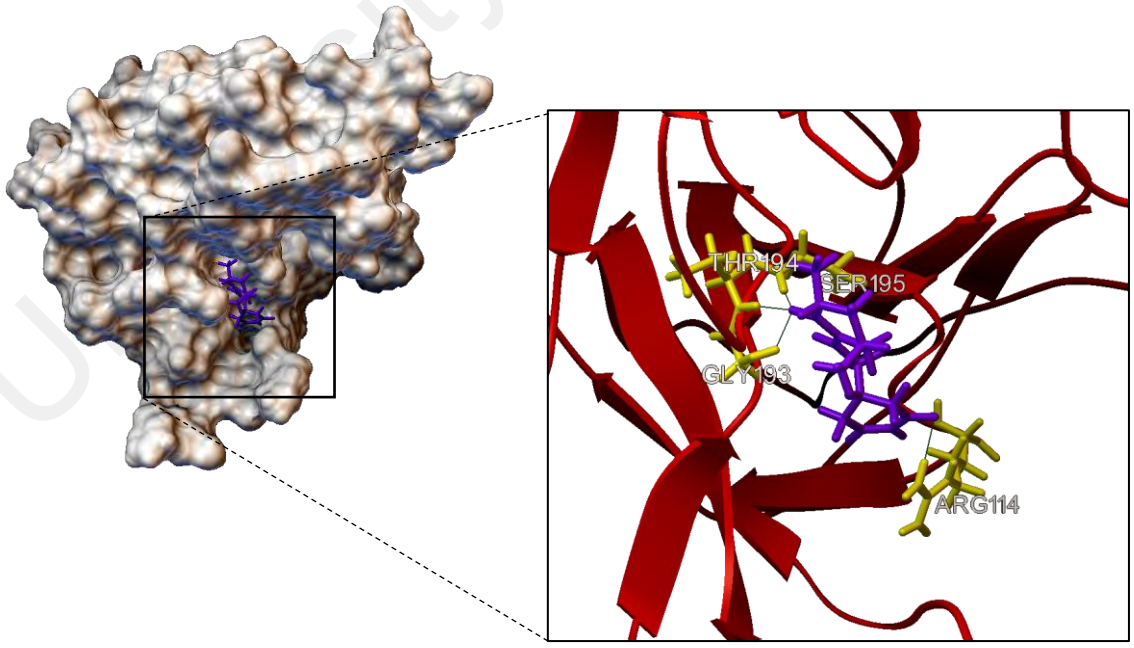


Figure 4.11: The predicted orientation of Lig567 based on its lowest binding energy to the active site of the 'closed' protease complex (Model_1). The zoomed-in image shows the hydrogen bonds interaction (green lines) between the protease residues (yellow) with Lig567 (purple). This image was generated via AutoDockTools.

CHAPTER 5: DISCUSSION

5.1 The Dengue Protease Complex Structures

Aforementioned, there are currently two distinct structural conformations of the dengue protease complex. The 'open' structural configuration was the first recorded and deposited crystal structure of the protease complex in the absence of ligand (Erbel et al., 2006). Meanwhile, the 'closed' structure was initially identified when a ligand was bounded on the protease complex producing a so called 'active' conformation (Noble et al., 2012). It is known that protein structural conformation can change drastically in the presence of ligand (Lionta et al., 2014), a phenomenon referred as the induced-fit effect. However, another study has recorded that even in the absence of ligands, the 'closed' configuration was still observed (Chen et al., 2014). In addition, a study expressing an unlinked protease complex revealed that majority of the expressed protease also existed in the 'closed' configuration whereas the rest were in the 'open' configuration (Kim et al., 2013). Thus, both of the conformations appeared to be the protease complex's natural configuration. On a side note, the non-peptide linker (Gly_4-Ser-Gly_4), that linked NS2B cofactor and NS3 protease domain (Figure 4.1) was claimed as essential for optimal catalytic activity in an early study (Leung et al., 2001). However, a recent study revealed that unlinked proteases complex was still enzymatically active and shared similar secondary structure with the linked complex (Li et al., 2014).

Previously, a majority of the *in silico* studies utilised the 'open' structure as it was the first and only recorded structure of the protease complex. However, current research begins to focus on the 'closed' structure as it is claimed to have a significant contribution towards the protease stability (Chen et al., 2014) and the ligand affinity towards the active site (Nitsche et al., 2014) due to the NS2B structural arrangement as

shown in Figure 4.2. Nevertheless, neglecting the 'open' structure may not be wise since it was one of the recorded protease complexes in its' natural configuration. Therefore, by utilising the 'open' and 'closed' protease structure, it will ensure maximal coverage of all of the possible dengue protease structural conformations; thus, would ultimately provide more robust and significant results. However, only those with good resolution (≤ 2.5 Å) were taken into consideration to be used as the protein of interest or target protein. The 4M9K was chosen since it was the latest 'open' structure deposited in PDB with the best resolution whereas the 3U1I was the only 'closed' structure with a fully resolved NS2B cofactor region.

Even though there have been a lot of intensive research that was done related to dengue virus, the availability of 3D structure for protease complex for all of the different serotypes remains limited. The currently available dengue protease complex structures are tabulated in Table 4.1 which showed that there was no representative for the fourth serotype and the representative for the first serotype contain a mutated structure. Therefore, this limits any comprehensive studies to be conducted on the protease complex structure of the four different dengue virus serotypes. The insufficient availability of the required 3D structures often present as a challenge especially in the structure-based drug design and discovery (Lounnas et al., 2013) whereas the reliability in homology modelling still remains subjective (Rodrigues et al., 2013).

Nevertheless, the availability of the dengue protease protein sequences of all the four serotypes in a secondary sequence database (RefSeq) allowed MSA to be conducted. A high total score from the MSA, with a maximum of 100, indicates that the subjected sequences shared high sequence similarity with one another (Di Tommaso et al., 2011). Initial MSA result of NS3 has a rather low total score (Appendix B) mainly because the NS3 sequences obtained from RefSeq also include the helicase domain, several other

regions, and sites, making the NS3 total sequence length to 619 residues. Meanwhile, the NS3 protein sequences of the selected protease complex ('open' and 'closed') obtained from PDB does not have them and only consisted of approximately 180 amino acid residues. Therefore, to focus on the protease domain, a second MSA which only include the protease domain sequences was conducted and revealed a high total score (Appendix C). The combined MSA results from NS2B and NS3 indicated that the four dengue serotypes would likely produce a protease complex configuration that resembles one another. This was further supported by several other studies which emphasised on the conservation of protease structure between the different DENV serotypes (Ayub et al., 2013; Oliveira et al., 2014). Hence, using a protease complex that originates from different serotypes may not be a major issue in search of promising dengue protease inhibitor.

The inability to capture certain region on the crystallised structure often comes as a limitation to fully optimise the deposited structure for further analysis. It was reported that approximately 69% of the deposited 3D structure in PDB contains missing residues (Djinovic-Carugo, & Carugo, 2015). Within the macromolecule structure particularly in protein, there are certain regions that are highly flexible, making it extremely difficult for crystallographers to clearly define that particular region opting them to assign it as missing instead. Therefore, to address the missing residues in the obtained structures, loop modelling via MODELLER was utilised. MODELLER generates several comparative models by using the full sequence provided within the PDB file as the template. The DOPE-based protocols for loop modelling were the derived version from the standard protocol which aimed to generate better quality loops. However, it was shown that in the case of modelling the dengue protease complex, the standard protocol fare better than the DOPE-based. Thus, with the favourable stereochemical quality of

the models generated from the standard protocol, the models were selected to be further applied in the subsequent procedure.

5.2 MD Simulations and Ensemble Protease Complexes

MD simulations offer comprehensive insights on the molecular behaviour and interactions at the atomic level in its natural environment (Durrant & McCammon, 2011). Moreover, it generates a range of possible structural conformations throughout the stipulated duration. Therefore, as the duration of simulation is prolonged, the number of generated structures also increased. This study was performed in a 10 ns simulation to obtain conformational ensemble structures of the protease complex. The duration may appear short but should fare better than the original stiff crystallised protein structure as pointed out by Amaro and Li (2010).

Replicates of the MD simulation run was expected to produce a slightly different outcome on each run as reported in other published studies (Mutt & Sowdhamini, 2016; Karami et al., 2016). The calculated changes in the RMSD of the Cα atoms in the protease complexes represent the flexibility of the protein throughout the MD simulations (Dagget & Levitt, 1992). Based on the results generated visualised in Figure 4.5A and Figure 4.6A, there was no tremendous fluctuation of RMSD readings during the MD simulations. This indicated that both of the 'open' and 'closed' protease complexes were stable throughout the simulations. The point of convergent or sometimes referred to equilibration mentioned in Section 4.2 may have been interpreted differently by others as Knapp and his colleagues (2011) have proven. Nevertheless, it is still remained as one of the common analysis in identifying structural fluctuations generated within the simulations' time period (Martinez, 2015). Moreover, the stability and validity of the complexes and system were further characterised by their low and stable total energy (Figures 4.5B and 4.6B).

Apart from RMSD, RMSF provides a measure of residue displacement in the protein averaged over the simulation period (Martinez, 2015). Hence, a large fluctuation of RMSF indicated a high flexibility property as shown in the area between the 60^{th} and 80^{th} residue in the 'open' protease complex (Figure 4.5C). Further inspection of this area revealed that it was within a loop region of the protein and away from the active site. Moreover, it also encompassed the non-peptide linker (Gly_4-Ser-Gly_4). Therefore, the high degree of flexibility in this region was expected due to known flexibility of the loop region (Subramani & Floudas, 2012). It was also noted that there are several peaks prior to the aforementioned region showed a higher fluctuation in the RMSF value compared to the remaining residues (Figure 4.5C). Since those peaks were located within the NS2B section of the protein, it represented the degree of flexibility possessed by the NS2B cofactor structure as reported by other studies (Melino et al., 2006; Phong et al., 2011). Meanwhile, in the 'closed' protease structure, the observed high fluctuation in the RMSF value (Figure 4.6C) was within the loop region located in close proximity to the NS2B section of the protein which would likely affect the ligand affinity towards the active site. Therefore, this highlights the importance of MD simulations towards identifying an effective drug (De Vivo et al., 2016) especially prior to virtual screening since it provides an insight of the protease dynamics and flexibility.

The flexibility of the 'open' and 'closed' proteases can further be seen in the models chosen based on the selected cluster as visualised in Figure 4.8 whereby models mostly differ at the loop regions. Apart from the flexibility shown in the NS2B cofactor structure that was discussed earlier, the flexibility of NS3 domain structure was also notable. Although there were no significant conformational changes occurred on the catalytic triad positions, the flexibility of its corresponding region would definitely affect the binding of ligands on the active site. A study conducted by Zuo et al. (2009) which analysed the dynamics of the dengue protease complex (PDB ID: 2FOM) in the

absence of ligand, showed similar protein fluctuation pattern observed in this study. In addition, they also pointed out that in the absence of NS2B, the NS3 domain was more flexible but rather unstable. In another word, the NS2B cofactor and its configuration played a significant role in the protease complex stability which may explain the higher and more frequent fluctuation observed at the NS3 region of the 'open' protease (Figure 4.5C) compared to the 'closed' protease (Figure 4.6C). Therefore, it can also be inferred that the NS3 domain structure in the protease complex does possess a certain degree of flexibility that would contribute to ligand affinity towards the protease active site.

5.3 Ensemble-based Virtual Screening Evaluation

In order to take into account the flexibility of a protein during the screening or docking procedure, an ensemble of representative structures obtained from MD simulations was utilised onto the molecular docking process. The different binding energy yield by a ligand towards the protease complexes in a different conformation (represented as models) demonstrated that a ligand affinity does get affected by the different structural conformation whereby some has a greater affinity towards a particular model compared to the rest. As an example, shown in Table 4.8, the binding energy of the control ligand in Model_8 was the lowest or most negative value (−6.38 kcal/mol) whereas, in Model_9, it has the highest or most positive value (–4.69 kcal/mol) within the 'open' protease models. Therefore, this stresses on how protein dynamics affect the binding of a ligand and was also emphasised by Amaro and Li (2010) for researchers and scientist to move beyond the traditional and widely accepted induced-fit theory.

The initial screening result (Table 4.6), which was aimed to identify promising compounds among the collected ligands had revealed that only a small number of the ligands (~2.1%) showed consistency in producing favourable binding energy on each of

the model for both of the protease complexes. The top 10 chosen ligands (Table 4.7) were among the ~2.1% to ensure that the ligands would not only have the capability to bind on both of the protease complexes ('open' and 'closed) but in every models as well. It was observed that the calculated binding energy in the second docking showed a similar pattern as the initial docking and only showed a slight difference in value. On a general overview based on Table 4.8 and Table 4.9, all of the ligands exhibited a stronger binding energy (more negative) in the 'closed' models than in the 'open' models. As predicted, the structural conformation exhibited by the NS2B cofactor in the 'closed' protease models played a major part towards ligand affinity as pointed out by Nitsche et al. (2014).

Further inspection in the protein-ligand interactions was conducted by using the first models to represent the 'open' and 'closed' protease complex. Summarised in Table 4.10 and Table 4.11, the strongest mean binding energy (more negative) was exhibited by Lig567 in both of the protease complexes which indicated that a strong bond was generated between the ligand and protease complex. This may be highly contributed to the low number of available torsions (two active torsions) in Lig567 whereas, the control ligand which has the highest number of torsions (24 active torsions) displayed the weakest mean binding energy. Thus, it can be inferred that based on the estimated binding energy generated on all of the chosen ligands (more negative than control), they too will be able to form stable protein-ligand complexes exhibited by the control ligand as reported by Noble et al. (2012).

The main aim in centralising the grid box on the catalytic triad was to achieve interactions between the protein and ligand at the active site of the protease. Therefore, it will increase the probability in disrupting the proteolytic activity of the protease and ultimately prevent the processing of the dengue polyprotein (Noble et al., 2010). The

molecular docking evaluation revealed that majority of the chosen ligands form a hydrogen bond with the His-113 residue specifically on the 'NE2' atom in the 'open' protease complex except for Lig566 (Table 4.10). This interaction between the ligand and histidine residue is vital since it will plausibly disrupt the catalytic residues interactions which requires the histidine residue to attract proton on the serine residue for the nucleophilic reaction and subsequently the proteolysis to take place. Apart from the hydrogen bond interactions, hydrophobic contacts between several residues on the protease such as Val-114, Leu-190, Phe-192, Ser-193, Pro-194, Thr-196, Tyr-212, Asn-214, and Tyr-223 with all of the chosen ligands were also observed which will further enhance the protein-ligand complex stability as mentioned by Patil et al. (2009).

The ligand (Lig566) that did not form a hydrogen bond with the His-113 residue of the catalytic triads on the 'open' protease complex forms a bond with the Ser-197 residue instead. The interaction between the ligands with the serine residue of the catalytic triad was predominantly observed in the 'closed' protease complex (Table 4.11). Similarly, the formation of hydrogen bonds with this residue (Ser-195 in the 'closed' protease complex) will interrupt the protease catalytic activity. It was observed that the protein-ligand interaction on the serine residue occurred at the 'HN' atom which was located in close proximity to the serine's 'OH' atom. The formed hydrogen bond on that atom ('HN') would possibly hinder the 'OH' atom protonation that otherwise would form a highly reactive alkoxide ion for the protease catalytic reaction. Hence, this will eventually inhibit the protease complex, achieving the primary purpose of the virtual screening.

Apart from the ligands preferability in forming a hydrogen bond with the Ser-195 residue, it was also notable that majority of the ligands formed more hydrogen bonds in the 'closed' protease complex (Table 4.11). These hydrogen bonds interaction with the

other nearby residues (Arg-114, Gly-193, and Thr-194), would ultimately stabilised the position of the ligands on the protease thus, providing a better opportunity for the ligands to occupy the active site. Moreover, hydrophobic interactions were also observed between the chosen ligands and the nearby protease complex residues (Val-112, Lys-191, Pro-192, Asn-212, and Tyr-221). However, there were two of the chosen ligands (Lig563 and Lig569) that did not form any hydrogen bond with the catalytic triads on the 'closed' protease complex rather, they formed hydrogen bonds with the Met-40 and Asn-212 residues. Upon further evaluation, the position of the mentioned residues was located near to the residues of the catalytic triads (His-111 and Asp-135) and the residues formed hydrophobic interactions with the ligands instead. These weak interactions between the catalytic triad and the ligand may not be sufficient in preventing the catalysis to occur. Thus, Lig563 and Lig569 may not be the suitable candidate as the lead-compound for developing dengue protease inhibitor.

In addition to the analysis of the binding energy and protein-ligand hydrogen bond interactions, the distance between the protein and ligand were also evaluated (Table 4.10 and Table 4.11). Overall, a majority of the selected ligands in both of the 'open' and 'closed' protease model formed shorter interactions with the catalytic triads when compared to the control ligand. Thus, the short distance would inevitably stabilise the ligand positions on the protease and strengthen the affinity. Moreover, in both of the protease models, the number of formed clusters during molecular docking for the selected ligands with the lowest binding energy was relatively higher than the number of the cluster generated for the control ligand. Therefore, this indicated that there is a high possibility of the selected ligands to form the predicted orientation and interactions with the protease complex and thus increasing their potential as promising dengue protease inhibitors.

CHAPTER 6: SUMMARY

In this study, virtual screening was applied to identify promising compounds that can be used as antiviral which inhibit the DENV protease. The advantages offered in the ensemble-based virtual screening enabled the identification of natural-based compounds that showed a good affinity towards all of the represented dynamics generated by the dengue protease complex. As a recap, the represented dynamics of the dengue protease complex were obtained from MD simulations whereas the ligand candidates were mined from a natural based database using the determined criteria. Finally, virtual screening was conducted via molecular docking procedure which predicted the ligand's binding affinity and position on the protease.

Analysis of the obtained results revealed that among the chosen ligands which were selected from the initial screening, eight of the ligands (Lig556, Lig557, Lig558, Lig562, Lig564, Lig565, Lig566, and Lig567) have shown favourable protein-ligand interactions on both of the protease complexes ('open' and 'closed'). The most important feature evaluated was the ligand interaction with the catalytic triad which is vital in determining the capability of the ligands to act as a potential inhibitor. Despite the promising potentials hold by these ligands, the obtained results were merely a prediction and exact effect of the ligands (antagonist or agonist) on the protease are not yet known.

In a nutshell, the aforementioned ligands would provide a good starting point as lead-compounds for further inspection and future *in vitro* experiments. The derivatives for each of the ligand can also be further refined to achieve an effective inhibitory activity of DENV protease in the quest for generating an antiviral for DENV.

REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX, 1–2*, 19–25.

Acharya, C., Coop, A., Polli, J. E., & MacKerell, A. D. (2011). Recent advances in ligand-based drug design: Relevance and utility of the conformationally sampled pharmacophore approach. *Current Computer-Aided Drug Design*, *7*(1), 10–22.

Achee, N. L., Gould, F., Perkins, T. A., Reiner, R. C., Morrison, A. C., Ritchie, S. A., … Scott, T. W. (2015). A critical assessment of vector control for dengue prevention. *PLoS Neglected Tropical Diseases*, *9*(5), e0003655.

Aguiar, M., Stollenwerk, N., & Halstead, S. B. (2016). The risks behind Dengvaxia recommendation. *The Lancet Infectious Diseases, 16*(8), 882–883.

Aguilera-Pesantes, D., Robayo, L. E., Mendez, P. E., Mollocana, D., Marrero-Ponce, Y, Torres, F. J., & Mendez, M. A. (2017). Discovering key residues of dengue virus NS2b-NS3-protease: New binding sites for antiviral inhibitors design. *Biochemical and Biophysical Research Communications, 492*(4), 631–642.

Amaro, R. E. & Li. W. W. (2010). Emerging methods for ensemble-based virtual screening. *Current Topics in Medicinal Chemistry, 10*(1), 3–13.

Anderson, C.R., Downs, W. G., & Hill, A. E. (1956). Isolation of dengue virus from a human being in Trinidad. *Science, 124*(3214), 224–225.

Arakaki, T. L., Fang, N. X., Fairlie, D. P., Young, P. R., & Martin, J. L. (2002). Catalytically active dengue virus NS3 protease forms aggregates that are separable by size exclusion chromatography. *Protein Expression and Purification, 25*(12), 241–247.

Araújo, H. R. C., Carvalho, D. O., Ioshino, R. S., Costa-da-Silva, A. L., & Capurro, M. L. (2015). *Aedes aegypti* control strategies in Brazil: Incorporation of new technologies to overcome the persistence of dengue epidemics. *Insects*, *6*(2), 576–594.

Arragain, L., Dupont-Rouzeyrol, M., O'Connor, O., Sigur, N., Grangeon, J. P, Huguon, E., … Descloux, E. (2017). Vertical transmission of dengue virus in the peripartum period and viral kinetics in newborns and breast milk: New data. *Journal of the Pediatric Infectious Disease Society, 6*(4), 324–331

.

Ayub, A., Ashfaq, U. A., Idrees, S., & Haque, A. (2013). Global consensus sequence development and analysis of dengue NS3 conserved domains. *BioResearch Open Access*, *2*(5), 392–396.

Azam, M. N., Yeasmin, M., Ahmed, N. U., & Chakraborty, H. (2016). Modelling occurrence of dengue cases in Malaysia. *Iranian Journal of Public Health*, *45*(11), 1511–1512.

Bäck, A. T. & Lundkvist, Å. (2013). Dengue viruses – An overview. *Infection Ecology & Epidemiology, 3*, 19839.

Banerjee, P., Erehman, J., Gohlke, B.-O., Wilhelm, T., Preissner, R., & Dunkel, M. (2015). Super Natural II—A database of natural products. *Nucleic Acids Research*, *43*(Database issue), D935–D939.

Basavannacharya, C. & Vasudevan, S. G. (2014). Suramin inhibits helicase activity of NS3 protein of dengue virus in fluorescence-based high throughput assay format. *Biochemical and Biophysical Research Communications, 453*(3), 539–544.

Benmansour, F., Trist, I., Coutard, B., Decroly, E., Querat, G., Brancale, A., & Barral, K. (2017). Discovery of novel dengue virus NS5 methyltransferase non-nucleoside inhibitors by fragment-based drug design. *European Journal of Medicinal Chemistry, 125*, 865–880.

Brabant, M., Baux, L., Casimir, R., Briand, J. P., Chaloin, O., Porceddu, M., … Jacotot, E. (2009). A flavivirus protein M-derived peptide directly permeabilizes mitochondrial membranes, triggers cell death and reduces human tumor growth in nude mice. *Apoptosis, 14*, 1190–1203.

Brault, J. B., Kudelko, M., Vidalain, P. O., Tangy, F., Desprès, P., & Pardigon, N. (2011). The interaction of flavivirus M protein with light chain Tctex-1 of human dynein plays a role in late stages of virus replication. *Virology, 417*, 369–378.

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics, 126*(1), 014101.

Byk, L. A. & Gamarnik, A. V. (2016). Properties and functions of the dengue virus capsid protein. *Annual Review of Virology, 3*(1), 263–281.

Calisher, C. H. (2005). Persistent emergence of dengue. *Emerging Infectious Diseases*, *11*(5), 738–739.

Carrington, L. B. & Simmons, C. P. (2014). Human to mosquito transmission of dengue viruses. *Frontiers in Immunology*, *5*, 290.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., … Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, *26*(16), 1668–1688.

Chandramouli, S., Joseph, J. S., Daudenarde, S., Gatchalian, J., Cornillez-Ty, & C., Kuhn, P. (2010). Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. *Journal of Virology, 84*(6), 3059–3067.

Chang, J., Schul, W., Butters, T. D., Yip, A., Liu, B., Goh, A., … Block, T. M. (2011). Combination of alpha-glucosidase inhibitor and ribavirin for the treatment of dengue virus infection *in vitro* and *in vivo*. *Antiviral Research*, *89*(1), 26–34.

Chen, C. Y. (2011). TCM Database@Taiwan: The world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS ONE*, *6*(1), e15939.

Chen, L. H. & Wilson, M. E. (2004). Transmission of dengue virus without a mosquito vector: Nosocomial mucocutaneous transmission and other routes of transmission. *Clinical Infectious Diseases, 39*, e56–e60.

Chen, W., Loscha, K. V., Nitsche, C., Graham, B., & Otting, G. (2014). The dengue virus NS2B-NS3 protease retains the closed conformation in the complex with BPTI. *Federation of European Biochemical Societies (FEBS) Letters, 588*(14), 2206–2211.

Chin, T. S. (2014). *Alternative treatments for dengue fever*. *The Star Online*. Retrieved January 4, 2018, from https://www.thestar.com.my/lifestyle/health/2014/04/13/alternative-treatments-for-dengue-fever/

Clyde, K., Kyle, J. L., & Harris, E. (2006). Recent advances in deciphering viral and host determinants dengue virus replication and pathogenesis. *Journal of Virology, 80*(23), 11418-11432.

Courageot, M.P., Frenkiel, M.P., Duarte Dos Santos, C., Deubel, V., & Desprès, P. (2000). α-Glucosidase inhibitors reduce dengue virus production by affecting the initial steps of virion morphogenesis in the endoplasmic reticulum. *Journal of Virology*, *74*(1), 564–572.

Cruz-Oliveira, C., Freire, J. M., Conceição, T. M., Higa, L. M., Castanho, M. A. R. B., & Da Poian, A. T. (2015). Receptors and routes of dengue virus entry into the host cells. *Federation of European Microbiological Societies* (*FEMS*) *Microbiology Reviews*, *39*(2), 155–170.

CureJoy. (2017). *Home remedies to treat dengue and ways to prevent it*. Retrieved January 1, 2018, from https://www.curejoy.com/content/home-remedies-to-treat-dengue-fever/

Daggett, V. & Levitt, M. (1992). A model of the molten globule state from molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America, 89*(11), 5142–5146.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089–10092.

De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry, 59*(9), 4035–4061.

Desai, V. H., Kumar, S. P., Pandya, H. A., & Solanki, H. A. (2015). Receptor-guided *de novo* design of dengue envelope protein inhibitors. *Applied Biochemistry and Biotechnology, 177*(4), 861–878.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., … Notredame, C. (2011). T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, *39*(Web Server issue), W13–W17.

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, *47*, 20–33.

Djinovic-Carugo, K. & Carugo, O. (2015). Missing strings of residues in protein crystal structures. *Intrinsically Disordered Proteins*, *3*(1), e1095697.

Dong, H., Ray, D., Ren, S., Zhang, B., Puig-Basagoiti, F., Takagi, Y., … Shi, P. Y. (2007). Distinct RNA elements confer specificity to flavivirus RNA cap methylation events. *Journal of Virolgy, 89*(9), 4412–4421.

Durrant. J. D. & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BioMed Central Biology, 9* (1), 71.

Dwivedi, V. D., Tripathi, I. P., Tripathi, R. C., Bharadwaj, S., & Mishra, S. K. (2017). Genomics, proteomics and evolution of dengue virus. *Briefings in Functional Genomics, 16*(4), 217–227.

Elshuber, S., Allison, S. L., Heinz, F. X., & Mandl, C. W. (2003). Cleavage of protein prM is necessary for infection of BHK-21 cells by tick-borne encephalitis virus FN1. *Journal of General Virology, 84*, 183–191.

Erbel, P., Schiering, N., D'Arcy, A., Renatus, M., Kroemer, M., Lim, S. P., … Hommel, U. (2006). Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nature Structural & Molecular Biology, 12*(4), 372–373.

Fischl, W. & Bartenschlager, R. (2013). High-throughput screening using dengue virus reporter genomes. In E. Y. Gong (Ed.), *Antiviral methods and protocols* (pp. 205–219). Totowa, NJ: Humana Press.

Forshey, B. M., Stoddard, S. T., & Morrison, A. C. (2016). Dengue viruses and lifelong immunity: Reevaluating the conventional wisdom. *The Journal of Infectious Diseases, 214*, 979–981.

Freisner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., … Shenkin, P. S. (2004). Glide: A new approach for rapid, accurate docking and scoring. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry, 47*(7), 1739–1749.

Gartz, N. G. (2004). Critical review of the vector status of *Aedes albopictus*. *Medical and Veterinary Entomology, 18*(3), 215.

Godói, I. P., Santos, A. S., Reis, E. A., Lemos, L. L. P., Brandão, C. M. R., Alvares, J., … Guerra Júnior, A. A. (2017). Consumer willingness to pay for dengue vaccine (CYD-TDV, Dengvaxia®) in Brazil; Implications for future pricing considerations. *Frontiers in Pharmacology*, *8*, 41.

Gong, E. Y., Kenens, H., Ivens, T., Dockx, K., Vermeiren, K., Vandercruyssen, G., … Kraus, G. (2013a). Expression and purification of dengue virus NS5 polymerase and development of a high-throughput enzymatic assay for screening inhibitors of dengue polymerase. In E. Y. Gong (Ed.), *Antiviral methods and protocols* (pp. 237–247). Totowa, NJ: Humana Press.

Gong, E. Y., Smets, A., Verheyen, N., Clynhens, M., Gustin, E., Lory, P., & Kraus, G. (2013b). A duplex real-time RT-PCR assay for profiling inhibitors of four dengue serotypes. In E. Y. Gong (Ed.), *Antiviral methods and protocols* (pp. 195–204). Totowa, NJ: Humana Press.

Greenwood, M. (2016). *Aggressive and human-loving Aedes aegypti mosquito spreading Zika and other diseases*. Retrieved December 20, 2017, from http://publichealth.yale.edu /news/article.aspx?id=12026

Gu, F. & Shi, P. Y. (2014). The challenges of dengue drug discovery and development. *Clinical Investigation, 4*(8), 683–685.

Gu, J., Gui, Y., Chen, L., Yuan, G., Lu, H.-Z., & Xu, X. (2013). Use of natural products as chemical library for drug discovery and network pharmacology. *PLoS ONE*, *8*(4), e62839.

Gubler, D. J. & Clark. G. G. (1995). Dengue/dengue hemorrhagic fever: The emergence of a global health problem. *Emerging Infectious Diseases, 1*(2), 55–57.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clinical Microbiology Reviews*, *11*(3), 480–496.

Gupta, D. (2016). Dengvaxia®: Prayers anwered or not? *Indian Journal of Community Health, 28*(4), 398–399.

Halstead, S. B. (2016). Licensed dengue vaccine: Public health conundrum and scientific challenge. *The American Journal of Tropical Medicine and Hygiene, 95*(4), 741–745.

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics, 4*(8), 17.

Harvey, A. L., Edrada-Ebel, R., & Quinn, R. J. (2015). The re-emergence of natural products for drug discovery in the genomics era. *Nature Reviews Drug Discovery, 14*, 111–129.

Heilman, J. M., Wolff, J. D., Beards, G. M., & Basden, B. J. (2014). Dengue fever: A wikipedia clinical review. *Open Medicine, 8*(4), e105.

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINC: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry, 18*(12), 1463–1472.

Hii, Y. L., Zaki, R. A., Aghamohammadi, N., & Rocklöv, J. (2016). Research on climate and dengue in Malaysia: A systematic review. *Current Environmental Health Reports*, *3*, 81–90.

Holmes, E. C., Bartley, L. M., & Garnett, G. P. (1998). The Emergence of dengue: Past present and future. *Biomedical Research Reports, 1*, 301–325.

Hughes, J., Rees, S., Kalindjian, S., & Philpott, K. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), 1239–1249.

Jamroz, M., Kolinski, A., & Kihara, D. (2016). Ensemble-based evaluation for protein structure models. *Bioinformatics, 32* (12), i314–i321.

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology, 267*(3), 727–748.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics, 79*(2), 926–935.

Karami, Y., Laine, E., & Carbone, A. (2016). Dissecting protein architecture with communication blocks and communicating segment pairs. *BioMed Central Bioinformatics*, *17*(Suppl 2), 13.

Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., Rao, B. G., Tung, R. D., & Navia, M. A. (1995). Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *Journal of The American Chemical Society, 117*(3), 1181–1182.

Kim, Y. M., Gayen, S., Kang, C., Joy, J., Huang, Q., Chen, A. S., … Keller, T. H. (2013). NMR Analysis of a novel enzymatically active unlinked dengue NS2B-NS3 protease complex. *The Journal of Biological Chemistry*, *288*(18), 12891–12900.

Knapp, B., Frantal, S., Cibena, M., Schreiner, W., & Bauer, P. (2011). Is an intuitive convergence definition of molecular dynamics simulation solely based on the root mean square deviation possible? *Journal of Computational Biology, 18*(8), 997–1005.

Koehn, F. E. & Carter, G. T. (2005). The evolving role of natural products in drug discovery. *Nature Reviews Drug Discovery, 4*(3), 206–220.

Kok, W. M. (2016). New developments in flavivirus drug discovery. *Expert Opinion on Drug Discovery, 11*(5), 433–445.

Kore, P. P., Mutha, M. M., Antre, R. V., Oswal, R. J., & Kshirsagar, S. S. (2012). Computer-aided drug design: An innovative tool for modelling. *Open Journal of Medicinal Chemistry, 2*, 139–148.

Kouretova, J., Hammamy, M. Z., Epp, A., Hardes, K., Kallis, S., Zhang, L., … Steinmetzer, T. (2017). Effects of NS2B-NS3 protease and furin inhibition on West Nile and dengue virus replication. *Journal of Enzyme Inhibition and Medicinal Chemistry, 32*(1), 712–721.

Kraemer, M. U. G., Sinka, M. E., Duda, K. A., Mylne, A., Shearer, F. M., Brady, O. J., … Hay, S. I. (2015). The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence. *Scientific Data, 2*, 1–8.

Kulczyński, M., Tomaszewski, M., Luniewski, M., & Olender, A. (2017). Air transport and the spread of infectious diseases. *World Scientific News, 76,* 123–135.

Lafrance, A. (2016). *The sneak-attack mosquito*. Retrieved November 30, 2017, from https://www.theatlantic .com/science/archive/2016/04/aedes-aegypti/479619/

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography, 26*, 283–291.

Lavecchia, A. & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: A critical review. *Current Medical Chemistry, 20* (23), 2839–2860.

Leelananda, S. P. & Lindert, S. (2016). Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry, 12*, 2694–2718.

Leung, D., Schroder, K., White, H., Fang, N. X., Stoermer, M. J., Abbenante, G., … Fairlie, D. P. (2001). Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. *Journal of Biological Chemistry, 276*(49), 45762–45771.

Li, H., Zhu, L., Hou, S., Yang, J., Wang, J., & Liu, J. (2014). An inhibitor model of BPTI to unlined dengue virus NS2B-NS3 protease. *Federation of European Biochemical Societies (FEBS) Letters, 588*, 2794–2799.

Lin, K.-H., Ali, A., Rusere, L., Soumana, D. I., Kurt Yilmaz, N., & Schiffer, C. A. (2017). Dengue virus NS2B/NS3 protease inhibitors exploiting the prime side. *Journal of Virology*, *91*(10), e00045–17.

Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: Principles, applications and recent advances. *Current Topics in Medicinal Chemistry, 14*, 1923–1938.

Lipinski, C. A. (2004). Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discovery Today: Technologies, 1*(4), 337–341.

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews, 1–3*, 3–25.

Lomenick, B., Olsen, R. W., & Huang, J. (2011). Identification of direct protein targets of small molecules. *American Chemical Society (ACS) Chemical Biology*, *6*(1), 34–46.

Loub, W. D., Farnsworth, N. R., Soejarto, D. D., & Quinn, M. L. (1985). NAPRALERT: Computer handling of natural product research data. *Journal of Chemical Information and Computer Science, 25*(2), 99–103.

Lounnas, V., Ritschel, T., Kelder, J., McGuire, R., Bywater, R. P., & Foloppe, N. (2013). Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Computational and Structural Biotechnology Journal*, *5*, e201302011.

Lutomia, J., Barrera, R., Makio, A., Mustiya, J., Koka, H., Owaka, S., … Sang, R. (2016). Dengue outbreak in Mombsa City, Kenya, 2013-2014: Entomologic investigation. *PLoS Neglected Tropical Diseases, 10*(10), e0004981.

Luz, P. M., Vanni, T., Medlock, J., Paltiel, A. D., & Galvani, A. P. (2011). Dengue vector control strategies in an urban setting: An economic modelling assessment. *Lancet*, *377*(9778), 1673–1680.

MacKerell, Jr., A. D., Feig, M., & Brooks, III, C. L. (2004). Extending the treatment of backbone energetic in protein force fields: Limitation of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry, 25*(11), 1400–1415.

Martínez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS ONE*, *10*(3), e0119264.

McMeniman, C. J., Lane, R. V., Cass, B. N., Fong, A. W. C., Sidhu, M., Wang, Y. F., & O'Neil, S. L. (2009). Stable introduction of a life-shortening *Wolbachia* infection into the mosquito *Aedes aegypti*. *Science, 323*(5910), 141–144.

Melino, S., Fucito, S., Campagna, A., Wrubi, F., Gamarnik, A., Cicero, D. O., & Paci, M. (2006). The active essential CFNS3d protein complex. *The Federation of European Biochemical Societies (FEBS) Journal, 273*(16), 3650–3662.

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design, 7*(2), 146–157.

Messina, J. P., Brady, O. J., Scott, T. W., Zou, C., Pigott, D. M., Duda, K. A., … Hay, S. I. (2014). Global spread of dengue virus types: Mapping the 70 year history. *Trends in Microbiology, 22*(3), 138–146.

Mir, M., Khurshid, R. & Aftab, R. (2014). Management of thrombocytopenia and flu-like symptoms in dengue patients with herbal water of *Euphorbia hirta*. *Journal of Ayub Medical College, 24*(3–4), 6–9.

Mirza, S. B., Salmas, R. E., Fatmi, M. Q., & Durdagi, S. (2016). Virtual screening of eighteen million compounds against dengue virus: Combined molecular docking and molecular dynamics simulations study. *Journal of Molecular Graphics and Modelling, 66*, 99–107.

Moreira, L. A., Iturbe-Omaetxe, I., Jeffery, J.A., Lu, G., Pyke, A. T., Hedges, L. M., … O'Neil, S. L. (2009). A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and Plasmodium. *Cell, 139*(7), 1268–1278.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). Autodock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry, 30*(16), 2785–2791.

Murphy, B. R. & Whitehead, S. S. (2011). Immune response to dengue virus and prospects for a vaccine. *Annual Review of Immunology, 29*, 587–619.

Murray, N. E. A., Quam, M. B., & Wilder-Smith, A. (2013). Epidemiology of dengue: Past, present and future prospects. *Clinical Epidemiology, 5*, 299–309.

Mutt, E. & Sowdhamini, R. (2016). Molecular dynamics simulations and structural analysis to decipher functional impact of a twenty residue insert in the ternary complex of *Mus musculus* TdT isoform. *PLoS ONE*, *11*(6), e0157286.

Muzari, M. O., Devine, G., Davis, J., Crunkhorn, B., van den Hurk, A., Whelan, P., … Ritchie, S. (2017). Holding back the tiger: Successful control program protects Australia from *Aedes albopictus* expansion. *PLoS Neglected Tropical Diseases, 11*(2), e0005286.

Nedjadi T., El-Kafrawy,S., Sohrab, S. S. Despres, P., Damanhouri, G., & Azhar, E. (2015). Tackling dengue fever: Current status and challenges. *Virology Journal, 12*, 212.

Nemésio, H., Palomares-Jerez, F., & Villalaín, J. (2011). The membrane-active regions of the dengue virus proteins C and E. *Biochimica et Biophysica Acta (BBA)-Biomembranes, 1808*(10), 2390–2402.

Nitsche, C. & Klein, C. D. (2013). Fluorimetric and HPLC-based dengue virus protease assays using a FRET substrate. In E. Y. Gong (Ed.), *Antiviral methods and protocols* (pp. 221–236). Totowa, NJ: Humana Press.

Nitsche, C., Holloway, S., Schirmeister, T., & Klein, C. D. (2014). Biochemistry and medicinal chemistry of the dengue virus protease. *Chemical Reviews, 114*(22), 11348–11381.

Noble, C. G., Chen, Y. L., Dong, H., Gu, F., Lim, S. P., Schul, W., … Shi, P. Y. (2010). Strategies for development of dengue virus inhibitors. *Antiviral Research, 85*(3), 450–462.

Noble, C. G., Seh, C. C., Chao, A. T., & Shi, P. Y. (2012). Ligand-bound structures of the dengue virus protease reveal the active conformation. *Journal of Virology, 86*(1), 438–446.

Normile, D. (2017). Safety concerns derail dengue vaccination program. *Science, 358*(6370), 1514–1515.

Normile. D. (2013). Tropical medicine. Surprising new dengue virus throws a spanner in disease control efforts. *Science, 342*(6157), 415.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics, 3,* 33.

Oliveira, A. S., Silva, M. L., Oliveira, A. F. C. S., Silva, C. C., Teixeira, R. R., & De Paula, S. O. (2014). NS3 and NS5 proteins: Important targets for anti-dengue drug design. *Journal of the Brazilian Chemical Society*, *25*(10), 1759–1769.

Orozco. M. (2014). A theoretical view of protein dynamics. *Chemical Society Reviews, 43*, 5051–5056.

Osanai C. H., Travassoss da Rosa, A. P., Tang, A. T., do Amaral, R. S., Passos, A. D., & Tauil, P. L. (1983). Dengue outbreak in Bao Vista, Roraima. *Journal of the São Paulo Institute of Tropical Medicine, 25*(1), 53–54.

Parida, M. M., Upadhyay, C., Pandya, G., & Jana, A. M. (2002) Inhibitory potential of neem (*Azadirachta indica Juss*) leaves on dengue virus type-2 replication. *Journal of Ethnopharmacology, 79*(2), 273–278.

Parrinello, M. & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics, 52*(12), 7182.

Patil, R., Das, S., Stanley, A., Yadav, L., Sudhakar, A., & Varma, A. K. (2009). Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathway of drug-designing. *PLos ONE, 5*(8), e12029.

Paupy, C., Dellate, H., Bagny, L., Corbel, V., & Fontenille, D. (2009). *Aedes albopictus*, an arbovirus vector: From the darkness to the light. *Microbes and Infection, 11*(14–15), 1177–1185.

Pedretti, A., Villa, L., & Vistoli, G. (2004). VEGA – An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. *Journal of Computer Aided Molecular Design, 18*(3), 167–173.

Peng, M., Watanabe, S., Chan, K. W. K., He, Q., Zhao, Y., Zhang, Z., … Li, G. (2017). Luteolin restricts dengue virus replication through inhibition of the proprotein convertase furin. *Antiviral Research, 143*, 176–185.

Perera, R. & Kuhn, R. J. (2008). Structural proteomics of dengue virus. *Current Opinion in Microbiology*, *11*(4), 369–377.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera – A visualization system for exploratory research and analysis. *Journal of Computational Chemistry, 25*(13), 1605–1612.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., … Schulten, K. (2005). Scalable Molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802.

Phong, W. Y., Moreland, N. J., Lim, S. P., Wen, S. P., Wen, D., Paradkar, P. N., & Vasudevan, S. G. (2011). Dengue protease activity: The structural integrity and interaction of NS2B with NS3 protease and its potential as a drug target. *Bioscience reports, 31*(5), 399–409.

Prommalikit, O. & Thisyakorn, U. (2015). Dengue virus virulence and disease severity. *The Southeast Asian Journal of Tropical Medicine and Public Health*, *46* (Supplement 1), 35–42.

Qi, R. F., Zhang, L., & Chi, C. W. (2008). Biological characteristics of dengue virus and potential targets for drug design. *Acta Biochimica et Biophysica Sinica, 40* (2), 91–101.

Quam, M. B., Sessions, O., Kamaraj, U. S., Rocklöv, J., & Wilder-Smith, A. (2016). Dissecting Japan's dengue outbreak in 2014. *The American Journal of Tropical Medicine and Hygiene*, *94*(2), 409–412.

Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard III, W. A., & Skiff, W. M. (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society, 114*(25), 10024–10035.

Rarey, M., Kramer B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology, 261*(3), 470–489.

Rezza, G. (2012). *Aedes albopictus* and the reemergence of dengue. *BioMed Central Public Health, 12*, 72.

Rico-Hesse, R. (2010). Dengue virus virulence and transmission determinants. *Current Topics in Microbiology and Immunology*, *338*, 45–55.

Rodenhuis-Zybert, I. A., Wilschut, J., & Smit, J. M. (2010). Dengue virus life cycle: Viral and host factors modulating infectivity. *Cellular and Molecular Life Sciences, 67*, 2773–2786.

Rodrigues, J. P., Melquiond, A. S., Karaca, E., Trellet, M., van Dijk, M., van Zundert, G. C., … Bonvin, A. M. (2013). Defining the limits of homology modelling in information-driven protein docking. *Proteins: Structure, Function, and Bioinformatics, 81*(12), 2119–2128.

Rose, P. W., Prlić, A., Bi, C., Bluhm, W. F., Christie, C. H., Dutta, S., … Burley, S. K. (2015). The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Research*, *43*(Database issue), D345–D356.

Russel, P. K., Buescher, E. L., McCown, J. M., & Ordoñez, J. (1966). *The American Journal of Tropical Medicine and Hygiene, 15*(4), 573–579.

Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology, 234*(3), 779–815.

Sánchez-Vargas, I., Scott, J. C., Poole-Smith, B. K., Franz, A. W. E., Barbosa-Solomieu V, Wilusz, J., … Blair, D. (2009). Dengue virus type 2 infections of *Aedes aegypti* are modulated by the mosquito's RNA interference pathway. *PLoS Pathogens, 5*(2), e1000299.

Sarala, N. & Paknikar, S. (2014). Papaya extract to treat dengue: A novel therapeutic option? *Annals of Medical and Health Sciences Research*, *4*(3), 320–324.

Sayce, A. C., Alonzi, D. S., Killingbeck, S. S., Tyrrell, B. E., Hill, M. L., Caputo, A. T., … Zitzmann, N. (2016). Iminosugars inhibit dengue virus production via inhibition of ER alpha-glucosidases – not glycolipid processing enzymes. *PLoS Neglected Tropical Diseases, 10*(3), e0004524.

Scitable. (2014a). *What is dengue fever?* Retrieved November 20, 2017, from https://www.nature.com/ scitable/topicpage/what-is-dengue-fever-22399100

Scitable. (2014b). *Dengue viruses*. Retrieved November 20, 2017, from https://www. nature.com/scitable /topicpage/dengue-viruses-22400925

Seal, R. (2015). *12 natural home remedies for treating dengue fever*. Retrieved December 27, 2017, from http://lmt-lss.com/12-natural-home-remedies-for-treating-dengue-fever/

Sharma, S. (2013). *Myths and facts about dengue*. Retrieved November 23, 2017, from http://indiatoday.into day.in/story/myths-and-facts-about-dengue/1/327080.html

Simmonds, P., Becher, P., Bukh, J., Gould, E. A., Meyers, G., Monath, T., … ICTV Report Consortium. (2017). ICTV virus taxonomy profile: *Flaviviridae*. *Journal of General Virology, 9*(8), 2–3.

Simmons, C. P., Farrar, J. J., van Vinh Chau, N., & Wills, B. (2012). Dengue. *New England Journal of Medicine, 366*(15), 1423–1432.

Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational Methods in Drug Discovery. *Pharmacological Reviews*, *66*(1), 334–395.

Solanki, H. A., Shankar, S. S., & Pandya, H. A. (2017). Dengue as a druggable disease: A review. *Journal of Fever, 1*(1), 1–19.

Sorin, E. J., Alvarado, W., Cao, S., Radcliffe, A., La, P., & An, Y. (2017). Ensemble molecular dynamics of a protein-ligand complex: Residual inhibitor entropy enhances drug potency in butyrylcholinesterase. *Bioenergetics*, *6*(1), 145.

Suaya, J. A., Shepard, D. S., Chang, M. S., Caram, M., Hoyer, S., Socheat, D., … Nathan, M. B. (2007). Cost-effectiveness of annual targeted larviciding campaigns in Cambodia against the dengue vector Aedes aegypti. *Tropical Medicine & International Health, 12*(9), 1026–1036.

Subramani, A., & Floudas, C. A. (2012). Structure prediction of loops with fixed and flexible stems. *The Journal of Physical Chemistry. B*, *116*(23), 6670–6682.

Sweeney, N. L., Hanson, A. M., Mukherjee, S., Ndjomou, J., Geiss, B. J., Steel, J. J., … Frick, D. N. (2015). Benzothiazole and pyrrolone flavivirus inhibitors targeting the viral helicase. *American Chemical Society Infectious Diseases, 1*(3), 140–148.

Szymański, P., Markowicz, M., & Mikiciuk-Olasik, E. (2012). Adaptation of high-throughput screening in drug discovery—Toxicological screening tests. *International Journal of Molecular Sciences*, *13*(1), 427–452.

Takagi, Y., Matsui, K., Nobori, H., Maeda, H., Sato, A. Kurosu, T., … Yoshida, Y. (2017). Discovery of novel cyclic peptide inhibitors of dengue virus NS2B-NS3 protease with antiviral activity. *Bioorganic & Medicinal Chemistry Letters, 27*(15), 3586–3590.

Talele, T. T., Khedkar, S. A., & Rigby, A. C. (2010). Successful applications of computer aided drug discovery: Moving drugs from concept to clinic. *Current Topics in Medicinal Chemistry, 10*(1), 124–141.

Thi, K. P., Viet, H. H., & Nguyen, H. M. (2016). Major resistant mechanism to insecticides of Aedes aegypti mosquito: A vector of dengue and Zika virus in Vietnam. *SM Tropical Medicine Journal, 1*(2), 1010.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461.

Valli, M., dos Santos, R. N., Figueira, L. D., Nakajima, C. H., Castro-Gamboa, I, Andricopulo, A. D., & Bolzani, V. S. (2013). Development of a natural products database from biodiversity of Brazil. *Journal of Natural Products, 76*(3), 439–444.

Verlet. L. (1967). Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review, 159*(1), 98–103.

Vincetti, P., Caporuscio, F., Kaptein, S., Gioiello, A., Mancino, V., Suzuki, Y., … Radi, M. (2015). Discovery of multitarget antivirals acting on both the dengue virus NS5-NS3 interaction and the host Src/Fyn kinases. *Journal of Medicinal Chemistry, 58*(12), 4964–4975.

Virginia Tech. (2015). *Mosquito sex-determining gene could help fight dengue fever*. Retrieved December 28, 2017, from www.sciencedaily.com/releases/2015/05/150521143815 .htm

World Health Organization. (2016). Dengue vaccine: WHO position paper – July 2016. *Weekly Epidemiological Report, 91*, 349–364.

World Health Organization. (2017). *Dengue fever – Burkina Faso*. *Disease Outbreak News*. Retrieved January 2, 2018, from http://www.who.int/csr/don/6-november-2017-dengue-burkina-faso/en/

World Health Organization. (2017). *Dengue fever – Sri Lanka*. *Disease Outbreak News*. Retrieved December 25, 2017, from http://www.who.int/csr/don/19-july-2017-dengue-sri-lanka/en/

Xie, T., Song, S., Li, S., Ouyang, L., Xia, L., & Huang, J. (2015). Review of natural product databases. *Cell Proliferation, 48*, 398–404.

Yacoub, S., Mongkolsapaya, J., & Screaton, G. (2016). Recent advances in understanding dengue. *F1000Research, 5*, Faculty Reviews–78.

Yildiz, M., Ghosh, S., Bell, J. A., Sherman, W., & Hardy, J. A. (2013). Allosteric inhibition of the NS2B-NS3 protease from dengue virus. *American Chemical Society (ACS) Chemical Biology*, *8*(12), 2744–2752.

Zeidler, J. D., Fernandes-Siqueira, L. O., Barbosa, G. M., & Da Poian, A. T. (2017). Non-canonical roles of dengue virus non-structural proteins. *Viruses, 9*, 42.

Zhao, H. & Caflisch, A. (2014). Molecular dynamics in drug design. *European Journal of Medicinal Chemistry, 91*, 4–14.

Zuo, Z., Liew, O. W., Chen, G., Chong, P. C. J., Lee, S. H., Chen, K., … Zhu, W. (2009). Mechanism of NS2B-mediated activation of NS3pro in dengue virus: Molecular dynamics simulations and bioassays. *Journal of Virology, 83*(2), 1060–1070.

APPENDIX A

MSA Result of NS2B Protein Sequences via T-Coffee

```
T-COFFEE, Version_11.00.d625267
CPU TIME:0 sec.
SCORE=97
*
 BAD AVG GOOD
*
DENV1_NS2B   :  98
DENV2_NS2B   :  98
DENV3_NS2B   :  98
DENV4_NS2B   :  98
4M9K_NS2B    :  93
3U1I_NS2B    :  96
cons         :   9

DENV1_NS2B   SWPLNEGIMAVGIVSILLSSLLKNDVPLAGPLIAGGMLIACYVISGSSADLSLEKAAEVSWEEEAEHSG
DENV2_NS2B   SWPLNEAIMAVGMVSILASSLLKNDIPMTGPLVAGGLLTVCYVLTGRSADLELERAADVKWEDQAEISG
DENV3_NS2B   SWPLNEGVMAVGLVSILASSLLRNDVPMAGPLVAGGLLIACYVITGTSADLTVEKAPDVTWEEEAEQTG
DENV4_NS2B   SWPLNEGIMAVGLVSLLGSALLKNDVPLAGPMVAGGLLLAAYVMSGSSADLSLEKAANVQWDEMADITG
4M9K_NS2B    ----------------GSHML-----------------------EADLELERAADVRWEEQAEISG
3U1I_NS2B    ----------------------------GPL--------------GSDLTVEKAADVTWEEEAEQTG

cons         :** :*:*.:* *:: *: :*

DENV1_NS2B   ASHNILVEVQDDGTMKIKDEERDDTLTILLKATLLAISGVYPMSIPATLFVWYFWQKKKQR
DENV2_NS2B   SSPILSITISEDGSMSIKNEEEEQTLTILIRTGLLVISGLFPVSIPITAAAWYLWEVKKQR
DENV3_NS2B   VSHNLMITVDDDGTMRIKDDETENILTVLLKTALLIVSGIFPYSIPATLLVWHTWQKQTQR
DENV4_NS2B   SSPIVEVKQDEDGSFSIRDVEETNMITLLVKLALITVSGLYPLAIPVTMTLWYMWQVKTQR
4M9K_NS2B    SSPILSITISEDGSMSIKNEEEEQTL----------------------------------
3U1I_NS2B    VSHNLMITVDDDGTMRIKDDETENIL----------------------------------

cons         *  : :  .:**:: *:: *  : :
```

APPENDIX B

MSA Result of Full NS3 Protein Sequences via T-Coffee

```
T-COFFEE, Version_11.00.d625267
CPU TIME:0 sec.
SCORE=78
*
 BAD AVG GOOD
*
RefSeq_D1    :  86
RefSeq_D2    :  76
RefSeq_D3    :  86
RefSeq_D4    :  82
4M9K_NS3     :  55
3U1I_NS3     :  55
cons         :   7

RefSeq_D1    ---------SGVLWD-------------TPSPPEVER-AVLD------------------DGIYRILQR
RefSeq_D2    -------------AG--------VLWDVPSPPPMGKAELEDG------------------AY--RIKQK
RefSeq_D3    ---------SGVLWD-------------VPSPPETQK-AELE------------------EGVYRIKQQ
RefSeq_D4    -----------------------------------------SGALWDVPSPAATKKAALSEGVYRIMQR
4M9K_NS3     --------------GGGGSGGGGAGVLWDVPSPPPVGKAELE------------------DGAYRIKQK
3U1I_NS3     GGGGSGGGGSGVLWD-------------VPSPPETQK-AELE------------------EGVYRIKQQ

cons                                                                         ** *:


RefSeq_D1    GLLGRSQVGVGVFQEGVFHTMWHVTRGAVLMYQGKRLEPSWASVKKDLISYGGGWRFQGSWNAGEE---
RefSeq_D2    GILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHKGK--RIEPSWADVKKDLISYGGGWKLEG---EWKEG
RefSeq_D3    GIFGKTQVGVGVQKEGVFHTMWHVTRGAVLTHNGKRLEPNWASVKKDLISYGGGWRLSAQWQKGEE---
RefSeq_D4    GLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMISYGGGWRLGDKW---DK-E-
4M9K_NS3     GILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHKGKRIEPSWADVKKDLISYGGGWKLEGEW---KE-G-
3U1I_NS3     GIFGKTQVGVGVQKEGVFHTMWHVTRGAVLTHNGKRLEPNWASVKKDLISYGGGWRLSAQW---QK-G-

cons         *::* :*:*.*:  **.**********:*: ::    . . : .  .    . *     .    .


RefSeq_D1    --VQVIAVEP---------GKNPKNVQTAPGTFKTPEGEVGA-----IALDFKPGTSGSPIVNREGKIV
RefSeq_D2    EE-----VQV---------LALEPGKNPRAVQTKP---GLFKTNAGTIGAVSLDFSPGTSGSPIIDKKG
RefSeq_D3    --VQVIAVEP---------GKNPKNFQTTPGTFQTTTGEIGA-----IALDFKPGTSGSPIINREGKVV
RefSeq_D4    EDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVT---LDFK-----PGTSGSPIINRKGKVIGLYGNG
4M9K_NS3     EEVQVLALEP---------GKNPRAVQTKPGLFKTNTGTIGA-----VSLDFSPGTSGSPIVDKKGKVV
3U1I_NS3     EEVQVIAVEP---------GKNPKNFQTMPGTFQTTTGEIGA-----IALDFKPGTSGSPIINREGKVV

cons                ::                  .      .             .         .


RefSeq_D1    GLYGNGVVTTSGTY------V---------SAIAQ--AKAS-QEGPLPEIEDEVFRKRNLTIMDLHPGS
RefSeq_D2    -------------KVVGLYGNGVVTRSGAYVSAIAQ--TEKSIE-DNPEIEDDIFRKRRLTIMDLHPGA
RefSeq_D3    GLYGNGVVTKNGGY------V---------SGIAQTNAEP---DGPTPELEEEMFKKRNLTIMDLHPGS
RefSeq_D4    VVTKSG-------------DY---------VSAITQ--AER-IGEPDYEVDEDIFRKKRLTIMDLHPGA
4M9K_NS3     GLYGNG-------VVTRSGAY---------VSAIAN--TEKSIE-DNPEIEDDIFRK------------
3U1I_NS3     GLYGNG-------VVTKNGGY---------VSGIA--Q-------------------------------

cons                                        .


RefSeq_D1    GKTRRYLPAIVREAIKRKLRTLVLAPTRVVASEMAEALKGMPIRYQTTAVKSEHTGKEIVDLMCHATFT
RefSeq_D2    GKTKRYLPAIVREAIKRGLRTLILAPTRVVAAEMEEALRGLPIRYQTPAIRAEHTGREIVDLMCHATFT
RefSeq_D3    GKTRKYLPAIVREAIKRRLRTLILAPTRVVAEMEEALKGLPIRYQTTATKSEHTGREIVDLMCHATFT
RefSeq_D4    GKTKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSEHTGREIVDLMCHATFT
4M9K_NS3     ---------------------------------------------------------------------
3U1I_NS3     ---------------------------------------------------------------------
cons
```

```
RefSeq_D1   MRLLSPVRVPNYNMIIMDEAHFTDPASIAARGYISTRVGMGEAAAIFMTATPPGSVEAFPQSNAVIQDE
RefSeq_D2   MRLLSPVRVPNYNLIIMDEAHFTDPASIAARGYISTRVEMGEAAGIFMTATPPGSRDPFPQSNAPIIDE
RefSeq_D3   MRLLSPVRVPNYNLIIMDEAHFTDPASIAARGYISTRVGMGEAAAIFMTATPPGTADAFPQSNAPIQDE
RefSeq_D4   TRLLSSTRVPNYNLIVMDEAHFTDPSSVAARGYISTRVEMGEAAAIFMTATPPGATDPFPQSNSPIEDI
4M9K_NS3    ------------------------------------------------------------------------
3U1I_NS3    ------------------------------------------------------------------------

cons

RefSeq_D1   ERDIPERSWNSGYDWITDFPGKTVWFVPSIKSGNDIANCLRKNGKRVVQLSRKTFDTEYQKTKNNDWDY
RefSeq_D2   EREIPERSWNSGHEWVTDFKGKTVWFVPSIKAGNDIAACLRKNGKKVIQLSRKTFDSEYVKTRTNDWDF
RefSeq_D3   ERDIPERSWNSGNEWITDFAGKTVWFVPSIKAGNDIANCLRKNGKKVIQLSRKTFDTEYQKTKLNDWDF
RefSeq_D4   EREIPERSWNTGFDWITDYQGKTVWFVPSIKAGNDIANCLRKSGKKVIQLSRKTFDTEYPKTKLTDWDF
4M9K_NS3    ------------------------------------------------------------------------
3U1I_NS3    ------------------------------------------------------------------------

cons

RefSeq_D1   VVTTDISEMGANFRADRVIDPRRCLKPVILKDGPERVILAGPMPVTVASAAQRRGRIGRNQNKEGDQYI
RefSeq_D2   VVTTDISEMGANFKAERVIDPRRCMKPVILTDGEERVILAGPMPVTHSSAAQRRGRIGRNPKNENDQYI
RefSeq_D3   VVTTDISEMGANFKADRVIDPRRCLKPVILTDGPERVILAGPMPVTAASAAQRRGRVGRNPQKENDQYI
RefSeq_D4   VVTTDISEMGANFRAGRVIDPRRCLKPVILPDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYV
4M9K_NS3    ------------------------------------------------------------------------
3U1I_NS3    --------------TNAEPDGPTPELEEE-----------------------------------------

cons

RefSeq_D1   YMGQPLNNDEDHAHWTEAKMLLDNINTPEGIIPALFEPEREKSAAIDGEYRLRGEARKTFVELMRRGDL
RefSeq_D2   YMGEPLENDEDCAHWKEAKMLLDNINTPEGIIPSMFEPEREKVDAIDGEYRLRGEARKTFVDLMRRGDL
RefSeq_D3   FTGQPLNNDEDHAHWTEAKMLLDNINTPEGIIPALFEPEREKSAAIDGEYRLKGESRKTFVELMRRGDL
RefSeq_D4   FSGDPLKNDEDHAHWTEAKMLLDNIYTPEGIIPTLFGPEREKTQAIDGEFRLRGEQRKTFVELMRRGDL
4M9K_NS3    ------------------------------------------------------------------------
3U1I_NS3    ------------------------------------------------------------------------

cons

RefSeq_D1   PVWLSYKVASEGFQYSDRRWCFDGERNNQVLEENMDVEIWTKEGERKKLRPRWLDARTYSDPLALREFK
RefSeq_D2   PVWLAYRVAAEGINYADRRWCFDGVKNNQILEENVEVEIWTKEGERKKLKPRWLDARIYSDPLALKEFK
RefSeq_D3   PVWLAHKVASEGIKYTDRKWCFDGQRNNQILEENMDVEIWTKEGEKKKLRPRWLDARTYSDPLALKEFK
RefSeq_D4   PVWLSYKVASAGISYEDREWCFTGERNNQILEENMEVEIWTREGEKKKLRPRWLDARVYADPMALKDFK
4M9K_NS3    ------------------------------------------------------------------------
3U1I_NS3    ------------------------------------------------------------------------

cons

RefSeq_D1   EFAAGRR
RefSeq_D2   EFAAGRK
RefSeq_D3   DFAAGRK
RefSeq_D4   EFASGRK
4M9K_NS3    -------
3U1I_NS3    -------

cons
```

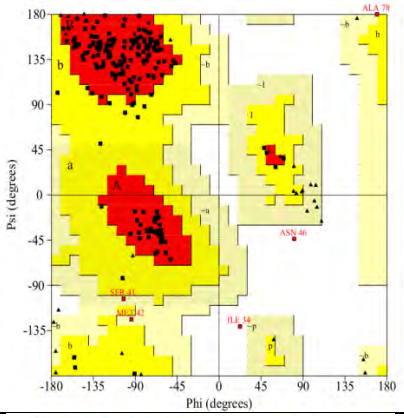
APPENDIX C

MSA Result of NS3 Protease Domain Sequences via T-Coffee

```
T-COFFEE, Version_11.00.d625267
CPU TIME:0 sec.
SCORE=99
*
 BAD AVG GOOD
*
DENV1_NS3pro    :  99
DENV2_NS3pro    :  99
DENV3_NS3pro    :  99
DENV4_NS3pro    :  99
4M9K_NS3        :  99
3U1I_NS3        :  99
cons            :   9

DENV1_NS3pro    ------------------------DGIYRILQRGLLGRSQVGVGVFQEGVFHTMWHVTRGAVLMYQ
DENV2_NS3pro    ------------------------EDGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHK
DENV3_NS3pro    ------------------------GVYRIKQQGIFGKTQVGVGVQKEGVFHTMWHVTRGAVLTHN
DENV4_NS3pro    ------------------------SEGVYRIMQRGLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHE
4M9K_NS3        -----AGVLWDVPSPPPVGKAELEDGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHK
3U1I_NS3        SGVLWD----VPSPPETQK-AELEEGVYRIKQQGIFGKTQVGVGVQKEGVFHTMWHVTRGAVLTHN

cons                     *  *** *:*::* :*:*.*:  **.**********:*: ::


DENV1_NS3pro    GKRLEPSWASVKKDLISYGGGWRFQGSWNAGEEVQVIAVEPGKNPKNVQTAPGTFKTPEGEVGAIA
DENV2_NS3pro    GKRIEPSWADVKKDLISYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGLFKTNAGTIGAVS
DENV3_NS3pro    GKRLEPNWASVKKDLISYGGGWRLSAQWQKGEEVQVIAVEPGKNPKNFQTTPGTFQTTTGEIGAIA
DENV4_NS3pro    TGRLEPSWADVRNDMISYGGGWRLGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVT
4M9K_NS3        GKRIEPSWADVKKDLISYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGLFKTNTGTIGAVS
3U1I_NS3        GKRLEPNWASVKKDLISYGGGWRLSAQWQKGEEVQVIAVEPGKNPKNFQTMPGTFQTTTGEIGAIA

cons            *:**.**.*::*:*******::  .*.  *:***:*:******: .** ** *:*  * :**::


DENV1_NS3pro    LDFKPGTSGSPIVNREGKIVGLYGNGVVTTSGTYVSAIAQ------------------
DENV2_NS3pro    LDFSPGTSGSPIIDKKGKVVGLYGNGVVTRSGAYVSAIAQ------------------
DENV3_NS3pro    LDFKPGTSGSPIINREGKVVGLYGNGVVTKNGGYVSGIAQ------------------
DENV4_NS3pro    LDFKPGTSGSPIINRKGKVIGLYGNGVVTKSGDYVSAITQ------------------
4M9K_NS3        LDFSPGTSGSPIVDKKGKVVGLYGNGVVTRSGAYVSAIANTEKSIEDNPEIEDDIFRK
3U1I_NS3        LDFKPGTSGSPIINREGKVVGLYGNGVVTKNGGYVSGIAQTNAEPDGPTPELEEE---

cons            ***.********:::::**::*********  .* ***.*::
```

APPENDIX D

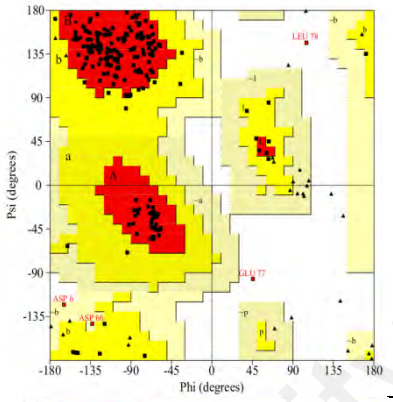
Ramachandran Plots of the Generated Models via DOPE and DOPEHR Protocols (MODELLER) for the 'Open' and 'Closed' Protease Complex



DOPE-based modelling ('Open' Structure)

Plot statistics

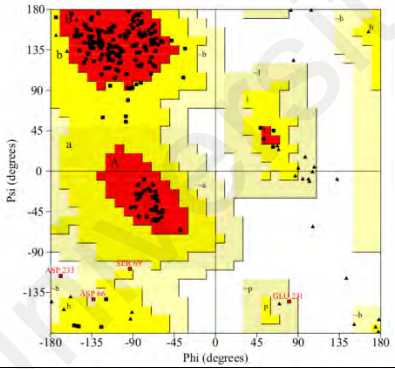
Residues in most favoured regions [A,B,L]	176	89.3%
Residues in additional allowed regions [a,b,l,p]	16	8.1%
Residues in generously allowed regions [~a,~b,~l,~p]	3	1.5%
Residues in disallowed regions	2	1.0%
Number of non-glycine and non-proline residues	197	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	37	
Number of proline residues	12	
Total number of residues	247	



DOPEHR-based modelling ('Open' Structure)

Plot statistics

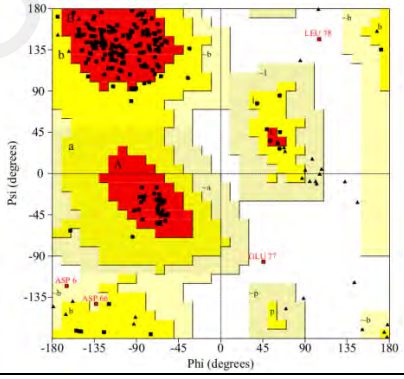
Residues in most favoured regions [A,B,L]	175	88.8%
Residues in additional allowed regions [a,b,l,p]	17	8.6%
Residues in generously allowed regions [~a,~b,~l,~p]	3	1.5%
Residues in disallowed regions	2	1.0%
Number of non-glycine and non-proline residues	197	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	37	
Number of proline residues	12	
Total number of residues	247	



DOPE-based modelling ('Closed' Structure)

Plot statistics

Residues in most favoured regions [A,B,L]	168	88.9%
Residues in additional allowed regions [a,b,l,p]	17	9.0%
Residues in generously allowed regions [~a,~b,~l,~p]	3	1.6%
Residues in disallowed regions	1	0.5%
Number of non-glycine and non-proline residues	189	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	39	
Number of proline residues	13	
Total number of residues	242	



DOPEHR-based modelling ('Closed' Structure)

Plot statistics

Residues in most favoured regions [A,B,L]	163	86.2%
Residues in additional allowed regions [a,b,l,p]	22	11.6%
Residues in generously allowed regions [~a,~b,~l,~p]	2	1.1%
Residues in disallowed regions	2	1.1%
Number of non-glycine and non-proline residues	189	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	39	
Number of proline residues	13	
Total number of residues	242	